Fundamentals of Photonics
Bahaa E. A. Saleh, Malvin Carl Teich

ISBNs: 0-471-83965-5 (Hardback); 0-471-2-1374-8 (Electronic)

CHAPTER

2

WAVE OPTICS

Christiaan Huygens (1629–1695) advanced several new concepts concerning the propagation of light waves.

Thomas Young (1773–1829) championed the wave theory of light and discovered the principle of optical interference.

Light propagates in the form of waves. In free space, light waves travel with a constant speed $c_o = 3.0 \times 10^8$ m/s (30 cm/ns or 0.3 mm/ps). The range of optical wavelengths contains three bands— ultraviolet (10 to 390 nm), visible (390 to 760 nm), and infrared (760 nm to 1 mm). The corresponding range of optical frequencies stretches from 3×10^{11} Hz to 3×10^{16} Hz, as illustrated in Fig. 2.0-1.

The **wave theory** of light encompasses the ray theory (Fig. 2.0-2). Strictly speaking, ray optics is the limit of wave optics when the wavelength is infinitesimally short. However, the wavelength need not actually be equal to zero for the ray-optics theory to be useful. As long as the light waves propagate through and around objects whose dimensions are much greater than the wavelength, the ray theory suffices for describing most phenomena. Because the wavelength of visible light is much shorter than the dimensions of the visible objects encountered in our daily lives, manifestations of the wave nature of light are not apparent without careful observation.

In this chapter, light is described by a scalar function, called the wavefunction, which obeys the wave equation. The precise physical meaning of the wavefunction is

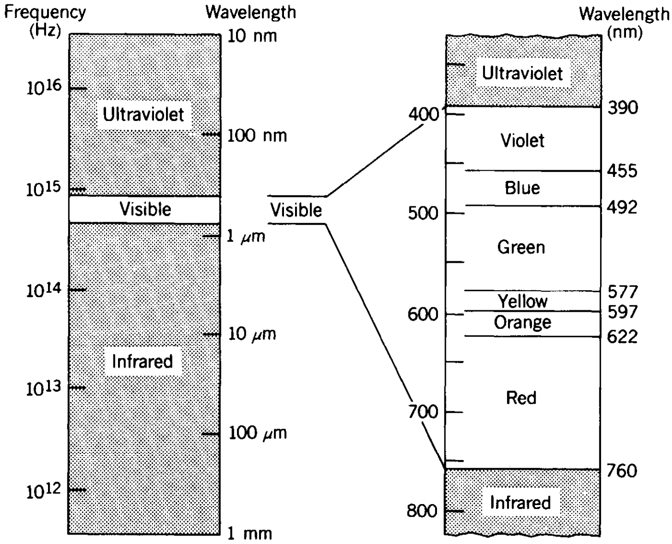


Figure 2.0-1 Optical frequencies and wavelengths.

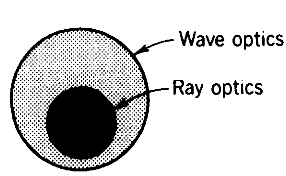


Figure 2.0-2 Wave optics encompasses ray optics. Ray optics is the limit of wave optics when the wavelength is very short.

not specified; it suffices to say at this point that it may represent any of the components of the electric or magnetic fields (as described in Chap. 5, which covers the electromagnetic theory of light). This, and a relation between the optical power density and the wavefunction, constitute the postulates of the scalar wave model of light, hereafter called **wave optics**. The consequences of these simple postulates are many and far reaching. Wave optics constitutes a basis for describing a host of optical phenomena that fall outside the confines of ray optics, including interference and diffraction, as demonstrated in this and the following two chapters.

Wave optics has its limitations. It is not capable of providing a complete picture of the reflection and refraction of light at the boundaries between dielectric materials, nor of explaining those optical phenomena that require a vector formulation, such as polarization effects. In Chap. 5 the electromagnetic theory of light is presented and the conditions under which scalar wave optics provides a good approximation to certain electromagnetic phenomena are elucidated.

This chapter begins with the postulates of wave optics (Sec. 2.1). In Secs. 2.2 to 2.5 we consider monochromatic waves, and polychromatic light is discussed in Sec. 2.6. Elementary waves, such as the plane wave and the spherical wave, are introduced in Sec. 2.2. Section 2.3 establishes that ray optics can be derived from wave optics. The transmission of optical waves through simple optical components such as mirrors, prisms, lenses, and gratings is examined in Sec. 2.4. Interference, an important manifestation of the wave nature of light, is the subject of Secs. 2.5 and 2.6.

