1 WAVE NATURE OF LIGHT

1.1 LIGHT WAVES IN A HOMOGENEOUS MEDIUM

A. Plane Electromagnetic Wave

The wave nature of light, quite aside from its photonic behavior, is well recognized by such phenomena as interference and diffraction. We can treat light as an electromagnetic (EM) wave with time-varying electric and magnetic fields, E_x and B_y, respectively, which are propagating through space in such a way that they are always perpendicular to each other and the direction of propagation z as illustrated in Figure 1.1. The simplest traveling wave is a sinusoidal wave that, for propagation along z, has the general mathematical form

Traveling wave along z

$$E_x = E_o \cos(\omega t - kz + \phi_o) \tag{1.1.1}$$

in which E_x is the electric field at position z at time t, k is the **propagation constant**[2] given by $2\pi/\lambda$, where λ is the wavelength, ω is the angular frequency, E_o is the amplitude of the wave, and ϕ_o is a phase constant, which accounts for the fact that at $t = 0$ and $z = 0$; E_x may or may not necessarily be zero depending on the choice of origin. The argument $(\omega t - kz + \phi_o)$ is called the **phase** of the wave and denoted by ϕ. Equation (1.1.1) describes a **monochromatic plane wave** of infinite extent traveling in the positive z direction as depicted in Figure 1.2. In any plane perpendicular to the direction of propagation (along z), the phase of the wave, according to Eq. (1.1.1), is constant, which means that the field in this plane is also constant. A surface over which the phase of a wave is constant at a given instant is referred to as a **wavefront**. A wavefront of a plane wave is obviously an infinite plane perpendicular to the direction of propagation as shown in Figure 1.2.

We know from electromagnetism that time-varying magnetic fields result in time-varying electric fields (Faraday's law) and vice versa. A time-varying electric field would set up a time-varying magnetic field with the same

[2] Some authors also call k the *wave number*. However, in spectroscopy, the wave number implies $1/\lambda$, reciprocal wavelength. To avoid any confusion, propagation constant would be preferred for k.

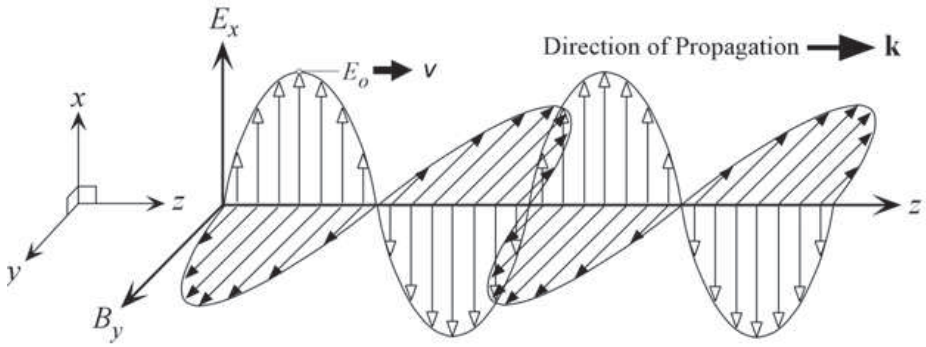


FIGURE 1.1 An electromagnetic wave in a homogenous and isotropic medium is a traveling wave that has time-varying electric and magnetic fields which are perpendicular to each other and the direction of propagation z. This is a snapshot at a given time of a particular harmonic or a sinusoidal EM wave. At a time δt later, a point on the wave, such as the maximum field, would have moved a distance $v\delta t$ in the z-direction.

frequency. According to electromagnetic principles,[3] a traveling electric field E_x as represented by Eq. (1.1.1) would always be accompanied by a traveling magnetic field B_y with the same wave frequency and propagation constant (ω and k) but the directions of the two fields would be orthogonal as in Figure 1.1. Thus, there is a similar traveling wave equation for the magnetic field component B_y. We generally describe the interaction of a light wave with a non-conducting matter (conductivity, $\sigma = 0$) through the electric field component E_x rather than B_y because it is the electric field that displaces the electrons in molecules or ions in the crystal and thereby gives rise to the polarization of matter. However, the two fields are linked, as in Figure 1.1, and there is an intimate relationship between them. The **optical field** refers to the electric field E_x.

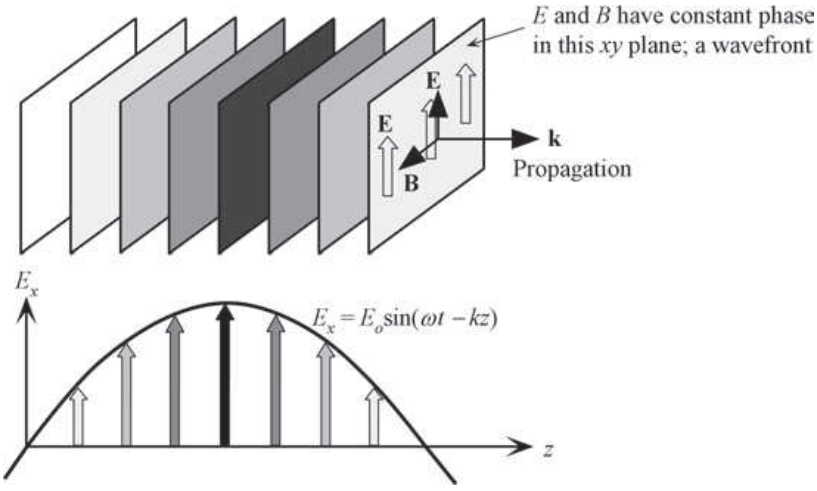


FIGURE 1.2 A plane EM wave traveling along z has the same E_x (or B_y) at any point in a given xy plane. All electric field vectors in a given xy plane are therefore in phase. The xy planes are of infinite extent in the x and y directions.

[3]Maxwell's equations formulate electromagnetic phenomena and provide relationships between the electric and magnetic fields and their space and time derivatives. We need only to use a few selected results from Maxwell's equation without delving into their derivations. The magnetic field B is also called the magnetic induction or magnetic flux density. The magnetic field *intensity* H and magnetic field B in a non-magnetic material are related by $B = \mu_o H$ in which μ_o is the absolute permeability of the medium.

We can also represent a traveling wave using the exponential notation since $\cos\phi = \mathsf{Re}\left[\exp(j\phi)\right]$ in which Re refers to the real part. We then need to take the real part of any complex result at the end of calculations. Thus, we can write Eq. (1.1.1) as

$$E_x(z, t) = \mathsf{Re}\left[E_o \exp(j\phi_o) \exp j(\omega t - kz)\right]$$

or

$$E_x(z, t) = \mathsf{Re}\left[E_c \exp j(\omega t - kz)\right] \tag{1.1.2}$$

Traveling wave along z

in which $E_c = E_o \exp(j\phi_o)$ is a complex number that represents the amplitude of the wave and includes the constant phase information ϕ_o. Note that in Eq. (1.1.2), $\exp j(\omega t - kz)$ represents $e^{j(\omega t - kz)}$.

We indicate the direction of propagation with a vector **k**, called the **wave vector** (or **propagation vector**), whose magnitude is the **propagation constant**, $k = 2\pi/\lambda$. It is clear that **k** is perpendicular to constant phase planes as indicated in Figure 1.2. Consider an electromagnetic wave that is propagating along some arbitrary direction **k**, as indicated in Figure 1.3. The electric field $E(\mathbf{r}, t)$ at an arbitrary point **r** is given by

$$E(\mathbf{r}, t) = E_o \cos(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_o) \tag{1.1.3}$$

Traveling wave in 3D with wave vector **k**

because the dot product $\mathbf{k}\cdot\mathbf{r}$ is along the direction of propagation similar to kz as indicated in Figure 1.3. The latter can be shown by drawing a plane that has the point **r** and is perpendicular to **k** as illustrated in Figure 1.3. The dot product is the product of **k** and the projection of **r** onto **k**, which is $\mathbf{r}'$ in Figure 1.3, so that $\mathbf{k}\cdot\mathbf{r} = kr'$. Indeed, if propagation is along z, $\mathbf{k}\cdot\mathbf{r}$ becomes kz. In general, if **k** has components k_x, k_y, and k_z along x, y, and z, then from the definition of the dot product, $\mathbf{k}\cdot\mathbf{r} = k_x x + k_y y + k_z z$.

The relationship between time and space for a given phase, ϕ for example, that corresponds to a maximum field, according to Eq. (1.1.1), is described by

$$\phi = \omega t - kz + \phi_o = \text{constant}$$

During a time interval δt, this constant phase (and hence the maximum field) moves a distance δz. The phase velocity of this wave is therefore $\delta z/\delta t$. Thus the **phase velocity** v is

$$v = \frac{dz}{dt} = \frac{\omega}{k} = \upsilon\lambda \tag{1.1.4}$$

Phase velocity

in which υ is the frequency ($\omega = 2\pi\upsilon$) of the EM wave. For an EM wave propagating in free space v is the speed of light in vacuum or c.

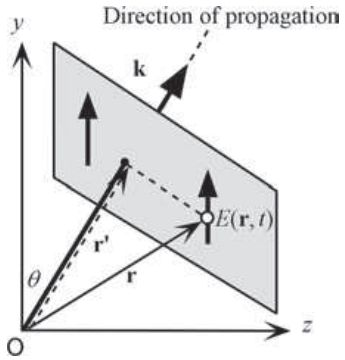


FIGURE 1.3 A traveling plane EM wave along a direction **k**.

We are often interested in the phase difference $\Delta\phi$ at a given time between two points on a wave (Figure 1.1) that are separated by a certain distance. If the wave is traveling along z with a wave vector k, as in Eq. (1.1.1), then the phase difference between two points separated by Δz is simply $k\Delta z$ since ωt is the same for each point. If this phase difference is 0 or multiples of 2π then the two points are in phase. Thus phase difference $\Delta\phi$ can be expressed as $k\Delta z$ or $2\pi\Delta z/\lambda$.

B. Maxwell's Wave Equation and Diverging Waves

Consider the plane EM wave in Figure 1.2. All constant phase surfaces are xy planes that are perpendicular to the z-direction. A cut of a plane wave parallel to the z-axis is shown in Figure 1.4 (a) in which the parallel dashed lines at right angles to the z-direction are wavefronts. We normally show wavefronts that are separated by a phase of 2π or a whole wavelength λ as in the figure. The vector that is normal to a wavefront surface at a point such as P represents the direction of wave propagation (**k**) at that point P. Clearly, the propagation vectors everywhere are all parallel and the plane wave propagates without the wave diverging; *the plane wave has no divergence.* The amplitude of the planar wave E_o does not depend on the distance from a reference point, and it is the same at all points on a given plane perpendicular to **k** (*i.e.*, independent of x and y). Moreover, as these planes extend to infinity there is infinite energy in the plane wave. A plane wave such as the one in Figure 1.4 (a) is an idealization that is useful in analyzing many wave phenomena. In reality, however, the electric field in a plane at right angles to **k** does not extend to infinity since the light beam would have a finite cross-sectional area and finite power. We would need an infinitely large EM source with infinite power to generate a perfect plane wave!

In practice there are many types of possible EM waves. These waves must obey a special wave equation that describes the time and space dependence of the electric field. In an isotropic and linear dielectric medium, the relative permittivity (ε_r) is the same in all directions and is independent of the electric field. The field E in such a medium obeys **Maxwell's EM wave equation**

Maxwell's wave equation

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} - \varepsilon_o \varepsilon_r \mu_o \frac{\partial^2 E}{\partial t^2} = 0 \qquad \textbf{(1.1.5)}$$

in which μ_o is the absolute permeability, ε_o is the absolute permittivity, and ε_r is the relative permittivity of the medium. Equation (1.1.5) assumes an isotropic medium (as discussed later) and

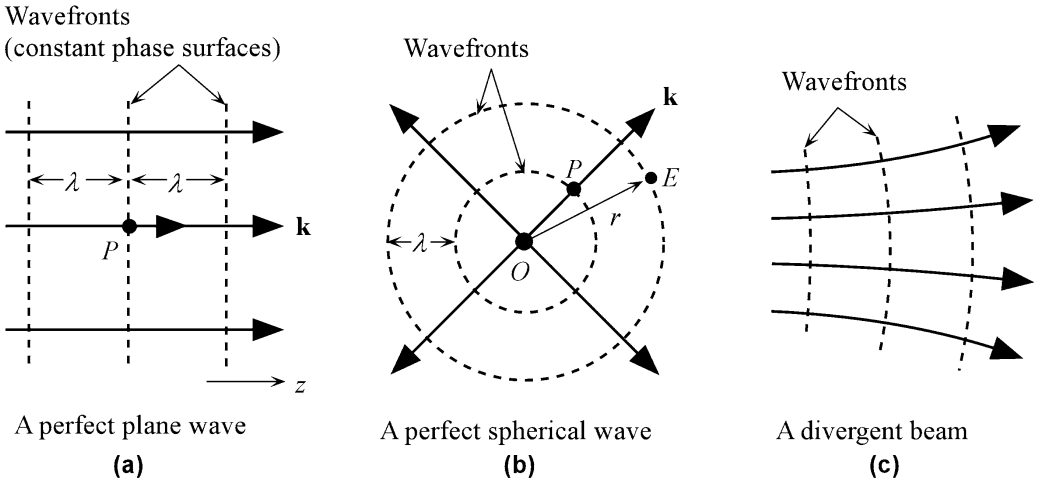


FIGURE 1.4 Examples of possible EM waves. (a) A perfect plane wave. (b) A perfect spherical wave. (c) A divergent beam.

that the conductivity of the medium is zero. To find the time and space dependence of the field, we must solve Eq. (1.1.5) in conjunction with the initial and boundary conditions. We can easily show that the plane wave in Eq. (1.1.1) satisfies Eq. (1.1.5). There are many possible waves that satisfy Eq. (1.1.5) that can therefore exist in nature.

A **spherical wave** is described by a traveling field that emerges from a point EM source and whose amplitude decays with distance r from the source. At any point r from the source, the field is given by

$$E = \frac{A}{r}\cos(\omega t - kr) \tag{1.1.6}$$

Spherical wave

in which A is a constant. We can substitute Eq. (1.1.6) into Eq. (1.1.5) to show that Eq. (1.1.6) is indeed a solution of Maxwell's equation (transformation from Cartesian to spherical coordinates would help). A cut of a spherical wave is illustrated in Figure 1.4 (b) where it can be seen that *wavefronts are spheres* centered at the point source O. The direction of propagation **k** at any point such as P is determined by the normal to the wavefront at that point. Clearly **k** vectors diverge out and, as the wave propagates, the constant phase surfaces become larger. **Optical divergence** refers to the angular separation of wave vectors on a given wavefront. The spherical wave has 360° of divergence in all planes through the point source. It is apparent that plane and spherical waves represent two extremes of wave propagation behavior from perfectly parallel to fully diverging wave vectors. They are produced by two extreme sizes of EM wave source: an infinitely large source for the plane wave and a point source for the spherical wave. In reality, an EM source is neither of infinite extent nor in point form, but would have a finite size and finite power. Figure 1.4 (c) shows a more practical example in which a light beam exhibits some inevitable divergence while propagating; the wavefronts are slowly bent away thereby spreading the wave. Light rays of geometric optics are drawn to be normal to constant phase surfaces (wavefronts). Light rays therefore follow the wave vector directions. Rays in Figure 1.4 (c) slowly diverge away from each other. The reason for favoring plane waves in many optical explanations is that, at a distance far away from a source, over a *small spatial region*, the wavefronts will appear to be plane even if they are actually spherical. Figure 1.4 (a) may be a small part of a huge spherical wave.

Many light beams, such as the output from a laser, can be described by assuming that they are **Gaussian beams**. Figure 1.5 illustrates a Gaussian beam traveling along the z-axis. The beam still has an $\exp j(\omega t - kz)$ dependence to describe propagation characteristics but the amplitude varies spatially away from the beam axis and also *along* the beam axis. Such a beam has similarities to that in Figure 1.4 (c); it slowly diverges[4] and is the result of radiation from a source of finite extent. The light intensity (*i.e.*, the radiation energy flow per unit area per unit time) distribution across the beam cross-section anywhere along z is Gaussian as shown in Figures 1.5 (b) and (c). The **beam diameter** $2w$ at any point z is defined in such a way that the cross-sectional area πw^2 at that point contains 86% of the beam power. Thus, the beam diameter $2w$ increases as the beam travels along z. The Gaussian beam shown in Figure 1.5 (a) starts from O with a finite width $2w_o$ where the wavefronts are parallel and then the beam slowly diverges as the wavefronts curve out during propagation along z. The finite width $2w_o$ where the wavefronts are parallel is called the **waist** of the beam; w_o is the **waist radius**

[4]The divergence is due to the self-diffraction of the beam—the beam is diffraction limited. Diffraction is covered later in this chapter. Further, the intensity of light will be defined quantitatively in Section 1.4. For the present discussion it represents the radiation energy flow per unit time per unit area.

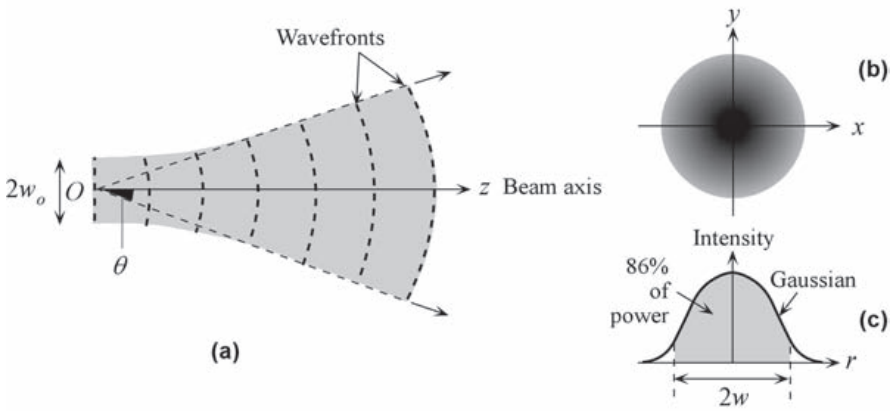


FIGURE 1.5 (a) Wavefronts of a Gaussian light beam. (b) Light intensity across a beam cross-section. (c) Light intensity vs. radial distance r from beam axis (z).

and $2w_o$ is the **spot size**. Far away from the source, the beam diameter $2w$ increases linearly with distance z. The increase in beam diameter $2w$ with z makes an angle 2θ at O, as shown in Figure 1.5, which is called the **beam divergence**. The greater the waist, the narrower the divergence. The two are related by

Far field Gaussian beam divergence

$$2\theta = \frac{4\lambda}{\pi(2w_o)} \tag{1.1.7}$$

It can be shown that the Gaussian beam is a solution of Maxwell's equations when the beam divergence is small,[5] and the intensity decays slowly with distance z. Suppose that we reflect the Gaussian beam back on itself (by using a spherical mirror that has the right curvature to match the incident wavefront) so that the beam is traveling in the $-z$ direction and converging toward O; simply reverse the direction of travel in Figure 1.5 (a). The wavefronts would be "straightening out," and at O they would be parallel again. The beam would still have the same finite diameter $2w_o$ (waist) at O. From then on, the beam again diverges out just as it did traveling in $+z$ direction as illustrated in Figure 1.6 (a). When we try to focus a Gaussian beam using a lens or a spherical mirror, as in Figure 1.6 (a), the beam cannot be brought a point but to a finite spot size $2w_o$. The relationship in Eq. (1.1.7) for the beam divergence can be used to find this minimum spot size $2w_o$ to which a Gaussian beam can be focused.

At a certain distance z_o from O, the beam diameter becomes $2^{1/2}(2w_o)$ as illustrated in Figure 1.6 (a). The distance z_o is called the **Rayleigh range**, and is given by

Gaussian beam Rayleigh range

$$z_o = \frac{\pi w_o^2}{\lambda} \tag{1.1.8}$$

The Rayleigh range is also known as the **depth of focus**. $2z_o$ is called the **confocal parameter**. The region along z far away from the Rayleigh region, $z \gg z_o$, is called the **far-field region**.

The width $2w$ of a Gaussian beam at a position along z increases with z and is given by

Gaussian beam width at distance

$$2w = 2w_o\left[1 + \left(\frac{z}{z_o}\right)^2\right]^{1/2} = 2w_o\left[1 + \left(\frac{z\lambda}{\pi w_o^2}\right)^2\right]^{1/2} \tag{1.1.9a}$$

[5] This is called the *paraxial approximation* in which the normals to the wavefronts, the wave vectors, make small angles with the z-axis.

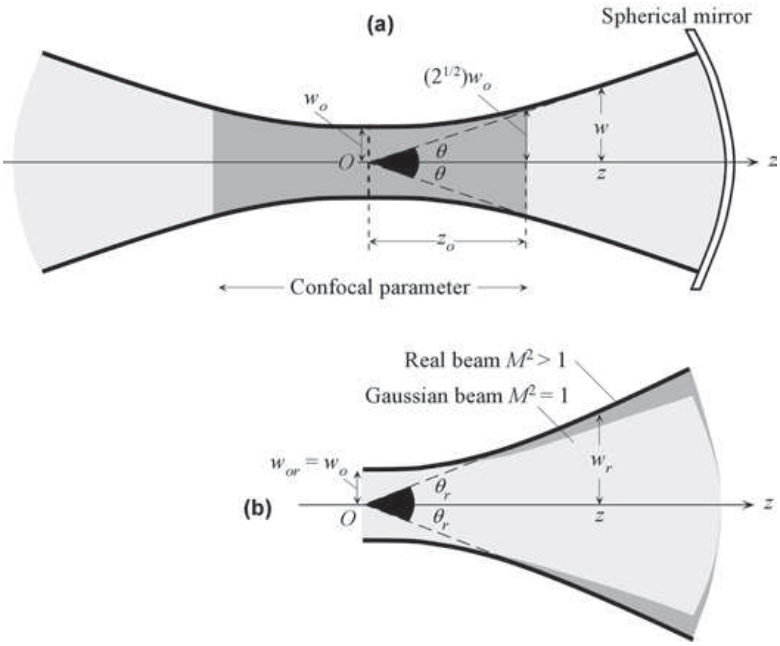


FIGURE 1.6 (a) Gaussian beam definitions. The region z_o is called the Rayleigh range and also the depth of focus. (b) Comparison of a real beam with $M^2 > 1$ with a Gaussian beam with $M^2 = 1$ and the same waist $2w_o$.

Far away from the Rayleigh range, for $z \gg z_o$, in the far-field region, the beam width increases *linearly* with z, that is,

$$2w \approx (2w_o)\frac{z}{z_o} \tag{1.1.9b}$$

Gaussian beam width at distance z

Notice that the product of the beam radius w_o (half the beam waist) and half the divergence angle θ from Eq. (1.1.7) is given by $w_o\theta = \lambda/\pi$, that is, it depends only on the wavelength and is a well-defined constant for a given wavelength. The product $w_o\theta$ is called the **beam parameter product**.

The Gaussian beam concept is so useful in photonics that a special quantity, called the M^2-factor, has been introduced to compare a given laser beam to an ideal Gaussian beam. The ***M*² factor** measures the deviation of the real laser beam from the Gaussian characteristics, in which $M^2 = 1$ for an ideal (theoretical) Gaussian beam shape. Suppose that $2\theta_r$ and $2w_{or}$ are the divergence and waist, respectively, of the *real* laser beam, and 2θ and $2w_o$ are those for the ideal Gaussian. The M^2 *factor* is defined by[6]

$$M^2 = \frac{w_{or}\theta_r}{w_o\theta} = \frac{w_{or}\theta_r}{(\lambda/\pi)} \tag{1.1.10}$$

M^2 factor definition

where we have used $w_o\theta = \lambda/\pi$ for an ideal Gaussian beam.

According to Eq. (1.1.10), M^2 is the ratio of the beam parameter product of the real beam to that of a Gaussian beam, and hence M^2 gauges the **beam quality** of the laser beam. For many lasers, M^2 is greater than unity, and can be as high as 10–30 in multimode lasers.

[6]Some authors define M, instead of M^2, as $M = \theta_r/\theta = w_{or}/w_o$. In addition, the reader should not be too concerned with the terms "multimode" and "single mode" at this point, except that they represent the types of radiation that is emitted from lasers.

For stable He-Ne single mode lasers, it is very close to unity. We can apply the above Gaussian equations to a real laser beam by replacing w_o with w_{or}/M^2. For example, the width at a distance z becomes

Beam width at distance z

$$2w = 2w_{or}\left[1 + \left(\frac{z\lambda M^2}{\pi w_{or}}\right)^2\right]^{1/2} \tag{1.1.11}$$

Far away from the Rayleigh range, $2w_r = M^2(2w)$, where $2w$ is the ideal Gaussian beam width at the same location. Suppose that we put a Gaussian beam with a waist w_o onto the real beam and adjust w_o to be the same as w_{or}, *i.e.*, $w_{or} = w_o$. Then, from Eq. (1.1.10) the divergence of the real beam is greater inasmuch as $\theta_r = M^2\theta$, which is shown in Figure 1.6 (b).

EXAMPLE 1.1.1 A diverging laser beam

Consider a He-Ne laser beam at 633 nm with a spot size of 1 mm. Assuming a Gaussian beam, what is the divergence of the beam? What are the Rayleigh range and the beam width at 25 m?

Solution

Using Eq. (1.1.7), we find

$$2\theta = \frac{4\lambda}{\pi(2w_o)} = \frac{4(633 \times 10^{-9}\,\text{m})}{\pi(1 \times 10^{-3}\,\text{m})} = 8.06 \times 10^{-4}\,\text{rad} = 0.046^\circ$$

The Rayleigh range is

$$z_o = \frac{\pi w_o^2}{\lambda} = \frac{\pi\left[(1 \times 10^{-3}\,\text{m})/2\right]^2}{(633 \times 10^{-9}\,\text{m})} = 1.24\,\text{m}$$

The beam width at a distance of 25 m is

$$2w = 2w_o\left[1 + (z/z_o)^2\right]^{1/2} = (1 \times 10^{-3}\,\text{m})\left\{1 + \left[(25\,\text{m})/(1.24\,\text{m})\right]^2\right\}^{1/2}$$
$$= 0.0202\,\text{m} \quad \text{or} \quad 20\,\text{mm}.$$

1.2 REFRACTIVE INDEX AND DISPERSION

When an EM wave is traveling in a dielectric medium, the oscillating electric field polarizes the molecules of the medium at the frequency of the wave. Indeed, the EM wave propagation can be considered to be the propagation of this polarization in the medium. The field and the induced molecular dipoles become coupled. The relative permittivity ε_r measures the ease with which the medium becomes polarized and hence it indicates the extent of interaction between the field and the induced dipoles. To find the nature of propagation of an EM wave in a dielectric medium, and hence the phase velocity, we have to solve Maxwell's equations in a dielectric medium. If we assume the medium is insulating, nonmagnetic, and also **isotropic**, that is the relative permittivity is independent of the direction of propagation of the EM wave and the optical field, then the solution becomes quite simple and leads to Maxwell's wave equation stated in Eq. (1.1.5). We can continue to represent the EM wave in a similar fashion to its propagation in vacuum but

we need to assign to it a new phase velocity v, and a new wavelength, both of which depend on ε_r. In a dielectric medium of relative permittivity ε_r, the phase velocity v is given by

$$v = \frac{1}{\sqrt{\varepsilon_r \varepsilon_o \mu_o}} \tag{1.2.1}$$

Phase velocity in a medium

It is important to use the relative permittivity at the frequency of operation in Eq. (1.2.1) since ε_r depends on the frequency. Typical frequencies that are involved in optoelectronic devices are in the infrared (including far infrared), visible, and UV, and we generically refer to these frequencies as *optical frequencies*; they cover a somewhat arbitrary range from roughly 10^{12} Hz to 10^{16} Hz.

For an EM wave traveling in free space, $\varepsilon_r = 1$ and $v_{\text{vacuum}} = 1/(\varepsilon_o \mu_o)^{1/2} = c = 3 \times 10^8\,\text{m}\,\text{s}^{-1}$, the velocity of light in vacuum. The ratio of the speed of light in free space to its speed in a medium is called the **refractive index** n of the medium, that is,

$$n = \frac{c}{v} = \sqrt{\varepsilon_r} \tag{1.2.2}$$

Definition of refractive index

If k is the propagation constant ($k = 2\pi/\lambda$) and λ is the wavelength, both in free space, then in the medium[7] $k_{\text{medium}} = nk$ and $\lambda_{\text{medium}} = \lambda/n$. Equation (1.2.2) is in agreement with our intuition that light propagates more slowly in a denser medium that has a higher refractive index. We should note that the frequency υ (or ω) remains the same.[8]

The refractive index of a medium is not necessarily the same in all directions. In noncrystalline materials such as glasses and liquids, the material structure is the same in all directions and n does not depend on the direction. The refractive index is then **isotropic**. In crystals, however, the atomic arrangements and interatomic bonding are different along different directions. Crystals, in general, have nonisotropic, or **anisotropic**, properties. Depending on the crystal structure, the relative permittivity ε_r is different along different crystal directions. This means that, in general, the refractive index n seen by a propagating electromagnetic wave in a crystal will depend on the value of ε_r along the direction of the oscillating electric field (*i.e.*, along the direction of polarization). For example, suppose that the wave in Figure 1.1 is traveling along the z-direction in a particular crystal with its electric field oscillating along the x-direction. If the relative permittivity along this x-direction is ε_{rx}, then $n_x = (\varepsilon_{rx})^{1/2}$. The wave therefore propagates with a phase velocity that is c/n_x. The variation of n with direction of propagation and the direction of the electric field depends on the particular crystal structure. With the exception of cubic crystals (such as diamond), all crystals exhibit a degree of optical anisotropy that leads to a number of important applications as discussed in Chapter 6. Typically, noncrystalline solids, such as glasses and liquids, and cubic crystals are **optically isotropic**; they possess only one refractive index for all directions.

Relative permittivity ε_r or the dielectric constant of materials, in general, depends on the frequency of the electromagnetic wave. The relationship $n = (\varepsilon_r)^{1/2}$ between the refractive index n and ε_r must be applied at the same frequency for both n and ε_r. The relative permittivity for many materials can be vastly different at high and low frequencies because different polarization

[7] On occasions, we will need to use k_o and λ_o for the free-space propagation constant and wavelength (as in the next section) and use k and λ for those values inside the medium. In each case, these quantities will be clearly defined to avoid confusion.

[8] We are accustomed to describing light in terms of its wavelength and often quote wavelengths in nm (or Å) in the visible and IR regions. However, there would be certain advantages to using the frequency instead of the wavelength, for example, terahertz (THz) instead of nm, one of which is that the frequency v does not change in the medium. (See Roger A. Lewis, *Am. J. Phys*, *79*, 341, 2011.)

TABLE 1.1 Low-frequency (LF) relative permittivity ε_r(LF) and refractive index n

Material	ε_r(LF)	$[\varepsilon_r(\text{LF})]^{1/2}$	n (at λ)	Comment
Si	11.9	3.44	3.45 (at 2.15 μm)	Electronic bond polarization up to optical frequencies
Diamond	5.7	2.39	2.41 (at 590 nm)	Electronic bond polarization up to UV light
GaAs	13.1	3.62	3.30 (at 5 μm)	Ionic polarization contributes to ε_r(LF)
SiO_2	3.84	2.00	1.46 (at 600 nm)	Ionic polarization contributes to ε_r(LF)
Water	80	8.9	1.33 (at 600 nm)	Dipolar polarization contributes to ε_r(LF), which is large

mechanisms operate at these frequencies.[9] At low frequencies all polarization mechanisms present can contribute to ε_r, whereas at optical frequencies only the electronic polarization can respond to the oscillating field. Table 1.1 lists the relative permittivity ε_r(LF) at low frequencies (*e.g.*, 60 Hz or 1 kHz as would be measured, for example, using a capacitance bridge in the laboratory) for various materials. It then compares $\left[\varepsilon_r(\text{LF})\right]^{1/2}$ with n.

For silicon and diamond there is an excellent agreement between $\left[\varepsilon_r(\text{LF})\right]^{1/2}$ and n. Both are covalent solids in which electronic polarization (electronic bond polarization) is the only polarization mechanism at low and high frequencies. Electronic polarization involves the displacement of light electrons with respect to positive ions of the crystal. This process can readily respond to the field oscillations up to optical or even ultraviolet frequencies.

For GaAs and SiO_2 $\left[\varepsilon_r(\text{LF})\right]^{1/2}$ is larger than n because at low frequencies both of these solids possess a degree of ionic polarization. The bonding is not totally covalent and there is a degree of ionic bonding that contributes to polarization at frequencies below far-infrared wavelengths.

In the case of water, the ε_r(LF) is dominated by orientational or dipolar polarization, which is far too sluggish to respond to high-frequency oscillations of the field at optical frequencies.

It is instructive to consider what factors affect n. The relative permittivity depends on the polarizability α per molecule (or atom) in the solid. (α is defined as the induced electric dipole moment per unit applied field.) The simplest and approximate expression for the relative permittivity is

$$\varepsilon_r \approx 1 + \frac{N\alpha}{\varepsilon_o}$$

in which N is the number of molecules per unit volume. Both the atomic concentration, or density, and polarizability therefore increase n. For example, glasses of given type but with greater density tend to have higher n.

The frequency or wavelength dependence of ε_r and hence n is called the **dispersion relation**, or simply dispersion. There are various theoretical and empirical models that describe the n vs. λ behavior. The Cauchy dispersion equation in its simplest form is given by[10]

Cauchy short form dispersion equation

$$n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \qquad \textbf{(1.2.3)}$$

[9] Chapters 7 and 9 in *Principles of Electronic Materials and Devices*, 3rd Edition, S. O. Kasap (McGraw-Hill, 2006) provides a semiquantitative description of the frequency dependence of ε_r and hence the wavelength dependence of n.

[10] Dispersion relations like the one in Eq. (1.2.3) are always in terms of the free-space wavelength λ. (It does not make sense to give them in terms of the actual wavelength in the medium.)

TABLE 1.2 Sellmeier and Cauchy coefficients

Sellmeier	A_1	A_2	A_3	λ_1 (μm)	λ_2 (μm)	λ_3 (μm)
SiO_2 (fused silica)	0.696749	0.408218	0.890815	0.0690660	0.115662	9.900559
86.5%SiO_2-13.5%GeO_2	0.711040	0.451885	0.704048	0.0642700	0.129408	9.425478
GeO_2	0.80686642	0.71815848	0.85416831	0.068972606	0.15396605	11.841931
Sapphire	1.023798	1.058264	5.280792	0.0614482	0.110700	17.92656
Diamond	0.3306	4.3356	–	0.1750	0.1060	–

Cauchy	Range of $h\nu$ (eV)	n_{-2} (eV^2)	n_0	n_2 (eV^{-2})	n_4 (eV^{-4})
Diamond	0.05–5.47	-1.07×10^{-5}	2.378	8.01×10^{-3}	1.04×10^{-4}
Silicon	0.002–1.08	-2.04×10^{-8}	3.4189	8.15×10^{-2}	1.25×10^{-2}
Germanium	0.002–0.75	-1.0×10^{-8}	4.003	2.2×10^{-1}	1.4×10^{-1}

Source: Sellmeier coefficients combined from various sources. Cauchy coefficients from D. Y. Smith *et al.*, *J. Phys. CM*, *13*, 3883, 2001.

where A, B, and C are material-specific constants. A more general Cauchy dispersion relation is of the form

Cauchy dispersion equation in photon energy

$$n = n_{-2}(h\nu)^{-2} + n_0 + n_2(h\nu)^2 + n_4(h\nu)^4 \tag{1.2.4}$$

where $h\nu$ is the photon energy, and n_0, n_{-2}, n_2, and n_4 are constants; values for diamond, Si, and Ge are listed in Table 1.2. The general Cauchy equation is usually applicable over a wide photon energy range.

Another useful dispersion relation that has been widely used, especially in optical fibers, is the Sellmeier equation given by

Sellmeier equation

$$n^2 = 1 + \frac{A_1\lambda^2}{\lambda^2 - \lambda_1^2} + \frac{A_2\lambda^2}{\lambda^2 - \lambda_2^2} + \frac{A_3\lambda^2}{\lambda^2 - \lambda_3^2} \tag{1.2.5}$$

where A_1, A_2, A_3 and λ_1, λ_2, λ_3 are constants, called **Sellmeier coefficients.**[11] Equation (1.2.5) turns out to be quite a useful semi-empirical expression for calculating n at various wavelengths if the Sellmeier coefficients are known. Higher terms involving A_4 and higher A coefficients can generally be neglected in representing n vs. λ behavior over typical wavelengths of interest. For example, for diamond, we only need the A_1 and A_2 terms. The Sellmeier coefficients are listed in various optical data handbooks.

EXAMPLE 1.2.1 Sellmeier equation and diamond

Using the Sellmeier coefficients for diamond in Table 1.2, calculate its refractive index at 610 nm (red light) and compare with the experimental quoted value of 2.415 to three decimal places.

[11] This is also known as the Sellmeier-Herzberger formula.

Solution

The Sellmeier dispersion relation for diamond is

$$n^2 = 1 + \frac{0.3306\lambda^2}{\lambda^2 - 175\,\text{nm}^2} + \frac{4.3356\lambda^2}{\lambda^2 - 106\,\text{nm}^2}$$

$$n^2 = 1 + \frac{0.3306(610\,\text{nm})^2}{(610\,\text{nm})^2 - (175\,\text{nm})^2} + \frac{4.3356(610\,\text{nm})^2}{(610\,\text{nm})^2 - (106\,\text{nm})^2} = 5.8308$$

So that

$$n = 2.4147$$

which is 2.415 to three decimal places and matches the experimental value.

EXAMPLE 1.2.2 Cauchy equation and diamond

Using the Cauchy coefficients for diamond in Table 1.2, calculate the refractive index at 610 nm.

Solution

At $\lambda = 610$ nm, the photon energy is

$$h\upsilon = \frac{hc}{\lambda} = \frac{(6.626 \times 10^{-34}\,\text{J s})(2.998 \times 10^{8}\,\text{m s}^{-1})}{(610 \times 10^{-9}\,\text{m})} \times \frac{1}{1.602 \times 10^{-19}\,\text{J eV}^{-1}} = 2.0325\,\text{eV}$$

Using the Cauchy dispersion relation for diamond with coefficients from Table 1.2,

$$\begin{aligned} n &= n_{-2}(h\upsilon)^{-2} + n_0 + n_2(h\upsilon)^2 + n_4(h\upsilon)^4 \\ &= (-1.07 \times 10^{-5})(2.0325)^{-2} + 2.378 + (8.01 \times 10^{-3})(2.0325)^2 \\ &\quad + (1.04 \times 10^{-4})(2.0325)^4 \\ &= 2.4140 \end{aligned}$$

which is slightly different than the value calculated in Example 1.2.1; one reason for the discrepancy is due to the Cauchy coefficients quoted in Table 1.2 being applicable over a wider wavelength range at the expense of some accuracy. Although both dispersion relations have four parameters, A_1, A_2, λ_1, λ_2 for Sellmeier and n_{-2}, n_0, n_2, n_4 for Cauchy, the functional forms are different.

1.3 GROUP VELOCITY AND GROUP INDEX

Since there are no perfect monochromatic waves in practice, we have to consider the way in which a group of waves differing slightly in wavelength will travel along the z-direction. Figure 1.7 shows how two perfectly harmonic waves of slight different frequencies $\omega - \delta\omega$ and $\omega + \delta\omega$ interfere to generate a periodic **wave packet** that contains an oscillating field at the mean frequency ω that is amplitude modulated by a slowly varying field of frequency $\delta\omega$. We are interested in the velocity of this wave packet. The two sinusoidal waves of frequencies $\omega - \delta\omega$ and $\omega + \delta\omega$ will propagate with propagation constants $k - \delta k$ and $k + \delta k$ respectively inside the material so that their sum will be

$$E_x(z, t) = E_o \cos\left[(\omega - \delta\omega)t - (k - \delta k)z\right] + E_o \cos\left[(\omega + \delta\omega)t - (k + \delta k)z\right]$$

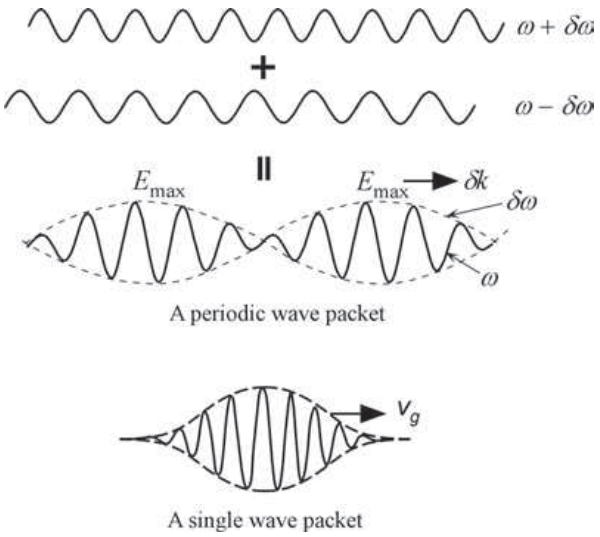


FIGURE 1.7 Two waves of slightly different wavelengths traveling in the same direction result in a wave packet that has an amplitude variation which travels at the group velocity.