2.1 POSTULATES OF WAVE OPTICS

The Wave Equation

Light propagates in the form of waves. In free space, light waves travel with speed c_o. A homogeneous transparent medium such as glass is characterized by a single constant, its refractive index n (≥ 1). In a medium of refractive index n, light waves travel with a reduced speed

$$\boxed{c = \frac{c_o}{n}.} \tag{2.1-1}$$

Speed of Light in a Medium

An optical wave is described mathematically by a real function of position $\mathbf{r} = (x, y, z)$ and time t, denoted $u(\mathbf{r}, t)$ and known as the **wavefunction**. It satisfies the wave equation,

$$\boxed{\nabla^2 u - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0,} \tag{2.1-2}$$

The Wave Equation

where ∇^2 is the Laplacian operator, $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. Any function satisfying (2.1-2) represents a possible optical wave.

Because the wave equation is linear, the **principle of superposition** applies; i.e., if $u_1(\mathbf{r}, t)$ and $u_2(\mathbf{r}, t)$ represent optical waves, then $u(\mathbf{r}, t) = u_1(\mathbf{r}, t) + u_2(\mathbf{r}, t)$ also represents a possible optical wave.

At the boundary between two different media, the wavefunction changes in a way that depends on the refractive indices. However, the laws that govern this change depend on the physical significance assigned to the wavefunction (i.e., the component of the electromagnetic field it represents), as discussed in Chap. 5.

The wave equation is approximately applicable to media with position-dependent refractive indices, provided that the variation is slow within distances of a wavelength. The medium is then said to be locally homogeneous. For such media, n in (2.1-1) and c in (2.1-2) are simply replaced by position-dependent functions $n(\mathbf{r})$ and $c(\mathbf{r})$, respectively.

Intensity, Power, and Energy

The optical **intensity** $I(\mathbf{r}, t)$, defined as the optical power per unit area (units of watts/cm^2), is proportional to the average of the squared wavefunction,

$$\boxed{I(\mathbf{r}, t) = 2\langle u^2(\mathbf{r}, t)\rangle.} \tag{2.1-3}$$

Optical Intensity

The operation $\langle \cdot \rangle$ denotes averaging over a time interval that is much longer than the time of an optical cycle, but much shorter than any other time of interest (the duration of a pulse of light, for example). The duration of an optical cycle is extremely short: 2×10^{-15} s = 2 fs for light of wavelength 600 nm, as an example. This concept is explained further in Sec. 2.6.

Although the physical meaning of the wavefunction $u(\mathbf{r}, t)$ has not been specified, (2.1-3) represents its connection with a physically measurable quantity—the optical intensity. There is some arbitrariness in the definition of the wavefunction and its relation to the intensity. Equation (2.1-3) could have, for example, been written without the factor 2 and the wavefunction scaled by a factor $\sqrt{2}$, so that the intensity remains the same. The choice of the factor 2 will later prove convenient, however.

The optical **power** $P(t)$ (units of watts) flowing into an area A normal to the direction of propagation of light is the integrated intensity

$$P(t) = \int_A I(\mathbf{r}, t)\, dA. \tag{2.1-4}$$

The optical **energy** (units of joules) collected in a given time interval is the time integral of the optical power over the time interval.

2.2 MONOCHROMATIC WAVES

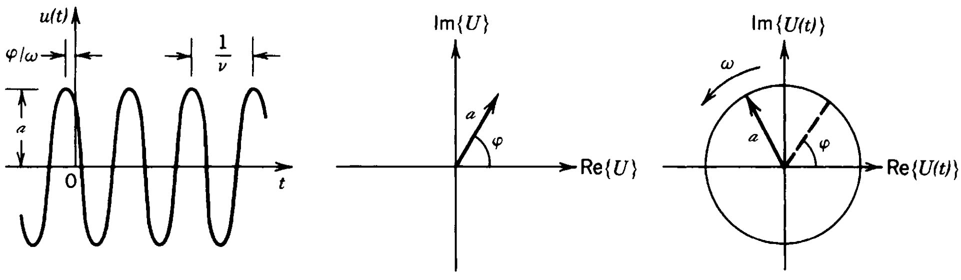
A monochromatic wave is represented by a wavefunction with harmonic time dependence,

$$u(\mathbf{r}, t) = a(\mathbf{r}) \cos[2\pi\nu t + \varphi(\mathbf{r})], \tag{2.2-1}$$

as shown in Fig. 2.2-1(a), where

$a(\mathbf{r})$ = amplitude
$\varphi(\mathbf{r})$ = phase
ν = frequency (cycles/s or Hz)
$\omega = 2\pi\nu$ = angular frequency (radians/s).

Both the amplitude and the phase are generally position dependent, but the wavefunc-



(a) (b) (c)

Figure 2.2-1 Representations of a monochromatic wave at a fixed position $\mathbf{r}$: (a) the wavefunction $u(t)$ is a harmonic function of time; (b) the complex amplitude $U = a \exp(j\varphi)$ is a fixed phasor; (c) the complex wavefunction $U(t) = U \exp(j2\pi\nu t)$ is a phasor rotating with angular velocity $\omega = 2\pi\nu$ radians/s.

tion is a harmonic function of time with frequency ν at all positions. The frequency of optical waves lies in the range 3×10^{11} to 3×10^{16} Hz, as depicted in Fig. 2.0-1.

A. Complex Representation and the Helmholtz Equation

Complex Wavefunction

It is convenient to represent the real wavefunction $u(\mathbf{r}, t)$ in (2.2-1) in terms of a complex function

$$U(\mathbf{r}, t) = a(\mathbf{r}) \exp[j\varphi(\mathbf{r})] \exp(j2\pi\nu t), \tag{2.2-2}$$

so that

$$u(\mathbf{r}, t) = \operatorname{Re}\{U(\mathbf{r}, t)\} = \tfrac{1}{2}[U(\mathbf{r}, t) + U^*(\mathbf{r}, t)]. \tag{2.2-3}$$

The function $U(\mathbf{r}, t)$, known as the **complex wavefunction**, describes the wave completely; the wavefunction $u(\mathbf{r}, t)$ is simply its real part. Like the wavefunction $u(\mathbf{r}, t)$, the complex wavefunction $U(\mathbf{r}, t)$ must also satisfy the wave equation,

$$\boxed{\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = 0.} \tag{2.2-4}$$

The Wave Equation

The two functions satisfy the same boundary conditions.

Complex Amplitude

Equation (2.2-2) may be written in the form

$$U(\mathbf{r}, t) = U(\mathbf{r}) \exp(j2\pi\nu t), \tag{2.2-5}$$

where the time-independent factor $U(\mathbf{r}) = a(\mathbf{r}) \exp[j\varphi(\mathbf{r})]$ is referred to as the **complex**

amplitude. The wavefunction $u(\mathbf{r}, t)$ is therefore related to the complex amplitude by

$$u(\mathbf{r}, t) = \operatorname{Re}\{U(\mathbf{r}) \exp(j2\pi\nu t)\} = \tfrac{1}{2}[U(\mathbf{r}) \exp(j2\pi\nu t) + U^*(\mathbf{r}) \exp(-j2\pi\nu t)]. \tag{2.2-6}$$

At a given position $\mathbf{r}$, the complex amplitude $U(\mathbf{r})$ is a complex variable [depicted in Fig. 2.2-1(b)] whose magnitude $|U(\mathbf{r})| = a(\mathbf{r})$ is the amplitude of the wave and whose argument $\arg\{U(\mathbf{r})\} = \varphi(\mathbf{r})$ is the phase. The complex wavefunction $U(\mathbf{r}, t)$ is represented graphically by a phasor rotating with angular velocity $\omega = 2\pi\nu$ radians/s [Fig. 2.2-1(c)]. Its initial value at $t = 0$ is the complex amplitude $U(\mathbf{r})$.

The Helmholtz Equation

Substituting $U(\mathbf{r}, t) = U(\mathbf{r}) \exp(j2\pi\nu t)$ into the wave equation (2.2-4), we obtain the differential equation

$$\boxed{(\nabla^2 + k^2)U(\mathbf{r}) = 0,} \tag{2.2-7}$$

Helmholtz Equation

called the **Helmholtz equation**, where

$$\boxed{k = \frac{2\pi\nu}{c} = \frac{\omega}{c}} \tag{2.2-8}$$

Wavenumber

is referred to as the **wavenumber**.