By using the trigonometric identity $\cos A + \cos B = 2\cos\left[\frac{1}{2}(A - B)\right]\cos\left[\frac{1}{2}(A + B)\right]$ we arrive at

$$E_x(z, t) = 2E_o \cos\left[(\delta\omega)t - (\delta k)z\right] \cos(\omega t - kz)$$

As illustrated in Figure 1.7, this represents a sinusoidal wave of frequency ω, which is amplitude modulated by a very slowly varying sinusoid of frequency $\delta\omega$. The system of waves, that is, the modulation, travels along z at a speed determined by the modulating term, $\cos\left[(\delta\omega)t - (\delta k)z\right]$. The maximum in the field occurs when $\left[(\delta\omega)t - (\delta k)z\right] = 2m\pi = \text{constant}$ (m is an integer), which travels with a velocity

$$\frac{\delta z}{\delta t} = \frac{\delta\omega}{\delta k}$$

or[12]

$$v_g = \frac{d\omega}{dk} \qquad \textbf{(1.3.1)}$$

Group velocity

where the velocity v_g is the **group velocity** of the waves, since it determines the speed of propagation of the maximum electric field along z. The group velocity represents the speed with which energy or information is propagated since it defines the speed of the envelope of the amplitude variation. The maximum electric field in Figure 1.7 advances with a velocity v_g whereas the phase variations in the electric field propagate at the phase velocity v.

The wave packet we generated by adding two slightly different harmonic waves is periodic with a period $2\pi/\delta\omega$. By adding many such harmonic waves with slightly different frequencies but with the right amplitudes, we can generate a *single wave packet* that is nonperiodic as illustrated in Figure 1.7. This wave packet travels with a group velocity v_g.

[12] It is quite common to replace small changes in expressions like $\delta z/\delta t$ by differentials, and write it as dz/dt, which we will often do in the future.

We know that in vacuum, $\omega = ck$ and the group velocity is

Group velocity in vacuum

$$v_g(\text{vacuum}) = \frac{d\omega}{dk} = c = \text{phase velocity} \tag{1.3.2}$$

In vacuum or air, the group velocity is the same as the phase velocity.

For an EM wave in a medium, k in Eq. (1.3.2) is the propagation constant inside the medium, which can be written $k = 2\pi n/\lambda_o$ where λ_o is the free pace wavelength. The group velocity then is not necessarily the same as the phase velocity v, which depends on ω/k and is given by c/n. The group velocity v_g, on the other hand, is $d\omega/dk$, which depends on how the propagation changes in the medium, δk, with the change in frequency $\delta\omega$, and $d\omega/dk$, is not necessarily the same as ω/k when the refractive index has a wavelength dependence. Suppose that the refractive index $n = n(\lambda_o)$ is a function of (free space) wavelength λ_o, perhaps being described by one of the expressions in Section 1.2. Its gradient would be $dn/d\lambda_o$. We can easily find the group velocity, as shown in Example 1.3.1, by first finding $d\omega$ and dk in terms of dn and $d\lambda_o$, and then using Eq. (1.3.2),

Group velocity in a medium

$$v_g(\text{medium}) = \frac{d\omega}{dk} = \frac{c}{n - \lambda_o\left(\dfrac{dn}{d\lambda_o}\right)} \tag{1.3.3}$$

This can be written as

Group velocity in a medium

$$v_g(\text{medium}) = \frac{c}{N_g} \tag{1.3.4}$$

in which

Group index

$$N_g = n - \lambda_o \frac{dn}{d\lambda_o} \tag{1.3.5}$$

is defined as the **group index of the medium**. Equation (1.3.5) defines the group refractive index N_g of a medium and determines the effect of the medium on the group velocity via Eq. (1.3.4). What is important in Eqs. (1.3.4) and (1.3.5) is the gradient of the refractive index, $dn/d\lambda_o$. If the refractive index is constant and independent of the wavelength, at least over the wavelength range of interest, then $N_g = n$; and the group and phase velocities are the same.

In general, for many materials the refractive index n and hence the group index N_g depend on the wavelength of light by virtue of ε_r being frequency dependent. Then, both the phase velocity v and the group velocity v_g depend on the wavelength and the medium is called a **dispersive medium**. The refractive index n and the group index N_g of pure SiO_2 (silica) glass are important parameters in optical fiber design in optical communications. Both of these parameters depend on the wavelength of light as shown in Figure 1.8. Around 1300 nm, N_g is minimum, which means that for wavelengths close to 1300 nm, N_g is wavelength independent. Thus, light waves with wavelengths around 1300 nm travel with the same group velocity and do not experience dispersion. This phenomenon is significant in the propagation of light in optical fibers as discussed in Chapter 2.

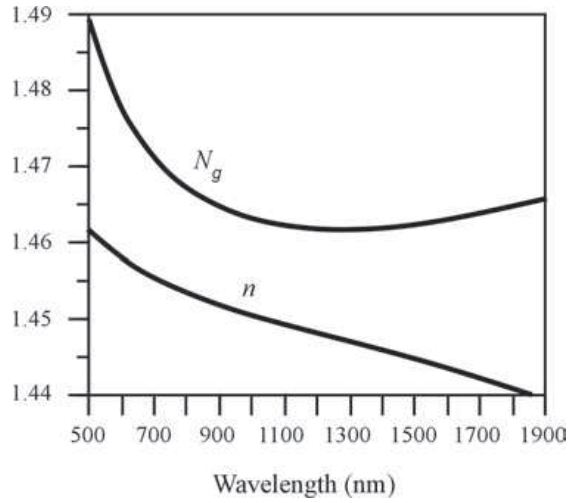


FIGURE 1.8 Refractive index n and the group index N_g of pure SiO_2 (silica) glass as a function of wavelength.

EXAMPLE 1.3.1 Group velocity

Consider two sinusoidal waves that are close in frequency, that is, waves of frequencies $\omega - \delta\omega$ and $\omega + \delta\omega$ as in Figure 1.7. Their propagation constant will be $k - \delta k$ and $k + \delta k$. The resultant wave will be

$$E_x(z, t) = E_o \cos\left[(\omega - \delta\omega)t - (k - \delta k)z\right] + E_o \cos\left[(\omega + \delta\omega)t - (k + \delta k)z\right]$$

By using the trigonometric identity $\cos A + \cos B = 2\cos\left[\frac{1}{2}(A - B)\right]\cos\left[\frac{1}{2}(A + B)\right]$ we arrive at

$$E_x(z, t) = 2E_o \cos\left[(\delta\omega)t - (\delta k)z\right]\cos(\omega t - kz)$$

As illustrated in Figure 1.7, this represents a sinusoidal wave of frequency ω, which is amplitude modulated by a very slowly varying sinusoid of frequency $\delta\omega$. The system of waves, that is, the modulation, travels along z at a speed determined by the modulating term, $\cos\left[(\delta\omega)t - (\delta k)z\right]$. The maximum in the field occurs when $\left[(\delta\omega)t - (\delta k)z\right] = 2m\pi =$ constant (m is an integer), which travels with a velocity

$$\frac{\delta z}{\delta t} = \frac{\delta\omega}{\delta k} \quad \text{or} \quad v_g = \frac{d\omega}{dk}$$

This is the group velocity of the waves, as stated in Eq. (1.3.1), since it determines the speed of propagation of the maximum electric field along z.[13]

EXAMPLE 1.3.2 Group velocity and index

Consider $\omega = 2\pi c/\lambda_o$ and $k = 2\pi n/\lambda_o$, where λ_o is the free-space wavelength. By finding expressions for $d\omega$ and dk in terms of dn and $d\lambda_o$ derive Eq. (1.3.4) for the group velocity v_g.

Solution

Differentiate $\omega = 2\pi c/\lambda_o$ to get $d\omega = -(2\pi c/\lambda_o^2)d\lambda_o$, and then differentiate $k = 2\pi n/\lambda_o$ to find

$$dk = 2\pi n(-1/\lambda_o^2)d\lambda_o + (2\pi/\lambda_o)\left(\frac{dn}{d\lambda_o}\right)d\lambda_o = -(2\pi/\lambda_o^2)\left(n - \lambda_o\frac{dn}{d\lambda_o}\right)d\lambda_o$$

[13] It is left as an exercise to show that the same result can be obtained by using a sine instead of cosine function for the waves.

We can now substitute for $d\omega$ and dk in Eq. (1.3.1),

$$v_g = \frac{d\omega}{dk} = \frac{-(2\pi c/\lambda_o^2)d\lambda_o}{-(2\pi/\lambda_o^2)\left(n - \lambda_o \dfrac{dn}{d\lambda_o}\right)d\lambda_o} = \frac{c}{n - \lambda_o \dfrac{dn}{d\lambda_o}}$$

EXAMPLE 1.3.3 Group and phase velocities

Consider a light wave traveling in a pure SiO_2 (silica) glass medium. If the wavelength of light is 1 μm and the refractive index at this wavelength is 1.450, what is the phase velocity, group index (N_g), and group velocity (v_g)?

Solution

The phase velocity is given by

$$v = c/n = (3 \times 10^8 \,\mathrm{m\,s^{-1}})/(1.450) = 2.069 \times 10^8 \,\mathrm{m\,s^{-1}}$$

From Figure 1.8, at $\lambda = 1\,\mu\mathrm{m}, N_g = 1.463$, so that

$$v_g = c/N_g = (3 \times 10^8 \,\mathrm{m\,s^{-1}})/(1.463) = 2.051 \times 10^8 \,\mathrm{m\,s^{-1}}$$

The group velocity is about ~0.9% smaller than the phase velocity.

1.4 MAGNETIC FIELD, IRRADIANCE, AND POYNTING VECTOR

Although we have considered the electric field component E_x of the electromagnetic wave, we should recall that the magnetic field (magnetic induction) component B_y always accompanies E_x in an EM wave propagation. In fact, if v is the phase velocity of an EM wave in an isotropic dielectric medium and n is the refractive index, then according to electromagnetism, at all times and anywhere in an EM wave,[14]

Fields in an EM wave

$$E_x = vB_y = \frac{c}{n} B_y \tag{1.4.1}$$

in which $v = (\varepsilon_o \varepsilon_r \mu_o)^{-1/2}$ and $n = \varepsilon_r{}^{1/2}$. Thus, the two fields are simply and intimately related for an EM wave propagating in an isotropic medium. Any process that alters E_x also intimately changes B_y in accordance with Eq. (1.4.1).

As the EM wave propagates in the direction of the wave vector **k** as shown in Figure 1.9, there is an energy flow in this direction. The wave brings with it electromagnetic energy. A small region of space in which the electric field is E_x has an energy density, that is, energy per unit volume, given by $(1/2)\varepsilon_o \varepsilon_r E_x^2$. Similarly, a region of space where the magnetic field is B_y has an energy density $(1/2\mu_o)B_y^2$. Since the two fields are related by Eq. (1.4.1), the energy densities in the E_x and B_y fields are the same.

Energy densities in an EM wave

$$\frac{1}{2}\varepsilon_o \varepsilon_r E_x^2 = \frac{1}{2\mu_o} B_y^2 \tag{1.4.2}$$

The total energy density in the wave is therefore $\varepsilon_o \varepsilon_r E_x^2$. Suppose that an ideal "energy meter" is placed in the path of the EM wave so that the receiving area A of this meter is

[14] This is actually a statement of Faraday's law for EM waves. In vector notation it is often expressed as $\omega \mathbf{B} = \mathbf{k} \times \mathbf{E}$.

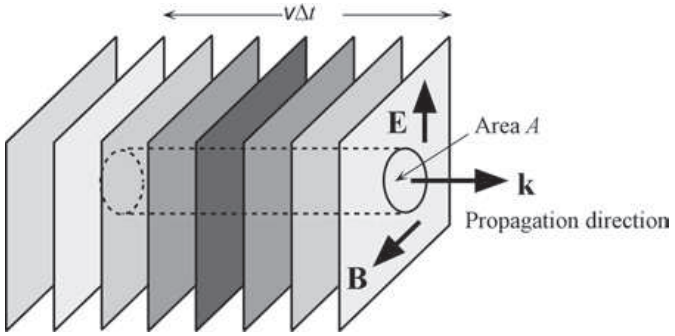


FIGURE 1.9 A plane EM wave traveling along **k** crosses an area A at right angles to the direction of propagation. In time Δt, the energy in the cylindrical volume $Av\Delta t$ (shown dashed) flows through A.

perpendicular to the direction of propagation. In a time interval Δt, a portion of the wave of spatial length $v\Delta t$ crosses A. Thus, a volume $Av\Delta t$ of the EM wave crosses A in time Δt. The energy in this volume consequently becomes received. If S is the EM power flow per unit area, then

$$S = \text{Energy flow per unit time per unit area}$$

giving

$$S = \frac{(Av\Delta t)(\varepsilon_o \varepsilon_r E_x^2)}{A\Delta t} = v\varepsilon_o \varepsilon_r E_x^2 = v^2 \varepsilon_o \varepsilon_r E_x B_y \tag{1.4.3}$$

In an isotropic medium, the energy flow is in the direction of wave propagation. If we use the vectors **E** and **B** to represent the electric and magnetic fields in the EM wave, then the wave propagates in a direction $\mathbf{E} \times \mathbf{B}$ because this direction is perpendicular to both **E** and **B**. The EM power flow per unit area in Eq. (1.4.3) can be written as,

$$\mathbf{S} = v^2 \varepsilon_o \varepsilon_r \mathbf{E} \times \mathbf{B} \tag{1.4.4}$$

Poynting vector

in which **S**, called the **Poynting vector**, represents the energy flow per unit time per unit area in a direction determined by $\mathbf{E} \times \mathbf{B}$ (direction of propagation). Its magnitude, power flow per unit area, is called the **irradiance**.[15]

The field E_x at the receiver location (say, $z = z_1$) varies sinusoidally, which means that the energy flow also varies sinusoidally. The irradiance in Eq. (1.4.3) is the **instantaneous irradiance**. If we write the field as $E_x = E_o \sin(\omega t)$ and then calculate the average irradiance by averaging S over one period we would find the **average irradiance**,

$$I = S_{\text{average}} = \frac{1}{2} v\varepsilon_o \varepsilon_r E_o^2 \tag{1.4.5}$$

Average irradiance (intensity)

Since $v = c/n$ and $\varepsilon_r = n^2$ we can write Eq. (1.4.5) as

$$I = S_{\text{average}} = \frac{1}{2} c\varepsilon_o n E_o^2 = (1.33 \times 10^{-3}) n E_o^2 \tag{1.4.6}$$

Average irradiance (intensity)

The instantaneous irradiance can be measured only if the power meter can respond more quickly than the oscillations of the electric field, and since this is in the optical frequencies range,

[15] The term *intensity* is widely used and interpreted by many engineers as power flow per unit area even though the strictly correct term is *irradiance*. Many optoelectronic data books simply use intensity to mean irradiance. By the way, the Poynting vector was named after John H. Poynting (1851–1914), an English physicist, who was a physics professor at the University of Birmingham (known as Mason Science College at the time).

all practical measurements invariably yield the average irradiance. This is because all detectors have a response rate much slower than the frequency of the wave.

The irradiance of an EM wave depends on the distance from the EM source. For an *ideal* plane wave, the irradiance at a distance r from the "infinite extended plane EM source" is independent of r. On the other hand, for a spherical wave, the irradiance drops as $1/r^2$ because the electric field drops as $1/r$. Suppose that the total power emitted by a point source placed at some point O is P_o as shown in Figure 1.10. Then the irradiance I at a distance r from O is

Irradiance of a spherical wave

$$I = \frac{P_o}{4\pi r^2} \tag{1.4.7}$$

For a Gaussian beam, the irradiance has a Gaussian distribution across the beam cross-section, as illustrated in Figure 1.5 (b) and (c), and also decreases as the beam propagates along z; the power in the beam becomes spread over larger and larger wavefront surfaces as the wave propagates along z. The irradiance I at a point z from O, and at a radial distance r from the beam axis (Figures 1.5 and 1.6) is given by

Irradiance of a Gaussian beam

$$I(z, r) = I_o\left(\frac{w_o}{w}\right)^2 \exp\left(-\frac{2r^2}{w^2}\right) \tag{1.4.8}$$

where w_o is the beam waist, w is the beam width at a distance z from O, and I_o is the maximum beam irradiance, which occurs at $z = 0$ when $w = w_o$. Since w depends on z through $2w = 2w_o\left[1 + (z/z_o)^2\right]^{1/2}$, the irradiance also depends on z and I decreases with z. At far away from the Rayleigh range, $z \gg z_o$, the irradiance on the beam axis is

Irradiance of a Gaussian beam on axis

$$I_{\text{axis}}(z) = I_o\frac{z_o^2}{z^2} \tag{1.4.9}$$

which shows that the decay of light intensity with distance is similar to that for a spherical wave. The radial dependence, of course, remains Gaussian.

The total optical power P_o is the EM power carried by a wave, and can be found by integrating the irradiance. For a Gaussian beam P_o and I_o are related by

Total power and irradiance of a Gaussian beam

$$P_o = \frac{1}{2}\left[I_o(\pi w_o^2)\right] \tag{1.4.10}$$

where it can be seen that the apparent "cross-sectional area" of the beam, πw_o^2, is used to multiply the maximum irradiance with a factor of half to yield the total optical power.

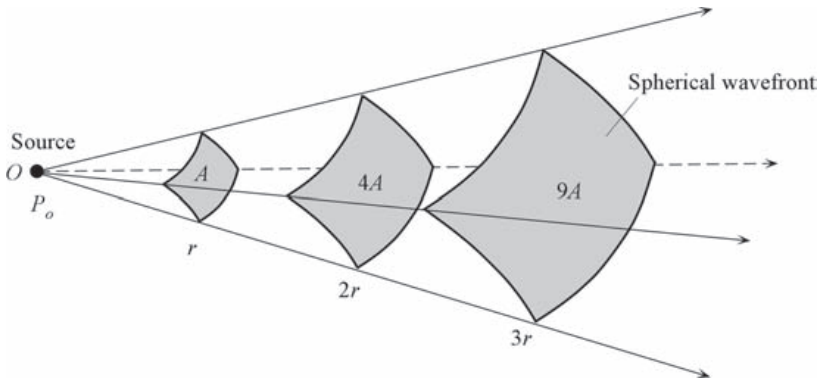


FIGURE 1.10 The irradiance of a spherical wave decreases with distance from the source because the area through which the power flows increases as r^2.

EXAMPLE 1.4.1 Electric and magnetic fields in light

The intensity (irradiance) of the red laser beam from a He-Ne laser at a certain location has been measured to be $1\,\text{mW}\,\text{cm}^{-2}$. What are the magnitudes of the electric and magnetic fields? What are the magnitudes if this beam were in a glass medium with a refractive index $n = 1.45$ and the irradiance was still $1\,\text{mW}\,\text{cm}^{-2}$?

Solution

Using Eq. (1.4.6) for the average irradiance, the field in air is

$$E_o = \sqrt{\frac{2I}{c\varepsilon_o n}} = \sqrt{\frac{2(1 \times 10^{-3} \times 10^4\,\text{W m}^{-2})}{(3 \times 10^8\,\text{m s}^{-1})(8.85 \times 10^{-12}\,\text{F m}^{-1})(1)}}$$

so that

$$E_o = 87\,\text{V m}^{-1} \quad \text{or} \quad 0.87\,\text{V cm}^{-1}$$

The corresponding magnetic field is

$$B_o = E_o/c = (87\,\text{V m}^{-1})/(3 \times 10^8\,\text{m s}^{-1}) = 0.29\ \mu\text{T}$$

If the same intensity were in a glass medium of $n = 1.45$, then

$$E_o(\text{medium}) = \sqrt{\frac{2I}{c\varepsilon_o n}} = \sqrt{\frac{2(1 \times 10^{-3} \times 10^4\,\text{W m}^{-2})}{(3 \times 10^8\,\text{m s}^{-1})(8.85 \times 10^{-12}\,\text{F m}^{-1})(1.45)}}$$

or

$$E_o(\text{medium}) = 72\ \text{V m}^{-1}$$

and

$$B_o(\text{medium}) = nE_o(\text{medium})/c = (1.45)(72\ \text{V m}^{-1})/(3 \times 10^8\,\text{m s}^{-1}) = 0.35\ \mu\text{T}$$

EXAMPLE 1.4.2 Power and irradiance of a Gaussian beam

Consider a 5 mW He-Ne laser that is operating at 633 nm, and has a spot size of 1 mm. Find the maximum irradiance of the beam and the axial (maximum) irradiance at 25 m from the laser.

Solution

The 5 mW rating refers to the total optical power P_o available, and 633 nm is the free-space output wavelength λ. Apply Eq. (1.4.10), $P_o = (1/2)\left[I_o(\pi w_o^2)\right]$,

$$5 \times 10^{-3}\,\text{W} = \frac{1}{2} I_o(\pi)\left(\frac{1}{2} \times 1 \times 10^{-3}\,\text{m}\right)^2$$

which gives

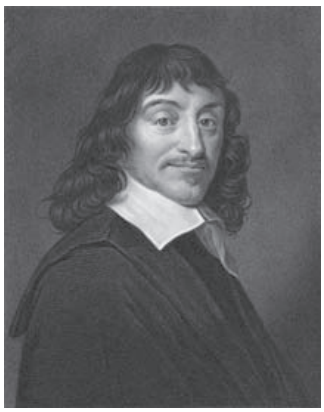
$$I_o = 1.273 \times 10^4\,\text{W m}^{-2} = 1.273\,\text{W cm}^{-2}$$

The Rayleigh range z_o was calculated previously as $z_o = \pi w_o^2/\lambda = 1.24\,\text{m}$ in Example 1.1.1. At $z = 25\,\text{m}$, the axial irradiance is

$$I_{\text{axis}} = (1.273 \times 10^4\,\text{W m}^{-2})\frac{(1.24\,\text{m})^2}{(25\,\text{m})^2} = 31.3\,\text{W m}^{-2} = 3.13\,\text{mW cm}^2$$

1.5 SNELL'S LAW AND TOTAL INTERNAL REFLECTION (TIR)

Willebrord Snellius (Willebrord Snel van Royen, 1580–1626) was a Dutch astronomer and a mathematician, who was a professor at the University of Leiden. He discovered his law of refraction in 1621 which was published by Réne Descartes in France 1637; it is not known whether Descartes knew of Snell's law or formulated it independently. *(Courtesy of AIP Emilio Segre Visual Archives, Brittle Books Collection.)*



René Descartes (1596–1650) was a French philosopher who was also involved with mathematics and sciences. He has been called the "Father of Modern Philosophy." Descartes was responsible for the development of Cartesian coordinates and analytical geometry. He also made significant contributions to optics, including reflection and refraction. *(Courtesy of Georgios Kollidas/Shutterstock.com.)*

We consider a traveling plane EM wave in a medium (1) of refractive index n_1 propagating toward a medium (2) with a refractive index n_2. Constant phase fronts are joined with broken lines and the wave vector $\mathbf{k}_i$ is perpendicular to the wavefronts as shown in Figure 1.11. When the wave reaches the plane boundary between the two media, a transmitted wave in medium 2 and a reflected wave in medium 1 appear. The transmitted wave is called the **refracted light**. The angles θ_i, θ_t, θ_r define the directions of the incident, transmitted, and reflected waves, respectively, with respect to the normal to the boundary plane as shown in Figure 1.11. The wave vectors of the reflected and transmitted waves are denoted as $\mathbf{k}_r$ and $\mathbf{k}_t$. Since both the incident and reflected waves are in the same medium, the magnitudes of $\mathbf{k}_r$ and $\mathbf{k}_i$ are the same, $k_r = k_i$.

Simple arguments based on constructive interference can be used to show that there can be only one reflected wave that occurs at an angle equal to the incidence angle. The two waves along A_i and B_i are in phase. When these waves are reflected to become waves A_r and B_r then they must still be in phase, otherwise they will interfere destructively and destroy each other. The only way the two waves can stay in phase is if $\theta_r = \theta_i$. All other angles lead to the waves A_r and B_r being out of phase and interfering destructively.

The refracted waves A_t and B_t are propagating in a medium of refracted index $n_2(<n_1)$ that is different than n_1. Hence the waves A_t and B_t have different velocities than A_i and B_i. We consider what happens to a wavefront such as AB, corresponding perhaps to the maximum field, as it propagates from medium 1 to 2. We recall that the points A and B on this front are always in phase. During the time it takes for the phase B on wave B_i to reach B', phase A on wave A_t has

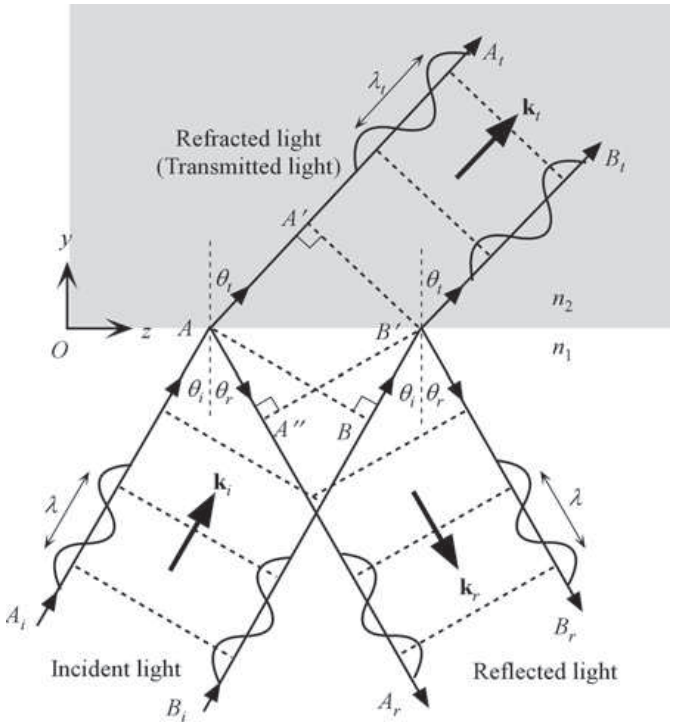


FIGURE 1.11 A light wave traveling in a medium with a greater refractive index ($n_1 > n_2$) suffers reflection and refraction at the boundary. (Notice that λ_t is slightly longer than λ.)

progressed to A'. The wavefront AB thus becomes the front $A'B'$ in medium 2. Unless the two waves at A' and B' still have the same phase, there will be no transmitted wave. A' and B' points on the front are in phase only for one particular transmitted angle, θ_t.

If it takes time t for the phase at B on wave B_i to reach B', then $BB' = v_1 t = ct/n_1$. During this time t, the phase A has progressed to A' where $AA' = v_2 t = ct/n_2$. A' and B' belong to the same front just like A and B so that AB is perpendicular to $\mathbf{k}_i$ in medium 1 and $A'B'$ is perpendicular to $\mathbf{k}_t$ in medium 2. From geometrical considerations, $AB' = BB'/\sin\theta_i$ and $AB' = AA'/\sin\theta_t$ so that

$$AB' = \frac{v_1 t}{\sin\theta_i} = \frac{v_2 t}{\sin\theta_t}$$

or

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \qquad \textbf{(1.5.1)}$$

Snell's Law

This is **Snell's law**,[16] which relates the angles of incidence and refraction to the refractive indices of the media.

If we consider the reflected wave, the wavefront AB becomes $A''B'$ in the reflected wave. In time t, phase B moves to B' and A moves to A''. Since they must still be in phase to constitute the reflected wave, BB' must be equal to AA''. Suppose it takes time t for the wavefront B to move to B' (or A to A''). Then, since $BB' = AA'' = v_1 t$, from geometrical considerations,

$$AB' = \frac{v_1 t}{\sin\theta_i} = \frac{v_1 t}{\sin\theta_r}$$

so that $\theta_i = \theta_r$. Angles of incidence and reflection are the same.

[16] Snell's law is known as Descartes's law in France as he was the first to publish it in his "Discourse on Method" in 1637.

When $n_1 > n_2$ then obviously the transmitted angle is greater than the incidence angle as apparent in Figure 1.11. When the refraction angle θ_t reaches 90°, the incidence angle is called the **critical angle** θ_c, which is given by

Total internal reflection (TIR)

$$\sin\theta_c = \frac{n_2}{n_1} \tag{1.5.2}$$

When the incidence angle θ_i exceeds θ_c then there is no transmitted wave but only a reflected wave. The latter phenomenon is called **total internal reflection (TIR)**. The effect of increasing the incidence angle is shown in Figure 1.12. It is the TIR phenomenon that leads to the propagation of waves in a dielectric medium surrounded by a medium of smaller refractive index as shown in Chapter 2. Although Snell's law for $\theta_i > \theta_c$ shows that $\sin\theta_t > 1$ and hence θ_t is an "imaginary" angle of refraction, there is however a wave called the evanescent wave, whose amplitude decays exponentially with distance into the second medium as discussed below. The wave exists only in the interface region from which the reflected wave emerges (not outside).

Snell's law can also be viewed as the **k**-vector of light parallel to the interface being continuous through the interface, that is, having the same value on both sides of the interface. In medium n_1, $\mathbf{k}_i$ parallel to the interface is $k_i\sin\theta_i$ or $kn_1\sin\theta_i$, where $k_i = kn_1$, and k is the magnitude of the wave vector in free space. In medium n_2, $\mathbf{k}_t$ parallel to the interface is $k_t\sin\theta_t$ or $kn_2\sin\theta_t$. If **k**'s component tangential to the interface remains constant, $kn_1\sin\theta_i = kn_2\sin\theta_t$, then we obtain Snell's law in Eq. (1.5.1). Put differently, Snell's law is equivalent to

Snell's Law

$$n\sin\theta = \text{constant through an interface between different media} \tag{1.5.3}$$

Snell's law of refraction and TIR play a very important role in many optoelectronic and photonic devices. A prism is a transparent optical component that can deflect a light beam as illustrated in Figure 1.13. There are two basic types of prism. In a **refracting prism**, the light deflection is caused by refractions whereas in a **reflecting prism** it is caused by one or more TIRs. (Some prisms such as composite prisms need both refraction and TIR to achieve their desired deflection.) The deflection δ depends not only on the incidence angle of the light beam on the prism, the prism material (n), and geometry, but also on the wavelength and the polarization state of the incident light. The reason is that the refractive index n of the prism material normally depends on the wavelength, and further, for certain materials (*e.g.*, quartz, calcite), it depends on the polarization state (direction of the electric field) of light as well.

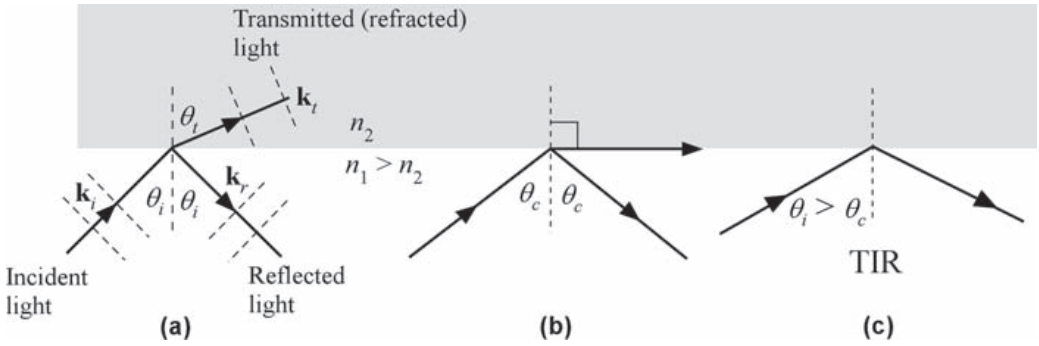


FIGURE 1.12 Light wave travelling in a more dense medium strikes a less dense medium. Depending on the incidence angle with respect to θ_c, which is determined by the ratio of the refractive indices, the wave may be transmitted (refracted) or reflected. (a) $\theta_i < \theta_c$ (b) $\theta_i = \theta_c$ (c) $\theta_i > \theta_c$ and total internal reflection. (Wavefronts are only indicated in (a).)

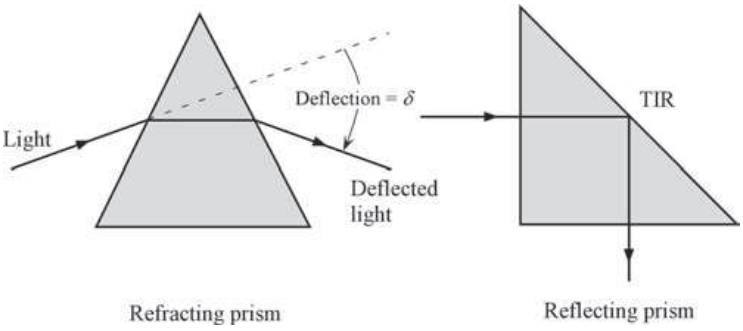


FIGURE 1.13 Basic types of prism: refracting and reflecting prisms.

EXAMPLE 1.5.1 Beam displacement

Lateral displacement of light, or **beam displacement**, occurs when a beam of light passes obliquely through a plate of transparent material, such as a glass plate. When a light beam is incident on a plate of transparent material of refractive index n, it emerges from the other side traveling parallel to the incident light but displaced from it by a distance d, called *lateral displacement*, as illustrated in Figure 1.14. Find the displacement d in terms of the incidence angle the plate thickness. What is d for a glass of $n = 1.600$, $d = 10$ mm if the incidence angle is 45°?

Solution

The displacement $d = BC = AB\sin(\theta_i - \theta_t)$. Further, $L/AB = \cos\theta_t$ so that combining these two equation we find

$$d = L\left[\frac{\sin(\theta_i - \theta_t)}{\cos\theta_t}\right]$$

We can expand $\sin(\theta_i - \theta_t) = \sin\theta_i\cos\theta_t - \cos\theta_i\sin\theta_t$, use $\cos\theta_t = \sqrt{1 - \sin^2\theta_t}$ and then apply Snell's law $n\sin\theta_t = n_o\sin\theta_i$ at the top surface to find

$$\frac{d}{L} = \sin\theta_i\left[1 - \frac{\cos\theta_i}{\sqrt{(n/n_o)^2 - \sin^2\theta_i}}\right]$$

which is maximum with $d = L$ when $\theta_i \simeq 90°$, glazing incidence. Substituting $n = 1.600, n_o = 1, \theta_i = 45°$, and $L = 10$ mm, we find, $d = 3.587$ mm. If the refractive index increases by 1%, $n = 1.616$, then $d = 3.630$ and the change in d is 0.043 mm or 43 μm, which can be measured electronically by using, for example, CCD or CMOS photodiode arrays.

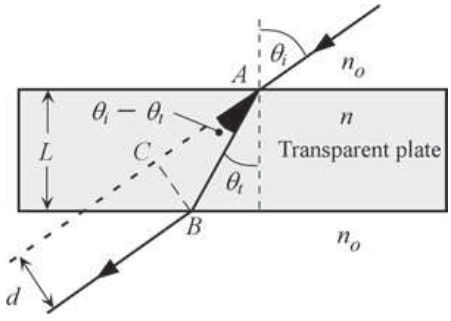


FIGURE 1.14 Lateral displacement of light passing obliquely through a transparent plate.

1.6 FRESNEL'S EQUATIONS

A. Amplitude Reflection and Transmission Coefficients (*r* and *t*)

Although the ray picture with constant phase wavefronts is useful in understanding refraction and reflection, to obtain the magnitude of the reflected and refracted waves and their relative phases, we need to consider the electric field in the light wave. The electric field in the wave must be perpendicular to the direction of propagation as shown in Figure 1.15. We can resolve the field E_i of the incident wave into two components, one in the plane of incidence, $E_{i,//}$, and the other perpendicular to the plane of incidence, $E_{i,\perp}$. The plane of incidence is defined as the plane containing the incident and the reflected rays, which in Figure 1.15 corresponds to the plane of the paper.[17] Similarly for both the reflected and transmitted waves, we will have field components parallel and perpendicular to the plane of incidence, that is, $E_{r,//}$, $E_{r,\perp}$, and $E_{t,//}$, $E_{t,\perp}$.

As apparent from Figure 1.15, the incident, transmitted, and reflected wave all have a wave vector component along the z-direction; that is, they have an effective velocity along z. The fields $E_{i,\perp}$, $E_{r,\perp}$, and $E_{t,\perp}$ are all perpendicular to the z-direction. These waves are called **transverse electric field** (TE) waves. On the other hand, waves with $E_{i,//}$, $E_{r,//}$, and $E_{t,//}$ have only their magnetic field components perpendicular to the z-direction, and these are called **transverse magnetic field** (TM) waves.

We will describe the incident, reflected, and refracted waves each by the exponential representation of a traveling wave:

Incident wave

$$E_i = E_{io}\exp j(\omega t - \mathbf{k_i}\cdot\mathbf{r}) \tag{1.6.1}$$

Reflected wave

$$E_r = E_{ro}\exp j(\omega t - \mathbf{k_r}\cdot\mathbf{r}) \tag{1.6.2}$$

Transmitted wave

$$E_t = E_{to}\exp j(\omega t - \mathbf{k_t}\cdot\mathbf{r}) \tag{1.6.3}$$

in which $\mathbf{r}$ is the position vector, the wave vectors $\mathbf{k_i}$, $\mathbf{k_r}$, and $\mathbf{k_t}$ describe the directions of the incident, reflected, and transmitted waves, and E_{io}, E_{ro}, and E_{to} are the respective amplitudes. Any phase changes such as ϕ_r and ϕ_t in the reflected and transmitted waves with respect to the phase of the incident wave are incorporated into the complex amplitudes, E_{ro} and E_{to}. Our objective is to find E_{ro} and E_{to} with respect to E_{io}.

We should note that similar equations can be stated for the magnetic field components in the incident, reflected, and transmitted waves but these will be perpendicular to the corresponding electric fields. The electric and magnetic fields anywhere on the wave must be perpendicular to each other as a requirement of electromagnetic wave theory. This means that with $E_{//}$ in the EM wave we have a magnetic field $B_\perp$ associated with it such that $B_\perp = (n/c)E_{//}$. Similarly, $E_\perp$ will have a magnetic field $B_{//}$ associated with it such that $B_{//} = (n/c)E_\perp$.

There are two useful fundamental rules in electromagnetism that govern the behavior of the electric and magnetic fields at a boundary between two dielectric media, which we can arbitrarily label as 1 and 2. These rules are called **boundary conditions**. The first states that the

---

[17] The definitions of the field components follow those of S. G. Lipson *et al.*, *Optical Physics*, 3rd Edition (Cambridge University Press, 1995) and Grant Fowles, *Introduction to Modern Optics*, 2nd Edition (Dover Publications, Inc., 1975), whose clear treatments of this subject are highly recommended. The majority of authors use a different convention, which leads to different signs later in the equations ("they [Fresnel's equations] must be related to the specific electric field directions from which they are derived." Eugene Hecht, *Optics*, 4th Edition (Addison Wesley, Pearson Education, 2002), p. 115.

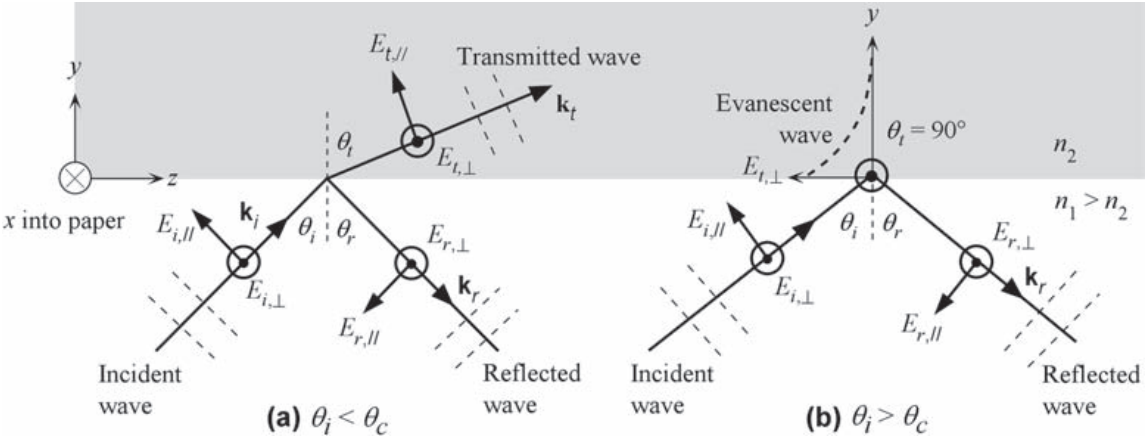


FIGURE 1.15 Light wave traveling in a more dense medium strikes a less dense medium. The plane of incidence is the plane of the paper and is perpendicular to the flat interface between the two media. The electric field is normal to the direction of propagation. It can be resolved into perpendicular ($\perp$) and parallel (//) components. (a) $\theta_i < \theta_c$ then some of the wave is transmitted into the less dense medium. Some of the wave is reflected. (b) $\theta_i > \theta_c$ then the incident wave suffers total internal reflection. There is a decaying evanescent wave into the n_2-medium.

electric field that is tangential to the boundary surface, $E_{\text{tangential}}$, must be continuous across the boundary from medium 1 to 2, that is, at the boundary, $y = 0$ in Figure 1.15,

$$E_{\text{tangential}}(1) = E_{\text{tangential}}(2) \tag{1.6.4}$$

Boundary condition

The second rule is that the tangential component of the magnetic field, $B_{\text{tangential}}$, to the boundary must likewise be continuous from medium 1 to 2, provided that the two media are non-magnetic (relative permeability, $\mu_r = 1$),

$$B_{\text{tangential}}(1) = B_{\text{tangential}}(2) \tag{1.6.5}$$

Boundary condition

Using the boundary conditions above for the fields at $y = 0$ and the relationship between the electric and magnetic fields, we can find the reflected and transmitted waves in terms of the incident wave. The boundary conditions can be satisfied only if the reflection and incidence angles are equal, $\theta_r = \theta_i$, and the angles for the transmitted and incident wave obey Snell's law, $n_1 \sin\theta_1 = n_2 \sin\theta_2$.