Optical Intensity

The optical intensity is determined by use of (2.1-3). When

$$\begin{aligned} 2u^2(\mathbf{r}, t) &= 2a^2(\mathbf{r}) \cos^2[2\pi\nu t + \varphi(r)] \\ &= |U(\mathbf{r})|^2\{1 + \cos(2[2\pi\nu t + \varphi(\mathbf{r})])\} \end{aligned} \tag{2.2-9}$$

is averaged over a time longer than an optical period, $1/\nu$, the second term of (2.2-9) vanishes, so that

$$\boxed{I(\mathbf{r}) = |U(\mathbf{r})|^2.} \tag{2.2-10}$$

Optical Intensity

Thus *the optical intensity of a monochromatic wave is the absolute square of its complex amplitude*. The intensity of a monochromatic wave does *not* vary with time.

Wavefronts

The wavefronts are the surfaces of equal phase, $\varphi(\mathbf{r}) =$ constant. The constants are often taken to be multiples of 2π, $\varphi(\mathbf{r}) = 2\pi q$, where q is an integer. The wavefront normal at position $\mathbf{r}$ is parallel to the gradient vector $\nabla\varphi(\mathbf{r})$ (a vector with components $\partial\varphi/\partial x$, $\partial\varphi/\partial y$, and $\partial\varphi/\partial z$ in a Cartesian coordinate system). It represents the direction at which the rate of change of the phase is maximum.

Summary

- A monochromatic wave of frequency ν is described by a *complex wavefunction* $U(\mathbf{r}, t) = U(\mathbf{r})\exp(j2\pi\nu t)$, which satisfies the wave equation.
- The *complex amplitude* $U(\mathbf{r})$ satisfies the Helmholtz equation; its magnitude $|U(\mathbf{r})|$ and argument $\arg\{U(\mathbf{r})\}$ are the *amplitude* and *phase* of the wave, respectively. The optical *intensity* is $I(\mathbf{r}) = |U(\mathbf{r})|^2$. The *wavefronts* are the surfaces of constant phase, $\varphi(\mathbf{r}) = \arg\{U(\mathbf{r})\} = 2\pi q$ (q = integer).
- The *wavefunction* $u(\mathbf{r}, t)$ is the real part of the complex wavefunction, $u(\mathbf{r}, t) = \mathrm{Re}\{U(\mathbf{r}, t)\}$. The wavefunction also satisfies the wave equation.

B. Elementary Waves

The simplest solutions of the Helmholtz equation in a homogeneous medium are the plane wave and the spherical wave.

The Plane Wave

The plane wave has complex amplitude

$$U(\mathbf{r}) = A\exp(-j\mathbf{k}\cdot\mathbf{r}) = A\exp\left[-j(k_x x + k_y y + k_z z)\right], \qquad (2.2\text{-}11)$$

where A is a complex constant called the **complex envelope** and $\mathbf{k} = (k_x, k_y, k_z)$ is called the **wavevector**. For (2.2-11) to satisfy the Helmholtz equation (2.2-7), $k_x^2 + k_y^2 + k_z^2 = k^2$, so that the magnitude of the wavevector $\mathbf{k}$ is the wavenumber k.

Since the phase $\arg\{U(\mathbf{r})\} = \arg\{A\} - \mathbf{k}\cdot\mathbf{r}$, the wavefronts obey $\mathbf{k}\cdot\mathbf{r} = k_x x + k_y y + k_z z = 2\pi q + \arg\{A\}$ (q = integer). This is the equation describing parallel planes perpendicular to the wavevector $\mathbf{k}$ (hence the name "plane wave"). These planes are separated by a distance $\lambda = 2\pi/k$, so that

$$\boxed{\lambda = \frac{c}{\nu},} \qquad (2.2\text{-}12)$$

Wavelength

where λ is called the **wavelength**. The plane wave has a constant intensity $I(\mathbf{r}) = |A|^2$ everywhere in space so that it carries infinite power. This wave is clearly an idealization since it exists everywhere and at all times.