Applying the boundary conditions above to the EM wave going from medium 1 to 2, the amplitudes of the reflected and transmitted waves can be readily obtained in terms of n_1, n_2, and the incidence angle θ_i alone.[18] These relationships are called **Fresnel's equations**. If we define $n = n_2/n_1$ as the relative refractive index of medium 2 to that of 1, then the **reflection and transmission coefficients** for $E_\perp$ are

$$r_\perp = \frac{E_{ro,\perp}}{E_{io,\perp}} = \frac{\cos\theta_i - \left[n^2 - \sin^2\theta_i\right]^{1/2}}{\cos\theta_i + \left[n^2 - \sin^2\theta_i\right]^{1/2}} \tag{1.6.6a}$$

Reflection coefficient

[18] These equations are readily available in any electromagnetism textbook. Their derivation from the two boundary conditions involves extensive algebraic manipulation, which we will not carry out here. The electric and magnetic field components on both sides of the boundary are resolved tangentially to the boundary surface and the boundary conditions are then applied. We then use such relations as $\cos\theta_t = \left[1 - \sin\theta_t\right]^{1/2}$ and $\sin\theta_t$ is determined by Snell's law, *etc.*

and

Transmission coefficient

$$t_{\perp} = \frac{E_{to,\perp}}{E_{io,\perp}} = \frac{2\cos\theta_i}{\cos\theta_i + \left[n^2 - \sin^2\theta_i\right]^{1/2}} \tag{1.6.6b}$$

There are corresponding coefficients for the $E_{//}$ fields with corresponding **reflection and transmission coefficients**, $r_{//}$ and $t_{//}$.

Reflection coefficient

$$r_{//} = \frac{E_{ro,//}}{E_{io,//}} = \frac{\left[n^2 - \sin^2\theta_i\right]^{1/2} - n^2\cos\theta_i}{\left[n^2 - \sin^2\theta_i\right]^{1/2} + n^2\cos\theta_i} \tag{1.6.7a}$$

Transmission coefficient

$$t_{//} = \frac{E_{to,//}}{E_{io,//}} = \frac{2n\cos\theta_i}{n^2\cos\theta_i + \left[n^2 - \sin^2\theta_i\right]^{1/2}} \tag{1.6.7b}$$

Further, the coefficients above are related by

Transmission coefficient

$$r_{//} + nt_{//} = 1 \quad \text{and} \quad r_{\perp} + 1 = t_{\perp} \tag{1.6.8}$$

The significance of these equations is that they allow the amplitudes and phases of the reflected and transmitted waves to be determined from the coefficients $r_{\perp}$, $r_{//}$, $t_{//}$, and $t_{\perp}$. For convenience we take E_{io} to be a real number so that phase angles of $r_{\perp}$ and $t_{\perp}$ correspond to the **phase changes** measured with respect to the incident wave. For example, if $r_{\perp}$ is a complex quantity then we can write this as $r_{\perp} = |r_{\perp}|\exp(-j\phi_{\perp})$ in which $|r_{\perp}|$ and $\phi_{\perp}$ represent the relative amplitude and phase of the reflected wave with respect to the incident wave for the field perpendicular to the plane of incidence. Of course, when $r_{\perp}$ is a real quantity, then a positive number represents no phase shift and a negative number is a phase shift of 180° (or π). As with all waves, a negative sign corresponds to a 180° phase shift. Complex coefficients can be obtained only from Fresnel's equations if the terms under the square roots become negative and this can happen only when $n < 1$ (or $n_1 > n_2$), and also when $\theta_i > \theta_c$, the critical angle. Thus, phase changes other than 0 or 180° occur only when there is total internal reflection. Fresnel's equations for normal incidence are greatly simplified. Putting $\theta_i = 0$ into Eqs. (1.6.6) and (1.6.7) we find

Normal incidence

$$r_{//} = r_{\perp} = \frac{n_1 - n_2}{n_1 + n_2} \quad \text{and} \quad t_{//} = t_{\perp} = \frac{2n_1}{n_1 + n_2} \tag{1.6.9}$$

Figure 1.16 (a) shows how the magnitudes of the reflection coefficients, $|r_{\perp}|$ and $|r_{//}|$, vary with the incidence angle θ_i for a light wave traveling from a more dense medium, $n_1 = 1.44$, to a less dense medium, $n_2 = 1.00$, as predicted by Fresnel's equations. Figure 1.16 (b) shows the changes in the phase of the reflected wave, $\phi_{\perp}$ and $\phi_{//}$, with θ_i. The critical angle θ_c as determined from $\sin\theta_c = n_2/n_1$ in this case is 44°. It is clear that for incidence close to normal (small θ_i), there is no phase change in the reflected wave. The reflection coefficient in Eq. (1.6.9) is a positive quantity for $n_1 > n_2$, which means that the reflected wave suffers no phase change. This is confirmed by $\phi_{\perp}$ and $\phi_{//}$ in Figure 1.16 (b). As the incidence angle increases, eventually $r_{//}$ becomes zero at an angle of about 35°. We can find this special incidence angle, labeled as θ_p, by solving the Fresnel equation (1.6.7a) for $r_{//} = 0$. The field in the reflected wave is then always perpendicular to the plane of incidence and hence well-defined as illustrated in

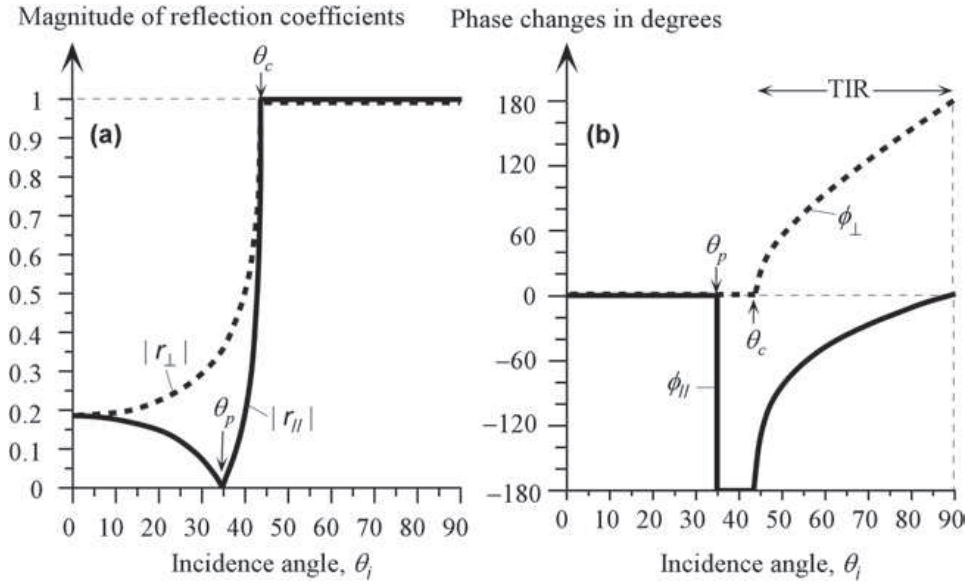


FIGURE 1.16 Internal reflection: (a) Magnitude of the reflection coefficients $r_{//}$ and $r_{\perp}$ vs. angle of incidence θ_i for $n_1 = 1.44$ and $n_2 = 1.00$. The critical angle is 44°. (b) The corresponding phase changes $\phi_{//}$ and $\phi_{\perp}$ vs. incidence angle θ_i.

Figure 1.17. This special angle is called the **polarization angle** or **Brewster's**[19] **angle** and from Eq. (1.6.7a) is given by

$$\tan\theta_p = \frac{n_2}{n_1} \tag{1.6.10}$$

Brewster's polarization angle

The reflected wave is then said to be **linearly polarized** because it contains *electric field oscillations that are contained within a well-defined plane*, which is perpendicular to the plane of incidence and also to the direction of propagation. Electric field oscillations in **unpolarized light**, on the other hand, can be in any one of infinite number of directions that are perpendicular

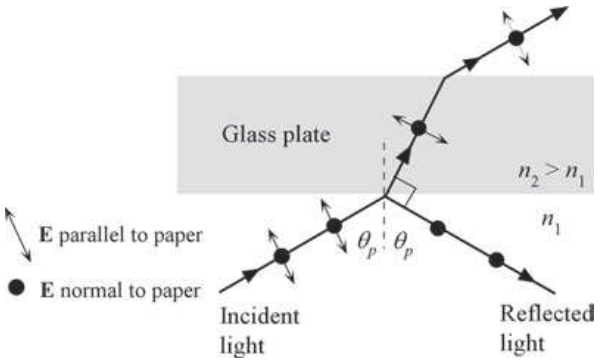


FIGURE 1.17 At the Brewster angle of incidence $\theta_i = \theta_p$, the reflected light contains only field oscillations normal to the plane of incidence (paper).

[19] After Sir David Brewster (1781–1868), a Scottish physicist who was educated in theology at the University of Edinburgh in Scotland. He became interested in the polarization properties of light from 1799 onwards, and reported some of his experiments in scientific journals, including the *Philosophical Transactions of London.*

to the direction of propagation. In linearly polarized light, however, the field oscillations are contained within a well-defined plane. Light emitted from many light sources, such as a tungsten light bulb or an LED diode, is unpolarized.[20] Unpolarized light can be viewed as a stream or collection of EM waves whose fields are randomly oriented in a direction that is perpendicular to the direction of light propagation.

For incidence angles greater than θ_p but smaller than θ_c, Fresnel's equation (1.6.7a) gives a negative number for $r_{//}$, which indicates a phase shift of 180° as shown in $\phi_{//}$ in Figure 1.16 (b). The magnitude of both $r_{//}$ and $r_{\perp}$ increases with θ_i as illustrated in Figure 1.16 (a). At the critical angle and beyond (past 44° in Figure 1.16), that is, when $\theta_i \geq \theta_c$, the magnitudes of both $r_{//}$ and $r_{\perp}$ go to unity so that the reflected wave has the same amplitude as the incident wave. The incident wave has suffered **total internal reflection**, TIR. When $\theta_i > \theta_c$, in the presence of TIR, both Eqs. (1.6.6) and (1.6.7) are complex quantities because then $\sin\theta_i > n$ and the terms under the square roots become negative. The reflection coefficients become complex quantities of the type $r_{\perp} = 1 \cdot \exp(-j\phi_{\perp})$ and $r_{//} = 1 \cdot \exp(-j\phi_{//})$ with the phase angles $\phi_{\perp}$ and $\phi_{//}$ being other than zero or 180°. The reflected wave therefore suffers phase changes, $\phi_{\perp}$ and $\phi_{//}$, in the components $E_{\perp}$ and $E_{//}$. These phase changes depend on the incidence angle, as illustrated in Figure 1.16 (b), and on n_1 and n_2.

Examination of Eq. (1.6.6) for $r_{\perp}$ shows that for $\theta_i > \theta_c$, we have $|r_{\perp}| = 1$, but the phase change $\phi_{\perp}$ is given by

Phase change in TIR

$$\tan\left(\tfrac{1}{2}\phi_{\perp}\right) = \frac{\left[\sin^2\theta_i - n^2\right]^{1/2}}{\cos\theta_i} \tag{1.6.11}$$

For the $E_{//}$ component, the phase change $\phi_{//}$ is given by

Phase change in TIR

$$\tan\left(\tfrac{1}{2}\phi_{//} + \tfrac{1}{2}\pi\right) = \frac{\left[\sin^2\theta_i - n^2\right]^{1/2}}{n^2\cos\theta_i} \tag{1.6.12}$$

We can summarize that in internal reflection ($n_1 > n_2$), the amplitude of the reflected wave from TIR is equal to the amplitude of the incident wave but its phase has shifted by an amount determined by Eqs. (1.6.11) and (1.6.12). The fact that $\phi_{//}$ has an additional π shift that makes $\phi_{//}$ negative for $\theta_i > \theta_c$ is due to the choice for the direction of the reflected optical field $E_{r,//}$ in Figure 1.15. This π shift can be ignored if we by simply invert $E_{r,//}$. (In many books Eq. (1.6.12) is written without the π-shift.)

The reflection coefficients in Figure 1.16 considered the case in which $n_1 > n_2$. When light approaches the boundary from the higher index side, that is $n_1 > n_2$, the reflection is said to be **internal reflection** and *at normal incidence there is no phase change*. On the other hand, if light approaches the boundary from the lower index side, that is $n_1 < n_2$, then it is called **external reflection**. Thus in external reflection light becomes reflected by the surface of an optically denser (higher refractive index) medium. There is an important difference between the two. Figure 1.18 shows how the reflection coefficients $r_{\perp}$ and $r_{//}$ depend on the incidence angle θ_i for external reflection ($n_1 = 1$ and $n_2 = 1.44$). At normal incidence, both coefficients are negative, which means that *in external reflection at normal incidence there is a phase shift of 180°*. Further, $r_{//}$ goes through zero at the **Brewster angle** θ_p given by Eq. (1.6.10). At this

[20] Even light from a tungsten light bulb or an LED has some polarization but this is usually negligibly small.

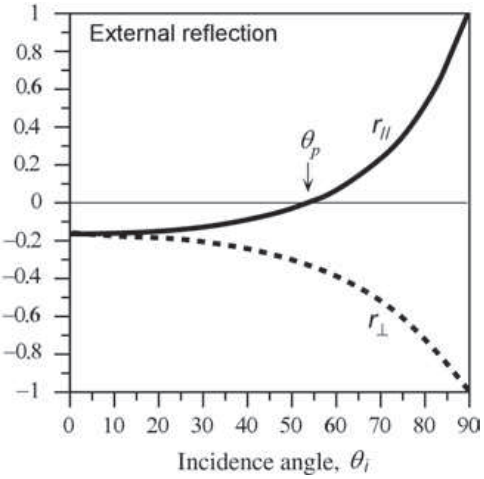


FIGURE 1.18 External reflection coefficients $r_{//}$ and $r_\perp$ vs. angle of incidence θ_i for $n_1 = 1.00$ and $n_2 = 1.44$.

angle of incidence, the reflected wave is polarized in the $E_\perp$ component only. Transmitted light in both internal reflection (when $\theta_i < \theta_c$) and external reflection does not experience a phase shift.

What happens to the transmitted wave when $\theta_i > \theta_c$? According to the boundary conditions, there must still be an electric field in medium 2; otherwise, the boundary conditions cannot be satisfied. When $\theta_i > \theta_c$, the field in medium 2 is a wave that travels near the surface of the boundary along the z direction as shown in Figure 1.19. The wave is called an **evanescent wave** and advances along z with its field decreasing as we move into medium 2; that is,

$$E_{t,\perp}(y, z, t) \propto e^{-\alpha_2 y} \exp j(\omega t - k_{iz} z) \tag{1.6.13}$$

Evanescent wave

in which $k_{iz} = k_i \sin\theta_i$ is the wave vector of the incident wave along the z-axis, and α_2 is an **attenuation coefficient** for the electric field penetrating into medium 2,[21]

$$\alpha_2 = \frac{2\pi n_2}{\lambda_o}\left[\left(\frac{n_1}{n_2}\right)^2 \sin^2\theta_i - 1\right]^{1/2} \tag{1.6.14}$$

Attenuation of evanescent wave

in which λ is the free-space wavelength. According to Eq. (1.6.13), the evanescent wave travels along z and has an amplitude that decays exponentially as we move from the boundary into medium 2 (along y). The field of the evanescent wave is e^{-1} in medium 2 when $y = 1/\alpha_2 = \delta$, which is called the **penetration depth**. It is not difficult to show that the evanescent wave is correctly predicted by Snell's law when $\theta_i > \theta_c$. The evanescent wave propagates along the boundary (along z) with the same speed as the z-component velocity of the incident and reflected waves. In Eqs. (1.6.1) and (1.6.2) we had assumed that the incident and reflected waves were *plane waves*, that is, of *infinite exten*t. If we were to extend the plane wavefronts on the reflected wave, these would cut the boundary as shown in Figure 1.19. The evanescent wave traveling along z can be thought of arising from these plane wavefronts at the boundary as in Figure 1.19. (Evanescent wave is important in light propagation in optical waveguides such as optical fibers.) If the incident wave is a narrow

[21] Normally the term *attenuation coefficient* refers to the attenuation of the irradiance but in this case it refers to the electric field.

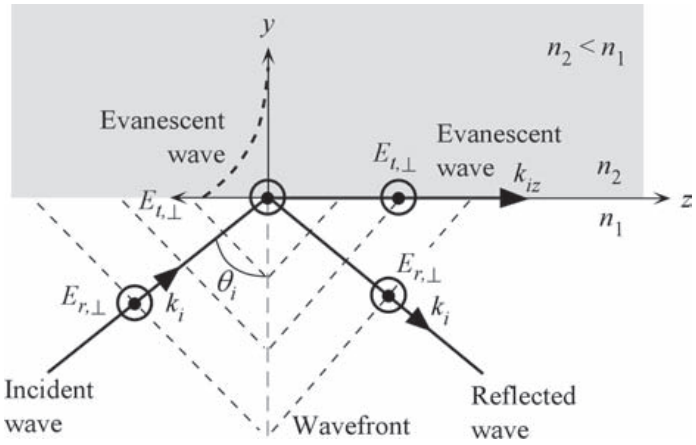


FIGURE 1.19 When $\theta_i > \theta_c$ for a *plane wave* that is reflected, there is an evanescent wave at the boundary whose magnitude decays into the n_2-medium.

beam of light (*e.g.*, from a laser pointer) then the reflected beam would have the same cross-section. There would still be an evanescent wave at the boundary, but it would exist only within the cross-sectional area of the reflected beam at the boundary.

B. Intensity, Reflectance, and Transmittance

It is frequently necessary to calculate the intensity or irradiance[22] of the reflected and transmitted waves when light traveling in a medium of index n_1 is incident at a boundary where the refractive index changes to n_2. In some cases we are simply interested in normal incidence where $\theta_i = 0°$. For example, in laser diodes light is reflected from the ends of an optical cavity where there is a change in the refractive index.

For a light wave traveling with a velocity v in a medium with relative permittivity ε_r, the light intensity I is defined in terms of the electric field amplitude E_o as

Light intensity or irradiance

$$I = \frac{1}{2} v \varepsilon_r \varepsilon_o E_o^2 \tag{1.6.15}$$

Here $\frac{1}{2}\varepsilon_r\varepsilon_o E_o^2$ represents the energy in the field per unit volume. When multiplied by the velocity v it gives the rate at which energy is transferred through a unit area. Since $v = c/n$ and $\varepsilon_r = n^2$ the intensity is proportional to nE_o^2.

Reflectance R measures the intensity of the reflected light with respect to that of the incident light and can be defined separately for electric field components parallel and perpendicular to the plane of incidence. The reflectances $R_\perp$ and $R_{//}$ are defined by

Reflectances

$$R_\perp = \frac{|E_{ro,\perp}|^2}{|E_{io,\perp}|^2} = |r_\perp|^2 \quad \text{and} \quad R_{//} = \frac{|E_{ro,//}|^2}{|E_{io,//}|^2} = |r_{//}|^2 \tag{1.6.16}$$

Although the reflection coefficients can be complex numbers that can represent phase changes, reflectances are necessarily real numbers representing intensity changes. Magnitude of a complex number is defined in terms of its product with its complex conjugate. For example, when $E_{ro,//}$ is a complex number then

$$|E_{ro,//}|^2 = (E_{ro,//})(E_{ro,//})^*$$

in which $(E_{ro,//})^*$ is the complex conjugate of $(E_{ro,//})$.

[22] Strictly the terms intensity and irradiance are not the same as mentioned in footnote 15.

From Eqs. (1.6.6a) and (1.6.7a) with normal incidence, these are simply given by

$$R = R_{\perp} = R_{//} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \tag{1.6.17}$$

Reflectance at normal incidence

Since a glass medium has a refractive index of around 1.5 this means that typically 4% of the incident radiation on an air–glass surface will be reflected back.

Transmittance T relates the intensity of the transmitted wave to that of the incident wave in a similar fashion to the reflectance. We must, however, consider that the transmitted wave is in a different medium and also that its direction with respect to the boundary is different from that of the incident wave by virtue of refraction. For normal incidence, the incident and transmitted beams are normal and the transmittances are defined and given by

$$T_{\perp} = \frac{n_2|E_{to,\perp}|^2}{n_1|E_{io,\perp}|^2} = \left(\frac{n_2}{n_1}\right)|t_{\perp}|^2 \quad \text{and} \quad T_{//} = \frac{n_2|E_{to,//}|^2}{n_1|E_{io,//}|^2} = \left(\frac{n_2}{n_1}\right)|t_{//}|^2 \tag{1.6.18}$$

or

$$T = T_{\perp} = T_{//} = \frac{4n_1n_2}{(n_1 + n_2)^2} \tag{1.6.19}$$

Transmittance at normal incidence

Further, the fraction of light reflected and fraction transmitted must add to unity. Thus $R + T = 1$.

When the light is incident at an angle θ_i, as in Figure 1.15, then the transmitted light has an angle θ_t with respect to normal. The corresponding transmittances are given by

$$T_{\perp} = \left(\frac{n_2\cos\theta_t}{n_1\cos\theta_i}\right)|t_{\perp}|^2 \quad \text{and} \quad T_{//} = \left(\frac{n_2\cos\theta_t}{n_1\cos\theta_i}\right)|t_{//}|^2 \tag{1.6.20}$$

Transmittance at oblique incidence

Each transmittance in Eq. (1.6.20) is with respect to the incident intensity in the corresponding polarization.

C. Goos-Hänchen Shift and Optical Tunneling

A light traveling in an optically more dense medium suffers total internal reflection when it is incident on a less dense medium at an angle of incidence greater than the critical angle ($\theta_i > \theta_c$) as shown in Figure 1.12 (c). Simple ray trajectory analysis gives the impression that the reflected ray emerges from the point of contact of the incident ray with the interface as in Figure 1.12 (c). However, careful optical experiments examining the incident and reflected beams have shown that the reflected wave appears to be laterally shifted from the point of incidence at the interface as illustrated in Figure 1.20. Although the angles of incidence and reflection are the same (as one expects from Fresnel's equation), the reflected beam, nonetheless, is laterally shifted and appears to be reflected from a *virtual plane inside the optically less dense medium*. The lateral shift is known as the **Goos-Haenchen shift**.

The lateral shift of the reflected beam can be understood by considering that the reflected beam experiences a phase change ϕ, as shown in Figure 1.16 (b), and that the electric field extends into the second medium roughly by a penetration depth $\delta = 1/\alpha_2$. We know that phase changes other than 0 or 180° occur only when there is total internal reflection. We can equivalently represent this phase change ϕ and the penetration into the second medium by shifting the reflected wave along the direction of propagation of the evanescent wave, that is, along z,

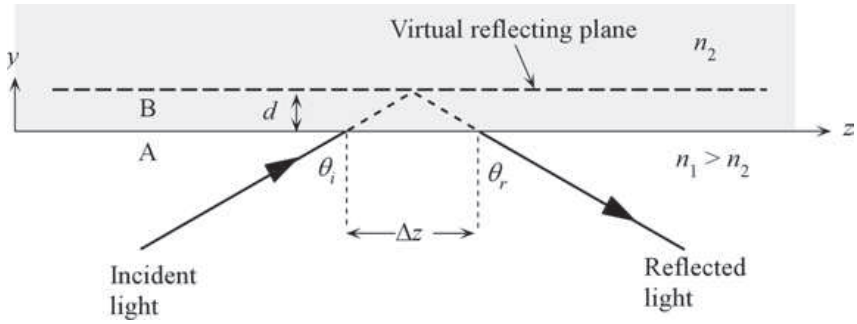


FIGURE 1.20 The reflected light beam in total internal reflection appears to have been laterally shifted by an amount Δz at the interface. It appears as though it is reflected from a virtual plane at a depth d in the second medium from the interface.

by an amount[23] Δz as shown in Figure 1.20. The lateral shift depends on the angle of incidence and the penetration depth. We can represent the reflection as if it were occurring from a virtual plane placed at some distance d from the interface (d is not the same as δ), then from simple geometric considerations, $\Delta z = 2d \tan \theta_i$. Actual d, however, is quite complicated to calculate, but its *order of magnitude* is the penetration depth δ, so that for light with $\lambda = 1$ μm incident at 85° at a glass–glass ($n_1 = 1.450$ and $n_2 = 1.430$) interface and suffering TIR, $\delta = 0.78$ μm, which means $\Delta z \sim 18$ μm.

Total internal reflection occurs whenever a wave propagating in an optically denser medium, such as in A in Figure 1.20, is incident at an angle greater than the critical angle at the interface AB with a medium B of lower refractive index. If we were to shrink the thickness d of medium B, as in Figure 1.21, we would observe that when B is sufficiently thin, an attenuated light beam emerges on the other side of B in C. This phenomenon in which an incident wave is partially transmitted through a medium where it is forbidden in terms of simple geometrical optics is called **optical tunneling** and is a consequence of the electromagnetic wave nature of light. It is due to the fact that the field of the evanescent wave penetrates into medium B and reaches the interface BC before it vanishes. The optical tunneling phenomenon is illustrated in Figure 1.21 and is referred to as **frustrated total internal reflection** (FTIR): the proximity of medium C

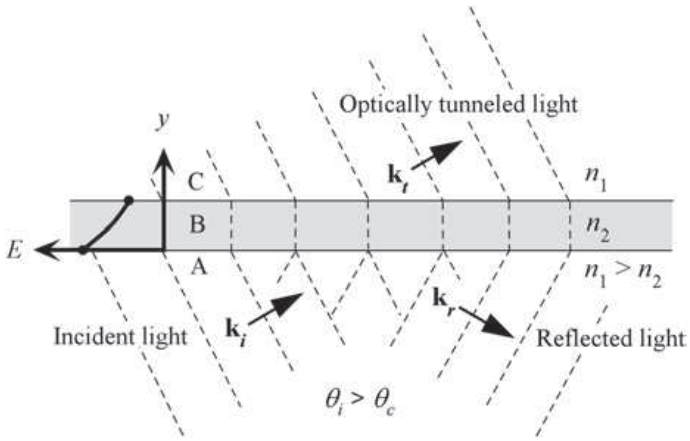


FIGURE 1.21 When medium B is thin, the field penetrates from the AB interface into medium B and reaches BC interface, and gives rise to a transmitted wave in medium C. The effect is the tunneling of the incident beam in A through B to C. The maximum field E_{max} of the evanescent wave in B decays in B along y but is finite at the BC boundary and excites the transmitted wave.

[23] The actual analysis of the Goos-Haenchen shift is more complicated; see for example J. E. Midwinter, *Optical Fibers for Transmission* (John Wiley and Sons, 1979), Ch. 3. d in Figure 1.20 is not simply the penetration depth δ. In fact, d is such that Δz represents the right phase change observed upon TIR.

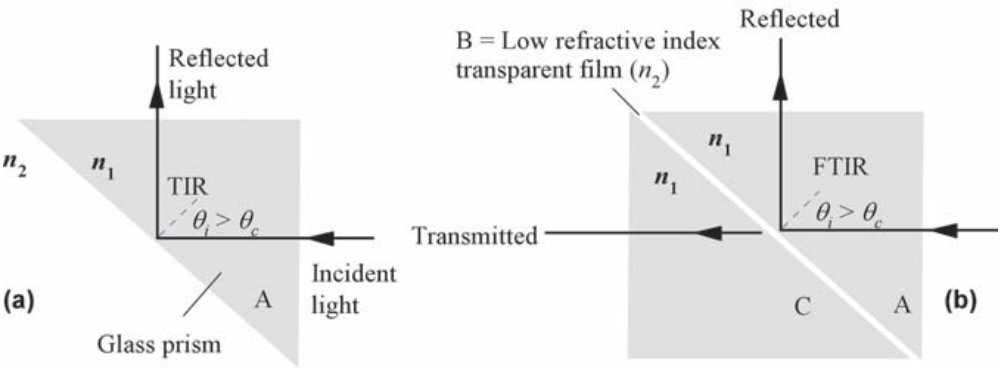
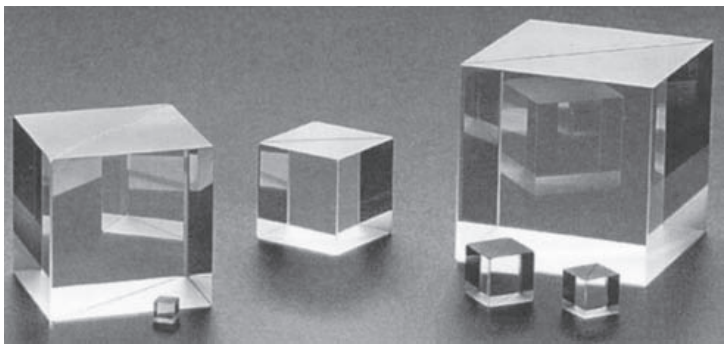


FIGURE 1.22 (a) A light incident at the long face of a glass prism suffers TIR; the prism deflects the light. (b) Two prisms separated by a thin low refractive index film forming a beam-splitter cube. The incident beam is split into two beams by FTIR.



Beam splitter cubes. *(Courtesy of CVI Melles Griot.)*

frustrates TIR. The transmitted wave in C carries some of the light intensity and thus the intensity of the reflected wave is reduced. Notice that the field of the evanescent wave in B decays and has a finite value at the BC interface, where the field oscillations are able to excite a transmitted wave.

Frustrated total internal reflection is utilized in beam splitters as shown in Figure 1.22. A light beam entering the glass prism A suffers TIR at the hypotenuse face ($\theta_i > \theta_c$ at the glass–air interface) and becomes reflected; the prism deflects the light as in Figure 1.22 (a). In the *beam splitter cube* in Figure 1.22 (b), two prisms, A and C, are separated by a thin film, B, of low refractive index. Some of the light energy is now tunneled through this thin film and transmitted into C and out from the cube. FTIR at the hypotenuse face of A leads to a transmitted beam and hence to the splitting of the incident beam into two beams. The extent of energy division between the two beams depends on the thickness of the thin layer B and its refractive index.

EXAMPLE 1.6.1 Reflection of light from a less dense medium (internal reflection)

A ray of light that is traveling in a glass medium of refractive index $n_1 = 1.450$ becomes incident on a less dense glass medium of refractive index $n_2 = 1.430$. Suppose that the free-space wavelength (λ) of the light ray is 1 μm.

(a) What should the minimum incidence angle for TIR be?
(b) What is the phase change in the reflected wave when $\theta_i = 85°$ and when $\theta_i = 90°$?
(c) What is the penetration depth of the evanescent wave into medium 2 when $\theta_i = 85°$ and when $\theta_i = 90°$?

Solution

(a) The critical angle θ_c for TIR is given by $\sin\theta_c = n_2/n_1 = 1.430/1.450$ so that $\theta_c = 80.47°$.

(b) Since the incidence angle $\theta_i > \theta_c$, there is a phase shift in the reflected wave. The phase change in $E_{r,\perp}$ is given by $\phi_\perp$. With $n_1 = 1.450$, $n_2 = 1.430$, and $\theta_i = 85°$

$$\tan\left(\tfrac{1}{2}\phi_\perp\right) = \frac{\left[\sin^2\theta_i - n^2\right]^{1/2}}{\cos\theta_i} = \frac{\left[\sin^2(85°) - \left(\dfrac{1.430}{1.450}\right)^2\right]^{1/2}}{\cos(85°)}$$

$$= 1.61449 = \tan\left[\tfrac{1}{2}(116.45°)\right]$$

so that the phase change is 116.45°. For the $E_{r,//}$ component, the phase change is

$$\tan\left(\tfrac{1}{2}\phi_{//} + \tfrac{1}{2}\pi\right) = \frac{\left[\sin^2\theta_i - n^2\right]^{1/2}}{n^2\cos\theta_i} = \frac{1}{n^2}\tan\left(\tfrac{1}{2}\phi_\perp\right)$$

so that

$$\tan\left(\tfrac{1}{2}\phi_{//} + \tfrac{1}{2}\pi\right) = (n_1/n_2)^2\tan(\phi_\perp/2) = (1.450/1.430)^2\tan\left(\tfrac{1}{2}116.45°\right)$$

which gives $\phi_{//} = -62.1°$. (*Note*: If we were to invert the reflected field, this phase change would be $+117.86°$)

We can repeat the calculation with $\theta_i = 90°$ to find $\phi_\perp = 180°$ and $\phi_{//} = 0°$.

Note that as long as $\theta_i > \theta_c$, the magnitude of the reflection coefficients are unity. Only the phase changes.

(c) The amplitude of the evanescent wave as it penetrates into medium 2 is

$$E_{t,\perp}(y) = E_{to,\perp}\exp(-\alpha_2 y)$$

We ignore the z-dependence, $\exp j(\omega t - k_z z)$, as it only gives a propagating property along z. The field strength drops to e^{-1} when $y = 1/\alpha_2 = \delta$, which is the *penetration depth*. The attenuation constant α_2 is

$$\alpha_2 = \frac{2\pi n_2}{\lambda_o}\left[\left(\frac{n_1}{n_2}\right)^2\sin^2\theta_i - 1\right]^{1/2}$$

that is,

$$\alpha_2 = \frac{2\pi(1.430)}{(1.0\times10^{-6}\,\text{m})}\left[\left(\frac{1.450}{1.430}\right)^2\sin^2(85°) - 1\right]^{1/2} = 1.28\times10^6\,\text{m}^{-1}$$

so the penetration depth is $\delta = 1/\alpha_2 = 1/(1.28\times10^6\,\text{m}) = 7.8\times10^{-7}\,\text{m}$, or 0.78 μm. For 90°, repeating the calculation we find $\alpha_2 = 1.5\times10^6\,\text{m}^{-1}$, so that $\delta = 1/\alpha_2 = 0.66$ μm. We see that the penetration is greater for smaller incidence angles. This will be an important consideration later in analyzing light propagation in optical fibers.

EXAMPLE 1.6.2 Reflection at normal incidence, and internal and external reflection

Consider the reflection of light at normal incidence on a boundary between a glass medium of refractive index 1.5 and air of refractive index 1.

(a) If light is traveling from air to glass, what is the reflection coefficient and the intensity of the reflected light?

(b) If light is traveling from glass to air, what is the reflection coefficient and the intensity of the reflected light?

Solution

(a) The light travels in air and becomes partially reflected at the surface of the glass that corresponds to external reflection. Thus $n_1 = 1$ and $n_2 = 1.5$. Then

$$r_{//} = r_{\perp} = \frac{n_1 - n_2}{n_1 + n_2} = \frac{1 - 1.5}{1 + 1.5} = -0.2$$

This is negative, which means that there is a 180° phase shift. The reflectance (R), which gives the fractional reflected power, is

$$R = r_{//}^2 = 0.04 \quad \text{or} \quad 4\%.$$

(b) The light travels in glass and becomes partially reflected at the glass–air interface that corresponds to internal reflection. Thus $n_1 = 1.5$ and $n_2 = 1$. Then

$$r_{//} = r_{\perp} = \frac{n_1 - n_2}{n_1 + n_2} = \frac{1.5 - 1}{1.5 + 1} = 0.2$$

There is no phase shift. The reflectance is again 0.04 or 4%. In both cases (a) and (b), the amount of reflected light is the same.

EXAMPLE 1.6.3 Reflection and transmission at the Brewster angle

A light beam traveling in air is incident on a glass plate of refractive index 1.50. What is the Brewster or polarization angle? What are the relative intensities of the reflected and transmitted light for the polarization perpendicular and parallel to the plane of incidence at the Brestwer angle of incidence?

Solution

Light is traveling in air and is incident on the glass surface at the polarization angle θ_p. Here $n_1 = 1, n_2 = 1.5$, $n = n_2/n_1 = 1.5$, and $\tan\theta_p = (n_2/n_1) = 1.5$ so that $\theta_p = 56.31°$. We now use Fresnel's equations to find the reflected and transmitted amplitudes. For the perpendicular polarization, from Eq. (1.6.6a),

$$r_{\perp} = \frac{\cos(56.31°) - \left[1.5^2 - \sin^2(56.31°)\right]^{1/2}}{\cos(56.31°) + \left[1.5^2 - \sin^2(56.31°)\right]^{1/2}} = -0.385$$

On the other hand, $r_{//} = 0$. The reflectances $R_{\perp} = |r_{\perp}|^2 = 0.148$ and $R_{//} = |r_{//}|^2 = 0$ so that the reflected light has no parallel polarization in the plane of incidence. Notice the negative sign in $r_{\perp}$, which indicates a phase change of π.

From Eqs. (1.6.6b) and (1.7.7b), the transmission coefficients are

$$t_{\perp} = \frac{2\cos(56.31°)}{\cos(56.31°) + \left[1.5^2 - \sin^2(56.31°)\right]^{1/2}} = 0.615$$

and

$$t_{//} = \frac{2(1.5)\cos(56.31°)}{(1.5)^2\cos(56.31°) + \left[1.5^2 - \sin^2(56.31°)\right]^{1/2}} = 0.667$$

Notice that $r_{//} + nt_{//} = 1$ and $r_{\perp} + 1 = t_{\perp}$, as we expect. To find the transmittance for each polarization, we need the refraction angle θ_t. From Snell's law, $n_1 \sin\theta_i = n_2 \sin\theta_t$, that is $(1)\sin(56.31°) = (1.5)\sin\theta_t$, we find $\theta_t = 33.69°$. Then, from Eq. (1.6.20),

$$T_{\perp} = \left[\frac{(1.5)\cos(33.69°)}{(1)\cos(56.31°)}\right](0.615)^2 = 0.852 \quad \text{and} \quad T_{//} = \left[\frac{(1.5)\cos(33.69°)}{(1)\cos(56.31°)}\right](0.667)^2 = 1$$

Clearly, light with polarization parallel to the plane of incidence has greater intensity. Note that $R + T = 1$ for both polarizations.

If we were to reflect light from a glass plate, keeping the angle of incidence at 56.3°, then the reflected light will be polarized with an electric field component perpendicular to the plane of incidence. The transmitted light will have the field greater in the plane of incidence; that is, it will be partially polarized. By using a stack of glass plates one can increase the polarization of the transmitted light. (This type of *pile-of-plates polarizer* was invented by Dominique F. J. Arago in 1812.)

1.7 ANTIREFLECTION COATINGS AND DIELECTRIC MIRRORS

Fresnel equations are routinely used in a number of applications in optoelectronics to design and fabricate optical coatings, that is, thin films, to reduce reflections and glare, and also in various components such as dielectric mirrors and filters. Section 1.14 on thin films optics provides a good example of their application for thin film coatings; but in this section we consider two practical applications in optoelectronics: antireflection (AR) coatings and dielectric mirrors.

A. Antireflection Coatings on Photodetectors and Solar Cells

When light is incident on the surface of a semiconductor, it becomes partially reflected. Partial reflection is an important consideration in solar cells where transmitted light energy into the semiconductor device is converted to electrical energy. The refractive index of Si is about 3.5 at wavelengths around 600–800 nm. Thus, the reflectance with $n_1(\text{air}) = 1$ and $n_2(\text{Si}) \approx 3.5$ is

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 = \left(\frac{1 - 3.5}{1 + 3.5}\right)^2 = 0.309$$

This means that 30% of the light is reflected and is not available for conversion to electrical energy; a considerable reduction in the efficiency of the solar cell.

However, we can coat the surface of the semiconductor device with a thin layer of a dielectric material, such as an a-$Si_{1-x}N_x$:H (amorphous hydrogenated silicon nitride based on silicon nitride, Si_3N_4, and x is typically 0.4–0.6), that has an intermediate refractive index. Figure 1.23

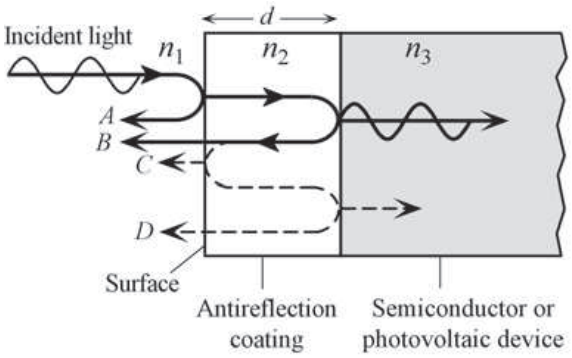


FIGURE 1.23 Illustration of how an antireflection (AR) coating reduces the reflected light intensity. The thickness d and the refractive index n_2 of the antireflection coefficient are such that the waves A and B have a phase difference of π and hence interfere destructively. There is no reflection.

illustrates how the thin dielectric coating, called an **antireflection (AR) coating**, reduces the reflected light intensity. In this case n_1(air) $= 1$, n_2(coating) ≈ 1.9, and n_3(Si) $= 3.5$. Light is first incident on the air/coating surface and some of it becomes reflected and this reflected wave is shown as A in Figure 1.23. Wave A has experienced a 180° phase change on reflection as this is an external reflection. The wave that enters and travels in the coating then becomes reflected at the coating/semiconductor surface. This wave, which is shown as B in Figure 1.23, also suffers a 180° phase change since $n_3 > n_2$. When wave B reaches A, it has suffered a total delay corresponding to traversing the thickness d of the coating twice. The phase difference is equivalent to $k_c(2d)$ in which $k_c = 2\pi/\lambda_c$ is the propagation constant in the coating and is given by $2\pi/\lambda_c$, where λ_c is the wavelength in the coating. Since $\lambda_c = \lambda/n_2$, where λ is the free-space wavelength, the phase difference $\Delta\phi$ between A and B is $(2\pi n_2/\lambda)(2d)$. To reduce the reflected light, A and B must interfere destructively and this requires the phase difference to be π or odd-multiples of π, $m\pi$ in which $m = 1, 3, 5, \ldots$ is an odd-integer. Thus

$$\left(\frac{2\pi n_2}{\lambda}\right)2d = m\pi \quad \text{or} \quad d = m\left(\frac{\lambda}{4n_2}\right) \tag{1.7.1}$$

Antireflection coating

Thus, the thickness of the coating must be odd multiples of the quarter wavelength in the coating and depends on the wavelength.