If the z axis is taken in the direction of the wavevector $\mathbf{k}$, then $U(\mathbf{r}) = A\exp(-jkz)$ and the corresponding wavefunction obtained from (2.2-6) is

$$u(\mathbf{r}, t) = |A|\cos\left[2\pi\nu t - kz + \arg\{A\}\right] = |A|\cos\left[2\pi\nu(t - z/c) + \arg\{A\}\right]. \qquad (2.2\text{-}13)$$

The wavefunction is therefore periodic in time with period $1/\nu$, and periodic in space with period $2\pi/k$, which is equal to the wavelength λ (see Fig. 2.2-2). Since the phase of the complex wavefunction, $\arg\{U(\mathbf{r}, t)\} = 2\pi\nu(t - z/c) + \arg\{A\}$, varies with time

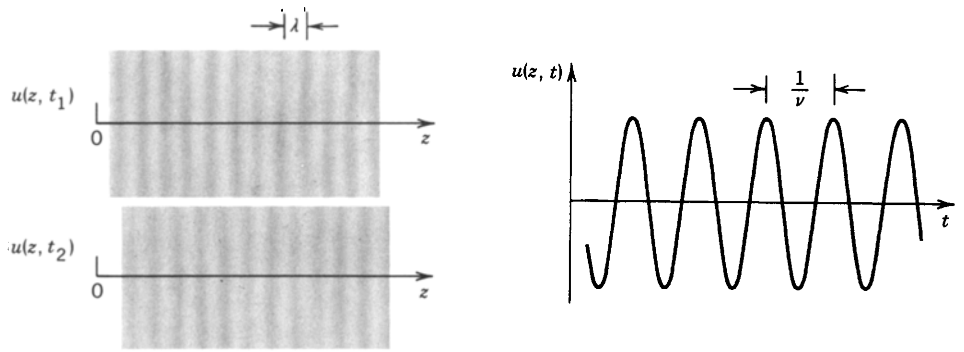


Figure 2.2-2 A plane wave traveling in the z direction is a periodic function of z with spatial period λ and a periodic function of t with temporal period $1/\nu$.

and position as a function of the variable $t - z/c$ (see Fig. 2.2-2), c is called the **phase velocity** of the wave.

In a medium of refractive index n, the phase velocity $c = c_o/n$ and the wavelength $\lambda = c/\nu = c_o/n\nu$, so that $\lambda = \lambda_o/n$ where $\lambda_o = c_o/\nu$ is the wavelength in free space. For a given frequency ν, the wavelength in the medium is reduced relative to that in free space by the factor n. As a consequence, the wavenumber $k = 2\pi/\lambda$ increases relative to that in free space ($k_o = 2\pi/\lambda_o$) by the factor n. In summary: *As a monochromatic wave propagates through media of different refractive indices its frequency remains the same, but its velocity, wavelength, and wavenumber are altered*:

$$\boxed{c = \frac{c_o}{n}, \qquad \lambda = \frac{\lambda_o}{n}, \qquad k = nk_o.} \tag{2.2-14}$$

The wavelengths shown in Fig. 2.0-1 are in free space ($n = 1$).

The Spherical Wave

Another simple solution of the Helmholtz equation is the spherical wave

$$U(\mathbf{r}) = \frac{A}{r}\exp(-jkr), \tag{2.2-15}$$

where r is the distance from the origin and $k = 2\pi\nu/c = \omega/c$ is the wavenumber. The intensity $I(\mathbf{r}) = |A|^2/r^2$ is inversely proportional to the square of the distance. Taking $\arg\{A\} = 0$ for simplicity, the wavefronts are the surfaces $kr = 2\pi q$ or $r = q\lambda$, where q is an integer. These are concentric spheres separated by a radial distance $\lambda = 2\pi/k$ that advance radially at the phase velocity c (Fig. 2.2-3).

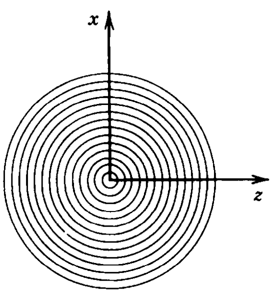


Figure 2.2-3 Cross section of the wavefronts of a spherical wave.

A spherical wave originating at the position $\mathbf{r}_0$ has a complex amplitude $U(\mathbf{r}) = (A/|\mathbf{r}-\mathbf{r}_0|)\exp(-jk|\mathbf{r}-\mathbf{r}_0|)$. Its wavefronts are spheres centered about $\mathbf{r}_0$. A wave with complex amplitude $U(\mathbf{r}) = (A/r)\exp(+jkr)$ is a spherical wave traveling inwardly (toward the origin) instead of outwardly (away from the origin).