To obtain a good degree of destructive interference between waves A and B, the two amplitudes must be comparable. (In fact, we have to consider multiple reflections as in Section 1.14, a subject of thin film optics) It turns out that we need $n_2 = (n_1 n_3)^{1/2}$. When $n_2 = (n_1 n_3)^{1/2}$, then the reflection coefficient between the air and coating is equal to that between the coating and the semiconductor. In this case we would need $(3.5)^{1/2}$ or 1.87. Thus, a-$Si_{1-x}N_x$:H is a good choice as an AR coating material on Si solar cells. Its refractive index can also be modified in thin film deposition by changing its composition x. Taking the wavelength to be 600 nm, $d = (600\text{ nm})/\left[4(1.9)\right] = 79.0$ nm or odd-multiples of d.

Once Eq. (1.7.1) is satisfied for an AR coating, it is not difficult to calculate the minimum reflectance obtainable from such a coating. However, we need to consider all reflected waves arising from multiple reflections in the thin film, such as C and D and so on, and we need to properly account for the magnitude of the field by multiplying the field with the right reflection coefficient at each reflection, and with the right transmission coefficient at each transmission. We add all the reflected fields, $A + B + C + D + \ldots$, and then square the total reflected field for the intensity. Once we have done all that, the minimum reflectance $R_{\min}$ is given by

$$R_{\min} = \left(\frac{n_2^2 - n_1 n_3}{n_2^2 + n_1 n_3}\right)^2 \tag{1.7.2}$$

Minimum and maximum reflectance

Clearly, the best choice is for $n_2 = (n_1 n_3)^{1/2}$; recall that $R_{\min}$ is only minimum at one wavelength that satisfies Eq. (1.7.1).

EXAMPLE 1.7.1 Antireflection coating on a photodetector

Consider an InGaAs photodetector for use at a wavelength of 1310 nm in optical communications. The refractive index of doped InGaAs that is used in the detector at 1310 nm is 3.61. What is the reflectance from its surface without an AR coating? What would be the ideal refractive index for an AR coating on InGaAs? What should be thickness of an AR coating that has a refractive index of 1.85. What is its reflectance?

Solution

We use $n_1 = 1$ for air, n_2 for the antireflection coating and $n_3 = 3.61$ for InGaAs,

Without an AR coating, the reflectance is

$$R = \left[(n_1 - n_3)/(n_1 + n_3)\right]^2 = \left[(1 - 3.61)/(1 + 3.61)\right]^2 = 0.32 \quad \text{or} \quad 32\%$$

The ideal AR coating would satisfy Eq. (1.7.1) and would have $n_2 = (n_1 n_3)^{1/2} = 1.90$.

From Eq. (1.7.1), the thickness of the AR layer should be

$$d = \lambda/(4n_2) = (1310 \text{ nm})/\left[4(1.85)\right] = 177 \text{ nm}$$

From Eq. (1.7.2), the minimum reflectance with the AR coating is

$$R_{\min} = \left(\frac{1.85^2 - (1)(3.61)}{1.85^2 + (1)(3.61)}\right)^2 = 7.1 \times 10^{-4}, \quad \text{or} \quad 0.07\%, \text{ very small}$$

B. Dielectric Mirrors and Bragg Reflectors

A **dielectric mirror** consists of a stack of dielectric layers of alternating refractive indices as schematically illustrated in Figure 1.24 (a), in which n_1 is greater than n_2. It is customary to write $n_1 = n_H$(high) and $n_2 - n_L$(low) to represent the $n_H n_L$ stack shown in the figure. The thickness of each layer is a quarter of wavelength or $\lambda_{\text{layer}}/4$ in which λ_{layer} is the wavelength of light in that layer, or λ_o/n where λ_o is the free-space wavelength at which the mirror is required to reflect the incident light and n is the refractive index of the layer. Reflected waves from the interfaces interfere constructively and give rise to a substantial reflected light over a band of wavelengths centered around λ_o as shown in Figure 1.24 (b). If there are sufficient numbers of layers, the reflectance can approach unity at the wavelength λ_o. Since n_1 (high) and n_2 (low) layers are used in pairs, the total number of such pairs of layers, or double layers, is denoted as N; as N increases, the reflectance also increases. The layers are coated, by vacuum deposition techniques, on a suitable substrate. The dielectric mirror illustrated in Figure 1.24 (a) is also known as a **quarter-wave dielectric stack**. Figure 1.24 (b) shows the typical reflectance vs. wavelength behavior of three dielectric mirrors with different n_1 to n_2 ratios and N values. The mirror has been designed to reflect at 1.55 μm.

The reflection coefficient r_{12} for light in layer 1 being reflected at the 1-2 boundary, from Eq. (1.6.9), is $r_{12} = (n_1 - n_2)/(n_1 + n_2)$ and is a *positive* number indicating no phase change. The reflection coefficient for light in layer 2 being reflected at the 2-1 boundary is $r_{21} = (n_2 - n_1)/(n_2 + n_1)$, which is $-r_{12}$ or *negative*, indicating a π phase change. Thus the reflection coefficient alternates in sign through the mirror. Consider two arbitrary waves, B and C, which are reflected at two consecutive interfaces. The two waves are therefore already out of phase by π due to reflections at the different boundaries. Further, wave B travels an additional distance that is twice $(\lambda_2/4)$ (the thickness of layer d_2) before reaching wave B and therefore experiences a phase change equivalent to $2(\lambda_2/4)$ or $\lambda_2/2$, that is π. The phase difference between B and C is then $\pi + \pi$ or 2π. Thus, waves B and C are in phase and *interfere constructively*. We can similarly show that waves C and D also interfere constructively and so on, so that all reflected waves from the consecutive boundaries interfere constructively. After several layers (depending on the n_1/n_2 ratio) the transmitted intensity will be very small and the reflected light intensity will be close to unity as indicated in Figure 1.24 (b). These dielectric

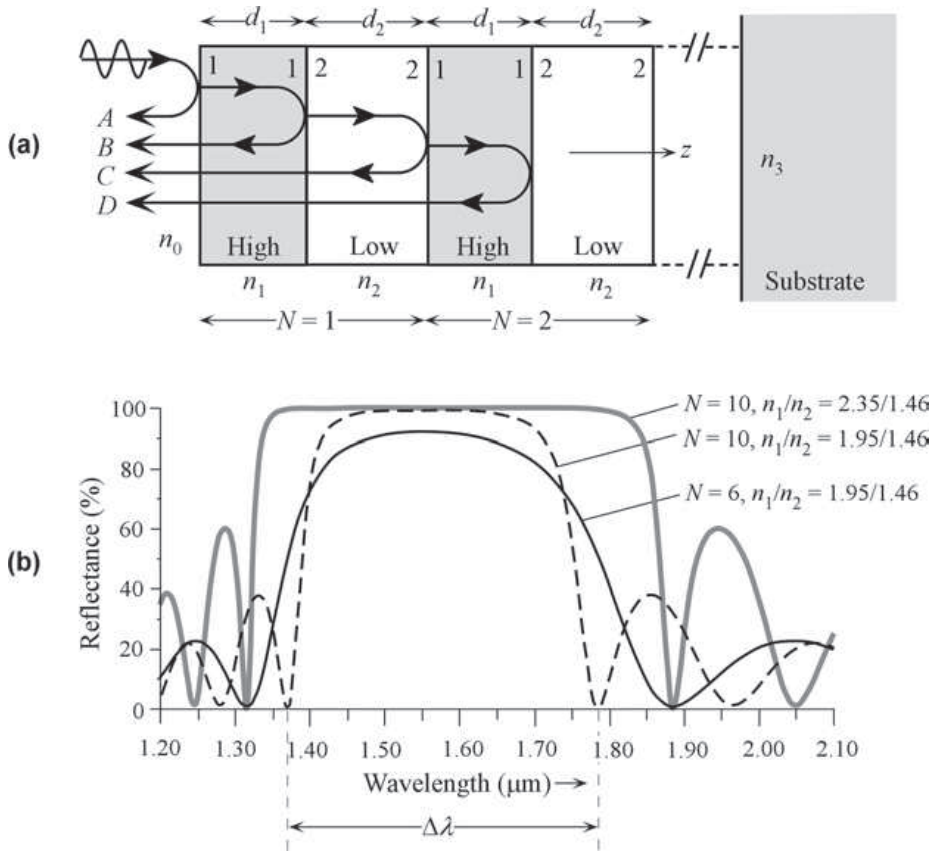


FIGURE 1.24 (a) Schematic illustration of the principle of the dielectric mirror with *many* high and low refractive index layers. Reflected waves *A*, *B*, *C*, *D*, and so on all interfere constructively if the layer thicknesses d_1 and d_2 are a quarter of a wavelength within the layer, that is $d_1 = \lambda/4n_1$ and $d_2 = \lambda/4n_2$, where λ is the free-space wavelength. The dielectric mirror is assumed to be coated on a substrate with an index n_3. (b) The reflectance of three different dielectric mirrors that have $N = 10$, $n_1/n_2 = 2.35/1.46$; $N = 10$, $n_1/n_2 = 1.95/1.46$; $N = 6$, $n_1/n_2 = 1.95/1.46$, $n_3 = 1.52$. (*Note*: $n(TiO_2) = 2.35$, $n(Si_3N_4) = 1.95$, $n(SiO_2) = 1.46$.)

mirrors are widely used in photonics, for example, in solid state lasers such as the vertical cavity surface emitting laser diode. Since the dielectric mirror has a periodic variation in the refractive index (the period being $d_1 + d_2$), similar to a diffraction grating, it is sometimes referred to as a **Bragg reflector**.[24] It is left as an exercise to show that if we interchange the high and low layers, $n_1 = n_L$ and $n_2 = n_H$, we obtain the same result.

As shown in Figure 1.24 (b), with sufficient number of double layers, the reflectance is almost unity over a band of wavelengths, $\Delta\lambda$. Conversely, the transmittance vanishes over the same wavelength range $\Delta\lambda$. This wavelength range in which the transmittance vanishes is called **reflectance bandwidth**; or the **stop band** for the transmitted light. As discussed in Section 1.17, the dielectric mirror in Figure 1.24 (a) is a one-dimensional photonic crystal in which there is a certain stop band within which there can be no propagation of waves along the z axis within the multilayer dielectric structure in Figure 1.24 (a).

[24] As we will see later, a periodic variation in the refractive index is actually a diffraction grating and is able to diffract or reflect a wave in a certain direction if the periodicity is right. The dielectric mirror is also called a one-dimensional Bragg grating structure. It is also a one-dimensional photonic crystal for large *N*.

Various dielectric mirrors, which are quarter wave dielectric stacks on Pyrex or Zerodur substrates. *(Courtesy of Newport.)*

There are two general observations from the reflectance spectra in Figure 1.24 (b). The reflectance R increases with N, the number of double layers used. R also increases with the refractive index ratio n_1/n_2; or put differently, with the index contrast. For a large number of layers, the bandwidth $\Delta\lambda$ of the dielectric mirror increases with the index contrast. The *maximum reflectance* R_N for N pairs of layers is given by

Maximum reflectance, dielectric mirror

$$R_N = \left[\frac{n_1^{2N} - (n_0/n_3)n_2^{2N}}{n_1^{2N} + (n_0/n_3)n_2^{2N}}\right]^2 \tag{1.7.3}$$

The bandwidth $\Delta\lambda$ when $2N$ is large (for near-unity reflectance) is given by

Reflectance bandwidth

$$\frac{\Delta\lambda}{\lambda_o} \approx (4/\pi)\arcsin\left(\frac{n_1 - n_2}{n_1 + n_2}\right) \tag{1.7.4}$$

which increases with the refractive index contrast as apparent from Figure 1.24 (b). If we interchange the low-high layers so that the high index layer is on the air side, there is very little change in the reflectance or the bandwidth in a highly reflecting dielectric mirror.

EXAMPLE 1.7.2 Dielectric mirror

Consider a dielectric mirror that has quarter wave layers consisting of Ta_2O_5 with $n_H = 1.78$ and SiO_2 with $n_L = 1.55$ both at 850 nm, the central wavelength at which the mirror reflects light. Suppose the substrate is Pyrex glass with an index $n_s = 1.47$ and the outside medium is air with $n_0 = 1$. Calculate the maximum reflectance of the mirror when the number N of double layers is 4 and 12. What would happen if you use TiO_2 with $n_H = 2.49$, instead of Ta_2O_5? What is the bandwidth for these two dielectric stacks when they are highly reflecting (with many pairs of layers)? Suppose we use a Si wafer as the substrate, what happens to the maximum reflectance?

Solution

We use $n_0 = 1$ for air, $n_1 = n_H = 1.78$, $n_2 = n_L = 1.55$, $n_3 = n_s = 1.47$, $N = 4$ in Eq. (1.7.3). Thus, for four pairs of layers, the maximum reflectance R_4 is

$$R_4 = \left[\frac{(1.78)^{2(4)} - (1/1.47)(1.55)^{2(4)}}{(1.78)^{2(4)} + (1/1.47)(1.55)^{2(4)}}\right]^2 = 0.40 \quad \text{or} \quad 40\%$$

If we repeat the calculation for 12 pairs of layers, we will find $R_{12} = 90.6\%$.

If we use TiO_2 with $n_1 = n_H = 2.49$, we would find $R_4 = 94.0\%$ and $R_{12} = 100\%$ (to two decimal places). Obviously the refractive index contrast is important; with the TiO_2-SiO_2 stack we only need four double layers to get roughly the same reflectance as from 12 pairs of layers of Ta_2O_5-SiO_2. If we interchange n_H and n_L in the 12-pair stack, that is, $n_1 = n_L$ and $n_2 = n_H$, the Ta_2O_5-SiO_2 reflectance falls to 80.8% but the TiO_2-SiO_2 stack is unaffected since it is already reflecting nearly all the light.

We can only compare bandwidths $\Delta\lambda$ for "infinite" stacks (those with $R \approx 100\%$). For the TiO_2-SiO_2 stack, Eq. (1.7.4) gives

$$\Delta\lambda \approx \lambda_o(4/\pi)\arcsin\left(\frac{n_2 - n_1}{n_2 + n_1}\right) = (850\text{ nm})(4/\pi)\arcsin\left(\frac{2.49 - 1.55}{2.49 + 1.55}\right) = 254\text{ nm}$$

On the other hand, for the Ta_2O_5-SiO_2 infinite stack, we get $\Delta\lambda = 74.8$ nm. As expected, $\Delta\lambda$ is narrower for the smaller contrast stack.

If we change the substrate to a silicon wafer with $n_3 = n_s = 3.50$, we would find that the Ta_2O_5-SiO_2 4-pair stack gives a reflectance of 68.5%, higher than before because the large index changes from n_L to n_s at the substrate interface provides further reflections.

1.8 ABSORPTION OF LIGHT AND COMPLEX REFRACTIVE INDEX

Generally when light propagates through a material it becomes *attenuated* in the direction of propagation as illustrated in Figure 1.25. We distinguish between *absorption* and *scattering* both of which gives rise to a loss of intensity in the regular direction of propagation. In **absorption**, the loss in the power in the propagating electromagnetic wave is due to the conversion of light energy to other forms of energy; for example, lattice vibrations (heat) during the polarization of the molecules of the medium or during the local vibrations of impurity ions driven by the optical field. The excitation of electrons from the valence band to the conduction band, or from impurities to the conduction band, in insulators and semiconductors would also absorb energy from the propagating radiation. Further, free electrons inside a medium with a finite conductivity can be drifted by the optical field in the radiation. As these electrons become scattered by lattice vibrations (or impurities) they will pass the energy they have acquired from the EM wave to lattice vibrations. There are other examples as well. In all cases, some of energy from the propagating light wave is absorbed and converted to other forms of energy.

On the other hand, **scattering** is a process by which the energy from a propagating EM wave is redirected as secondary EM waves in various directions away from the original direction of propagation. (This is discussed later in this chapter) The **attenuation coefficient** α is defined as the fractional decrease in the irradiance I of a wave per unit distance along the direction of propagation z

$$\alpha = -\frac{dI}{Idz} \tag{1.8.1}$$

Definition of attenuation coefficient

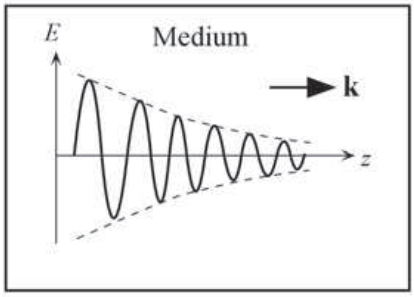


FIGURE 1.25 Attenuation of a traveling wave in a medium results in the decay of its amplitude.

When the irradiance decreases, dI/dz is negative, and the attenuation coefficient is a positive number. If the attenuation of the wave is due to absorption only, then α is the **absorption coefficient**.

It is instructive to consider what happens when a monochromatic light wave such as

Lossless propagation

$$E = E_o \exp j(\omega t - kz) \tag{1.8.2}$$

is propagating in a dielectric medium. The electric field E in Eq. (1.8.2) is either parallel to x or y since propagation is along z. As the wave travels through the medium, the molecules become polarized. This polarization effect is represented by the *relative permittivity* ε_r of the medium. If there were no losses in the polarization process, then the relative permittivity ε_r would be a real number and the corresponding refractive index $n = \varepsilon_r^{1/2}$ would also be a real number. However, we know that there are always some losses in all polarization processes. For example, when the ions of an ionic crystal are displaced from their equilibrium positions by an alternating electric field and made to oscillate, some of the energy from the electric field is coupled and converted to lattice vibrations (called phonons). These losses are generally accounted by describing the whole medium in terms of a **complex relative permittivity (or dielectric constant)** ε_r, that is,

Complex dielectric constant

$$\varepsilon_r = \varepsilon_r' - j\varepsilon_r'' \tag{1.8.3}$$

where the real part ε_r' determines the polarization of the medium with losses ignored and the imaginary part ε_r'' describes the losses in the medium.[25] For a lossless medium, obviously $\varepsilon_r = \varepsilon_r'$. The loss ε_r'' depends on the frequency of the wave and usually peaks at certain natural (resonant) frequencies involved in the absorption process.

An EM wave that is traveling in a medium and experiencing attenuation due to absorption can be generally described by a **complex propagation constant** k, that is,

Complex propagation constant

$$k = k' - jk'' \tag{1.8.4}$$

where k' and k'' are the real and imaginary parts. If we put Eq. (1.8.4) into Eq. (1.8.2) we will find the following:

Attenuated propagation

$$E = E_o \exp(-k''z) \exp j(\omega t - k'z) \tag{1.8.5}$$

The amplitude decays exponentially while the wave propagates along z. The **real** k' part of the complex propagation constant (wave vector) describes the propagation characteristics (*e.g.,* phase velocity $v = \omega/k'$). The **imaginary** k'' part describes the rate of attenuation along z. The irradiance I at any point along z is

$$I \propto |E|^2 \propto \exp(-2k''z)$$

so that the rate of change in the irradiance with distance is

Imaginary part k′

$$dI/dz = -2k''I \tag{1.8.6}$$

where the negative sign represents attenuation. Clearly $\alpha = 2k''$

---

[25] See, for example, S. O. Kasap, *Principles of Electronic Materials and Devices*, 3rd Edition (McGraw-Hill, 2006), Ch. 7. Further, some books use $\varepsilon_r = \varepsilon_r' + j\varepsilon_r''$ instead of Eq. (1.8.3), with a positive imaginary part, but the latter normally refers to an applied field that has a time dependence of the form $\exp(-j\omega t)$.

Suppose that k_o is the propagation constant in vacuum. This is a real quantity as a plane wave suffers no loss in free space. The **complex refractive index** N with a real part n and imaginary part K is defined as the ratio of the complex propagation constant in a medium to propagation constant in free space.

$$\mathrm{N} = n - jK = k/k_o = (1/k_o)\left[k' - jk''\right] \tag{1.8.7}$$

Complex refractive index

that is,

$$n = k'/k_o \quad \text{and} \quad K = k''/k_o$$

The real part n is simply and generally called the **refractive index** and K is called the **extinction coefficient**. In the absence of attenuation

$$k'' = 0, k = k' \quad \text{and} \quad \mathrm{N} = n = k/k_o = k'/k_o$$

We know that in the absence of loss, the relationship between the refractive index n and the relative permittivity ε_r is $n = \varepsilon_r^{1/2}$. This relationship is also valid in the presence of loss except that we must use a complex refractive index and complex relative permittivity, that is

$$\mathrm{N} = n - jK = \sqrt{\varepsilon_r} = \sqrt{\varepsilon_r' - j\varepsilon_r''} \tag{1.8.8}$$

Complex refractive index

By squaring both sides we can relate n and K directly to ε_r' and ε_r''. The final result is

$$n^2 - K^2 = \varepsilon_r' \quad \text{and} \quad 2nK = \varepsilon_r'' \tag{1.8.9}$$

Complex refractive index

Optical properties of materials are typically reported either by showing the frequency dependences of n and K or ε_r' and ε_r''. Clearly we can use Eq. (1.8.9) to obtain one set of properties from the other.

Figure 1.26 shows the real (n) and imaginary (K) parts of the complex refractive index of CdTe as a function of wavelength in the infrared region to highlight their behavior around a resonance absorption phenomenon, when the energy transfer is maximum from the EM wave to the material. Both ionic and electronic polarizations contribute to n (≈ 3.3) at low frequencies whereas only electronic polarization contributes to n (≈ 2.6) at high frequencies. The extinction coefficient peaks at about 72 μm when the EM wave oscillations are efficiently coupled to the *lattice vibrations*,

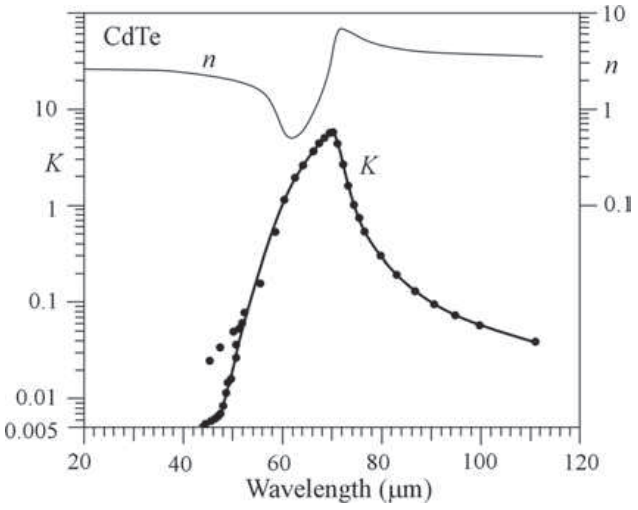


FIGURE 1.26 Optical properties of CdTe as a function of wavelength in the infrared region.

that is, the vibrations of chains of Cd^{2+} and Te^{2-} ions in the crystal,[26] so that energy is passed from the EM wave to these lattice vibrations. This type of absorption in which energy is passed to lattice vibrations is called **lattice** or **Reststrahlen absorption**. Notice that K has a clear peak whereas n shows a response that has a maximum, a minimum, and an inflection point (an S-like shape).

If we know the frequency dependence of the real part ε_r' of the relative permittivity of a material, we can also determine the frequency dependence of the imaginary part ε_r''; and vice versa. This may seem remarkable but it is true provided that we know the frequency dependence of either the real or imaginary part over as wide a range of frequencies as possible (ideally from dc to infinity) and the material is *linear*, that is, it has a relative permittivity that is independent of the applied field; the polarization response must be linearly proportional to the applied field.[27] The relationships that relate the real and imaginary parts of the relative permittivity are called **Kramers-Kronig relations**, which involve integral transformations. If $\varepsilon_r'(\omega)$ and $\varepsilon_r''(\omega)$ represent the frequency dependences of the real and imaginary parts, then one can be determined from the other. Similarly, if we know the wavelength dependence of n (or K), over as a wide wavelength range as possible, we can determine the wavelength dependence of K (or n) using Kramers-Kronig relations for n–K.

It is instructive to mention that the reflection and transmission coefficients that we derived above were based in using a real refractive index, that is neglecting losses. We can still use the reflection and transmission coefficients if we simply use the complex refractive index N instead of n. For example, consider a light wave traveling in free space incident on a material at normal incidence ($\theta_i = 90°$). The reflection coefficient is now

Reflection coefficient

$$r = \frac{1 - n + jK}{1 + n - jK} \tag{1.8.10}$$

The reflectance is then

Reflectance

$$R = \left|\frac{n - jK - 1}{n - jK + 1}\right|^2 = \frac{(n-1)^2 + K^2}{(n+1)^2 + K^2} \tag{1.8.11}$$

which reduce to the usual forms when the extinction coefficient $K = 0$.

The optical properties n and K can be determined by measuring the reflectance from the surface of a material as a function of polarization and the angle of incidence (based on Fresnel's equations).

EXAMPLE 1.8.1 Complex refractive index of InP

An InP crystal has a refractive index (real part) n of 3.549 and an extinction coefficient K of 0.302 at a wavelength of 620 nm (photon energy of 2 eV). Calculate the absorption coefficient α of InP at this wavelength and the reflectance of the air–InP crystal surface.

Solution

The absorption coefficient is

$$\alpha = 2k'' = 2k_o K = 2\left[(2\pi)/(620 \times 10^{-9}\ \text{m})\right](0.302) = 6.1 \times 10^6\ \text{m}^{-1}$$

[26] In physics, these would be *optical phonons.* The EM wave would be interacting with optical phonons, and passing its energy onto these phonons.

[27] In addition, the material system should be passive—contain no sources of energy.

The reflectance R is given by

$$R = \frac{(n-1)^2 + K^2}{(n+1)^2 + K^2} = \frac{(3.549-1)^2 + (0.302)^2}{(3.549+1)^2 + (0.302)^2} = 0.317 \quad \text{or} \quad 31.7\%$$

EXAMPLE 1.8.2 Reflectance of CdTe around resonance absorption

CdTe is also an optical material used in various applications such as lenses, wedges, prisms, beam splitters, antireflection coatings, and optical windows operating typically in the infrared region up to 25 μm. It is used as an optical material for low power CO_2 laser applications. Figure 1.26 shows the infrared refractive index n and the extinction coefficient K of CdTe. Calculate the absorption coefficient α and the reflectance R of CdTe at the Reststrahlen peak, and also at 50 μm and at 100 μm. What is your conclusion?

Solution

At the Reststrahlen peak, $\lambda \approx 70$ μm, $K \approx 6$, and $n \approx 4$, so that the corresponding free-space propagation constant is

$$k_o = 2\pi/\lambda = 2\pi/(70 \times 10^{-6}\,\text{m}) = 9.0 \times 10^4\,\text{m}^{-1}$$

The absorption coefficient α is $2k''$ as in Eq. (1.8.6) so that

$$\alpha = 2k'' = 2k_oK = 2(9.0 \times 10^4\,\text{m}^{-1})(6) = 1.08 \times 10^6\,\text{m}^{-1}$$

which corresponds to an *absorption depth* $1/\alpha$ of about 0.93 μm. The reflectance is

$$R = \frac{(n-1)^2 + K^2}{(n+1)^2 + K^2} = \frac{(4-1)^2 + 6^2}{(4+1)^2 + 6^2} = 0.74 \quad \text{or} \quad 74\,\%$$

At $\lambda = 50$ μm, we first read $K \approx 0.02$, and $n \approx 2$ from Figure 1.26. Repeating the above calculations we get $\alpha \approx 5.08 \times 10^3\,\text{m}^{-1}$, and $R = 0.11$ or 11%. There is a sharp increase in the reflectance from 11% to 72% as we approach the Reststrahlen peak.

At $\lambda = 100$ μm, we first read $K \approx 0.06$, and $n \approx 3.5$ from Figure 1.26. Repeating the above calculations, we find $\alpha = 7.5 \times 10^3\,\text{m}^{-1}$, and $R = 0.31$ or 31%, which is again smaller than the peak reflectance. R is maximum around the Reststrahlen peak. Those materials that strongly absorb light at a given wavelength inside the bulk of the sample will also strongly reflect light at the same wavelength when that light is incident on the surface of the sample.

1.9 TEMPORAL AND SPATIAL COHERENCE

When we represent a traveling EM wave by a pure sinusoidal wave, for example by

$$E_x = E_o \sin(\omega_o t - k_o z) \tag{1.9.1}$$

A sinusoidal wave

with a well-defined angular frequency $\omega_o = 2\pi\upsilon_o$ and a propagation constant k_o, we are assuming that the wave extends infinitely over all space and exists at all times as illustrated in Figure 1.27 (a) inasmuch as a sine function extends periodically over all values of its argument. Such a sine wave is *perfectly coherent* because all points on the wave are predictable. **Perfect coherence** is therefore understood to mean that we can predict the phase of any portion of the wave from any other portion of the wave. Temporal coherence measures the extent to which two points,

such as P and Q, separated in time at a given location in space can be correlated; that is, one can be reliably predicted from the other. At a given location, for a pure sine wave as in Figure 1.27 (a), any two points such as P and Q separated by any time interval are always correlated because we can predict the phase of one (Q) from the phase of the other (P) for any temporal separation.

Any time-dependent arbitrary function $f(t)$ can be represented by a sum of pure sinusoidal waves with varying frequencies, amplitudes, and phases. The **spectrum** of a function $f(t)$ represents the amplitudes of various sinusoidal oscillations that constitute the function. All these pure sine waves are added, with the right amplitude and phase, to make up $f(t)$. Mathematically, the spectrum of $f(t)$ is found by taking the Fourier transform of $f(t)$.[28] We need only one sinusoidal wave at a well-defined frequency $\upsilon_o = \omega_o/2\pi$ to make up Eq. (1.9.1) as illustrated in Figure 1.27 (a).

A pure sine wave is an idealization far from reality and in practice a wave can exist only over a finite time duration Δt, which corresponds to a finite wave train of length $l = c\Delta t$ as illustrated in Figure 1.27 (b). This duration Δt may be the result of the radiation emission process, modulation of the output from a laser, or some other process (indeed, in practice, the amplitude will not be constant over Δt). It is clear that we can only correlate points in the wave train within the duration or over the spatial extent $l = c\Delta t$. This wave train has a coherence time Δt and a coherence length $l = c\Delta t$. Since it is not a perfect sine wave, it contains a number of different frequencies in its spectrum. A proper calculation shows that most of the significant frequencies that constitute this finite wave train lie, as expected, centered around υ_o over a range $\Delta\upsilon$ as shown in Figure 1.27 (b). The spread $\Delta\upsilon$ is the spectral width of the wave train and depends on the temporal coherence length Δt. In this particular case of a truncated sinusoidal wave, $\Delta\upsilon = 2/\Delta t$.

In general, the radiation will not be a perfect truncated sine wave, and the exact relationship between $\Delta\upsilon$ and Δt will depend on the envelope function of the oscillations, which depends on the

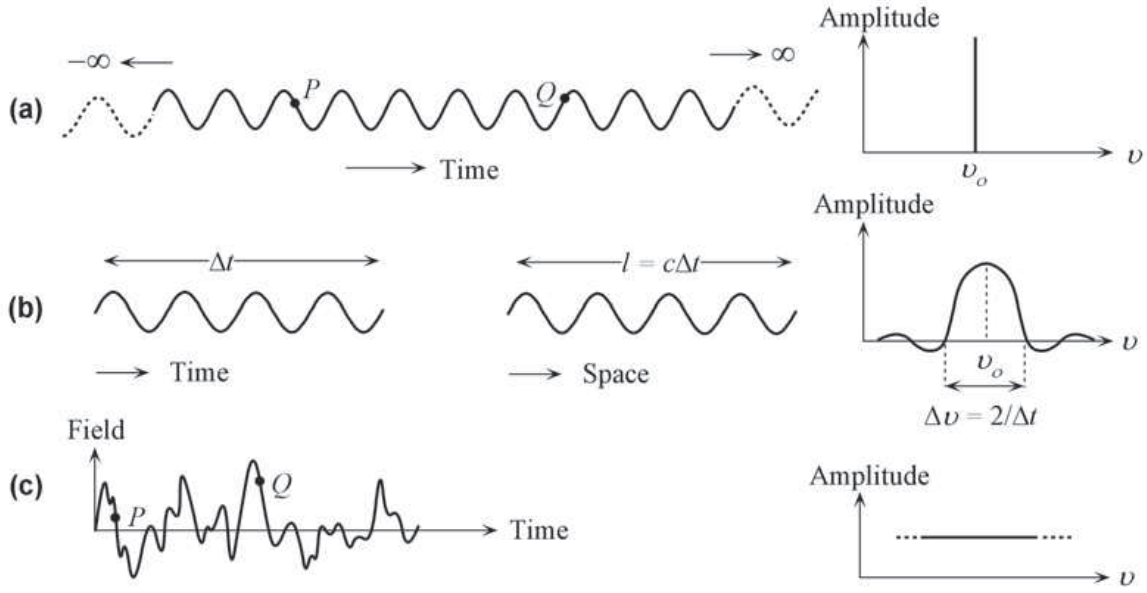


FIGURE 1.27 (a) A sine wave is perfectly coherent and contains a well-defined frequency υ_o. (b) A finite wave train lasts for a duration Δt and has a length l. Its frequency spectrum extends over $\Delta\upsilon = 2/\Delta t$. It has a coherence time Δt and a coherence length l. (c) White light exhibits practically no coherence.

[28] Fourier transformation of $f(t)$ involves integrating $f(t)\exp(-j2\pi\upsilon t)$ over time and produces a function $F(\upsilon)$ that depends on υ. $F(\upsilon)$ is called the **spectrum** of $f(t)$, and represents the pure harmonic waves (such as sine waves) that we need to reconstruct $f(t)$. $F(\upsilon)$ specifies the amplitudes and phases of these harmonics.

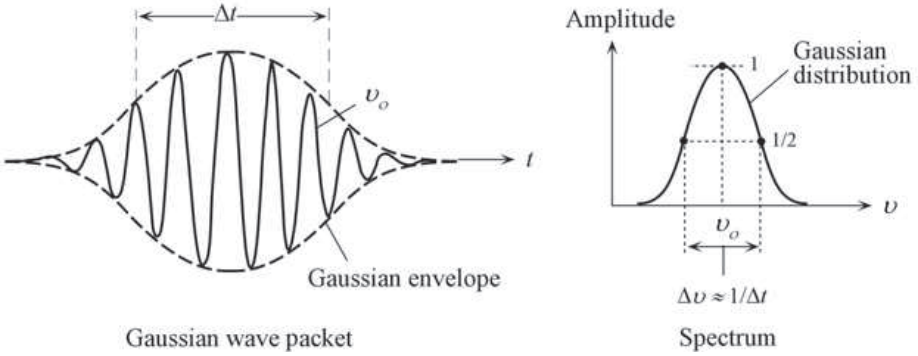


FIGURE 1.28 A Gaussian wave packet which has sinusoidal oscillations at a frequency υ_o and a Gaussian envelope (amplitude variation) over time. Its coherence time Δt between half maximum points is Δt. Its frequency spectrum is Gaussian, centered at the oscillation frequency υ_o, and extends over $\Delta\upsilon \approx 1/\Delta t$.

way the oscillations were generated or emitted. For example, if one takes the envelope of the oscillations to be Gaussian as shown in Figure 1.28 so that the wave is a *Gaussian wave packet*, then the Fourier transform will also be a Gaussian, centered at the oscillation frequency υ_o. The coherence length and coherence time then usually refer to some suitable widths of the respective Gaussian envelopes. If $\Delta\upsilon$ and Δt are the full widths at half maximum (FWHM) spreads in the frequency and time domains, as in Figure 1.28, then approximately[29]

Spectral width and coherence time

$$\Delta\upsilon \approx \frac{1}{\Delta t} \tag{1.9.2}$$

Equation (1.9.2) is usually known as the **bandwidth theorem**. Although a Gaussian wave packet is useful in representing finite-length light wave packets, it still needs to be suitably truncated to be able to represent more practical pulses since a Gaussian function extends over all times. This truncation does not limit the use of Eq. (1.9.2) since it is meant to apply approximately to any wave packet. Although the exact relationship in Eq. (1.9.2) depends on the shape of the wave packet, Eq. (1.9.2) is still widely used for approximately relating $\Delta\upsilon$ and Δt for various wave packets.

Coherence and spectral width are therefore intimately linked. For example, the orange radiation at 589 nm (both D-lines together) emitted from a sodium lamp has spectral width $\Delta\upsilon \approx 5 \times 10^{11}$ Hz. This means that its coherence time is $\Delta t \approx 2 \times 10^{-12}$ s or 2 ps, and its coherence length is 6×10^{-4} m or 0.60 mm. On the other hand, the red lasing emission from a He-Ne laser operating in multimode has a spectral width around 1×10^9 Hz, which corresponds to a coherence length of 30 cm. Furthermore, a continuous wave laser operating in a single mode will have a very narrow linewidth and the emitted radiation will perhaps have a coherence length of several hundred meters. Typically light waves from laser devices have substantial coherence lengths and are therefore widely used in wave-interference studies and applications such as interferometry, holography, and laser Doppler anemometry. Suppose that standing at one location in space we measure the field vs. time behavior shown in Figure 1.27 (c) in which the zero crossing of the signal occurs randomly. Given a point P on this "waveform," we cannot predict the "phase" or the signal at any other point Q. Thus P and Q are not in any way correlated for any temporal separation except Q coinciding with P (or being very close to it by an infinitesimally short time interval). There is no coherence in this "white" light signal and the signal essentially represents white noise;

[29] FWHM is the width of the Gaussian function between the half maximum points (see Appendix A). The bandwidth theorem for a Gaussian wave train is actually $\Delta\upsilon\Delta t = 0.88$ which can be derived by using Fourier transforms.

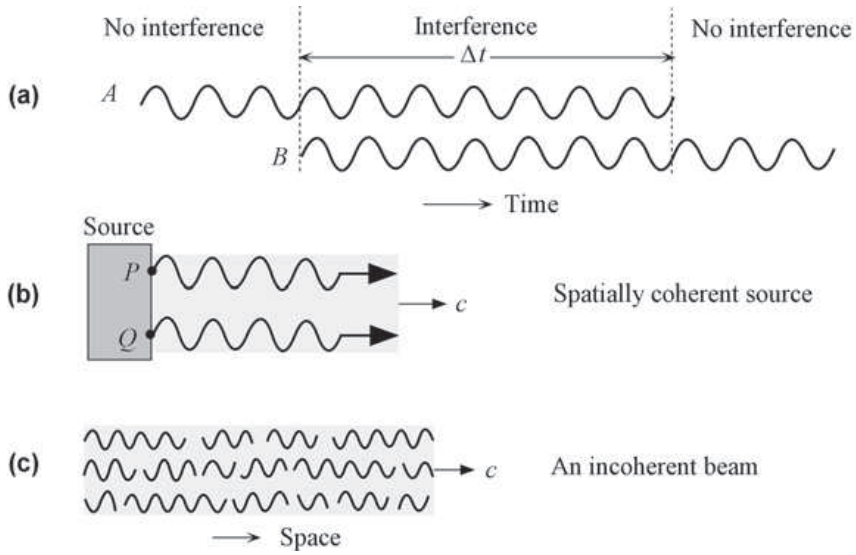


FIGURE 1.29 (a) Two waves can only interfere over the time interval Δt. (b Spatial coherence involves comparing the coherence of waves emitted from different locations on the source. (c) An incoherent beam.

its spectrum typically contains a wide range of frequencies. White light, like an ideal sine wave, is an idealization because it assumes all frequencies are present in the light beam. However, radiation emitted from atoms typically has a central frequency and a certain spectral width, that is, a degree of coherence. The light in the real world lies between (a) and (c) in Figure 1.27.

Coherence between two waves is related to the extent of correlation between two waves. The waves A and B in Figure 1.29 (a) have the same frequency υ_o but they coincide only over the time interval Δt and hence they can only give rise to interference phenomena over this time. They therefore have **mutual temporal coherence** over the time interval Δt. This situation can arise, for example, when two identical wave trains each of coherence length l travel different optical paths. When they arrive at the destination, they can interfere only over a space portion $c\Delta t$. Since interference phenomena can occur only for waves that have mutual coherence, interference experiments, such as Young's double slit experiments, can be used to measure mutual coherence between waves.

Spatial coherence describes the extent of coherence between waves radiated from different locations on a light source as shown in Figure 1.29 (b). If the waves emitted from locations P and Q on the source are in phase, then P and Q are spatially coherent. A spatially coherent source emits waves that are in phase over its entire emission surface. These waves, however, may have partial temporal coherence. A light beam emerging from a spatially coherent light source will hence exhibit spatial coherence across the beam cross-section, that is, the waves in the beam will be in phase over coherence length $c\Delta t$ in which Δt is the temporal coherence. A mostly incoherent beam will contain waves that have very little correlation with each other. The incoherent beam in Figure 1.29 (c) contains waves (across the beam cross-section) whose phases change randomly at random times. (Note, however, that there may be a very short time interval over which there is a little bit of temporal coherence.) The quantitative analysis of coherence requires mathematical techniques based on correlation functions and may be found in more advanced textbooks.