Fresnel Approximation of the Spherical Wave; The Paraboloidal Wave

Let us examine a spherical wave originating at $\mathbf{r} = \mathbf{0}$ at points $\mathbf{r} = (x, y, z)$ sufficiently close to the z axis but far from the origin, so that $(x^2 + y^2)^{1/2} \ll z$. The paraxial approximation of ray optics (see Sec. 1.2) would be applicable were these points the endpoints of rays beginning at the origin. Denoting $\theta^2 = (x^2 + y^2)/z^2 \ll 1$, we use an approximation based on the Taylor series expansion

$$r = (x^2 + y^2 + z^2)^{1/2} = z(1 + \theta^2)^{1/2} = z\left(1 + \frac{\theta^2}{2} - \frac{\theta^4}{8} + \cdots\right)$$

$$\approx z\left(1 + \frac{\theta^2}{2}\right) = z + \frac{x^2 + y^2}{2z}.$$

Substituting $r = z + (x^2 + y^2)/2z$ into the phase, and $r = z$ into the magnitude of $U(\mathbf{r})$ in (2.2-15), we obtain

$$\boxed{U(\mathbf{r}) \approx \frac{A}{z}\exp(-jkz)\exp\left[-jk\frac{x^2 + y^2}{2z}\right].} \tag{2.2-16}$$

Fresnel Approximation of a Spherical Wave

A more accurate value of r was used in the phase since the sensitivity to errors of the phase is greater. This is called the **Fresnel approximation**. It plays an important role in simplifying the theory of transmission of optical waves through apertures (**diffraction**), as discussed in Chap. 4.

The complex amplitude in (2.2-16) may be viewed as representing a plane wave $A\exp(-jkz)$ modulated by the factor $(1/z)\exp[-jk(x^2 + y^2)/2z]$, which involves a phase $k(x^2 + y^2)/2z$. This phase factor serves to bend the planar wavefronts of the plane wave into paraboloidal surfaces (Fig. 2.2-4), since the equation of a paraboloid of revolution is $(x^2 + y^2)/z$ = constant. Thus the spherical wave is approximated by a **paraboloidal wave**. When z becomes very large, the phase in (2.2-16) approaches kz and the magnitude varies slowly with z, so that the spherical wave eventually resembles the plane wave $\exp(-jkz)$, as illustrated in Fig. 2.2-4.

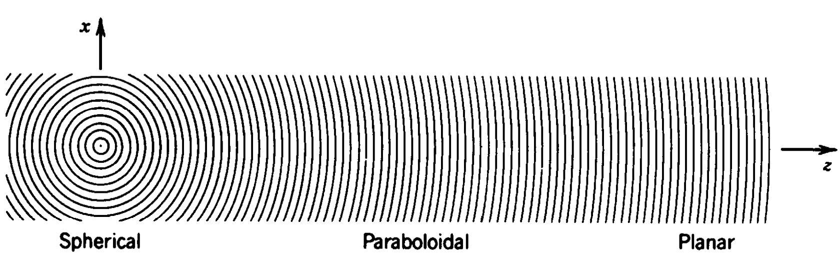


Figure 2.2-4 A spherical wave may be approximated at points near the z axis and sufficiently far from the origin by a paraboloidal wave. For very far points, the spherical wave approaches the plane wave.

The condition of validity of the Fresnel approximation is *not* simply that $\theta^2 \ll 1$. Although the third term of the series expansion, $\theta^4/8$, may be very small in comparison with the second and first terms, when multiplied by kz it may become comparable to π. The approximation is therefore valid when $kz\theta^4/8 \ll \pi$, or $(x^2 + y^2)^2 \ll 4z^3\lambda$. For points (x, y) lying within a circle of radius a centered about the z axis, the validity condition is $a^4 \ll 4z^3\lambda$ or

$$\frac{N_F \theta_m^2}{4} \ll 1, \tag{2.2-17}$$

where $\theta_m = a/z$ is the maximum angle and

$$\boxed{N_F = \frac{a^2}{\lambda z}} \tag{2.2-18}$$

Fresnel Number

is known as the **Fresnel number**.

EXERCISE 2.2-1

Validity of the Fresnel Approximation. Determine the radius of a circle within which a spherical wave of wavelength $\lambda = 633$ nm, originating at a distance 1 m away, may be approximated by a paraboloidal wave. Determine the maximum angle θ_m and the Fresnel number N_F.