EXAMPLE 1.9.1 Coherence length of LED light

Light emitting diodes (LEDs) emit light that is seemingly of a well-defined color, or wavelength. However, they have much broader spectral widths than lasers. For example, a red LED emitting at 650 nm will have a spectral line width of 22 nm. Find the coherence time and length of the emitted radiation.

Solution

The frequency and wavelength are related through $\upsilon = c/\lambda$, so that differentiating the latter we can find the frequency width $\Delta\upsilon$ from the wavelength width $\Delta\lambda$

$$\frac{\Delta\upsilon}{\Delta\lambda} \approx \left|\frac{d\upsilon}{d\lambda}\right| = \left|-\frac{c}{\lambda^2}\right|$$

so that

$$\Delta\upsilon = \Delta\lambda(c/\lambda^2) = (22 \times 10^{-9}\,\text{m})(3 \times 10^{8}\,\text{m s}^{-1})/(650 \times 10^{-9}\,\text{m})^2$$

$$= 1.562 \times 10^{13}\,\text{Hz}$$

Thus, the coherence time is

$$\Delta t \approx 1/\Delta\upsilon = 1/(1.562 \times 10^{13}\,\text{Hz}) = 6.40 \times 10^{-14}\,\text{s} \quad \text{or} \quad 64.0\,\text{fs}$$

The coherence length is

$$l_c = c\Delta t = 1.9 \times 10^{-5}\,\text{m} \quad \text{or} \quad 19\ \text{microns}$$

The above very short coherence length explains why LEDs are not used in interferometry.

1.10 SUPERPOSITION AND INTERFERENCE OF WAVES

Optical interference involves the superposition of two or more electromagnetic waves in which the electric field vectors are added; the fields add vectorially. The waves are assumed to be nearly monochromic, and have to have the same frequency. Two waves can only interfere if they exhibit *mutual temporal coherence* as in Figure 1.29 (a) at a point in space where they interact. Indeed, interference phenomena can be used to infer on the mutual coherence of the waves. When two waves with the same frequency with fields $\mathbf{E}_1$ and $\mathbf{E}_2$ interfere, they generate a resultant field $\mathbf{E}$ that corresponds to the superposition of **individual fields**, that is, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. Consider two linearly polarized plane waves that originate from O_1 and O_2, as schematically shown in Figure 1.30, so that the field oscillations at some arbitrary point of interest P is given by

$$\mathbf{E}_1 = \mathbf{E}_{o1}\sin(\omega t - kr_1 - \phi_1) \quad \text{and} \quad \mathbf{E}_2 = \mathbf{E}_{o2}\sin(\omega t - kr_2 - \phi_2) \qquad \textbf{(1.10.1)}$$

where r_1 and r_2 are the distances from O_1 and O_2 to P. These waves have the same ω and k. Due to the process that generates the waves, there is a constant phase difference between them given

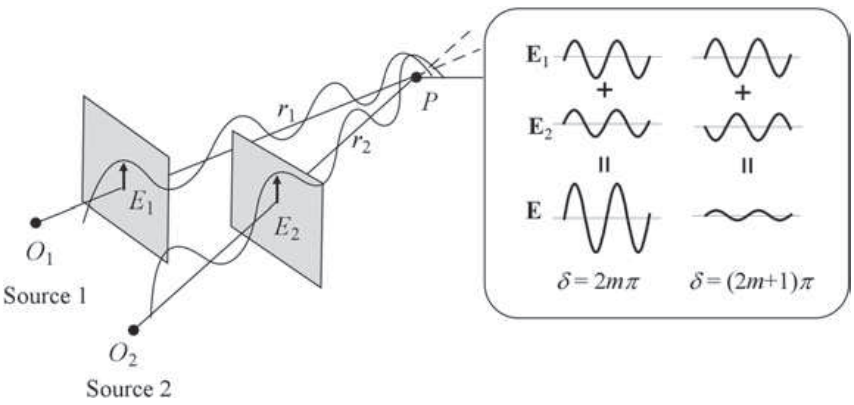


FIGURE 1.30 Interference of two mutually coherent waves of the same frequency originating from sources O_1 and O_2. We examine the resultant at P. The resultant field $\mathbf{E}$ depends on the phase angle δ which depends on the optical path difference $k(r_2 - r_1)$.

by $\phi_2 - \phi_1$. The resultant field at P will be the sum of these two waves, that is, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. Its irradiance depends on the time average of $\mathbf{E} \cdot \mathbf{E}$, that is, $\overline{\mathbf{E} \cdot \mathbf{E}}$, so that

$$\overline{\mathbf{E} \cdot \mathbf{E}} = \overline{(\mathbf{E}_1 + \mathbf{E}_2) \cdot (\mathbf{E}_1 + \mathbf{E}_2)} = \overline{\mathbf{E}_1^2} + \overline{\mathbf{E}_2^2} + 2\overline{\mathbf{E}_1 \cdot \mathbf{E}_2}$$

It is clear that the interference effect is in the $2\overline{\mathbf{E}_1 \cdot \mathbf{E}_2}$ term. We can simply the above equation a little further by assuming $\mathbf{E}_{o1}$ and $\mathbf{E}_{o2}$ are parallel with magnitudes E_{o1} and E_{o2}. Further, irradiance of the interfering waves are $I_1 = \frac{1}{2} c\varepsilon_o E_{o1}^2$ and $I_2 = \frac{1}{2} c\varepsilon_o E_{o2}^2$ so that the resultant irradiance is given by the sum of individual irradiances, I_1 and I_2, and has an additional third term I_{21}, that is,

Interference

$$I = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos\delta \tag{1.10.2}$$

Interference for mutually coherent beams

where the last term is usually written as $2(I_1 I_2)^{1/2} \cos\delta = I_{21}$, and δ is a phase difference given by

$$\delta = k(r_2 - r_1) + (\phi_2 - \phi_1) \tag{1.10.3}$$

Since we are using nearly monochromatic waves, $(\phi_2 - \phi_1)$ is constant, and the interference therefore depends on the term $k(r_2 - r_1)$, which represents the phase difference between the two waves as a result of the *optical path difference* between the waves. As we move point P, $k(r_2 - r_1)$ will change because the optical path difference between the two waves will change; and the interference will therefore also change.

Suppose $(\phi_2 - \phi_1) = 0$, the two waves are emitted from a spatially coherent source. Then, if the path difference $k(r_2 - r_1)$ is 0, 2π or a multiple of 2π, that is, $2m\pi$, $m = 0, \pm 1, \pm 2 \ldots$, then the interference intensity I will be maximum; such interference is defined as **constructive interference**. If the path difference $k(r_2 - r_1)$ is π or 3π or an odd multiple of π, $(2m + 1)\pi$, then the waves will be 180° out of phase, and the interference intensity will be minimum; such interference is defined as **destructive interference**; both constructive and destructive intensity are shown in Figure 1.30. The maximum and minimum irradiances are given by

Constructive and destructive interference

$$I_{\max} = I_1 + I_2 + 2(I_1 I_2)^{1/2} \quad \text{and} \quad I_{\min} = I_1 + I_2 - 2(I_1 I_2)^{1/2} \tag{1.10.4}$$

If the interfering beams have equal irradiances, then $I_{\max} = 4I_1$ and $I_{\min} = 0$.

It is important to emphasize that we have considered the interference of two nearly monochromatic waves that exhibited mutual temporal and spatial coherence. If the waves do not have any mutual coherence, that is, they are incoherent, then we cannot simply superimpose the electric fields as we did above, and the detector at P, which averages the measurement over its response time, will register an irradiance that is simply the sum of individual irradiances,

Irradiance of two superimposed incoherent beams

$$I = I_1 + I_2 \tag{1.10.5}$$

One of the most described interference experiments is **Young's two slit experiment** that generates an interference fringe. In the modern version of this, a coherent beam of light, as available from a laser, is incident on two parallel slits S_1 and S_2. There is a screen far away from the slits on which the waves emanating from the slits interfere, as shown in Figure 1.31. The result is an interference pattern that is composed of light and dark regions, corresponding to $I_{\max}$ and $I_{\min}$. Since S_1 and S_2 are excited by the same wavefront, they emit coherent waves, and we can take $\phi_2 - \phi_1 = 0$. Consider a point P at a distance y on the screen. The phase difference δ at P is then $k(r_2 - r_1)$. As we move P along y, $\delta = k(r_2 - r_1)$ changes and the irradiance on the screen goes through minima and maxima with distance y, following

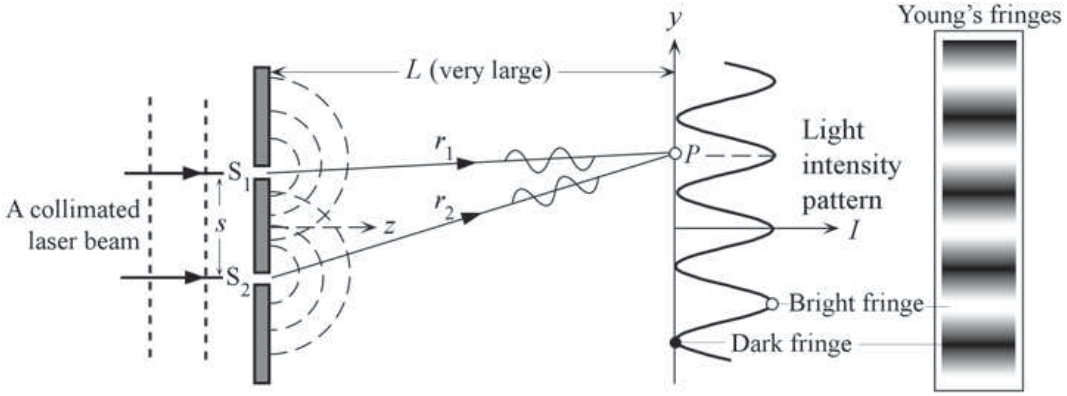


FIGURE 1.31 Young's two slit experiment. Slits S_1 and S_2, separated by s are illuminated at the same time by coherent (nearly monochromatic) collimated laser beam. The irradiance at the screen shows bright and dark fringes due to the interference of waves emanating from the two slits. The screen is assumed to be far away at a distance L from the slits ($L >> s$).

Eq. (1.10.2), generating dark and light bands or fringes. The maxima occur when $\delta = 2m\pi$, and minima when $\delta = (2m + 1)\pi$, where $m = 0, \pm 1, \pm 2$. It is not difficult to show that, if the screen is far away, $(r_2 - r_1) \simeq (s/L)y$ in which L is the distance from the slits to the screen, and s is the separation of the slits, so that δ is proportional to y. Assuming equal irradiances are emitted from S_1 and S_2, $I_1 + I_2 = I_o$, the brightness on the screen changes with y periodically as

$$I = I_o\{1 + \cos[(s/L)ky]\} \qquad \textbf{(1.10.6)}$$ *Interference*

The resulting periodic interference pattern on the screen, as shown in Figure 1.31, is often called **Young's interference fringes**. In a better treatment one needs to consider not only the coherence length of the waves from S_1 and S_2, but also the diffraction that takes place at each slit due to the finite width of the slit. (See diffraction in Section 1.12.)

1.11 MULTIPLE INTERFERENCE AND OPTICAL RESONATORS

Charles Fabry (1867–1945), left, and Alfred Perot (1863–1925), right, were the first French physicists to construct an optical cavity for interferometry. *(Perot:* The Astrophysical Journal*, Vol. 64, November 1926, p. 208, courtesy of the American Astronomical Society. Fabry: Courtesy of Library of Congress Prints and Photographs Division, Washington, DC 20540, USA.)*

An electrical resonator such as a parallel inductor-capacitor (LC) circuit allows only electrical oscillations at the resonant frequency f_o (determined by L and C) within a narrow bandwidth around f_o. Such an LC circuit thereby stores energy at the same frequency. We know that it also acts as a filter at the resonant frequency f_o, which is how we tune in our favorite radio stations. An **optical resonator** is the optical counterpart of the electrical resonator, storing energy or filtering light only at certain frequencies (wavelengths).

When two flat mirrors are perfectly aligned to be parallel as in Figure 1.32 (a) with free space between them, light wave reflections between the two mirrors M_1 and M_2 lead to constructive and destructive interference of these waves within the cavity. Waves reflected from M_1 traveling toward the right interfere with waves reflected from M_2 traveling toward the left. The result is a series of allowed **stationary or standing EM waves** in the cavity as in Figure 1.32 (b) (just like the stationary waves of a vibrating guitar string stretched between two fixed points). Since the electric field at the mirrors (assume metal coated) must be zero, we can only fit in an integer number m of half-wavelengths, $\lambda/2$, into the cavity length L,

Fabry–Perot cavity modes

$$m\left(\frac{\lambda}{2}\right) = L; \quad m = 1, 2, 3, \ldots \tag{1.11.1}$$

Each particular allowed λ, labeled as λ_m, satisfying Eq. (1.11.1) for a given m defines a **cavity mode** as depicted in Figure 1.32 (b). Inasmuch as light frequency υ and wavelength λ are related by $\upsilon = c/\lambda$, the corresponding frequencies υ_m of these modes are the **resonant frequencies** of the cavity

Cavity resonant frequencies

$$\upsilon_m = m\left(\frac{c}{2L}\right) = m\upsilon_f; \quad \upsilon_f = c/(2L) \tag{1.11.2}$$

in which υ_f is the lowest frequency corresponding to $m = 1$, the fundamental mode, and also the frequency separation of two neighboring modes, $\Delta\upsilon_m = \upsilon_{m+1} - \upsilon_m = \upsilon_f$. It is known as the **free spectral range**. Figure 1.32 (c) illustrates schematically the intensity of the allowed modes as a function of frequency. If there are no losses from the cavity, that is, the mirrors are perfectly reflecting, then the peaks at frequencies υ_m defined by Eq. (1.11.2) would be sharp lines. If the mirrors are not perfectly reflecting so that some radiation escapes from the cavity, then the mode peaks are not as sharp and have a finite width, as indicated in Figure 1.32 (c). It is apparent that this simple optical cavity serves to "store" radiation energy only at certain frequencies and it is called a **Fabry–Perot optical resonator**. The optical resonator does not have to have two

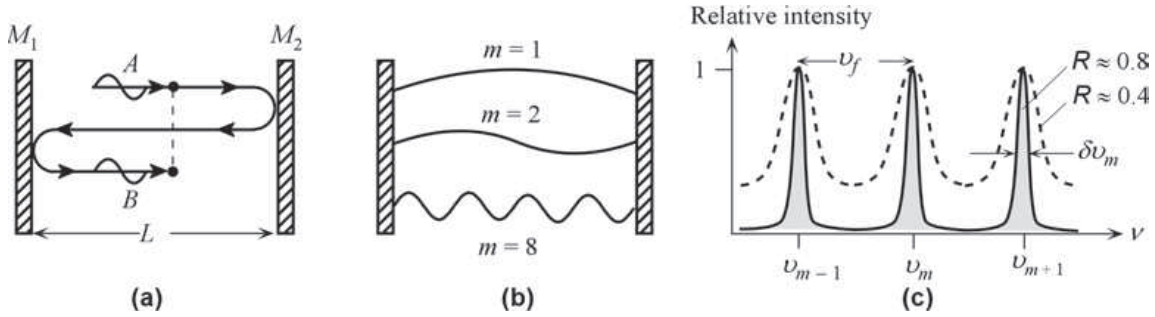


FIGURE 1.32 Schematic illustration of the Fabry–Perot optical cavity and its properties. (a) Reflected waves interfere. (b) Only standing EM waves, modes, of certain wavelengths are allowed in the cavity. (c) Intensity vs. frequency for various modes. R is the mirror reflectance and lower R means higher loss from the cavity.

mirrors and free space between them as shown in Figure 1.32 (a). It can be a solid medium, such as a dielectric (*e.g.,* glass) plate or a rod, whose ends are used to reflect light. The ends of the solid can even have thin film coatings, or dielectric mirrors, to enhance the reflection. Optical resonators are also called **etalons**.

Consider an arbitrary wave such as A traveling toward the right at some instant as shown in Figure 1.32 (a). After one round trip this wave would be again traveling toward the right but now as wave B, which has a phase difference and a different magnitude due to non-perfect reflections. If the mirrors M_1 and M_2 are identical with a reflection coefficient of *magnitude* r, then B has one round-trip phase difference of $k(2L)$ and a magnitude r^2 (two reflections) with respect to A. When A and B interfere, the result is

$$A + B = A + Ar^2 \exp(-j2kL)$$

Of course, just like A, B will continue on and will be reflected twice, and after one round trip it would be going toward the right again and we will now have three waves interfering and so on. After infinite round-trip reflections, the resultant field E_{cavity} is due to infinite such interferences.

$$\begin{aligned} E_{\text{cavity}} &= A + B + \ldots \\ &= A + Ar^2 \exp(-j2kL) + Ar^4 \exp(-j4kL) + Ar^6 \exp(-j6kL) + \ldots \end{aligned}$$

The sum of this geometric series is easily evaluated as

$$E_{\text{cavity}} = \frac{A}{1 - r^2 \exp(-j2kL)}$$

Once we know the field in the cavity we can calculate the intensity $I_{\text{cavity}} \propto |E_{\text{cavity}}|^2$. Further, we can use *reflectance* $R = r^2$ to further simplify the expression. The final result after algebraic manipulation is

$$I_{\text{cavity}} = \frac{I_o}{(1 - R)^2 + 4R \sin^2(kL)} \qquad \textbf{(1.11.3)}$$

Cavity intensity

in which $I_o \propto A^2$ is the original intensity. The intensity in the cavity is maximum $I_{\max}$ whenever $\sin^2(kL)$ in the denominator of Eq. (1.11.3) is zero, which corresponds to (kL) being $m\pi$, in which m is an integer. Thus, the intensity vs. k, or equivalently, the intensity vs. frequency spectrum, peaks whenever $kL = m\pi$, as in Figure 1.32 (c). These peaks are located at $k = k_m$ that satisfy $k_m L = m\pi$, which leads directly to Eqs. (1.11.1) and (1.11.2) that were derived intuitively. For those resonant k_m values, Eq. (1.11.3) gives

$$I_{\max} = \frac{I_o}{(1 - R)^2}; \quad k_m L = m\pi \qquad \textbf{(1.11.4)}$$

Maximum cavity intensity

A smaller mirror reflectance R means more radiation loss from the cavity, which affects the intensity distribution in the cavity. We can show from Eq. (1.11.3) that smaller R values result in broader mode peaks and a smaller difference between the minimum and maximum intensity in the cavity as schematically illustrated in Figure 1.32 (c). The **spectral width**[30] $\delta\upsilon_m$ of the Fabry–Perot etalon is the full width at half maximum (FWHM) of an individual mode

[30] The spectral width is also called the fringe width and m the fringe order.

intensity, as defined in Figure 1.32 (c). It can be calculated in a straightforward fashion when $R > 0.6$ from

Spectral width and Finesse

$$\delta\upsilon_m = \frac{\upsilon_f}{F}; \quad F = \frac{\pi R^{1/2}}{1 - R} \tag{1.11.5}$$

in which F is called the **finesse** of the resonator, which increases as losses decrease (R increases). Large finesses lead to sharper mode peaks. Finesse is the ratio of mode separation ($\Delta\upsilon_m$) to spectral width ($\delta\upsilon_m$).

We can also define a quality factor Q for the optical resonant cavity in a similar fashion to defining a *Q-factor* for an *LC* oscillator, that is,

Q-factor and Finesse

$$\text{Quality factor, } Q = \frac{\text{Resonant frequency}}{\text{Spectral width}} = \frac{\upsilon_m}{\delta\upsilon_m} = mF \tag{1.11.6}$$

The Q-factor is a measure of the frequency selectiveness of a resonator; the higher the Q-factor, the more selective the resonator, or narrower the spectral width. It is also a measure of the energy stored in the resonator per unit energy dissipated (due to losses such as from the reflecting surfaces) per cycle of oscillation.

The Fabry–Perot optical cavities are widely used in laser, interference filter, and spectroscopic applications. Consider a light beam that is incident on a Fabry–Perot cavity as in Figure 1.33. The optical cavity is formed by partially transmitting and reflecting plates. Part of the incident beam enters the cavity. We know that only special cavity modes are allowed to exist in the cavity since other wavelengths lead to destructive interference. Thus, if the incident beam has a wavelength corresponding to one of the cavity modes, it can sustain oscillations in the cavity and hence lead to a transmitted beam. The output light is a fraction of the light intensity in the cavity and is proportional to Eq. (1.11.3). Commercial interference filters are based on this principle except that they typically use two cavities in series formed by dielectric mirrors (a stack of quarter wavelength layers); the structure is more complicated than in Figure 1.33. Further, adjusting the cavity length L provides a "tuning capability" to scan different wavelengths.

Equation (1.11.3) describes the intensity of the radiation in the cavity. The intensity of the transmitted radiation in Figure 1.33 can be calculated, as above, by considering that each time a wave is reflected at the right mirror, a portion of it is transmitted, and that these transmitted waves can interfere only constructively to constitute a transmitted beam when $kL = m\pi$. Intuitively, if I_{incident} is the incident light intensity, then a fraction $(1 - R)$ of this would enter

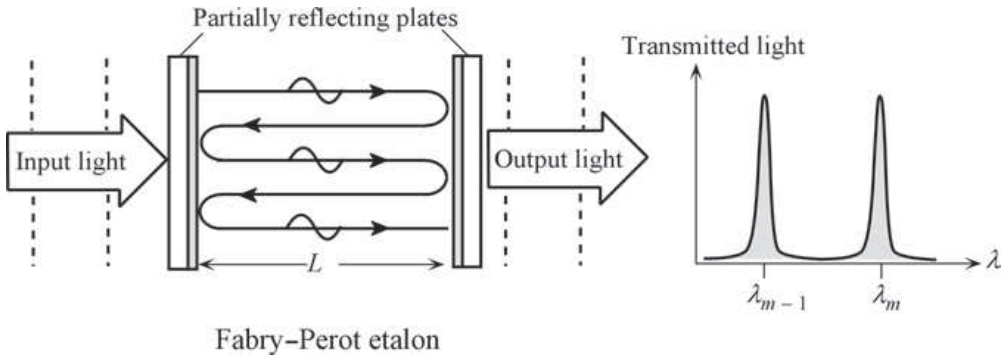


FIGURE 1.33 Transmitted light through a Fabry–Perot optical cavity.

Left: Fused silica etalon. Right: A 10 GHz air spaced etalon with 3 zerodur spacers. *(Courtesy of Light Machinery Inc.)*

the cavity to build up into I_{cavity} in Eq. (1.11.3), and a fraction $(1 - R)$ of I_{cavity} would leave the cavity as the transmitted intensity $I_{\text{tansmitted}}$. Thus,

$$I_{\text{transmitted}} = I_{\text{incident}} \frac{(1 - R)^2}{(1 - R)^2 + 4R\sin^2(kL)} \tag{1.11.7}$$

Transmitted mode intensities

which is again maximum just as for I_{cavity} whenever $kL = m\pi$ as shown in terms of wavelength in Figure 1.33.[31]

The ideas above can be readily extended to a medium with a refractive index n by using nk for k or λ/n for λ where k and λ are the free-space propagation constant and wavelength, respectively. Equations (1.11.1) and (1.11.2) become

$$m\left(\frac{\lambda}{2n}\right) = L \quad m = 1, 2, 3, \ldots \tag{1.11.8}$$

Fabry–Perot cavity modes in a medium

$$\upsilon_m = m\left(\frac{c}{2nL}\right) = m\upsilon_f; \quad \upsilon_f = c/(2nL) \tag{1.11.9}$$

Cavity resonant frequencies

Further, if the angle of incidence θ at the etalon face is not normal, then we can resolve k to be along the cavity axis; that is use $k\cos\theta$ instead of k in the discussions above.

The two mirrors in Figure 1.32 (a) were assumed to have the same reflectance R. Suppose that R_1 and R_2 are the reflectances of the mirrors M_1 and M_2. Then, we can continue to use the above equations by using an average geometric reflectance, that is $R = (R_1R_2)^{1/2}$.

EXAMPLE 1.11.1 Resonator modes and spectral width of a semiconductor Fabry–Perot cavity

Consider a Fabry–Perot optical cavity made of a semiconductor material with mirrors at its ends. (The mirrors have been obtained by coating the end of the semiconductor crystal.) The length of the semiconductor, and hence the cavity, is 250 μm and mirrors at the ends have a reflectance of 0.90. Calculate the cavity mode nearest to the free-space wavelength of 1310 nm. Calculate the separation of the modes, finesse, spectral width of each mode in frequency and wavelength, and the Q-factor.

[31] The term on the right multiplying I_{incident} in Eq. (1.11.7) is usually known as the *Airy function.*

Solution

The wavelength of radiation inside the cavity is λ/n, where n is the refractive index of the medium, which is 3.6, and λ is the free-space wavelength. We need to use Eq. (1.11.8), so that the mode number for the wavelength 1310 nm is

$$m = \frac{2nL}{\lambda} = \frac{2(3.6)(250 \times 10^{-6})}{(1310 \times 10^{-9})} = 1374.05$$

which must be an integer (1374) so that the actual mode wavelength is

$$\lambda_m = \frac{2nL}{m} = \frac{2(3.6)(250 \times 10^{-6})}{(1374)} = 1310.04 \text{ nm}$$

Thus, for all practical purposes the mode wavelength is 1310 nm.

The mode frequency is $\upsilon_m = c/\lambda_m$ so that the separation of the modes, from Eq. (1.11.9), is

$$\Delta\upsilon_m = \upsilon_f = \frac{c}{2nL} = \frac{(3 \times 10^8)}{2(3.6)(250 \times 10^{-6})} = 1.67 \times 10^{11} \text{ Hz} \quad \text{or} \quad 167 \text{ GHz}$$

The finesse is

$$F = \frac{\pi R^{1/2}}{1 - R} = \frac{\pi 0.90^{1/2}}{1 - 0.90} = 29.8$$

and the spectral width of each mode is

$$\delta\upsilon_m = \frac{\upsilon_f}{F} = \frac{1.67 \times 10^{11}}{29.8} = 5.59 \times 10^9 \text{ Hz} \quad \text{or} \quad 5.59 \text{ GHz}$$

The mode spectral width $\delta\upsilon_m$ will correspond to a certain spectral wavelength width $\delta\lambda_m$. The mode wavelength $\lambda_m = 1310$ nm corresponds to a mode frequency $\upsilon_m = c/\lambda_m = 2.29 \times 10^{14}$ Hz. Since $\lambda_m = c/\upsilon_m$, we can differentiate this expression to relate small changes in λ_m and υ_m,

$$\delta\lambda_m = \delta\left(\frac{c}{\upsilon_m}\right) = \left|-\frac{c}{\upsilon_m^2}\right|\delta\upsilon_m = \frac{(3 \times 10^8)}{(2.29 \times 10^{14})^2}(5.59 \times 10^9) = 3.2 \times 10^{-11} \text{ m} \quad \text{or} \quad 0.032 \text{ nm}$$

The Q-factor is

$$Q = mF = (1374)(29.8) = 4.1 \times 10^4$$

An optical cavity like this is used in so-called Fabry–Perot semiconductor laser diodes, and will be discussed further in Chapter 4.

1.12 DIFFRACTION PRINCIPLES

A. Fraunhofer Diffraction

An important property of waves is that they exhibit diffraction effects; for example, sound waves are able to bend (deflect around) corners and a light beam can similarly "bend" around an obstruction (though the bending may be very small). Figure 1.34 shows an example of a collimated light beam passing through a circular aperture (a circular opening in an opaque screen). The passing beam is found to be divergent and to exhibit an intensity pattern that has bright and dark rings, called *Airy rings*.[32] The passing beam is said to be diffracted and its light intensity pattern is called

[32] Sir George Airy (1801–1892), Astronomer Royal of Britain from 1835 to1881.

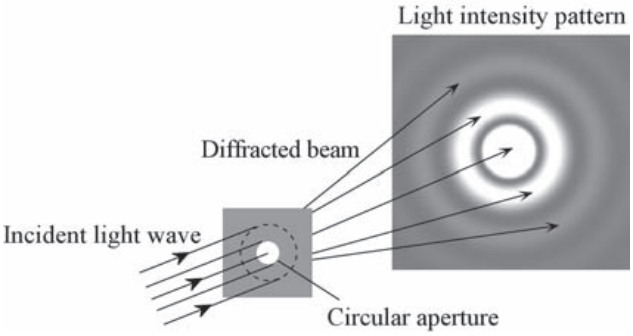


FIGURE 1.34 A collimated light beam incident on a small circular aperture becomes diffracted and its light intensity pattern after passing through the aperture is a diffraction pattern with circular bright rings (called Airy rings). If the screen is far away from the aperture, this would be a Fraunhofer diffraction pattern.

a **diffraction pattern**. Clearly, the light pattern of the diffracted beam does *not* correspond to the geometric shadow of the circular aperture. Diffraction phenomena are generally classified into two categories. In **Fraunhofer diffraction**,[33] the incident light beam is a plane wave (a collimated light beam) and the observation or detection of the light intensity pattern (by placing a photographic screen, *etc.*) is done far away from the aperture so that the waves received also look like plane waves. Inserting a lens between the aperture and the photographic screen enables the screen to be closer to the aperture. In **Fresnel diffraction**, the incident light beam and the received light waves are not plane waves but have significant wavefront curvatures. Typically, the light source and the photographic screen are both close to the aperture so that the wavefronts are curved. Fraunhofer diffraction is by far the most important; and it is mathematically easier to treat.

Diffraction can be understood in terms of the interference of multiple waves emanating from the aperture in the obstruction.[34] We will consider a plane wave incident on a one-dimensional slit of length *a.* According to the Huygens-Fresnel principle,[35] *every unobstructed point of a wavefront, at a given instant in time, serves as a source of spherical secondary waves (with the same frequency as that of the primary wave). The amplitude of the optical field at any point beyond is the superposition of all these wavelets (considering their amplitudes and relative phases).* Figures 1.35 (a) and (b) illustrates this point pictorially showing that, when the plane wave reaches the aperture, points in the aperture become sources of coherent spherical secondary waves. These spherical waves interfere to constitute the new wavefront (the new wavefront is the envelope of the wavefronts of these secondary waves). These spherical waves can interfere constructively not just in the forward direction as in (a) but also in other appropriate directions, as in (b), giving rise to the observed bright and dark patterns (rings for a circular aperture) on the observation screen.

We can divide the unobstructed width a of the aperture into a very large number N of coherent "point sources" each of extent $\delta y = a/N$ (obviously δy is sufficiently small to be nearly a point), as in Figure 1.36 (a). Since the aperture a is illuminated uniformly by the plane wave, the strength (amplitude) of each point source would be proportional to $a/N = \delta y$. Each would be a source of spherical waves. In the forward direction ($\theta = 0$), they would all be in phase and constitute a forward wave, along the z-direction. But they can also be in phase at some angle θ to the z-direction and hence give rise to a diffracted wave along this direction. We will evaluate the intensity of the received wave at a point on the screen as the *sum of all*

[33] Joseph von Fraunhofer (1787–1826) was a German physicist who also observed the various dark lines in Sun's spectrum due to hydrogen absorption.

[34] "No one has been able to define the difference between interference and diffraction satisfactorily" [R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, 1963)].

[35] Eugene Hecht, *Optics*, 4th Edition (Pearson Education, 2002), Ch. 10, p. 444.

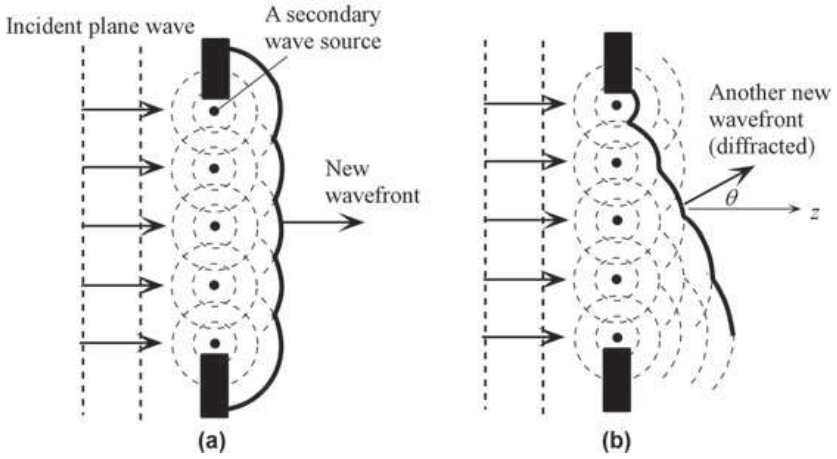


FIGURE 1.35 (a) Huygens-Fresnel principle states that each point in the aperture becomes a source of secondary waves (spherical waves). The spherical wavefronts are separated by λ. The new wavefront is the envelope of all these spherical wavefronts. (b) Another possible wavefront occurs at an angle θ to the z-direction, which is a diffracted wave.

waves arriving from all point sources in the aperture. The screen is far away from the aperture so the waves arrive almost parallel at the screen (alternatively a lens can be used to focus the diffracted parallel rays to form the diffraction pattern).

Consider an arbitrary direction θ, and consider the phase of the emitted wave (Y) from an arbitrary point source at y with respect to the wave (A) emitted from source at $y = 0$ as shown in Figure 1.36 (a). If k is the propagation constant, $k = 2\pi/\lambda$, the wave Y is out of phase with respect to A by $ky\sin\theta$. Thus the wave emitted from the point source at y has a field δE,

$$\delta E \propto (\delta y)\exp(-jky\sin\theta) \tag{1.12.1}$$

All of these waves from point sources from $y = 0$ to $y = a$ interfere at the screen at a point P that makes an angle θ at the slit, and the resultant field at the screen at this point P is

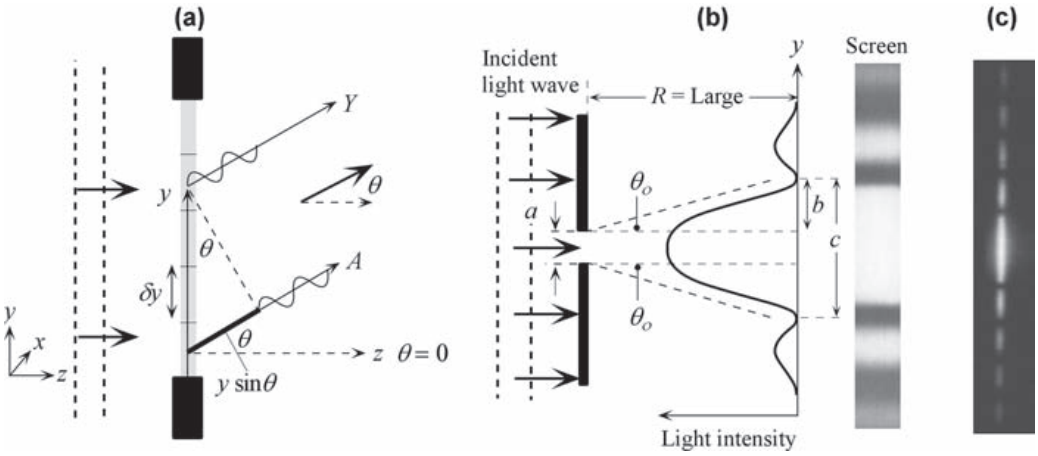


FIGURE 1.36 (a) The aperture has a finite width a along y, but it is very long along x so that it is a one-dimensional slit. The aperture is divided into N number of point sources each occupying δy with amplitude proportional to δy since the slit is excited by a plane electromagnetic wave. (b) The intensity distribution in the received light at the screen far away from the aperture: *the diffraction pattern.* Note that the slit is very long along x so that there is no diffraction along this dimension. The incident wave illuminates the whole slit. (c) Typical diffraction pattern using a laser pointer on a single slit. The difference from the pattern in (b) is due to the finite size of the laser pointer beam along x that is smaller than the length of the slit.

their sum. Because the screen is far away, a point on the screen is at the same distance from anywhere in the aperture. This means that all the spherical waves from the aperture experience the same phase change and decrease in amplitude in reaching the screen. This simply scales δE at the screen by an amount that is the same for all waves coming from the aperture. Thus, the resultant field $E(\theta)$ at point P at the screen is

$$E(\theta) = C\int_{y=0}^{y=a} \delta y \exp(-jky\sin\theta) \tag{1.12.2}$$

in which C is a constant. Integrating Eq. (1.12.2) we get

$$E(\theta) = \frac{Ce^{-j\frac{1}{2}ka\sin\theta}\, a\sin\left(\frac{1}{2}ka\sin\theta\right)}{\frac{1}{2}ka\sin\theta}$$

The light intensity I at a point at P at the screen is proportional to $|E_\theta|^2$, and thus

$$I(\theta) = \left[\frac{C'a\sin\left(\frac{1}{2}ka\sin\theta\right)}{\frac{1}{2}ka\sin\theta}\right]^2 = I(0)\,\mathrm{sinc}^2(\beta); \quad \beta = \tfrac{1}{2}(ka\sin\theta) \tag{1.12.3}$$

Single slit diffraction equation

in which C' is a constant and β is a convenient new variable representing θ, and sinc ("sink") is a function that is defined by $\mathrm{sinc}\,(\beta) = \sin(\beta)/(\beta)$.

If we were to plot Eq. (1.12.3) as a function of θ at the screen we would see the intensity (diffraction) pattern schematically depicted in Figure 1.36 (b). First, observe that the pattern has bright and dark regions, corresponding to constructive and destructive interference of waves emanating from the aperture. Second, the center bright region is wider than the aperture width a, which mean that the transmitted beam must be *diverging*. The zero intensity occurs when, from Eq. (1.12.3),

$$\sin\theta = \frac{m\lambda}{a}; \quad m = \pm 1, \pm 2, \ldots \tag{1.12.4}$$

Zero intensity points

The angle θ_o for the first zero, corresponding to $m = \pm 1$, is given by $\theta_o = \pm\lambda/a$, where we assumed that the divergence is small (usually the case) so that $\sin\theta_o \approx \theta_o$. Thus, the divergence $\Delta\theta$, the angular spread, of the diffracted beam is given by

$$\Delta\theta = 2\theta_o \approx \frac{2\lambda}{a} \tag{1.12.5}$$

Divergence from single slit of width a

A light wave at a wavelength 1300 nm, diffracted by a slit of width $a = 100$ μm (about the thickness of this page), has a divergence $\Delta\theta$ of about 1.5°. From Figure 1.36 (b), it is apparent that, using geometry, we can easily calculate the width c of the central bright region of the intensity pattern, given θ_o from Eq. (1.12.5) and the distance R of the screen from the aperture.

The diffraction patterns from two-dimensional apertures such as rectangular and circular apertures are more complicated to calculate but they use the same principle based on the multiple interference of waves emitted from all point sources in the aperture. The diffraction pattern of a rectangular aperture is shown in Figure 1.37. It involves the multiplication of two individual single slit (sinc) functions, one slit of width a along the horizontal axis, and the other of width b along the vertical axis. (Why is the diffraction pattern wider along the horizontal axis?)

The diffraction pattern from a circular aperture, known as **Airy rings**, was shown in Figure 1.34, and can be roughly visualized by rotating the intensity pattern in Figure 1.36 (b)

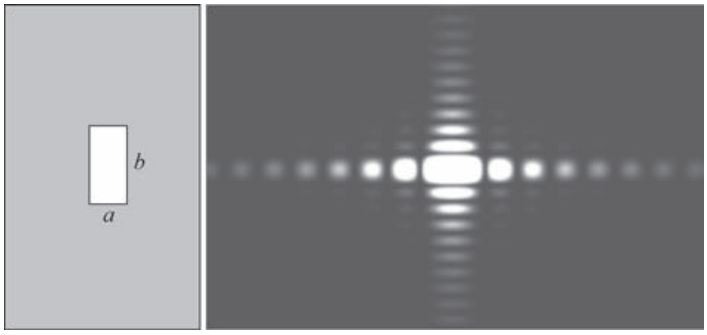


FIGURE 1.37 The rectangular aperture of dimensions $a \times b$ on the left gives the diffraction pattern on the right (b is twice a).

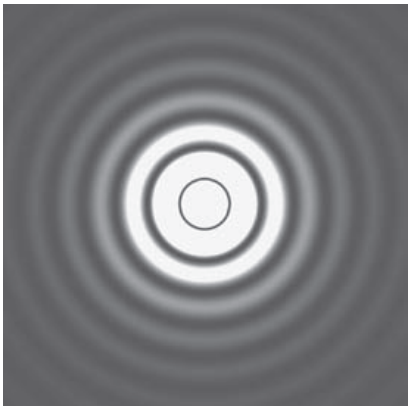
about the z-axis. We can, as we did for the single slit, sum all waves emanating from every point in the circular aperture, taking into account their relative phases when they arrive at the screen to obtain the actual intensity pattern at the screen. The result is that the diffraction pattern is a Bessel function of the first kind,[36] and not a simply rotated sinc function. The central white spot is called the **Airy disk**; its radius corresponds to the radius of the first dark ring. We can still use Figures 1.36 (a) and (b) to imagine how diffraction occurs from a circular aperture by taking this as a cut through the aperture so that a is now the diameter of the aperture, denoted as D. The angular position θ_o of the first dark ring, as defined as in Figure 1.36 (b), is determined by the diameter D of the aperture and the wavelength λ, and is given by

Angular radius of Airy disk

$$\sin\theta_o = 1.22\frac{\lambda}{D} \tag{1.12.6}$$

The divergence angle from the aperture center to the Airy disk circumference is $2\theta_o$. If R is the distance of the screen from the aperture, then the radius of the Airy disk, approximately b, can be calculated from the geometry in Figure 1.36 (b), which gives $b/R = \tan\theta_o \approx \theta_o$. If a lens is used to focus the diffracted light waves onto a screen, then $R = f$, focal length of the lens.

It is worth commenting on the Gaussian beam at this point. Suppose we now examine a Gaussian beam with a waist $2w_o$ that is the same as the aperture size D. The far field



Diffraction pattern far away from a circular aperture.

Diffraction pattern far away from a square aperture.

[36] Bessel functions are special mathematical functions, which can be looked up in mathematics handbooks. They are used in various engineering problems.

half-divergence angle θ of this Gaussian beam would be $\theta = (2/\pi)(\lambda/D)$ or $0.637(\lambda/D)$ as in Eq. (1.1.7). The Gaussian beam has a smaller divergence than the diffracted beam from a circular aperture. The difference is due the fact that each point in the circular aperture emits with the same intensity because the aperture is illuminated by a plane wave. If we were to change the emission intensity within the aperture to follow a Gaussian distribution, we would see a Gaussian beam as the "diffracted beam." The Gaussian beam is a self-diffracted beam and has the smallest divergence for a given beam diameter.

EXAMPLE 1.12.1 Resolving power of imaging systems

Consider what happens when two neighboring point light sources are examined through an imaging system with an aperture of diameter D (this may even be a lens). The two sources have an angular separation of $\Delta\theta$ at the aperture. The aperture produces a diffraction pattern of the sources S_1 and S_2, as shown in Figure 1.38. As the points get closer, their angular separation becomes narrower and the diffraction patterns overlap more. According to the **Rayleigh criterion**, the two spots are just resolvable when the principal maximum of one diffraction pattern coincides with the minimum of the other, which is given by the condition

$$\sin(\Delta\theta_{\min}) = 1.22\frac{\lambda}{D} \tag{1.12.7}$$

Angular limit of resolution

The human eye has a pupil diameter of about 2 mm. What would be the minimum angular separation of two points under a green light of 550 nm and their minimum separation if the two objects are 30 cm from the eye? The image will be a diffraction pattern in the eye, and is a result of waves in this medium. If the refractive index $n \approx 1.33$ (water) in the eye, then Eq. (1.12.7) is

$$\sin(\Delta\theta_{\min}) = 1.22\frac{\lambda}{nD} = 1.22\frac{(550 \times 10^{-9}\ \text{m})}{(1.33)(2 \times 10^{-3}\ \text{m})}$$

giving

$$\Delta\theta_{\min} = 0.0145^\circ$$

Their minimum separation s would be

$$s = 2L\tan(\Delta\theta_{\min}/2) = 2(300\ \text{mm})\tan(0.0145^\circ/2) = 0.076\ \text{mm} = 76\ \mu\text{m}$$

which is about the thickness of a human hair (or this page).

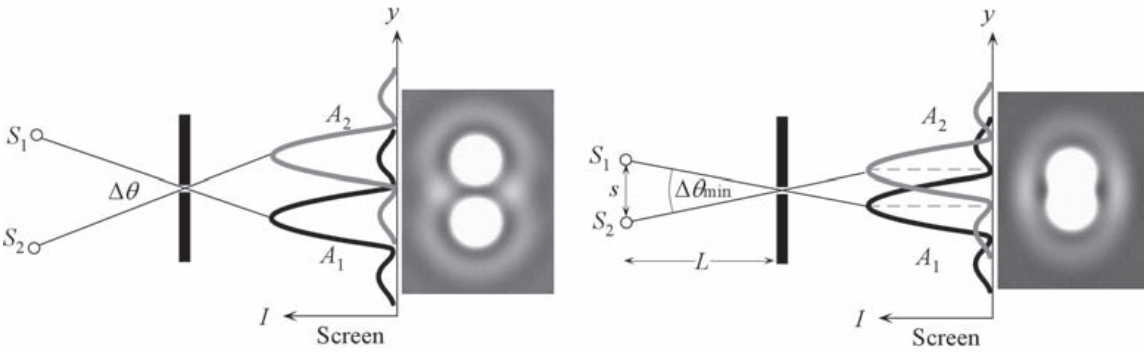


FIGURE 1.38 Resolution of imaging systems is limited by diffraction effects. As points S_1 and S_2 get closer, eventually the Airy patterns overlap so much that the resolution is lost. The Rayleigh criterion allows the minimum angular separation of two of the point sources to be determined. (Schematic illustration inasmuch as the side lobes are actually much smaller than the center peak.)

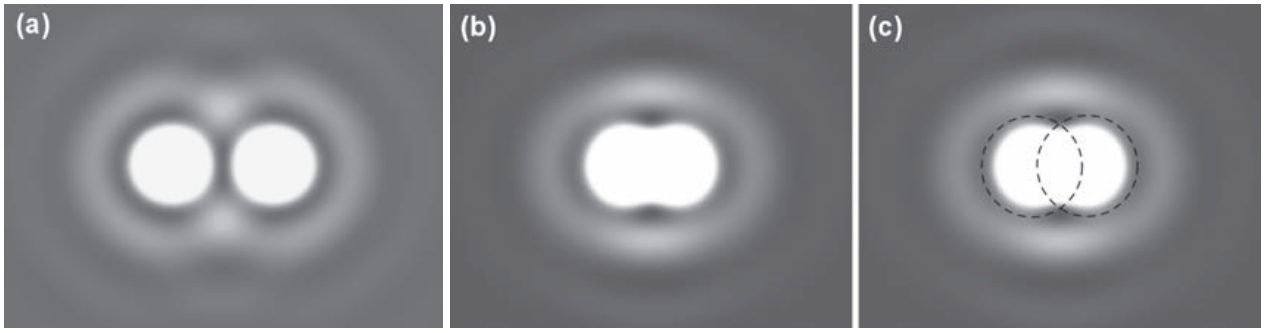


Image of two-point sources captured through a small circular aperture. (a) The two points are fully resolved since the diffraction patterns of the two sources are sufficiently separated. (b) The two images are near the Rayleigh limit of resolution. (c) The first dark ring of one image passes through the center of the bright Airy disk of the other. (Approximate.)

B. Diffraction Grating

A **diffraction grating** in its simplest form is an optical device that has a periodic series of slits in an opaque screen as shown in Figure 1.39 (a). An incident beam of light is diffracted in certain well-defined directions that depend on the wavelength λ and the grating properties. Figure 1.39 (b) shows a typical intensity pattern in the diffracted beam for a finite number of slits. There are "strong beams of diffracted light" along certain directions (θ) and these are labeled according to their occurrence: zero-order (center), first-order, either side of the zero-order, and so on. If there are an infinite number of slits then the diffracted beams have the same intensity. In reality, any periodic variation in the refractive index would serve as a diffraction grating and we will discuss other types later. As in Fraunhofer diffraction we will assume that the observation screen is far away, or that a lens is used to focus the diffracted parallel rays on to the screen (the lens in the observer's eye does it naturally).

We will assume that the incident beam is a plane wave so that the slits become coherent (synchronous) sources. Suppose that the width a of each slit is much smaller than the separation

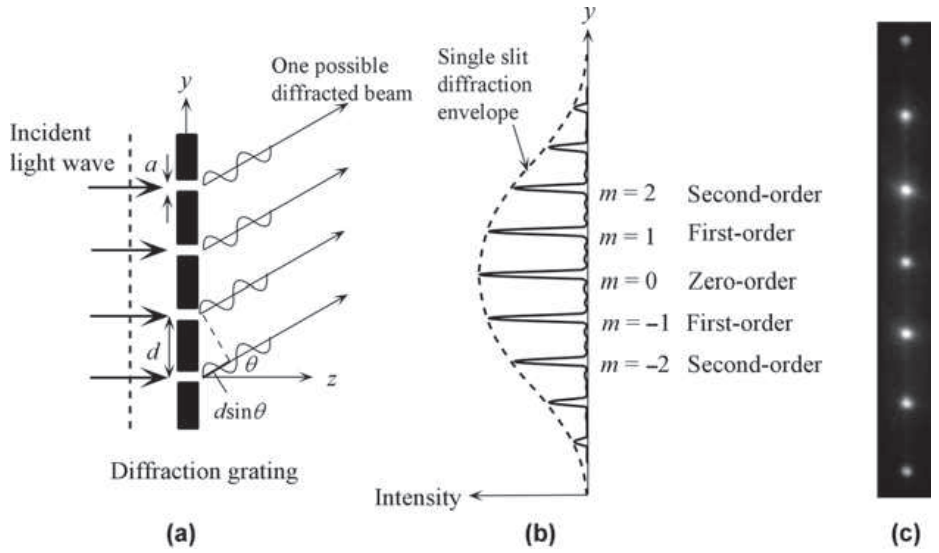


FIGURE 1.39 (a) A diffraction grating with N slits in an opaque screen. Slit periodicity is d and slit width is a; $a \ll d$. (b) The diffracted light pattern. There are distinct, that is diffracted, beams in certain directions (schematic). (c) Diffraction pattern obtained by shining a beam from a red laser pointer onto a diffraction grating. The finite size of the laser beam results in the dot pattern. (The wavelength was 670 nm, red, and the grating has 200 lines per inch.)

d of the slits as shown in Figure 1.39 (a). Waves emanating at an angle θ from two neighboring slits are out of phase by an amount that corresponds to an optical path difference $d\sin\theta$. Obviously, all such waves from pairs of slits will interfere constructively when this is a multiple of the whole wavelength

$$d\sin\theta = m\lambda; \quad m = 0, \pm 1, \pm 2, \ldots \tag{1.12.8}$$

Grating equation

which is the well-know **grating equation**, also known as the **Bragg**[37] **diffraction condition**. The value of m defines the diffraction order; $m = 0$ being zero-order, $m = \pm 1$ being first-order, *etc.* If the grating in Figure 1.39 (a) is in a medium of refractive index n, that is, the incident and diffracted beams are all in the same medium of index n, then we should use λ/n for the wavelength in Eq. (1.12.8), where λ is the free-space wavelength, that is, $d\sin\theta = m\lambda/n$.

The problem of determining the actual intensity of the diffracted beam is more complicated as it involves summing all such waves at the observer and, at the same time, including the diffraction effect of each individual narrow slit. With a smaller than d as in the Figure 1.39 (a), the amplitude of the diffracted beam is modulated by the diffraction amplitude of a single slit since the latter is spread substantially, as illustrated in Figure 1.39 (b). It is apparent that the diffraction grating provides a means of deflecting an incoming light by an amount that depends on its wavelength—the reason for their use in *spectroscopy.*

The diffraction grating in Figure 1.40 (a) is a **transmission grating**. The incident and diffracted beams are on opposite sides of the grating. Typically, parallel thin grooves on a glass plate would serve as a transmission grating as in Figure 1.40 (a). A **reflection grating** has the incident beam and the diffracted beams on the same side of the device as in Figure 1.40 (b). The surface of the device has a periodic reflecting structure, easily formed by etching parallel grooves in a metal film, *etc.* The reflecting unetched surfaces serve as synchronous secondary sources that interfere along certain directions to give diffracted beams of zero-order, first-order, *etc.* Among transmission gratings, it is customary to distinguish between *amplitude gratings* in which the transmission amplitude is modulated, and so-called *phase gratings* where only the refractive index is modulated, without any losses.

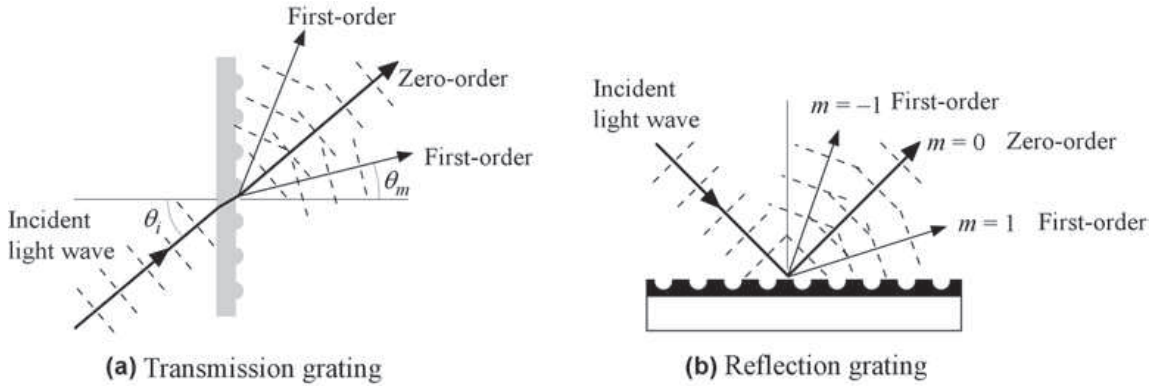


FIGURE 1.40 (a) Ruled periodic parallel scratches on a glass serve as a *transmission grating.* (The glass plate is assumed to be very thin.) (b) A *reflection grating.* An incident light beam results in various "diffracted" beams. The zero-order diffracted beam is the normal reflected beam with an angle of reflection equal to the angle of incidence.

[37] William Lawrence Bragg (1890–1971), Australian-born British physicist, won the Nobel Prize with his father, William Henry Bragg, for his "famous equation" when he was only 25 years old.

When the incident beam is not normal to the diffraction grating, then Eq. (1.12.8) must be modified. If θ_i is the angle of incidence with respect to the normal to the grating, then the diffraction angle θ_m for the m-th mode is given by

Grating equation

$$d(\sin\theta_m - \sin\theta_i) = m\lambda; \quad m = 0, \pm 1, \pm 2, \ldots \tag{1.12.9}$$

The same equation can be used for transmission and reflection gratings provided that we define the angles θ_i and θ_m as positive on either side of the normal as in Figure 1.40 (b).[38] Consider a grating with N slits. The slit width is a (very narrow), and d is the periodicity as before. The detector is at a distance L, far away from the grating. While the periodicity in the slits gives rise to the diffracted beams, the diffraction at each narrow slit defines the envelope of the diffracted intensities as shown in Figure 1.39 (b). If the incident plane wave is normal to the grating, the intensity distribution along y at the screen is given by

Grating diffraction pattern

$$I(y) = I_o\left[\frac{\sin\left(\frac{1}{2}k_y a\right)}{\frac{1}{2}k_y a}\right]^2\left[\frac{\sin\left(\frac{1}{2}Nk_y d\right)}{N\sin\left(\frac{1}{2}k_y d\right)}\right]^2 \tag{1.12.10}$$

where k_y is the scattering wave vector defined by $k_y = (2\pi/\lambda)(y/L) = (2\pi/\lambda)\sin\theta$, and I_o is the maximum intensity along $\theta = 0$. The second term represents the oscillations in the intensity due to interference from different slits. The first term is the envelope of the diffraction pattern, and is the diffraction pattern of a single slit.

The **resolvance** or the **resolving power** R of a diffraction grating is its ability to be able to separate out adjacent wavelengths. If $\lambda_2 - \lambda_1 = \Delta\lambda$ is the minimum wavelength separation that can be measured, as determined by the Rayleigh criterion (the maximum of the intensity distribution at λ_1 is at the first minimum of the intensity distribution at λ_2), and λ is the average wavelength $(1/2)(\lambda_1 + \lambda_2)$ in $\Delta\lambda$, then the **resolving power** is defined by

Resolving power

$$R = \lambda/\Delta\lambda \tag{1.12.11}$$

The separation $\Delta\lambda$ is also called the spectral resolution. If N grooves on a grating are illuminated and the order of diffraction is m, the theoretical resolving power is given simply by $R = mN$. The resolving power is also called the **chromatic resolving power** since it refers to the separation of wavelengths.

Diffraction gratings are widely used in spectroscopic applications because of their ability to provide light deflection that depends on the wavelength. In such applications, the undiffracted light that corresponds to the zero-order beam (Figure 1.40) is clearly not desirable because it wastes a portion of the incoming light intensity. Is it possible to shift this energy to a higher order? Robert William Wood (1910) was able to do so by ruling grooves on glass with a controlled shape as in Figure 1.41 (a) where the surface is angled periodically with a spatial period d. The diffraction condition in Eq. (1.12.9) applies with respect to the normal to the grating plane, whereas the first-order reflection corresponds to reflection from the flat surface, which is at an angle γ. Thus it is possible to "blaze" one of the higher orders (usually $m = 1$) by appropriately choosing γ. Most modern diffraction gratings are of this type. If the angle of incidence is θ_i with respect to the grating normal, then specular reflection occurs at an angle $(\gamma + \theta_i)$ with respect to the face normal and $(\gamma + \theta_i) + \gamma$ with respect to the grating normal. This reflection at $(\gamma + \theta_i) + \gamma$ should occur at diffraction angle θ_m so that

Blazing angle

$$2\gamma = \theta_m - \theta_i \tag{1.12.12}$$

[38] Some books use $d(\sin\theta_m + \sin\theta_i) = m\lambda$ for a transmission grating but the angles become positive on the incidence side and negative on the transmitted side with respect to the normal. It is a matter of sign convention.

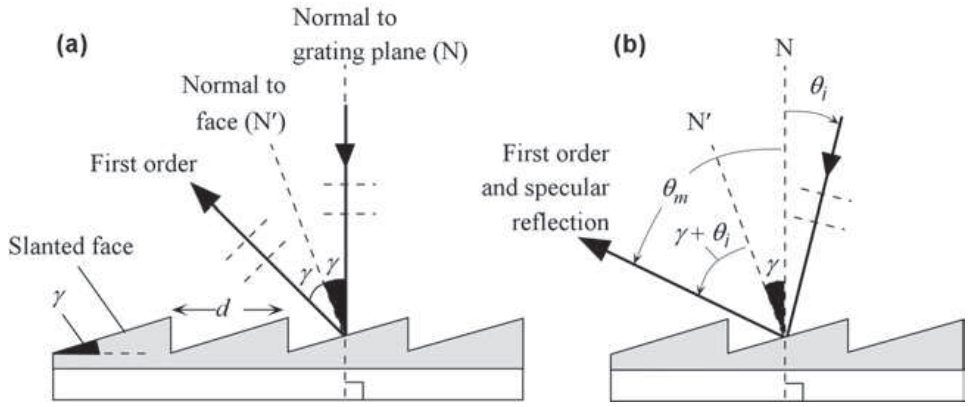


FIGURE 1.41 (a) A blazed grating. Triangular grooves have been cut into the surface with a periodicity d. The side of a triangular groove makes an angle γ to the plane of the diffraction angle. For normal incidence, the angle of diffraction must be 2γ to place the specular reflection on the diffracted beam. (b) When the incident beam is not normal, the specular reflection will coincide with the diffracted beam when $(\gamma + \theta_i) + \gamma = \theta_m$.

William Lawrence Bragg (1890–1971), Australian-born British physicist, won the Nobel Prize with his father, William Henry Bragg, for his "famous equation" when he was only 25 years old. *(SSPL via Getty Images.)*

"The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them."

EXAMPLE 1.12.2 A reflection grating

Consider a reflection grating with a period d that is 10 μm as in Figure 1.42 (a). Find the diffracted beams if a collimated light wave of wavelength 1550 nm is incident on the grating at an angle of 45° to its normal. What should be the blazing angle γ if we were to use a blazed grating with the same periodicity? What happens to the diffracted beams if the periodicity is reduced to 2 μm?

Solution

If we put $m = 0$ in Eq. (1.12.9) we would find the zero-order diffraction, which is at an angle 45°, as expected, and shown in Figure 1.42 (a). The general Bragg diffraction condition is

$$d(\sin\theta_m - \sin\theta_i) = m\lambda$$

so that

$$(10\ \mu\text{m})(\sin\theta_m - \sin(45°) = (+1)(1.55\ \mu\text{m})$$

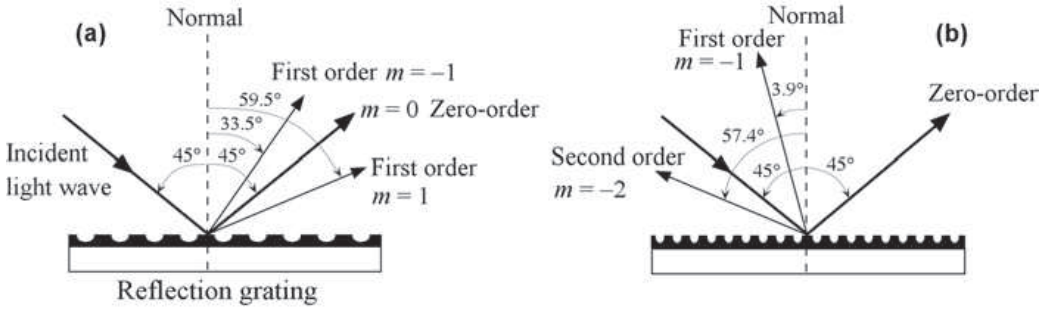


FIGURE 1.42 A light beam is incident at an angle 45° to the normal on a reflection grating. (a) The grating periodicity is 10 μm. (b) The periodicity is 2 μm.

and

$$(10\ \mu\text{m})(\sin\theta_m - \sin(45°) = (-1)(1.55\ \mu\text{m})$$

Solving these two equations, we find $\theta_m = 59.6°$ for $m = 1$, and $\theta_m = 33.5°$ for $m = -1$.

Consider Figure 1.41 (b) in which the secular reflection from the grooved surface coincides with the m_{th} order diffraction when $2\gamma = \theta_m - \theta_i$. Thus

$$\gamma = (1/2)(\theta_m - \theta_i) = (1/2)(59.6° - 45°) = 7.3°$$

Suppose that we reduce d to 2 μm. Recalculating the above we find $\theta_m = -3.9°$ for $m = -1$ and imaginary for $m = +1$. Further, for $m = -2$, there is a second-order diffraction beam at $-57.4°$. Both are shown in Figure 1.42 (b). It is left as an exercise to show that if we increase the angle of incidence, for example, $\theta_i = 85°$ on the first grating, the diffraction angle for $m = -1$ increases from 33.5° to 57.3° and the other diffraction peak ($m = 1$) disappears.

Additional Topics

1.13 INTERFEROMETERS

An **interferometer** is an optical instrument that uses the wave-interference phenomena to produce interference fringes (*e.g.*, dark and bright bands or rings) which can be used to measure the wavelength of light, surface flatness, or small distances. A nearly monochromatic light wave is split into two coherent waves traveling two different paths, and then the two waves are brought together and made to interfere on a screen (or a detector array); the result is an interference pattern as in the Young's fringes in Figure 1.31. In some interferometers, the intensity of the resultant interference is measured at one location, and this intensity is monitored due to changes in one of the optical path lengths. Even a small change in the optical path, distance × refractive index, nL, can cause a measurable shift in the diffraction pattern or a displacement in the fringes, which can be used to infer on nL. There are many types of interferometers.

The **Fabry–Perot interferometer** is a Fabry–Perot cavity–based interferometer that produces an interference ring pattern (bright and dark rings) when illuminated from a broad monochromatic light source as illustrated in Figure 1.43. The resonator consists of two parallel flat glass plates facing each other and separated by an adjustable spacer. A piezoelectric transducer can provide small changes in the spacing between the plates. The inside surface of the glass plates are coated to enhance reflections within the cavity. (Dielectric mirrors can also be used on the inner surfaces.)

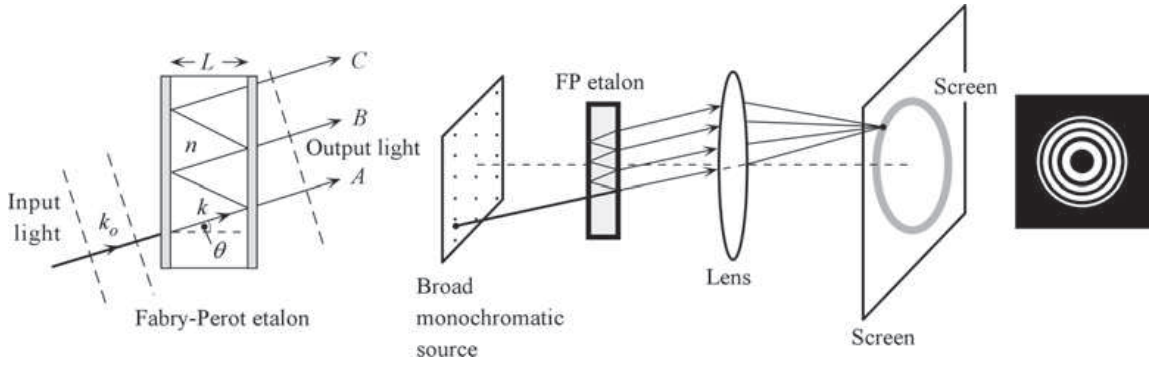


FIGURE 1.43 Fabry–Perot interferometer.

The light entering the cavity will experience multiple reflections inside the cavity. The transmitted waves *A*, *B*, *C*, *etc.*, are focused by a lens onto a point *P* on a screen. If *A*, *B*, *C*, *etc.*, are in phase then the point *P* will be a bright spot. Suppose that *L* is the separation of the plates, θ is the angle at which refracted rays inside the cavity are propagating with respect to the normal, $k = 2\pi n/\lambda$ is the propagation constant in the cavity (λ is the free-space wavelength), and *n* is the refractive index of the medium inside the cavity. We can then show that the phase difference between A and B is $2kL\cos\theta$, which must be $2m\pi$, where *m* is an integer, for constructive interference. Recall that for normal incidence this was $2kL$. Thus, *P* is bright if

$$2nL\cos\theta = m\lambda; \quad m = 0, \pm 1, \ldots \tag{1.13.1}$$

All such points with the same θ lie on a circle about the etalon axis. It is apparent that the interference pattern consists of dark and bright rings. The interference ring diameter depends on the wavelength and the optical separation *nL* (refractive index × distance) of the plates of the etalon. The interferometer can be used in spectroscopic applications such as the measurement of the source wavelength.

In Figure 1.44, showing a **Mach–Zehnder interferometer**, a coherent light beam is split by a beam splitter into two paths at *O*. These beams are then reflected by two mirrors M_1 and M_2 onto a second beam splitter where they are combined at *C* to give rise to an output light beam onto a detector *D*. The resultant field at *C* depends on the phase difference between the waves traveling the paths *OAC* and *OBC*. We can easily monitor the changes in *n* or the length *d* of a sample by placing it into one of the paths, such as in *OAC*. If $k_o = 2\pi/\lambda$ is the propagation constant in vacuum and $k = 2\pi n/\lambda$ is that in the sample, then the phase difference Φ between the weaves arriving at *C* is[39]

$$\Phi = k_o(\overline{OAC} - d) + kd - k_o\overline{OBC}$$

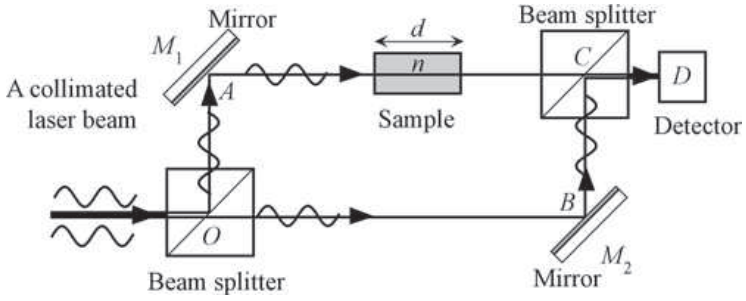


FIGURE 1.44 Mach–Zehnder interferometer.

[39] $\overline{OAC}$ stands for the distance from *O* to *A* to *C*.

The change $\Delta\Phi$ in the phase angle due changes in the sample would be

$$\Delta\Phi = \Delta(kd) = \frac{2\pi}{\lambda}\Delta(nd)$$

Thus, the change in the detected signal represents the change in *nd.* On the other hand, if we can somehow modulate n or d, we have the means of modulating the intensity at the detector. The *Mach–Zehnder modulators* used in communications rely on changing n in a suitable crystal (such as lithium niobate) by the application of an electric field through the electro-optic effect as discussed later in the book.

1.14 THIN FILM OPTICS: MULTIPLE REFLECTIONS IN THIN FILMS

There is much interest in the reflection of light from, and transmission through, a thin film coating of a transparent material (called an optical coating) on a substrate such as a semiconductor crystal; or even on glass as in antireflection coated lenses. The simplest case is shown in Figure 1.45 where light traveling in a medium of refractive index n_1 is incident on a thin film coating of index n_2 on a substrate of index n_3. There are multiple reflections in the thin film, and the problem is almost identical to the Fabry–Perot optical cavity in which there are similar multiple interferences.

Suppose the thickness of the coating is *d.* For simplicity, we will assume normal incidence. The phase change in traversing the coating thickness *d twice* is $\phi = 2 \times (2\pi/\lambda)n_2 d$ where λ is the free-space wavelength. The wave has to be multiplied by $\exp(-j\phi)$ to account for this phase difference in crossing the thickness twice. The reflection and transmission coefficients for the present n_1-n_2-n_3 system are given by,

$$r_1 = r_{12} = \frac{n_1 - n_2}{n_1 + n_2} = -r_{21}, \quad r_2 = r_{23} = \frac{n_2 - n_3}{n_2 + n_3}, \tag{1.14.1a}$$

and

$$t_1 = t_{12} = \frac{2n_1}{n_1 + n_2}, \quad t_1' = t_{21} = \frac{2n_2}{n_1 + n_2}, \quad t_2 = t_{23} = \frac{2n_2}{n_2 + n_3} \tag{1.14.1b}$$

where

$$1 - t_1 t_1' = r_1^2 \tag{1.14.1c}$$

The first reflected beam, $A_1 = A_0 \times r$, the second is $A_2 = A_o \times t_1 \times t_1' \times r_2 \times e^{-j\phi}$, and so on.

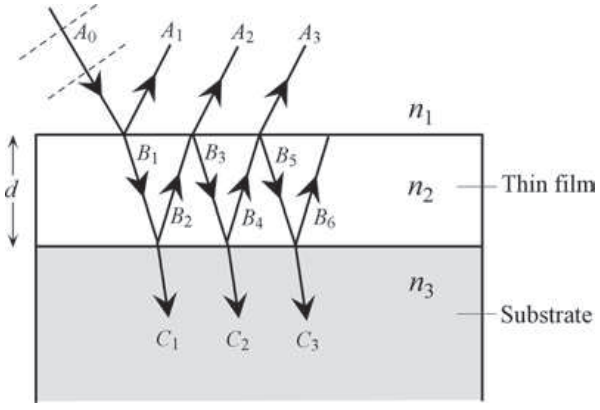


FIGURE 1.45 Light traveling in a medium of refractive index n_1 is incident on a thin film coating of index n_2 on a substrate of index n_3.

The amplitude of the reflected beam is

$$A_{\text{reflected}} = A_1 + A_2 + A_3 + A_4 + \ldots$$

that is,

$$A_{\text{reflected}}/A_0 = r_1 + t_1 t_1' r_2 e^{-j\phi} - t_1 t_1' r_1 r_2^2 e^{-j2\phi} + t_1 t_1' r_1^2 r_2^3 e^{-j3\phi} + \ldots \qquad \textbf{(1.14.2)}$$

which is a geometric series. Using Eq. (1.14.1c), Eq. (1.14.2) can be conveniently summed to obtain the overall reflection coefficient r

$$r = \frac{r_1 + r_2 e^{-j\phi}}{1 + r_1 r_2 e^{-j\phi}} \qquad \textbf{(1.14.3)}$$

Thin film reflection coefficient

Similarly, we can sum for the amplitude of the transmitted beam as

$$C_{\text{transmitted}} = C_1 + C_2 + C_3 + \ldots$$

that is,

$$C_{\text{transmitted}}/A_0 = t_1 t_2 e^{-j\phi/2} - t_1 t_2 r_1 r_2 e^{-j3\phi/2} + t_1 t_2 r_1^2 r_2^2 e^{-j5\phi/2} + \ldots \qquad \textbf{(1.14.4)}$$

which is a geometric series that sums to

$$t = \frac{t_1 t_2 e^{-j\phi/2}}{1 + r_1 r_2 e^{-j\phi}} \qquad \textbf{(1.14.5)}$$

Thin film transmission coefficient

Equations (1.14.3) and (1.14.5) describe the reflected and transmitted waves. The reflectance and transmittance are then

$$R = |r|^2 \quad T = (n_3/n_1)|t|^2 \qquad \textbf{(1.14.6)}$$

Thin film reflectance and transmittance

Figure 1.46 (a) shows R and T as a function of ϕ for $n_1 < n_2 < n_3$. Clearly, R is minimum, and T is maximum, whenever $\phi = \pi \times$ (odd number) or $\phi = 2(2\pi/\lambda)n_2 d = \pi(2m + 1)$, where $m = 0, 1, 2, \ldots$ The latter leads to

$$d = \frac{\lambda}{4n_2}(2m + 1); \quad m = 0, 1, 2, \ldots \qquad \textbf{(1.14.7)}$$

Thickness for minimum reflection

which is the thickness required to minimize the reflection, and maximize the transmission of light when n_2 is intermediate between n_1 and n_3. The oscillations in R and T with ϕ (*e.g.*, as the wavelength is scanned) are sometime referred to as an **interference fringes** in wavelength.

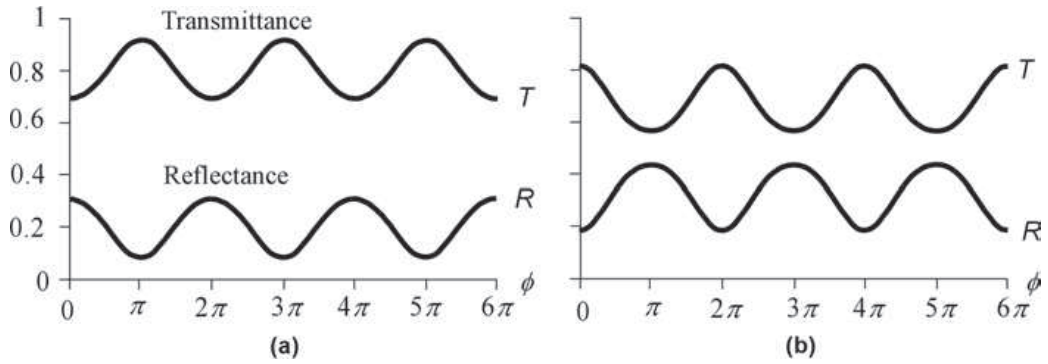


FIGURE 1.46 (a) Reflectance R and transmittance T vs. $\phi = 2n_2 d/\lambda$, for a thin film on a substrate where $n_1 = 1$ (air), $n_2 = 2.5$, $n_3 = 3.5$, and $n_1 < n_2 < n_3$. (b) R and T vs. ϕ for a thin film on a substrate where $n_1 = 1$ (air), $n_2 = 3.5$, $n_3 = 2.5$, and $n_2 > n_3 > n_1$.

Figure 1.46 (b) represents the reflectance and transmittance vs. ϕ of light through a thin layer of high index material on a low index substrate where $n_1 < n_3 < n_2$; for example, a semiconductor film on a glass substrate. Notice the difference between the two cases, especially the locations of the maxima and minima. (Why is there a difference?)

The phase change ϕ, of course, depends on three factors, d, n_2, and λ. The minimum and maximum reflectances in Figure 1.46 at those particular ϕ values can be found by using Eqs. (1.14.3) and (1.14.7). For $n_1 < n_2 < n_3$ as in Figure 1.46 (a)

Minimum and maximum reflectance

$$R_{\min} = \left(\frac{n_2^2 - n_1 n_3}{n_2^2 + n_1 n_3}\right)^2; \quad R_{\max} = \left(\frac{n_3 - n_1}{n_3 + n_1}\right)^2 \tag{1.14.8}$$

and the transmittance can be found from $R + T = 1$. When $n_1 < n_3 < n_2$ then $R_{\min}$ and $R_{\max}$ equations are interchanged. While $R_{\max}$ appears to be independent from n_2, the index n_2 is nonetheless still involved in determining maximum reflection inasmuch as R reaches $R_{\max}$ when $\phi = 2(2\pi/\lambda)n_2 d = \pi(2m)$; when $\phi = \pi \times$ (even number).

In transmission spectra measurements, a spectrophotometer is used to record the transmittance of a light beam as a function of wavelength through a sample. If the sample is a thin film on a substrate, there will be multiple reflections and interferences in the thin film, and the measured transmittance will exhibit maxima and minima as in Figure 1.46 as the wavelength (or ϕ) is scanned. The locations of the maxima and minima, with the knowledge of the substrate index (n_3), can be used to find d and n.

EXAMPLE 1.14.1 Thin film optics

Consider a semiconductor device with $n_3 = 3.5$ that has been coated with a transparent optical film (a dielectric film) with $n_2 = 2.5$, $n_1 = 1$ (air). If the film thickness is 160 nm, find the minimum and maximum reflectances and transmittances and their corresponding wavelengths in the visible range. (Assume normal incidence.)

Solution

This case corresponds to Figure 1.46 (a). Minimum reflectance $R_{\min}$ occurs at $\phi = \pi$ or odd multiple of π, and maximum reflectance $R_{\max}$ at $\phi = 2\pi$ or an integer multiple of 2π. From Eq. (1.14.8) we have

$$R_{\min} = \left(\frac{n_2^2 - n_1 n_3}{n_2^2 + n_1 n_3}\right)^2 = \left(\frac{2.5^2 - (1)(3.5)}{2.5^2 + (1)(3.5)}\right)^2 = 0.080 \quad \text{or} \quad 8.0\%$$

and

$$R_{\max} = \left(\frac{n_3 - n_1}{n_3 + n_1}\right)^2 = \left(\frac{3.5 - 1}{3.5 + 1}\right)^2 = 0.31 \quad \text{or} \quad 31\%$$

Corresponding transmittances are,

$$T_{\max} = 1 - R_{\min} = 0.92 \quad \text{or} \quad 92\%$$

and

$$T_{\min} = 1 - R_{\max} = 0.69 \quad \text{or} \quad 69\%$$

Without the thin film coating, the reflectance would be 31%, the maximum reflectance.

Given $\phi = 2dn_2(2\pi/\lambda)$, and $d = 160$ nm, and the fact that the minimum reflectance corresponds to $\phi =$ odd integer $\times\ \pi$, we choose $\phi = 3\pi$ (by trial and error). The wavelength λ_{min} is then given by

$$\lambda_{min} = 4\pi dn_2/\phi = 4\pi(160\text{ nm})(2.5)/(3\pi) = 533\text{ nm (green)}$$

which is in the visible. The maximum reflectance in the visible occurs when $\phi = 4\pi$, giving

$$\lambda_{max} = 4\pi dn_2/\phi = 4\pi(160\text{ nm})(2.5)/(4\pi) = 400\text{ nm (violet)}$$

1.15 MULTIPLE REFLECTIONS IN PLATES AND INCOHERENT WAVES

The interference of light waves in a thin film takes place because the waves have much longer coherence lengths than the thickness of the film so that the waves exhibit mutual coherence. We can add the electric field and interference leads to bright and dark fringes. A film would be considered too thick if it does not exhibit any interference phenomena due to the coherence length of the waves being shorter than the thickness. Such cases easily arise when we pass light through a transparent (or partially transparent) plate or when the light source is incoherent. In such cases, we cannot add the electric field to find the reflected and transmitted light irradiances. The overall reflectance and transmittance would be independent of the round-trip phase change ϕ inside the plate and would not exhibit the behavior in Figure 1.46. We have to use Eq. (1.10.5).

Consider a light beam of unit intensity that is passed through a thick plate of transparent material of index n_2 in a medium of index n_1 as in Figure 1.47. The first transmitted light intensity into the plate is $(1 - R)$, and the first transmitted light out is $(1 - R) \times (1 - R)$ or $(1 - R)^2$. However, there are internal reflections as shown, so that the second transmitted light is $(1 - R) \times R \times R \times (1 - R) = R^2(1 - R)^2$, so that the transmitted intensity through the plate is

$$T_{plate} = (1 - R)^2 + R^2(1 - R)^2 + R^4(1 - R)^2 + \ldots = (1 - R)^2\left[1 + R^2 + R^4 + \ldots\right]$$

or

$$T_{plate} = \frac{(1 - R)^2}{1 - R^2} \qquad \textbf{(1.15.1)}$$

Transmittance for a thick plate or incoherent light

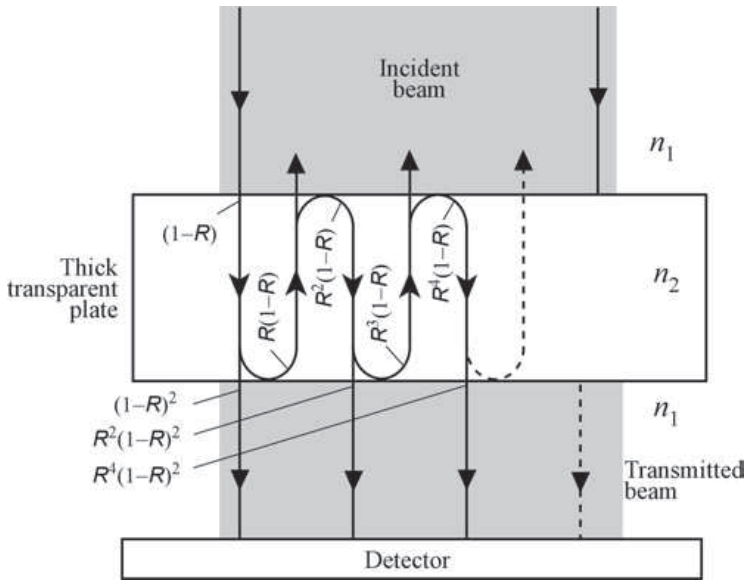


FIGURE 1.47 Transmitted and reflected light through a slab of material in which there is no interference.