C. Paraxial Waves

A wave is said to be paraxial if its wavefront normals are paraxial rays. One way of constructing a paraxial wave is to start with a plane wave $A \exp(-jkz)$, regard it as a "carrier" wave, and modify or "modulate" its complex envelope A, making it a slowly varying function of position $A(\mathbf{r})$ so that the complex amplitude of the modulated wave becomes

$$U(\mathbf{r}) = A(\mathbf{r}) \exp(-jkz). \tag{2.2-19}$$

The variation of $A(\mathbf{r})$ with position must be slow within the distance of a wavelength $\lambda = 2\pi/k$, so that the wave approximately maintains its underlying plane-wave nature.

The wavefunction $u(\mathbf{r}, t) = |A(\mathbf{r})| \cos[2\pi\nu t - kz + \arg\{A(\mathbf{r})\}]$ of a paraxial wave is sketched in Fig. 2.2-5(a) as a function of z at $t = 0$ and $x = y = 0$. This is a sinusoidal function of z with amplitude $|A(0, 0, z)|$ and phase $\arg\{A(0, 0, z)\}$ that vary slowly with z. Since the change of the phase $\arg\{A(x, y, z)\}$ is small within the distance of a wavelength, the planar wavefronts, $kz = 2\pi q$, of the carrier plane wave bend only slightly, so that their normals are paraxial rays [Fig. 2.2-5(b)].

The Paraxial Helmholtz Equation

For the paraxial wave (2.2-19) to satisfy the Helmholtz equation (2.2-7), the complex envelope $A(\mathbf{r})$ must satisfy another partial differential equation obtained by substituting (2.2-19) into (2.2-7). The assumption that $A(\mathbf{r})$ varies slowly with respect to z signifies that within a distance $\Delta z = \lambda$, the change ΔA is much smaller than A itself;

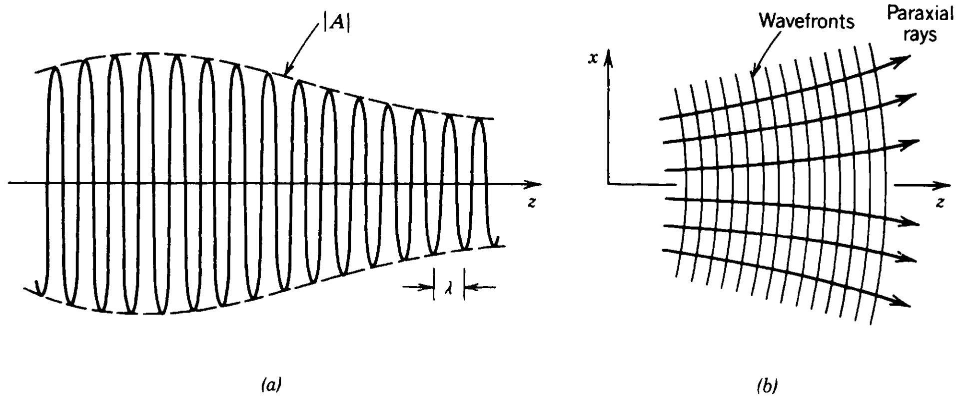


Figure 2.2-5 (a) The magnitude of a paraxial wave as a function of the axial distance z. (b) The wavefronts and wavefront normals of a paraxial wave.

i.e., $\Delta A \ll A$. This inequality of complex variables applies to the magnitudes of the real and imaginary parts separately. Since $\Delta A = (\partial A/\partial z)\,\Delta z = (\partial A/\partial z)\lambda$, it follows that $\partial A/\partial z \ll A/\lambda = Ak/2\pi$, and therefore

$$\frac{\partial A}{\partial z} \ll kA. \tag{2.2-20}$$

Similarly, the derivative $\partial A/\partial z$ varies slowly within the distance λ, so that $\partial^2 A/\partial^2 z \ll k\,\partial A/\partial z$, and therefore

$$\frac{\partial^2 A}{\partial z^2} \ll k^2 A. \tag{2.2-21}$$

Substituting (2.2-19) into (2.2-7) and neglecting $\partial^2 A/\partial z^2$ in comparison with $k\,\partial A/\partial z$ or $k^2 A$, we obtain

$$\boxed{\nabla_T^2 A - j2k\frac{\partial A}{\partial z} = 0,} \tag{2.2-22}$$

Paraxial Helmholtz Equation

where $\nabla_T^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse Laplacian operator.