Transmittance for a thick plate for incoherent light

By substituting for R in terms of the indices, we can write this as

$$T_{\text{plate}} = \frac{2n_1n_2}{n_1^2 + n_2^2} \tag{1.15.2}$$

For example, for a glass plate of $n_2 = 1.60$ in air ($n_1 = 1$), $T_{\text{plate}} = 89.9\%$ while the simple transmittance through an n_1–n_2 interface would give 94.7%. The overall reflectance is $1 - T_{\text{plate}}$ so that

Transmittance for a thick plate or incoherent beam of light

$$R_{\text{plate}} = \frac{(n_1 - n_2)^2}{n_1^2 + n_2^2} \tag{1.15.3}$$

One of the simplest ways to determine the refractive index of a plate is to measure the transmittance T_{plate} in Eq. (1.15.1), from which we can calculate n_2.

1.16 SCATTERING OF LIGHT

When a light beam propagates in a medium in which there are small particles or inhomogeneities, such as local changes in the refractive index of the medium, some of the power in the beam is radiated away from the direction of propagation, that is some of the power becomes scattered. **Scattering** is a process by which some of the power in a propagating electromagnetic wave is redirected as secondary EM waves in various directions away from the original direction of propagation as illustrated in Figure 1.48 (a). There are a number of scattering processes, which are usually classified in terms of the size of the scattering particles in relation to the wavelength of light that is scattered. In **Rayleigh scattering**, the scattering particle size, or the scale of inhomogeneities in a medium, is much smaller than the wavelength of light. The intensity of the scattered light at an angle θ to the original beam depends on the scattering process; the Rayleigh scattering case is shown in Figure 1.48 (b) in which the scattering is not spherically symmetric.

Consider what happens when a propagating wave encounters a molecule, an impurity in a crystal or a small dielectric particle (or region), which is smaller than the wavelength of light. The electric field in the wave polarizes the particle by displacing the lighter electrons with respect to the heavier positive nuclei. The electrons in the molecule couple and oscillate with the electric field in the wave (ac electronic polarization). The oscillation of charge "up" and "down," or the oscillation of the induced dipole, radiates EM waves all around the molecule as illustrated in Figure 1.48 (a). We should remember that an oscillating charge is like an alternating current which always radiates EM waves (like an antenna). The net effect is that the incident

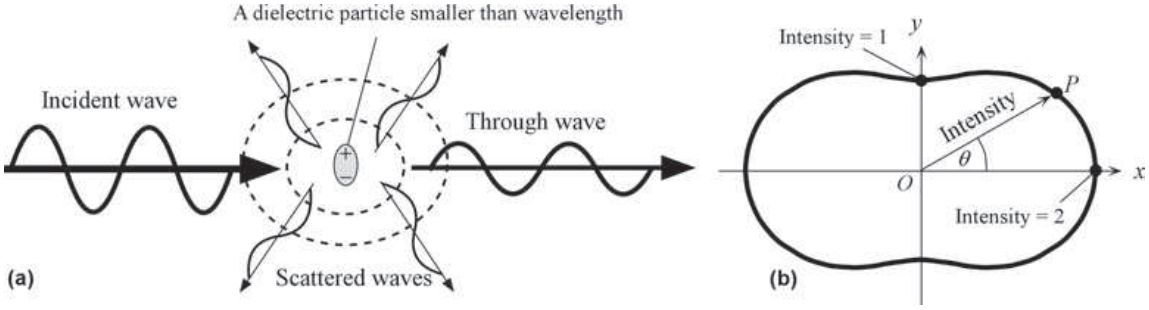


FIGURE 1.48 (a) Rayleigh scattering involves the polarization of a small dielectric particle or a region that is much smaller than the light wavelength. The field forces dipole oscillations in the particle (by polarizing it) which leads to the emission of EM waves in "many" directions so that a portion of the light energy is directed away from the incident beam. (b) A polar plot of the dependence of the intensity of the scattered light on the angular direction θ with respect to the direction of propagation x in Rayleigh scattering (in a polar plot, the radial distance OP is the intensity).

Lord Rayleigh (John William Strutt) was an English physicist (1877–1919) and a Nobel Laureate (1904) who made a number of contributions to wave physics of sound and optics. He formulated the theory of scattering of light by small particles and the dependence of scattering on $1/\lambda^4$ circa 1871. Then, in a paper in 1899 he provided a clear explanation on why the sky is blue. Ludvig Lorentz, around the same time, and independently, also formulated the scattering of waves from a small dielectric particle, though it was published in Danish (1890).[40] *(© Mary Evans Picture Library/Alamy.)*

wave becomes partially reradiated in different directions and hence loses intensity in its original direction of propagation. (We may think of the process as the particle absorbing some of the energy via electronic polarization and reradiating it in different directions.) It may be thought that the scattered waves constitute a spherical wave emanating from the scattering molecule, but this is not generally the case as the re-emitted radiation depends on the shape and polarizability of the molecule in different directions. We assumed a small particle so that at any time the field has no spatial variation through the particle, whose polarization then oscillates with the electric field oscillation. Whenever the size of the scattering region, whether an inhomogeneity, a small particle, a molecule, or a defect in a crystal, is much smaller than the wavelength λ of the incident wave, the scattering process is generally termed **Rayleigh scattering.** Typically, the particle size is smaller than one-tenth of the light wavelength.

Rayleigh scattering of light propagating in a glass medium is of particular interest in photonics because it results in the attenuation of the transmitted light pulses in optical fibers. The glass structure is such that there are small random spatial variations in the refractive index about some average value. There are therefore local fluctuations in the relative permittivity and polarizability, which effectively act if there are small inhomogeneities in the medium. These dielectric inhomogeneities arise from fluctuations in the relative permittivity that is part of the intrinsic glass structure. As the fiber is drawn by freezing a liquid-like flow, random thermodynamic fluctuations in the composition and structure that occur in the liquid state become frozen into the solid structure. Consequently, the glass fiber has small fluctuations in the relative permittivity which leads to Rayleigh scattering. A small inhomogeneous region acts like a small dielectric particle and scatters the propagating wave in different directions. Nothing can be done to eliminate Rayleigh scattering in glasses as it is part of their intrinsic structure.

It is apparent that the scattering process involves electronic polarization of the molecule or the dielectric particle. We know that this process couples most of the energy at ultraviolet frequencies where the dielectric loss due to electronic polarization is maximum and the loss is due to EM wave radiation. Therefore, as the frequency of light increases, the scattering becomes more severe. In other words, *scattering decreases with increasing wavelength.* The intensity of the scattered radiation is proportional to $1/\lambda^4$. For example, blue light, which has a shorter wavelength than red light, is scattered more strongly by air molecules. When we look at the sun directly, it appears yellow because the blue light has been scattered in the direct light more than

[40] Pedro Lilienfeld's "A Blue Sky History", in *Optics and Photonics News*, *15*(6), 32, 2004, is highly recommended; it also provides the original references.

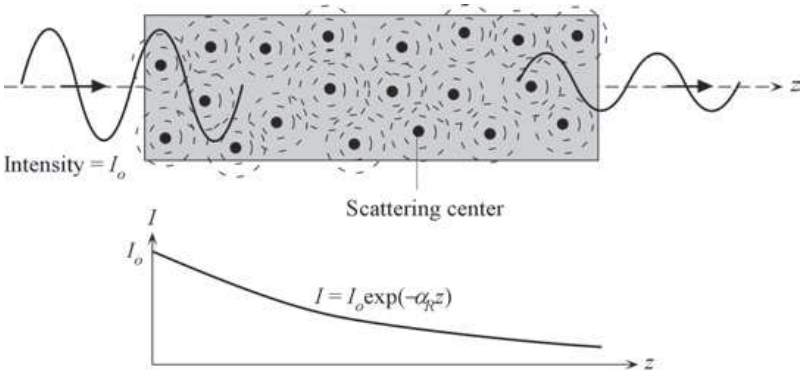


FIGURE 1.49 When a light beam propagates through a medium in which there are small particles, it becomes scattered as it propagates and loses power in the direction of propagation. The light becomes attenuated.

the red light. When we look at the sky in any direction but the sun, our eyes receive scattered light which appears blue; hence the sky is blue. At sunrise and sunset, the rays from the sun have to traverse the longest distance through the atmosphere and have the most blue light scattered, which gives the sun its red color at these times.

The intensity of a light beam in a medium with small particles (or inhomogeneities) decreases as the beam propagates due to Rayleigh scattering from these small particles as illustrated in Figure 1.49. The intensity at a position z inside the medium from the radiation receiving face is given by

Attenuation by scattering

$$I = I_o \exp(-\alpha_R z) \tag{1.16.1}$$

where α_R is the attenuation coefficient due to Rayleigh scattering. α_R depends on the concentration of scattering particles N, their radius a, wavelength λ and the mismatch between the refractive index n of the scattering spheres and the index n_o of the medium

Rayleigh attenuation coefficient

$$\alpha_R \propto N \cdot a^6 \cdot \frac{1}{\lambda^4} \cdot \left(\frac{n^2 - n_o^2}{n^2 + n_o^2} \right)^2 \tag{1.16.2}$$

Equation (1.16.2) has been written in terms of products to identify those factors that affect α_R.

Mie scattering refers to the scattering of light from scatterers that have dimensions comparable with, or greater than, the wavelength of light. For example, Mie scattering would occur for light scattering from long organic molecules in a solution, or scattering from various particulate pollutants (as in smog) in the atmosphere, including dust particles. The scatterers are assumed to have a refractive index significantly different from that of the surrounding medium. The scattering depends on the ratio of the scattering particle diameter to the wavelength of light, and favors scattering in the forward direction. The dependence on the wavelength is weaker than Rayleigh scattering.

The scattering of light from various particles suspended in a liquid is usually referred to as **turbidity**. For example, small solid particles suspended in water will scatter light and would make the water appear cloudy. The quality of water is often monitored for small particles by measuring the turbidity of water.

1.17 PHOTONIC CRYSTALS

The dielectric mirror, or the Bragg reflector, in Figure 1.24 is a stack of alternating layers of high and low refractive index material in a periodic manner. The one-dimensional periodic variation in the refractive index n in Figure 1.24 represents one of the simplest periodic structures within

a general class of optical materials called **photonic crystals**. A **photonic crystal** (PC) is a material that has been structured to possess a periodic modulation of the refractive index n, just like in Figure 1.24, so that the structure influences the propagation and confinement of light within it. The periodicity can be in one- (1D), two- (2D), or three-dimensional (3D); Figures 1.50 (a)–(c) illustrates 1D, 2D, and 3D simple photonic crystals as examples. In fact, quite complicated structures can be constructed that have very interesting optical properties. The dielectric mirror or the Bragg reflector can be viewed as one of the simplest 1D photonic crystals.[41] For the structure to influence the propagation of the EM wave, it has to diffract the wave, which means that the scale of periodic variations must be on the wavelength scale. As apparent from Figure 1.24, the 1D PC has a band of frequencies over which it reflects the light and, conversely, there is a **stop band** over which no transmission is possible through the dielectric stack. There is a band of frequencies that represent waves that are not allowed to go through this periodic structure in the direction of refractive index variation, along z in Figure 1.50; this band is called an **optical** or **photonic bandgap**.

The periodic variation in n in Figure 1.50 is normally assumed to extend indefinitely, whereas in practice, the PCs have a finite size, for example, a certain number of layers in the dielectric mirror, not infinite. As in normal crystals, the periodic structures in Figure 1.50 have a **unit cell**, which repeats itself to generate the whole lattice—that is, the whole crystal structure. For the 1D PC in Figure 1.50 (a), for example, two adjacent layers, n_1n_2, form the unit cell. We can move this unit cell along z by a distance Λ, the **period** (or **periodicity**), many times to generate the whole 1D photonic crystal.

The periodicity of a photonic crystal implies that any property at a location z will be the same at $z \pm \Lambda$, $z \pm 2\Lambda$ and so on; that is, there is *translational symmetry* along z (in 1D).

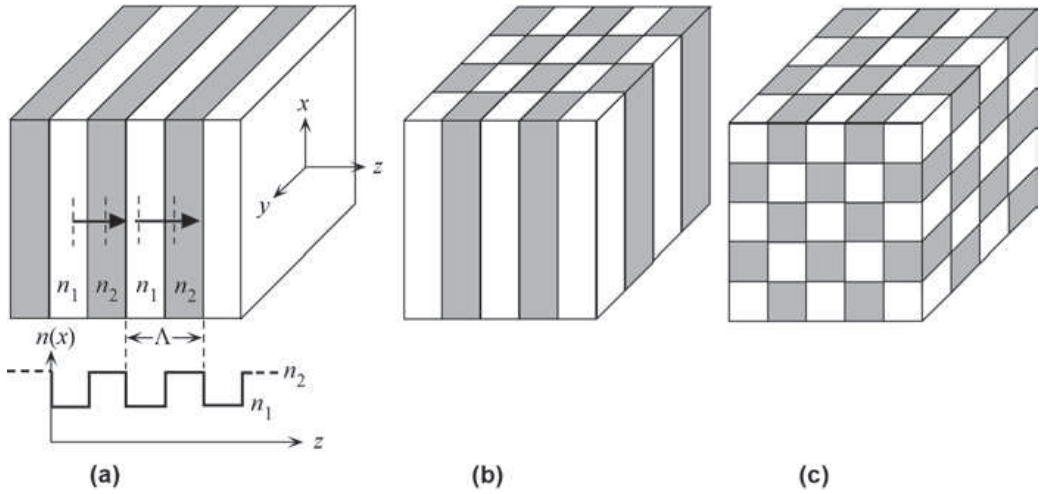


FIGURE 1.50 Photonic crystals in (a) 1D, (b) 2D, and (c) 3D, D being the dimension. Gray and white regions have different refractive indices and may not necessarily be the same size. Λ is the periodicity. The 1D photonic crystal in (a) is the well-known Bragg reflector, a dielectric stack.

[41] The propagation of light through such a one-dimensional (1D) periodic variation of n has been well-known, dating back to the early work of Lord Rayleigh in 1887. Eli Yablonovitch has suggested that the name "photonic crystal" should really only apply to 2D and 3D periodic structures with a large dielectric (refractive index) difference. (E. Yablonovitch, "Photonic crystals: What's in a name?," *Opt. Photon. News*, *18*, 12, 2007.) Nonetheless, the dielectric mirror in Figure 1.50 (a) is often considered as the simplest 1D photonic crystal to derive the concept of a "photonic bandgap," essentially a stop band.

The EM waves that are allowed to propagate along z through the periodic structure are called the **modes** of the photonic crystal. They have a special waveform that must bear the periodicity of the structure, and are called **Bloch waves**. Such a wave for the field E_x, for example, has the form $E_x(z, t) = A(z)\exp(-jkz)$, which represents a traveling wave along z and $A(z)$ is an amplitude function that has the periodicity of the structure, that is, it is periodic along z with a period Λ. $A(z)$ depends on the periodic refractive index function $n(z)$. As we will see in Chapter 3, the electron motion in a semiconductor crystal is also described by Bloch waves (electron wavefunctions).

As in the case of the dielectric mirror, the 1D PC has a band of frequencies over which there can be no propagation along z. In a homogeneous medium of refractive index n, the relationship between the frequency ω and the propagation constant k is simple, that is, $c/n = \omega/k$, where k is the propagation constant inside the medium The dispersion behavior of the medium, that is, ω vs. k, is a straight line with a slope c/n. The dispersion characteristic of a 1D PC for waves along z in Figure 1.50 (a) is shown in Figure 1.51 (a). We notice several important characteristics. At low frequencies (long wavelengths), the waves propagate as if they are in a homogenous medium with a constant phase velocity (dashed straight line). As expected, there is a band of frequencies $\Delta\omega = \omega_2 - \omega_1$ (between S_1 and S_2) over which no propagation along z is allowed, which is a **photonic bandgap** along z. The whole ω vs. k curves are periodic in k with a period $2\pi/\Lambda$. The point P is equivalent to P' because $k = k' + (2\pi/\Lambda)$. We only need to consider k-vales from $-\pi/L$ to π/L. This region is called the **first Brillouin zone**.

At low frequencies or long wavelengths (small k values) in Figure 1.51 (a), the wavelength is so much longer than the variations in n that the propagating wave experiences essentially some average refractive index, n_{av}, that is $(n_1 + n_2)/2$ if the n_1 and n_2 layers have the same thickness, and propagates through the structure as if the structure was a homogenous medium with some effective refractive index, n_{av}. Its phase velocity is ω/k, which is c/n_{av}, and its group velocity, the slope of the ω vs. k characteristic, is the same as c/n_{av}, in this region (ignoring the wavelength dependence of n_1 and n_2). As the wavelength decreases (ω increases), partial reflections of the waves from the boundaries become important and interfere with propagation. At sufficiently small wavelengths, diffraction becomes important as all these partially reflected waves interfere with each other and give rise to significant reflection. Eventually, a critical wavelength (corresponding to ω_1) is reached where the waves become fully diffracted or reflected backwards. A backward traveling (in $-z$ direction) wave experiences the same reflection. These forward and backward diffracted waves give rise to a standing wave in the structure; indeed, only the latter can exist and waves cannot propagate, that is, travel freely.

Eli Yablonovitch at the University of California at Berkeley, and Sajeev John (shown later in this chapter) at the University of Toronto, carried out the initial pioneering work on photonic crystals. Eli Yablonovitch has suggested that the name "photonic crystal" should apply to 2D and 3D periodic structures with a large dielectric (refractive index) difference. (E. Yablonovitch, "Photonic crystals: What's in a name?," *Opt. Photon. News*, *18*, 12, 2007.) Their original papers were published in the same volume of *Physical Review Letters* in 1987. According to Eli Yablonovitch, "Photonic Crystals are semiconductors for light." *(Courtesy of Eli Yablonovitch.)*

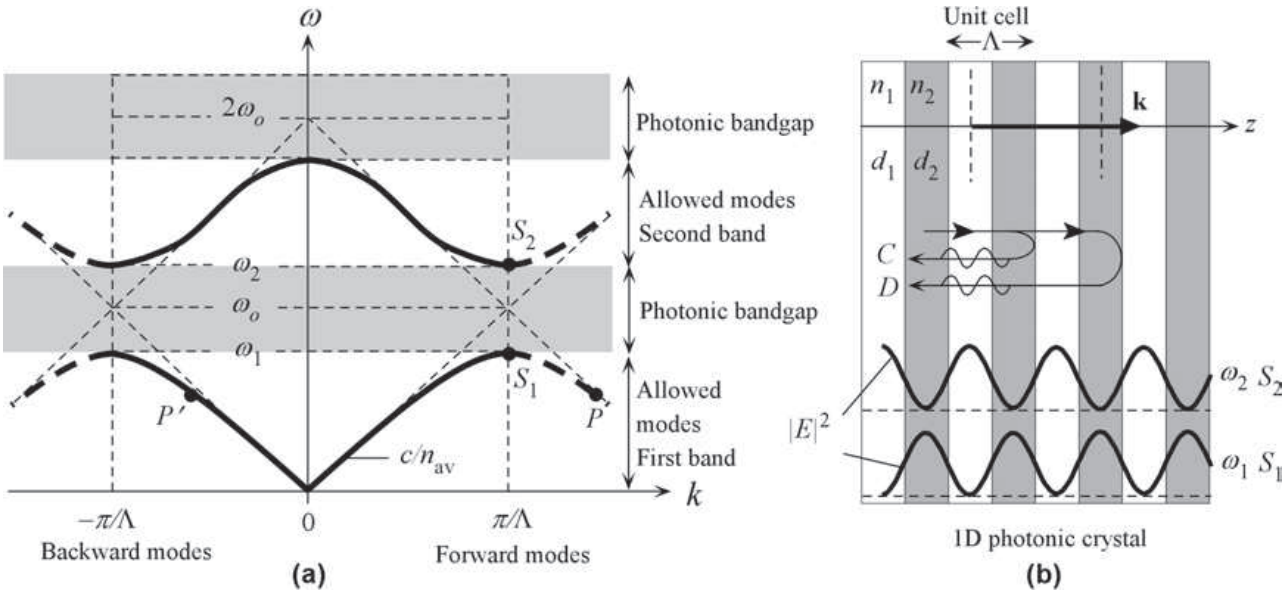


FIGURE 1.51 (a) Dispersion relation, ω vs. k, for waves in a 1D PC along the z-axis. There are allowed modes and forbidden modes. Forbidden modes occur in a band of frequencies called a photonic bandgap. (b) The 1D photonic crystal corresponding to (a), and the corresponding points S_1 and S_2 with their stationary wave profiles at ω_1 and ω_2.

Consider what happens when reflections such as C and D from two successive unit cells interfere constructively and give rise to a backward diffracted (reflected) wave as illustrated in Figure 1.51 (b). Reflections from two successive unit cells must be in phase for full reflection. This means that the phase difference between C and D $(2\pi/\lambda)(2n_1d_1 + 2n_2d_2)$ must be $2m\pi$, where $m = 1, 2, \ldots$ is an integer. If we define $n_{\text{av}} = (n_1d_1 + n_2d_2)/\Lambda$, and $k = n_{\text{av}}(2\pi/\lambda)$ then reflection occurs when $2k\Lambda = 2m\pi$, that is, $k = m\pi/\Lambda$. These are the wave vectors or propagation constants of the waves that cannot be propagated. The waves suffer Bragg reflection in 1D. What happens to these diffracted waves?

The diffracted waves in $+z$ and $-z$ directions set up a standing wave in the structure. If we write the two reflected waves in opposite directions as $A\exp(jkz)$ and $A\exp(-jkz)$ they would add as $A\exp(jkz) \pm A\exp(-jkz)$, which shows that there two possibilities S_1 and S_2. One of them, S_1, has most of its energy inside the n_2, high refractive index layers, and hence has a lower frequency ω_1. S_2 has most of its energy in the n_1 layers and has a higher frequency ω_2; these are shown in Figure 1.51 (b). There are no waves in the $\omega_2 - \omega_1$ interval, which we know as the **photonic bandgap** (PBG). The bandgap $\Delta\omega = \omega_2 - \omega_1$ increases linearly with the index difference $\Delta n = n_2 - n_1$.

It is clear that the 1D crystal exhibits a photonic bandgap for light propagation along z. It should also be apparent that there would be no PBGs for propagation along the x and y direction along which there are no periodic n-variations. Since we can resolve any **k**-vector along x, y, and z directions, overall there is no *net* PBG for light propagation in a 1D PC. The PBG for a 1D crystal is called a **pseudo PBG**.

The above ideas can be readily extended to 2D and 3D periodic structures. Again, in principle, there is no full PBG in the 2D PC. In the case of 3D periodic structures, there would be photonic bandgaps along x, y, and z directions, and for difference polarizations of the electric field. If the refractive index contrast and the periodicity in the 3D structure are such that these photonic bandgaps overlap in *all directions* and for *all polarizations* of light, as schematically depicted in Figure 1.52, the overlap frequency range $\Delta\omega$ becomes a **full photonic bandgap** in all directions for all polarizations of light—no light can propagate through this structure over this

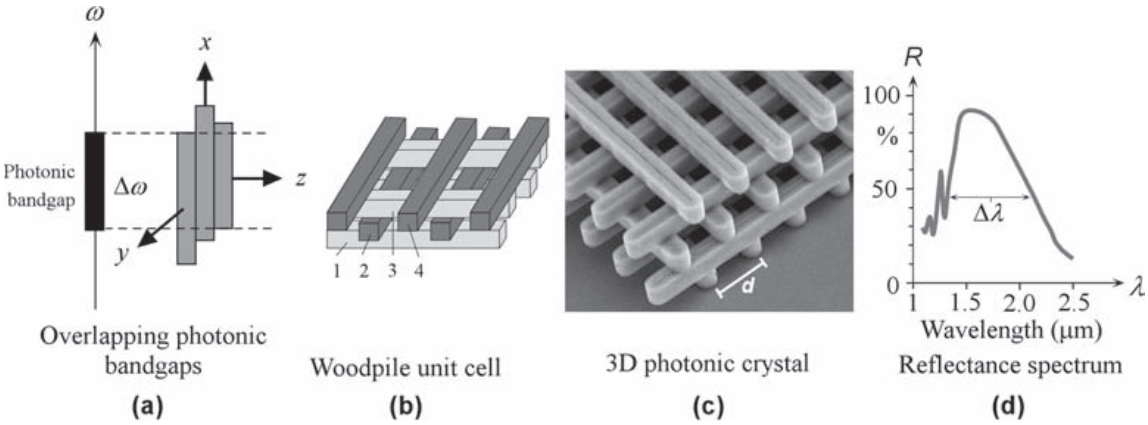
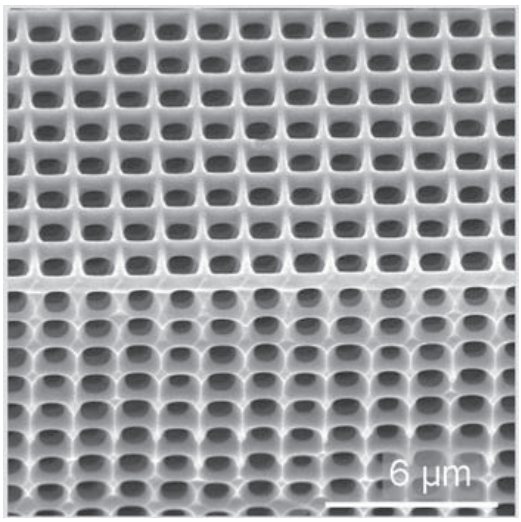
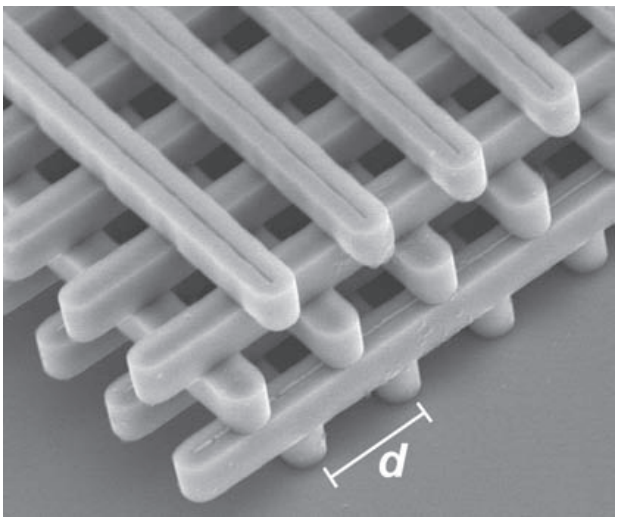


FIGURE 1.52 (a) The photonic bandgaps along *x, y,* and *z* overlap for all polarizations of the field, which results in a full photonic bandgap $\Delta\omega$ (an intuitive illustration). (b) The unit cell of a *woodpile* photonic crystal. There are 4 layers, labeled 1–4 in the figure, with each later having parallel "rods." The layers are at right angles to each other. Notice that layer 3 is shifted with respect to 1, and 4 with respect to 2. (c) An SEM image of a woodpile photonic crystal based on polycrystalline Si; the rod-to-rod pitch *d* is on the micron scale. (*Courtesy of Sandia National Laboratories.*) (d) The optical reflectance of a woodpile photonic crystal showing a photonic bandgap between 1.5 μm and 2 μm. The photonic crystal is similar to that in (c) with five layers and $d \approx 0.65$ μm. (*Source:* The reflectance spectrum was plotted using the data appearing in Fig. 3 in S-Y. Lin and J.G. Fleming, *J. Light Wave Technol.*, *17*, 1944, 1999.)

frequency range, $\Delta\omega$. The structure then has a **full photonic bandgap**. As it turns out not *any* 3D periodic structure results in a full photonic bandgap. Only certain 3D periodic structures allow full photonic bandgaps to develop. One such structure is the *wood pile* periodic structure shown in Figure 1.52 (b). The unit cell has four layers of rods. The rods are parallel in each layer and the layers are



An SEM image of a 3D photonic crystal made from porous silicon in which the lattice structure is close to being simple cubic. The silicon squares, the unit cells, are connected at the edges to produce a cubic lattice. This 3D PC has a photonic bandgap centered at 5 μm and about 1.9 μm wide. *(Courtesy of Max-Planck Institute for Microstructure Physics.)*



An SEM image of a 3D photonic crystal that is based on the *wood pile* structure. The rods are polycrystalline silicon. Although five layers are shown, the unit cell has four layers, *e.g.*, the four layers starting from the bottom layer. Typical dimensions are in microns. In one similar structure with rod-to-rod pitch $d = 0.65$ μm with only a few layers, the Sandia researchers were able to produce a photonic bandgap $\Delta\lambda$ of 0.8 μm centered around 1.6 μm within the telecommunications band. *(Courtesy of Sandia National Laboratories.)*

at 90° to each other. An SEM image of a woodpile photonic crystal is shown in Figure 1.52 (c). The optical reflectance, that is, 1 – transmittance, of this woodpile PC is shown in Figure 1.52 (d). It is apparent that there is a photonic bandgap, a stop band, over a range of frequencies, or wavelengths, $\Delta\lambda$, around 3 μm. The width and the location of the reflectance or the transmittance band depends on the structure of the photonic crystal, that is, the periodicity, unit cell structure and refractive index contrast. The colors of certain butterflies and insects arise not from pigments or colorants but from a photonic crystal effect with the right periodicity.

An important practical aspect of PCs, analogous to semiconductor crystals, is the importance of defects. Point and line defects that occur in normal crystals also occur in PCs, as illustrated in Figure 1.53. They are intentionally introduced to endow the PC structure with certain optical properties. A defect is a discontinuity in the periodicity of the PC. For example, if we upset the periodicity by removing a unit cell, we create a so-called **point defect**. This void can act as an *optical cavity*, trapping the EM radiation within the cavity as illustrated in Figure 1.53. We can, of course, remove a group of unit cells, or modify the refractive index over a few unit cells, which would create an optical microcavity. Defects introduce localized electromagnetic modes with frequencies within the photonic bandgap. Defects can tightly confine a mode of light in a very small cavity volume. We can also enhance the refractive index locally, which would also classify as a point defect.

Line defects are formed when a long row of unit cells are missing, or the refractive index stays constant over a long line in the crystal. Such a line in which the index is constant allows propagating EM modes within the photonic bandgap. Since EM waves can propagate within and along the line defect, the line defect acts as an **optical wave guide**, guiding the radiation along its length as shown in Figure 1.53. The EM wave cannot spread into the perfect photonic crystal region since, in this region, the frequency falls into the stop band.

One very important property of photonic crystals is their ability to *suppress* or *inhibit spontaneous emission*. We can understand this effect intuitively by considering a 3D PC made from a semiconductor. In spontaneous emission, an electron falls from the conduction band

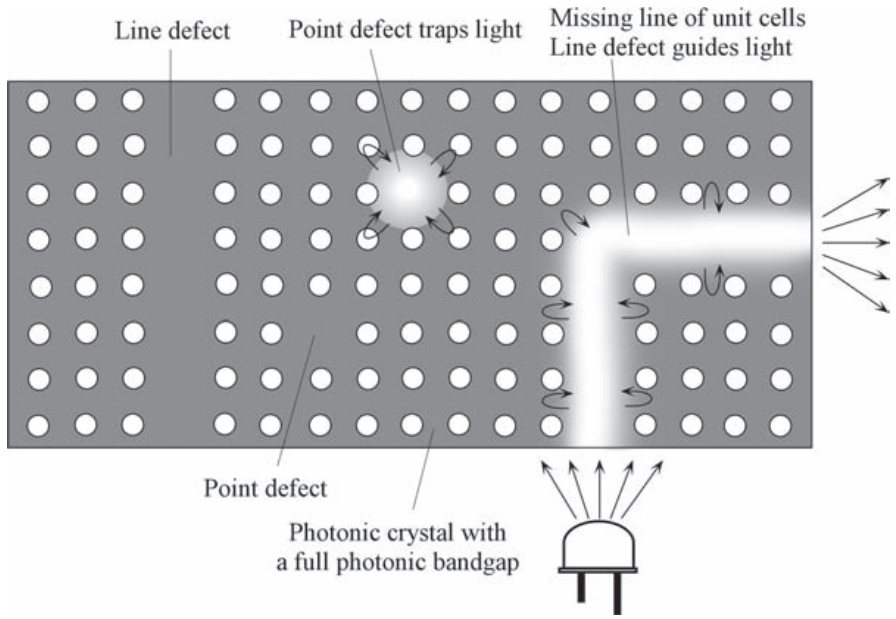


FIGURE 1.53 Schematic illustration of point and line defects in a photonic crystal. A point defect acts as an optical cavity, trapping the radiation. Line defects allow the light to propagate along the defect line. The light is prevented from dispersing into the bulk of the crystal since the structure has a full photonic bandgap. The frequency of the propagating light is in the bandgap, that is, in the stop band.

Sajeev John, at the University of Toronto, along with Eli Yablonovitch (shown earlier in the chapter) carried out the initial pioneering work in the development of the field of photonics crystals. Sajeev John was able to show that it is possible to trap light in a similar way the electron is captured, that is localized, by a trap in a semiconductor. Defects in photonic crystals can confine or localize electromagnetic waves; such effects have important applications in quantum computing and integrated photonics. *(Courtesy of Sajeev John.)*

to the valence band spontaneously and emits a photon of energy $\hbar\omega$ that corresponds to the bandgap energy E_g However, if the photon frequency $\hbar\omega$ falls into the bandgap of the PC, then this photon is not allowed to propagate or "exist" in the structure. It is prevented from being emitted—the photons have no place to go.

Photonic crystals have also been shown to exhibit a so-called *superprism effect*. Under appropriate conditions, the dispersion of light by a prism-shaped photonic crystal is considerably enhanced over that corresponding to a homogeneous prism having the same average refractive index as the photonic crystal. This originates from the strong curvature of the ω–k curve near the edge of the Brillouin Zone as apparent in Figure 1.51 (a). The latter may also be viewed as a high refractivity variation with wavelength.

Questions and Problems

1.1 Maxwell's wave equation and plane waves

(a) Consider a traveling sinusoidal wave of the form $E_x = E_o\cos(\omega t - kz + \phi_o)$. The latter can also be written as $E_x = E_o\cos\left[k(vt - z) + \phi_o\right]$, where $v = \omega/k$ is the velocity. Show that this wave satisfies Maxwell's wave equation, and show that $v = 1/(\mu_o\varepsilon_o\varepsilon_r)^{1/2}$.

(b) Consider a traveling function of any shape, even a very short delta pulse, of the form $E_x = f\left[k(vt - z)\right]$, where f is any function, which can be written is $E_x = f(\phi)$, $\phi = k(vt - z)$. Show that this traveling function satisfies Maxwell's wave equation. What is its velocity? What determines the form of the function f?

1.2 Propagation in a medium of finite small conductivity An electromagnetic wave in an isotropic medium with a dielectric constant ε_r and a finite conductivity σ and traveling along z obeys the following equation for the variation of the electric field E perpendicular to z

$$\frac{d^2E}{dz^2} - \varepsilon_o\varepsilon_r\mu_o\frac{\partial^2E}{\partial t^2} = \mu_o\sigma\frac{\partial E}{\partial t} \qquad \textbf{(P1.1)}$$

Show that one possible solution is a plane wave whose amplitude decays exponentially with propagation along z, that is, $E = E_o\exp(-\alpha' z)\exp\left[j(\omega t - kz)\right]$. Here $\exp(-\alpha' z)$ causes the envelope of the amplitude to decay with z (attenuation) and $\exp\left[j(\omega t - kz)\right]$ is the traveling wave portion. Show that in a medium in which α is small, the wave velocity and the attenuation coefficient of the field are given by

$$v = \frac{\omega}{k} = \frac{1}{\sqrt{\mu_o\varepsilon_o\varepsilon_r}} \quad \text{and} \quad \alpha' = \frac{\sigma}{2\varepsilon_o cn}$$

where n is the refractive index ($n = \varepsilon_r^{1/2}$). What is the attenuation coefficient α that describes the decay of the light intensity? (Metals with high conductivities are excluded.)

1.3 Point light source What is the irradiance measured at a distance of 1 m and 2 m from a 1 W light point source?

1.4 Gaussian beam Estimate the divergence and Rayleigh range of a Gaussian beam from a He-Ne Laser with $\lambda = 633$ nm and a beam width of 1.00 nm at $z = 0$. After traversing 10 m through vacuum, what will the beam width be?

1.5 Gaussian beam in a cavity with spherical mirrors Consider an optical cavity formed by two aligned spherical mirrors facing each other as shown in Figure 1.54. Such an optical cavity is called a *spherical mirror*

resonator, and is most commonly used in gas lasers. Sometimes, one of the reflectors is a plane mirror. The two spherical mirrors and the space between them form an optical resonator because only certain light waves with certain frequencies can exist in this optical cavity. The radiation inside a spherical mirror cavity is a *Gaussian beam*. The actual or particular Gaussian beam that fits into the cavity is that beam whose wavefronts at the mirrors match the curvature of the mirrors. Consider the symmetric resonator shown in Figure 1.54 in which the mirrors have the same radius of curvature R. When a wave starts at A, its wavefront is the same as the curvature of A. In the middle of the cavity it has the minimum width and at B the wave again has the same curvature as B. Such a wave in the cavity can replicate itself (and hence exist in the cavity) as it travels between the mirrors provided that it has right beam characteristics, that is the right curvature at the mirrors. The radius of curvature R of a Gaussian beam wavefront at a distance z along its axis is given by

$$R(z) = z\left[1 + (z_o/z)^2\right]; \quad z_o = \pi w_o^2/\lambda \text{ is the Rayleigh range}$$

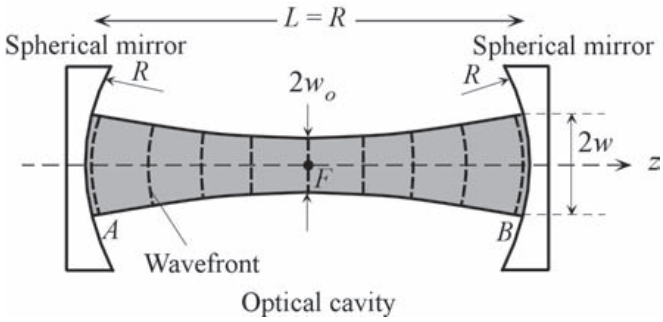


FIGURE 1.54 Two spherical mirrors reflect waves to and from each other. The optical cavity contains a Gaussian beam. This particular optical cavity is symmetric and confocal; the two focal points coincide at F.

Consider a confocal symmetric optical cavity in which the mirrors are separated by $L = R$.

(a) Show that the cavity length L is $2z_o$, that is, it is the same as twice the Rayleigh range, which is the reason the latter is called the **confocal length**.

(b) Show that the waist of the beam $2w_o$ is fully determined only by the radius of curvature R of the mirrors, and given by

$$2w_o = (2\lambda R/\pi)^{1/2}$$

(c) If the cavity length $L = R = 50$ cm, and $\lambda = 633$ nm, what is the waist of the beam at the center and also at the mirrors?

1.6 Cauchy dispersion equation Using the Cauchy coefficients and the general Cauchy equation, calculate refractive index of a silicon crystal at wavelengths of 200 μm and at 2 μm, over two orders of magnitude wavelength change. What is your conclusion?

1.7 Sellmeier dispersion equation Using the Sellmeier equation and the coefficients, obtain a graph of the refractive index of fused silica (SiO_2) versus its wavelength in the range of 500 nm to 1550 nm.

1.8 Sellmeier dispersion equation The Sellmeier dispersion coefficient for pure silica (SiO_2) and 86.5% SiO_2-13.5% GeO_2 are given in Table 1.2. Write a program on your computer or calculator, or use a math software package or even a spreadsheet program (*e.g.*, Excel) to obtain the refractive index n as a function of λ from 0.5 μm to 1.8 μm for both pure silica and 86.5% SiO_2-13.5% GeO_2. Obtain the group index, N_g, vs. wavelength for both materials and plot it on the same graph. Find the wavelength at which the material dispersion, defined as the derivative of the group velocity with respect to the wavelength, becomes zero in each material.

1.9 The Cauchy dispersion relation for zinc selenide ZnSe is a II–VI semiconductor and a very useful optical material used in various applications such as optical windows (especially high power laser windows), lenses, and prisms. It transmits over 0.50–19 μm. n in the 1–11 μm range is described by a Cauchy expression of the form

$$n = 2.4365 + \frac{0.0485}{\lambda^2} + \frac{0.0061}{\lambda^4} - 0.0003\lambda^2$$

in which λ is in μm. What are the n_{-2}, n_0, n_2 and n_4 coefficients? What is ZnSe's refractive index n and group index N_g at 5 μm?

1.10 Refractive index, reflection, and the Brewster's angle

(a) Consider light of free-space wavelength 1300 nm traveling in pure silica medium. Calculate the phase velocity and group velocity of light in this medium. Is the group velocity ever greater than the phase velocity?

(b) What is the Brewster angle (the polarization angle θ_p) and the critical angle (θ_c) for total internal reflection when the light wave traveling in this silica medium is incident on a silica–air interface. What happens at the polarization angle?

(c) What is the reflection coefficient and reflectance at normal incidence when the light beam traveling in the silica medium is incident on a silica–air interface?

(d) What is the reflection coefficient and reflectance at normal incidence when a light beam traveling in air is incident on an air–silica interface? How do these compare with part (c) and what is your conclusion?

1.11 Snell's law and lateral beam displacement What is the lateral displacement when a laser beam passes through two glass plates, each of thickness 1 mm, with refractive indices of $n_1 = 1.570$ and $n_2 = 1.450$ respectively, if the angle of incidence is 45°?

1.12 Snell's law and lateral beam displacement An engineer wants to design a refractometer (an instrument for measuring the refractive index) using the lateral displacement of light through a glass plate. His initial experiments involve using a plate of thickness L, and measuring the displacement of a laser beam when the angle of incidence θ_i is changed, for example, by rotating (tilting) the sample. For $\theta_i = 40°$ he measures a displacement of 0.60 mm, and when $\theta_i = 80°$ he measures 1.69 mm. Find the refractive index of the plate and its thickness. (*Note:* You need to solve a nonlinear equation for n numerically.)

1.13 Snell's law and prisms Consider the prism shown in Figure 1.55 that has an apex angle $\alpha = 60°$. The prism has a refractive index of n and it is in air.

(a) What are Snell's law at interfaces at A (incidence and transmittance angles of θ_i and θ_t) and B (incidence and transmittance angles of θ_i' and θ_t')?

(b) Total deflection $\delta = \delta_1 + \delta_2$ where $\delta_1 = \theta_i - \theta_t$ and $\delta_2 = \theta_t' - \theta_i'$. Now, $\beta + \theta_i' + \theta_t = 180°$ and $\alpha + \beta = 180°$. Find the deflection of the beam for an incidence angle of 45° for the following three colors at which n is known: Blue, $n = 1.4634$ at $\lambda = 486.1$ nm; yellow, $n = 1.4587$ at $\lambda = 589.2$ nm; red, $n = 1.4567$ at $\lambda = 656.3$ nm. What is the separation in distance between the rays if the rays are projected on a screen 1 m away.

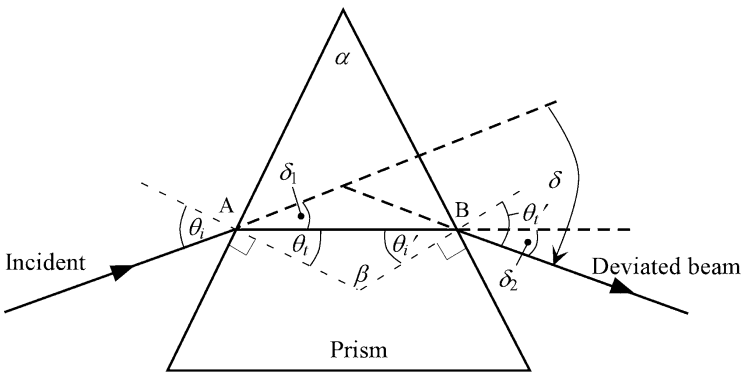


FIGURE 1.55 A light beam is deflected by a prism through an angle δ. The angle of incidence is θ_i. The apex angle of the prism is α.

1.14 Fermat's principle of least time Fermat's principle of least time in simple terms states that *when light travels from one point to another it takes a path that has the shortest time*. In going from a point A in some medium with a refractive index n_1 to a point B in a neighboring medium with refractive index n_2 as in Figure 1.56, the light path AOB involves refraction at O that satisfies Snell's law. The time it takes to travel from A to B is minimum only for the path AOB such that the incidence and refraction angles θ_i and θ_t satisfy Snell's law. Let's draw a straight line from A to B cutting the x-axes at O'. The line $AO'B$ will be our reference line and we will place the origin of x and y coordinates at O'. Without invoking Snell's law, we will vary point O along the x-axis (hence OO' is a variable labeled x), until the time it takes to travel AOB is minimum, and thereby derive Snell's law. The time t it takes for light to travel from A to B through O is

$$t = \frac{AO}{c/n_1} + \frac{OB}{c/n_2} = \frac{\left[(x_1 - x)^2 + y_1^2\right]^{1/2}}{c/n_1} + \frac{\left[(x_2 + x)^2 + y_2^2\right]^{1/2}}{c/n_2} \qquad \textbf{(P1.2)}$$

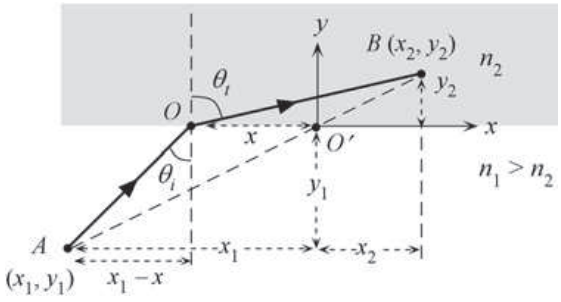


FIGURE 1.56 Consider a light wave traveling from point $A\,(x_1, y_2)$ to $B\,(x_1, y_2)$ through an arbitrary point O at a distance x from O'. The principle of least time from A to B requires that O is such that the incidence and refraction angles obey Snell's law.

The incidence and transmittance angles are given by

$$\sin\theta_i = \frac{x_1 - x}{\left[(x_1 - x)^2 + y_1^2\right]^{1/2}} \quad \text{and} \quad \sin\theta_i = \frac{(x_2 + x)}{\left[(x_2 + x)^2 + y_2^2\right]^{1/2}} \tag{P1.3}$$

Differentiate Eq. (P1.2) with respect to x to find the condition for the "least time" and then use Eq. (P1.3) in this condition to derive Snell's law.

Pierre de Fermat (1601–1665) was a French mathematician who made many significant contributions to modern calculus, number theory, analytical geometry, and probability. *(Courtesy of Mary Evans Picture Library/Alamy.)*

1.15 Antireflection (AR) coating

(a) A laser beam of wavelength 1550 nm from air is launched to a single mode optical fiber with a core refractive index $n_1 = 1.45$. Estimate the refractive index and thickness of film required for an anti-reflecting coating on this fiber.

(b) A Ge photodiode is designed to operate at 1550 nm, and it is required to have AR coatings to minimize reflected light. Two possible materials are available for AR coating: SiO_2 with a refractive index of 1.46, and TiO_2 with a refractive index of 2.2. Which would be better suited? What would be the thickness for the AR coating on this photodiode? The refractive index of Ge is about 4.

(c) Consider a Ge photodiode that is designed for operation around 1200 nm. What are the best AR coating refractive index and thickness if the refractive index of Ge is about 4.0?

1.16 Single- and double-layer antireflection V-coating For a single-layer AR coating of index n_2 on a material with index $n_3 (> n_2 > n_1)$, as shown in Figure 1.57 (a), the minimum reflectance at normal incidence is given by

$$R_{\min} = \left[\frac{n_2^2 - n_1 n_3}{n_2^2 + n_1 n_3}\right]^2$$

when the reflections $A, B, \ldots$ all interfere as destructively as possible. $R_{\min} = 0$ when $n_2 = (n_1 n_3)^{1/2}$. The choice of materials may not always be the best for a single-layer AR coating. *Double-layer AR coatings*, as shown in

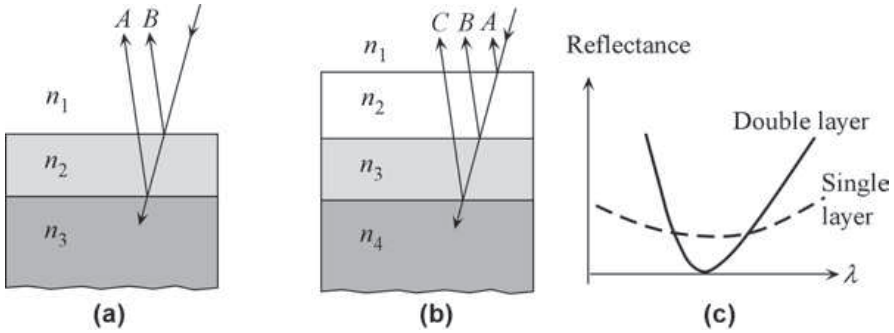


FIGURE 1.57 (a) A single-layer AR coating. (b) A double-layer AR coating and (c) its V-shaped reflectance spectrum over a wavelength range.

Figure 1.57 (b), can achieve lower and sharper reflectance at a specified wavelength as in Figure 1.57 (c). To reduce the reflection of light at the n_1–n_4 interface, two layers n_2 and n_3, each quarter wavelength in the layer (λ/n_2 and λ/n_3) are interfaced between n_1 and n_4. The reflections *A*, *B,* and *C* for normal incidence result in a minimum reflectance given by

$$R_{\min} = \left[\frac{n_3^2 n_1 - n_4 n_2^2}{n_3^2 n_1 + n_4 n_2^2}\right]^2$$

The double-layer reflectance vs. wavelength behavior usually has a V-shape, and such coatings are called *V-coatings.*

(a) Show that double-layer reflectance vanishes when

$$(n_2/n_3)^2 = n_1/n_4$$

(b) Consider an InGaAs, a semiconductor crystal with an index 3.8, for use in a photodetector. What is the reflectance without any AR coating?

(c) What is the reflectance when InGaAs is coated with a thin AR layer of Si_3N_4? Which material in Table 1.3 would be ideal as an AR coating?

TABLE 1.3 Typical AR materials and their approximate refractive indices over the visible wavelengths

	MgF_2	SiO_2	Al_2O_3	CeF_3	Sb_2O_3	Si_3N_4	SiO	ZrO_2	ZnS	TiO_2	CdS
n	1.38	1.46	1.65	1.65	1.9–2.1	1.95	2.0	2.05	2.35	2.35	2.60

(d) What two materials would you choose to obtain a V-coating? Note: The choice of an AR coating also depends on the technology involved in depositing the AR coating and its effects on the interface states between the AR layer and the semiconductor. $Si_{1-x}N_x$ is a common AR coating on devices inasmuch as it is a good passive dielectric layer, its deposition technology is well established and changing its composition (x) changes its index.

1.17 Single-, double-, and triple-layer antireflection coatings Figure 1.58 shows the reflectance of an uncoated glass, and glass that has a single- (1), double- (2) and triple- (3) layer AR coatings? The coating details are in the figure caption. Each layer in single- and double-layer AR coatings has a thickness of $\lambda/4$, where λ is the wavelength in the layer. The triple-layer AR layer has three coatings with thicknesses $\lambda/4$, $\lambda/2$, and $\lambda/4$. Can you qualitatively explain the results by using interference? What applications would need single-, double-, and triple-layer coatings?

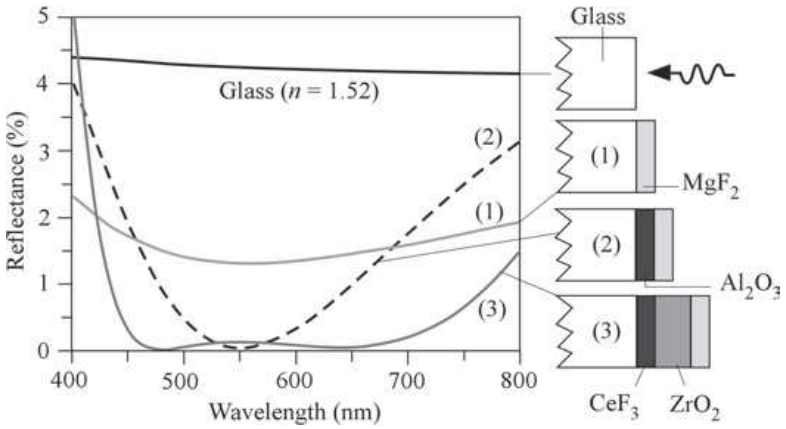


FIGURE 1.58 Reflectance vs. wavelength for a glass plate, $n = 1.52$, with and without AR coatings. (1) Single-layer AR coating is a quarter wavelength ($\lambda/4$) thick MgF_2, $n = 1.38$. (2) Double-layer coating is $\lambda/4$ thick MgF_2 and $\lambda/4$ thick Al_2O_3, $n = 1.69$. (3) Triple-layer coating is $\lambda/4$ thick MgF_2, $\lambda/2$ thick ZrO_2, $n = 2.05$, and a $\lambda/4$ thick CeF_3, $n = 1.64$. (*Source:* Plotted from data appearing in Figure 2.2 in S. Chattopadhyay *et al.*, *Mater. Sci. Engin. R*, *69*, 1, 2010.)

1.18 Reflection at glass–glass and air–glass interfaces A ray of light that is traveling in a glass medium of refractive index $n_1 = 1.460$ becomes incident on a less dense glass medium of refractive index $n_2 = 1.430$. Suppose that the free-space wavelength of the light ray is 850 nm.

(a) What should the minimum incidence angle for TIR be?

(b) What is the phase change in the reflected wave when the angle of incidence $\theta_i = 85°$ and when $\theta_i = 90°$?

(c) What is the penetration depth of the evanescent wave into medium 2 when $\theta_i = 85°$ and when $\theta_i = 90°$?

(d) What is the reflection coefficient and reflectance at normal incidence ($\theta_i = 0°$) when the light beam traveling in the glass medium ($n = 1.460$) is incident on a glass–air interface?

(e) What is the reflection coefficient and reflectance at normal incidence when a light beam traveling in air is incident on an air–glass ($n = 1.460$) interface? How do these compare with part (d) and what is your conclusion?

1.19 Dielectric mirror A dielectric mirror is made up of a quarter wave layer of GaAs with $n_H = 3.38$ and AlAs with $n_L = 3.00$ at around 1550 nm. The light is incident on the mirror from another semiconductor of refractive index $n_0 = 3.40$. Find out the number of pairs of layers N needed to get 90% reflectance. Find out the bandwidth of the reflected light.

1.20 TIR and polarization at water–air interface

(a) Given that the refractive index of water is about 1.33, what is the polarization angle for light traveling in air and reflected from the surface of the water?

(b) Consider a diver in sea pointing a flashlight towards the surface of the water. What is the critical angle for the light beam to be reflected from the water surface?

1.21 Reflection and transmission at a semiconductor–semiconductor interface A light wave with a wavelength of 890 nm (free-space wavelength) that is propagating in GaAs becomes incident on AlGaAs. The refractive index of GaAs is 3.60, that of AlGaAs is 3.30.

(a) Consider normal incidence. What are the reflection and transmission coefficients and the reflectance and transmittance? (From GaAs into AlGaAs.)

(b) What is the Brewster angle (the polarization angle θ_p) and the critical angle (θ_c) for total internal reflection for the wave in (a); the wave that is traveling in GaAs and incident on the GaAs–AlGaAs interface?

(c) What is the reflection coefficient and the phase change in the reflected wave when the angle of incidence $\theta_i = 79°$?

(d) What is the penetration depth of the evanescent wave into medium 2 when $\theta_i = 79°$ and when $\theta_i = 89°$? What is your conclusion?

1.22 Phase changes on TIR Consider a light wave of wavelength 870 nm traveling in a semiconductor medium (GaAs) of refractive index 3.60. It is incident on a different semiconductor medium (AlGaAs) of refractive index

3.40, and the angle of incidence is 80°. Will this result in total internal reflection? Calculate the phase change in the parallel and perpendicular components of the reflected electric field.

1.23 Fresnel's equations Fresnel's equations are sometimes given as follows:

$$r_\perp = \frac{E_{ro,\perp}}{E_{io,\perp}} = \frac{n_1\cos\theta_i - n_2\cos\theta_t}{n_1\cos\theta_i + n_2\cos\theta_t}$$

$$r_{//} = \frac{E_{ro,//}}{E_{io,//}} = \frac{n_1\cos\theta_t - n_2\cos\theta_i}{n_1\cos\theta_t + n_2\cos\theta_i}$$

$$t_\perp = \frac{E_{to,\perp}}{E_{io,\perp}} = \frac{2n_1\cos\theta_i}{n_1\cos\theta_i + n_2\cos\theta_t}$$

and

$$t_{//} = \frac{E_{to,//}}{E_{io,//}} = \frac{2n_1\cos\theta_i}{n_1\cos\theta_t + n_2\cos\theta_i}$$

Show that these reduce to Fresnel's equation given in Eqs. (1.6.6) and (1.6.7).

Using Fresnel's equations, find the reflection and transmission coefficients for normal incidence and show that

$$r_\perp + 1 = t_\perp \quad \text{and} \quad r_{//} + nt_{//} = 1$$

where $n = n_2/n_1$.

1.24 Fresnel's equations Consider a light wave traveling in a glass medium with an index $n_1 = 1.440$ and it is incident on the glass–air interface. Using Fresnel's equations only, that is, Eqs. (1.6.6a) and (1.6.6b), calculate the reflection coefficients $r_\perp$ and $r_{//}$ and hence reflectances $R_\perp$ and $R_{//}$ for (a) $\theta_i = 25°$ and (b) $\theta_i = 50°$. In the case of $\theta_i = 50°$, find the phase change $\phi_\perp$ and $\phi_{//}$ from the reflection coefficients by writing $r = |r|\exp(-j\phi)$. Compare $\phi_\perp$ and $\phi_{//}$ from $r_\perp$ and $r_{//}$ calculations with those calculated from Eqs. (1.6.11) and (1.6.12).

1.25 Goos-Haenchen phase shift A ray of light which is traveling in a glass medium (1) of refractive index $n_1 = 1.460$ becomes incident on a less dense glass medium (2) of refractive index $n_2 = 1.430$. Suppose that the free-space wavelength of the light ray is 850 nm. The angle of incidence $\theta_i = 85°$. Estimate the lateral Goos-Haenchen shift in the reflected wave for the perpendicular field component. Recalculate the Goos-Haenchen shift if the second medium has $n_2 = 1$ (air). What is your conclusion? Assume that the virtual reflection occurs from a virtual plane in medium B at a distance d that is roughly the same as the penetration depth. Note that d actually depends on the polarization, the direction of the field, but we will ignore this dependence.

1.26 Evanescent wave Total internal reflection of a plane wave from a boundary between a more dense medium (1) n_1 and a less dense medium (2) n_2 is accompanied by an evanescent wave propagating in medium 2 near the boundary. Find the functional form of this wave and discuss how its magnitude varies with the distance into medium 2.

1.27 TIR and FTIR

(a) By considering the electric field component in medium B in Figure 1.21, explain how you can adjust the amount of transmitted light through a thin layer between two higher refractive index media.

(b) What is the critical angle at the hypotenuse face of a beam splitter cube made of glass with $n_1 = 1.6$ and having a thin film of liquid with $n_2 = 1.3$. Can you use 45° prisms with normal incidence?

(c) Explain how a light beam can propagate along a layer of material between two different media as shown in Figure 1.59 (a). Explain what the requirements are for the indices n_1, n_2, n_3. Will there be any losses at the reflections?

(d) Consider the prism coupler arrangement in Figure 1.59 (b). Explain how this arrangement works for coupling an external light beam from a laser into a thin layer on the surface of a glass substrate. Light is then propagated inside the thin layer along the surface of the substrate. What is the purpose of the adjustable coupling gap?

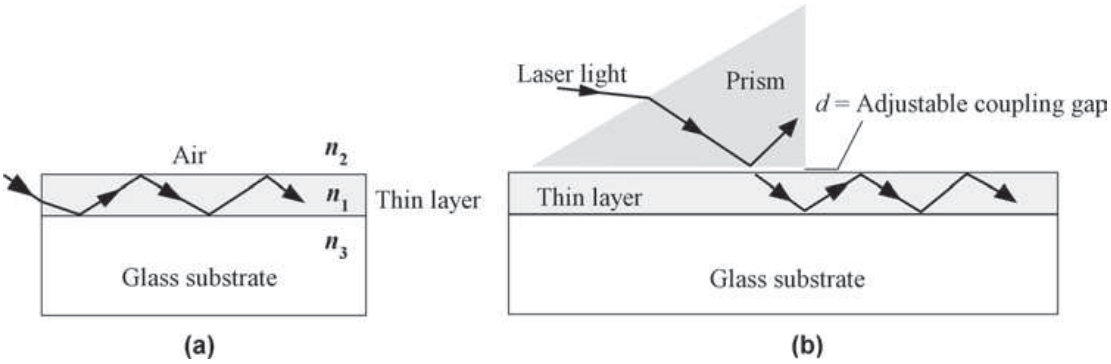


FIGURE 1.59 (a) Light propagation along an optical guide. (b) Coupling of laser light into a thin layer—optical guide—using a prism. The light propagates along the thin layer.

1.28 Complex refractive index and dielectric constant The complex refractive index $N = n - jK$ can be defined in terms of the *complex relative permittivity* $\varepsilon_r = \varepsilon_{r1} - j\varepsilon_{r2}$ as

$$\mathrm{N} = n - jK = \varepsilon_r^{1/2} = (\varepsilon_{r1} - j\varepsilon_{r2})^{1/2}$$

Show that

$$n = \left[\frac{(\varepsilon_{r1}^2 + \varepsilon_{r2}^2)^{1/2} + \varepsilon_{r1}}{2}\right]^{1/2} \quad \text{and} \quad K = \left[\frac{(\varepsilon_{r1}^2 + \varepsilon_{r2}^2)^{1/2} - \varepsilon_{r1}}{2}\right]^{1/2}$$

1.29 Complex refractive index Spectroscopic ellipsometry measurements on a germanium crystal at a photon energy of 1.5 eV show that the real and imaginary parts of the complex relative permittivity are 21.56 and 2.772, respectively. Find the complex refractive index. What is the reflectance and absorption coefficient at this wavelength? How do your calculations match with the experimental values of $n = 4.653$ and $K = 0.298$, $R = 0.419$ and $\alpha = 4.53 \times 10^6\,\mathrm{m}^{-1}$?

1.30 Complex refractive index Figure 1.26 shows the infrared extinction coefficient K of CdTe. Calculate the absorption coefficient α and the reflectance R of CdTe at 60 μm and 80 μm.

1.31 Refractive index and attenuation in the infrared region—Reststrahlen absorption Figure 1.26 shows the refractive index n and the extinction coefficient K as a function of wavelength λ in the infrared for a CdTe crystal due to lattice absorption, called Reststrahlen absorption. It results from the ionic polarization of the crystal induced by the optical field in the light wave. The relative permittivity ε_r due to positive (Cd^{2+}) and negative (Te^{2-}) ions being made to oscillate by the optical field about their equilibrium positions is given in its simplest form by

$$\varepsilon_r = \varepsilon_r' - j\varepsilon_r'' = \varepsilon_{rH} + \frac{\varepsilon_{rH} - \varepsilon_{rL}}{\left(\dfrac{\omega}{\omega_T}\right)^2 - 1 + j\dfrac{\gamma}{\omega_T}\left(\dfrac{\omega}{\omega_T}\right)} \qquad \textbf{(P1.4)}$$

where ε_{rL} and ε_{rH} are the relative permittivity at low (L) and high (H) frequencies, well below and above the infrared peak, γ is a loss coefficient characterizing the rate of energy transfer from the EM wave to lattice vibrations (phonons), and ω_T is a transverse optical lattice vibration frequency that is related to the nature of bonding between the ions in the crystal. Table 1.4 provides some typical values for CdTe and GaAs. Equation (P1.4) can be used to obtain a reasonable approximation to the infrared refractive index n and extinction coefficient K due to Reststrahlen absorption. (a) Consider CdTe, and plot n and K vs. λ from 40 μm to 90 μm and compare with the experimental results in Figure 1.26 in terms of the peak positions and the width of the extinction coefficient peak. (b) Consider

TABLE 1.4 Ionic polarization resonance parameters for CdTe and GaAs

	ε_{rL}	ε_{rH}	$\omega_T(\mathrm{rad\ s^{-1}})$	$\gamma(\mathrm{rad\ s^{-1}})$
CdTe	10.20	7.10	2.68×10^{13}	0.124×10^{13}
GaAs	13.0	11.0	5.07×10^{13}	0.045×10^{13}

GaAs, and plot n and K vs. λ from 30 μm to 50 μm. (c) Calculate n and K for GaAs at $\lambda = 38.02$ μm and compare with the experimental values $n = 7.55$ and $K = 0.629$. (You might want to use a logarithmic scale for K.)

1.32 Coherence length A narrow band pass filter transmits wavelengths in the range 5000 ± 0.5 A°. If this filter is placed in front of a source of white light, what is the coherence length of the transmitted light?

1.33 Spectral widths and coherence

(a) Suppose that frequency spectrum of a radiation emitted from a source has a central frequency υ_o and a spectral width $\Delta\upsilon$. The spectrum of this radiation in terms of wavelength will have a central wavelength λ_o and a spectral width $\Delta\lambda$. Clearly, $\lambda_o = c/\upsilon_o$. Since $\Delta\lambda \ll \lambda_o$ and $\Delta\upsilon \ll \upsilon_o$, using $\lambda = c/\upsilon$, show that the line width $\Delta\lambda$ and hence the coherence length l_c are

$$\Delta\lambda = \Delta\upsilon\frac{\lambda_o}{\upsilon_o} = \Delta\upsilon\frac{\lambda_o^2}{c} \quad \text{and} \quad l_c = c\Delta t = \frac{\lambda_o^2}{\Delta\lambda}$$

(b) Calculate $\Delta\lambda$ for a lasing emission from a He-Ne laser that has $\lambda_o = 632.8$ nm and $\Delta\upsilon \approx 1.5$ GHz. Find its coherence time and length.

1.34 Coherence lengths Find the coherence length of the following light sources:

(a) An LED emitting at 1550 nm with a spectral width 150 nm;
(b) A semiconductor laser diode emitting at 1550 nm with a spectral width 3 nm;
(c) A quantum well semiconductor laser diode emitting at 1550 nm with a spectral width of 0.1 nm;
(d) A multimode He-Ne laser with a spectral frequency width of 1.5 GHz;
(e) A specially designed single mode and stabilized He-Ne laser with a spectral width of 100 MHz.

1.35 Fabry–Perot optical cavity Consider an optical cavity formed between two identical mirrors, each with reflectance = 0.97. The refractive index of the medium enclosed between the mirrors is 1. Find out the minimum length of the optical cavity which can resolve spectral lines of a sodium lamp with line width $\Delta\lambda = 0.6$ nm and $\Delta\lambda = 589.3$ nm. Further, estimate the mode separation in frequency and wavelength. What are the finesse F and Q factors for this cavity?

1.36 Fabry–Perot optical cavity from a ruby crystal Consider a ruby crystal of diameter 1 cm and length 10 cm. The refractive index is 1.78. The ends have been silvered and the reflectances are 0.99 and 0.95 each. What is the nearest mode number that corresponds to a radiation of wavelength 694.3 nm? What is the actual wavelength of the mode closest to 694.3 nm? What is the mode separation in frequency and wavelength? What are the finesse F and Q factor for the cavity?

1.37 Fabry–Perot optical cavity spectral width Consider an optical cavity of length 40 cm. Assume the refractive index is 1, and use Eq. (1.11.3) to plot the peak closest to 632.8 nm for 4 values of $R = 0.99, 0.90, 0.75$ and 0.6. For each case find the spectral width $\delta\lambda_m$, the finesse F and Q. How accurate is Eq. (1.11.5) in predicting $\delta\lambda_m$. (You may want to use a graphing software for this problem.)

1.38 Diffraction A collimated beam of light of wavelength 632.8 nm is incident on a circular aperture of 250 μm. Find out the divergence of the transmitted beam. Obtain the diameter of the transmitted beam at a distance of 10 m. What would be the divergence if the aperture is a single slit of width 250 μm?

1.39 Diffraction intensity Consider diffraction from a uniformly illuminated circular aperture of diameter D. The far field diffraction pattern is given by a Bessel function of the first kind and first order, J_1, and the intensity at a point P on the angle θ_i with respect to the central axis through the aperture is

$$I(\gamma) = I_o\left(\frac{2J_1(\gamma)}{\gamma}\right)^2$$

where I_o is the maximum intensity, $\gamma = (1/2)kD\sin\theta$ is a variable quantity that represents the angular position θ on the screen as well as the wavelength ($k = 2\pi/\lambda$) and the aperture diameter D. $J_1(\gamma)$ can be calculated from

$$J_1(\gamma) = \frac{1}{\pi}\int_0^{\pi}\cos(\alpha - \gamma\sin\alpha)d\alpha$$

where α is an integration variable. Using numerical integration, or a suitable mathematics software program, plot $\left[J_1(\gamma)/\gamma\right]$ vs. γ for $\gamma = 0 - 8$ and confirm that zero-crossings occur at $\gamma = 3.83, 7.02$ and the maxima at $\gamma = 0, 5.14$. What is the intensity ratio of the first bright ring (at $\gamma = 5.14$) to that at the center of the Airy disk ($\gamma = 0$)? (You can use a very small γ instead of zero for the center intensity calculation.) Using the first zero at $\gamma = 3.83$, verify Eq. (1.12.5), $\sin\theta_o = 1.22\lambda/D$, where θ_o is the angular position of the first dark ring, as defined in Figure 1.36 (b).

George Bidell Airy (1801–1892, England). George Airy was a professor of astronomy at Cambridge and then the Astronomer Royal at the Royal Observatory in Greenwich, England. *(© Mary Evans Picture Library/Alamy.)*

1.40 Bragg diffraction A reflection grating is made on the surface of a semiconductor with a periodicity of 0.5 μm. If light of wavelength 1.55 μm is incident at an angle of 88° to the normal, find out the diffracted beam.

1.41 Diffraction grating for WDM Consider a transmission diffraction grating. Suppose that we wish to use this grating to separate out different wavelengths of information in a WDM signal at 1550 nm. (WDM stands of wavelength division multiplexing.) Suppose that the diffraction grating has a periodicity of 2 μm. The angle of incidence is 0° with respect to the normal to the diffraction grating. What is the angular separation of the two wavelength component s at 1.550 μm and 1.540 μm? How would you increase this separation?

1.42 A monochromator Consider an incident beam on a reflection diffraction grating as in Figure 1.60. Each incident wavelength will result in a diffracted wave with a different diffraction angle. We can place a small slit and allow only one diffracted wave λ_m to pass through to the photodetector. The diffracted beam would consist of wavelengths in the incident beam separated (or fanned) out after diffraction. Only one wavelength λ_m will be diffracted favorably to pass through the slit and reach the photodetector. Suppose that the slit width is $s = 0.1$ mm, and the slit is at a distance $R = 5$ cm from the grating. Suppose that the slit is placed so that it is at right angles to the incident beam: $\theta_i + \theta_m = \pi/2$. The grating has a corrugation periodicity of 1 μm.

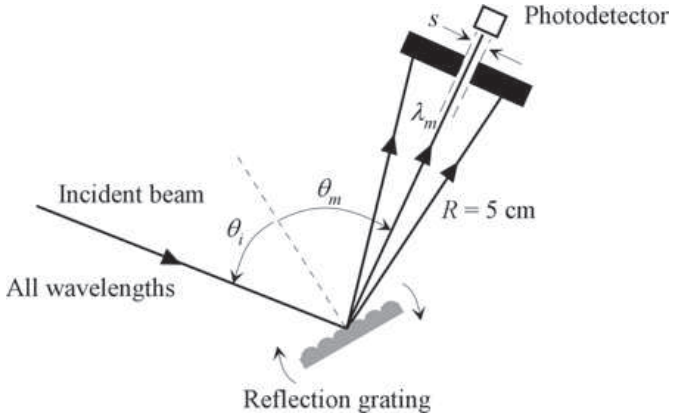


FIGURE 1.60 A monochromator based on using a diffraction grating.

(a) What is the range of wavelengths that can be captured by the photodetector when we rotate the grating from $\theta_i = 1°$ to 40°?

(b) Suppose that $\theta_i = 15°$. What is the wavelength that will be detected? What is the resolution, that is, the range of wavelengths that will pass through the slit? How can you improve the resolution? What would be the advantage and disadvantage in decreasing the slit width s?

1.43 Thin film optics Consider light incident on a thin film on a substrate, and assume normal incidence for simplicity.

(a) Consider a thin soap film in air, $n_1 = n_3 = 1, n_2 = 1.40$. If the soap thickness $d = 1\ \mu$m, plot the reflectance vs. wavelength from 0.35 μm to 0.75 μm, which includes the visible range. What is your conclusion?

(b) MgF_2 thin films are used on glass plates for the reduction of glare. Given that $n_1 = 1, n_2 = 1.38$, and $n_3 = 1.60$ (n for glass depends on the type of glass but 1.6 is a reasonable value), plot the reflectance as a function of wavelength from 0.35 μm to 0.75 μm for a thin film of thickness 0.10 μm. What is your conclusion?

1.44 Thin film optics Consider a glass substrate with $n_3 = 165$ that has been coated with a transparent optical film (a dielectric film) with $n_2 = 2.50, n_1 = 1$ (air). If the film thickness is 500 nm, find the minimum and maximum reflectances and transmittances and their corresponding wavelengths in the visible range for normal incidence. (Assume normal incidence.) Note that the thin n_2-film is not an AR coating, and for $n_1 < n_3 < n_2$,

$$R_{\max} = \left(\frac{n_2^2 - n_1 n_3}{n_2^2 + n_1 n_3}\right)^2 \quad \text{and} \quad R_{\min} = \left(\frac{n_3 - n_1}{n_3 + n_1}\right)^2$$

1.45 Thin film optics Consider light incident on a thin film on a substrate, and assume normal incidence for simplicity. Plot the reflectance R and transmittance T as a function of the phase change ϕ from $\phi = -4\pi$ to $+4\pi$ for the following cases

(a) Thin soap film in air, $n_1 = n_3 = 1, n_2 = 1.40$. If the soap thickness $d = 1\ \mu$m, what are the maxima and minima in the reflectance in the visible range?

(b) A thin film of MgF_2 on a glass plate for the reduction of glare, where $n_1 = 1, n_2 = 1.38$, and $n_3 = 1.70$ (n for glass depends on the type of glass but 1.7 is a reasonable value.) What should be the thickness of MgF2 for minimum reflection at 550 nm?

(c) A thin film of semiconductor on glass where $n_1 = 1, n_2 = 3.5$, and $n_3 = 1.55$.

1.46 Transmission through a plate Consider the transmittance of light through a partially transparent glass plate of index n in which light experiences attenuation (either by absorption or scattering). Suppose that the plate is in a medium of index n_o, the reflectance at each n–n_o interface is R and the attenuation coefficient is α.

(a) Show that

$$T_{\text{plate}} = \frac{(1 - R)^2 e^{-\alpha d}}{(1 - R^2) e^{-2\alpha d}}$$

(b) If T is the transmittance of a glass plate of refractive index n in a medium of index n_o show that, in the absence of any absorption in the glass plate,

$$n/n_o = T^{-1} + (T^{-2} - 1)^{1/2}$$

if we neglect any losses in the glass plate.

(c) If the transmittance of a glass plate in air has been measured to be 89.96%. What is its refractive index? Do you think this is a good way to measure the refractive index?

1.47 Scattering Consider Rayleigh scattering. If the incident light is unpolarized, the intensity I_s of the scattered light a point at a distance r at an angle θ to the original light beam is given by

$$I_s \propto \frac{1 - \cos^2\theta}{r^2}$$

Plot a polar plot of the intensity I_s at a fixed distance r from the scatter as we change the angle θ around the scatterer. In a polar plot, the radial coordinate (OP in Figure 1.48 (b)) is I_s. Construct a contour plot in the xy plane in which a contour represents a constant intensity. You need to vary r and θ or x and y such that I_s remains constant. Note $x = r\cos\theta$ and $y = r\sin\theta, \theta = \arctan(y/x), r = (x^2 + y^2)^{1/2}$.

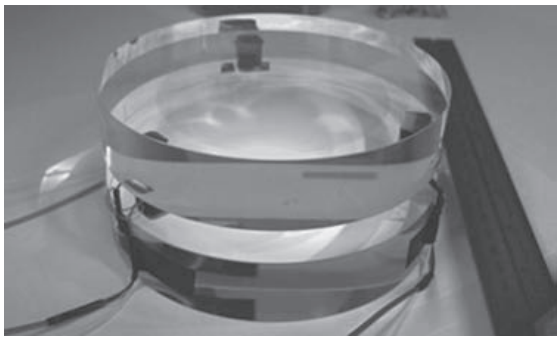
1.48 One-dimensional photonic crystal (a Bragg mirror) The 1D photonic crystal in Figure 1.50 (a), which is essentially a Bragg reflector, has the dispersion behavior shown in Figure 1.51 (a). The stop-band $\Delta\omega$ for normal incidence and for all polarizations of light is given by (R. H. Lipson and C. Lu, *Eur. J. Phys.*, *30*, S33, 2009)

$$\frac{\Delta\omega}{\omega_o} = (4/\pi)\arcsin\left(\frac{n_2 - n_1}{n_2 + n_1}\right)$$

where $\Delta\omega$ is the stop band, ω_o is the center frequency defined in Figure 1.51 and n_2 and n_1 are the high and low refractive indices. Calculate the lowest stop band in terms of photon energy in eV, and wavelength 1550 nm for a 1D photonic crystal structure with $n_1 d_1 = n_2 d_2 = \lambda/4$, made up of: (i) Si ($n_{Si} = 3.5$) and SiO_2 ($n_{SiO_2} = 1.445$) pairs, and (ii) Si_3N_4 ($n_{Si_3N_4} = 2.0$) and SiO_2 pairs.

1.49 Photonic crystals Concepts have been borrowed from crystallography, such as a unit cell, to define a photonic crystal. What is the difference between a unit cell used in a photonic crystal and that used in a real crystal? What is the size limit on the unit cell of a photonic crystal? Is the refractive index a microscopic or a macroscopic concept? What is the assumption on the refractive index?

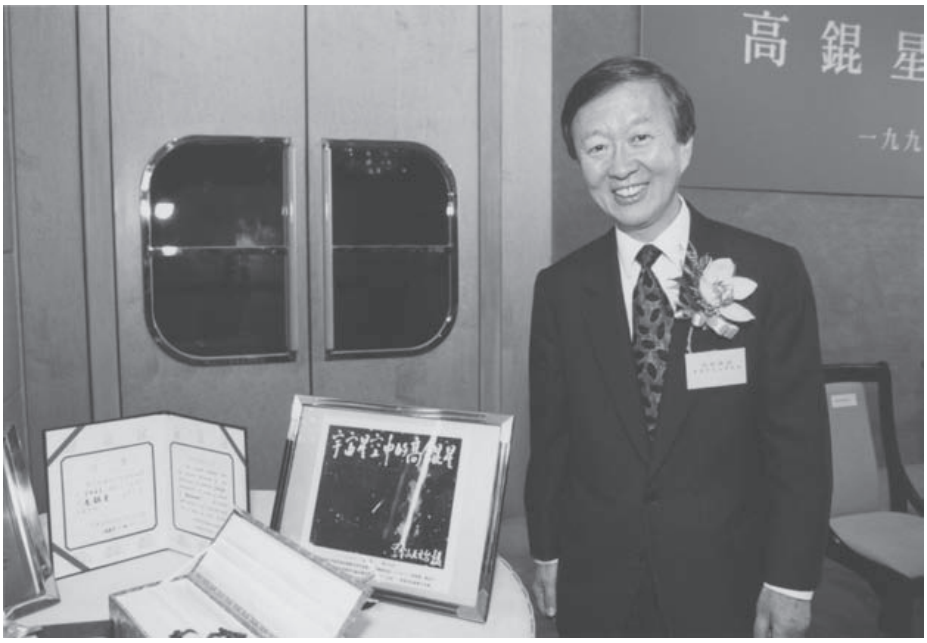
A scanning Fabry–Perot interferometer (Model SA200), used as a spectrum analyzer, which has a free spectral range of 1.5 GHz, a typical finesse of 250, spectral width (resolution) of 7.5 MHz. The cavity length is 5 cm. It uses two concave mirrors instead of two planar mirrors to form the optical cavity. A piezoelectric transducer is used to change the cavity length and hence the resonant frequencies. A voltage ramp is applied through the coaxial cable to the piezoelectric transducer to scan frequencies. *(Courtesy of Thorlabs.)*



This is a tunable large aperture (80 mm) etalon with two end plates that act as reflectors. The end plates have been machined to be flat to wavelength/110. There are three piezoelectric transducers that can tilt the end plates and hence obtain perfect alignment. *(Courtesy of Light Machinery.)*

The introduction of optical fiber systems will revolutionize the communications network. The low-transmission loss and the large bandwidth capability of the fiber systems allow signals to be transmitted for establishing communications contacts over large distances with few or no provisions of intermediate amplification.

—Charles K. Kao[1]



Charles Kao at the naming ceremony of Minor Planet (3463) "Kaokuen" by Nanjing's Purple Mountain Observatory in July 1996. Charles Kao and his colleagues carried out the early experiments on optical fibers at the Standard Telecommunications Laboratories Ltd. (the research center of Standard Telephones and Cables) at Harlow in the United Kingdom, during the 1960s. He shared the Nobel Prize in 2009 in Physics with Willard Boyle and George Smith for "groundbreaking achievements concerning the transmission of light in fibers for optical communication." In a milestone paper with George Hockam published in the IEE Proceedings in 1966 they predicted that the intrinsic losses of glass optical fibers could be much lower than 20 dB km^{-1}, which would allow their use in long-distance telecommunications. Today, optical fibers are used not only in telecommunications but also in various other technologies such as instrumentation and sensing. From 1987 to his retirement in 1996, Professor Kao was the vice-chancellor of the Chinese University of Hong Kong. *(Courtesy of the Chinese University of Hong Kong.)*

[1] Charles K. Kao (one of the pioneers of glass fibers for optical communications) *Optical Fiber Systems: Technology, Design, and Applications* (McGraw-Hill, 1982), p. 1.