Equation (2.2-22) is the **slowly varying envelope approximation** of the Helmholtz equation. We shall simply call it the **paraxial Helmholtz equation.** It is a partial differential equation that resembles the Schrödinger equation of quantum physics.

The simplest solution of the paraxial Helmholtz equation is the paraboloidal wave (Exercise 2.2-2), which is the paraxial approximation of the spherical wave. The most interesting and useful solution, however, is the **Gaussian beam,** to which Chap. 3 is devoted.

EXERCISE 2.2-2

The Paraboloidal Wave and the Gaussian Beam. Verify that a paraboloidal wave with the complex envelope $A(\mathbf{r}) = (A_0/z)\exp[-jk(x^2 + y^2)/2z]$ [see (2.2-16)] satisfies the paraxial Helmholtz equation (2.2-22). Show that the wave with complex amplitude $A(\mathbf{r}) =$

$[A/q(z)] \exp[-jk(x^2+y^2)/2q(z)]$, where $q(z) = z + jz_0$ and z_0 is a constant, also satisfies the paraxial Helmholtz equation. This wave, called the Gaussian beam, is the subject of Chap. 3. Sketch the intensity of the Gaussian beam in the plane $z = 0$.

*2.3 RELATION BETWEEN WAVE OPTICS AND RAY OPTICS

We proceed to show that ray optics is the limit of wave optics when the wavelength $\lambda_o \to 0$. Consider a monochromatic wave of free-space wavelength λ_o in a medium with refractive index $n(\mathbf{r})$ that varies sufficiently slowly with position so that the medium may be regarded as locally homogeneous. We write the complex amplitude in the form

$$U(\mathbf{r}) = a(\mathbf{r}) \exp[-jk_o S(\mathbf{r})], \tag{2.3-1}$$

where $a(\mathbf{r})$ is its magnitude, $-k_o S(\mathbf{r})$ its phase, and $k_o = 2\pi/\lambda_o$ is the wavenumber. We assume that $a(\mathbf{r})$ varies sufficiently slowly with $\mathbf{r}$, so that it may be regarded as constant within the distance of a wavelength λ_o.

The wavefronts are the surfaces $S(\mathbf{r}) =$ constant and the wavefront normals point in the direction of the gradient ∇S. In the neighborhood of a given position $\mathbf{r}_0$, the wave can be locally regarded as a plane wave with amplitude $a(\mathbf{r}_0)$ and wavevector $\mathbf{k}$ with magnitude $k = n(\mathbf{r}_0)k_o$ and direction parallel to the gradient vector ∇S at $\mathbf{r}_0$. A different neighborhood exhibits a local plane wave of different amplitude and different wavevector.

In Chap. 1 it was shown that the optical rays are normal to the equilevel surfaces of a function $S(\mathbf{r})$ called the eikonal (see Sec. 1.3C). We therefore associate the local wavevectors (wavefront normals) in wave optics with the rays of ray optics and recognize that the function $S(\mathbf{r})$, which is proportional to the phase of the wave, is nothing but the eikonal of ray optics (Fig. 2.3-1). This association has a formal mathematical basis, as will be demonstrated subsequently. With this analogy, ray optics can serve to determine the approximate effects of optical components on the wavefront normals, as illustrated in Fig. 2.3-1.

The Eikonal Equation

Substituting (2.3-1) into the Helmholtz equation, (2.2-7) provides

$$k_o^2\left[n^2 - |\nabla S|^2\right] a + \nabla^2 a - jk_o[2\nabla S \cdot \nabla a + a\nabla^2 S] = 0, \tag{2.3-2}$$

where $a = a(\mathbf{r})$ and $S = S(\mathbf{r})$. The real and imaginary parts of the left-hand side of (2.3-2) must both vanish. Equating the real part to zero and using $k_o = 2\pi/\lambda_o$, we obtain

$$|\nabla S|^2 = n^2 + \left(\frac{\lambda_o}{2\pi}\right)^2 \nabla^2 a / a. \tag{2.3-3}$$

The assumption that a varies slowly over the distance λ_o means that $\lambda_o^2 \nabla^2 a / a \ll 1$,

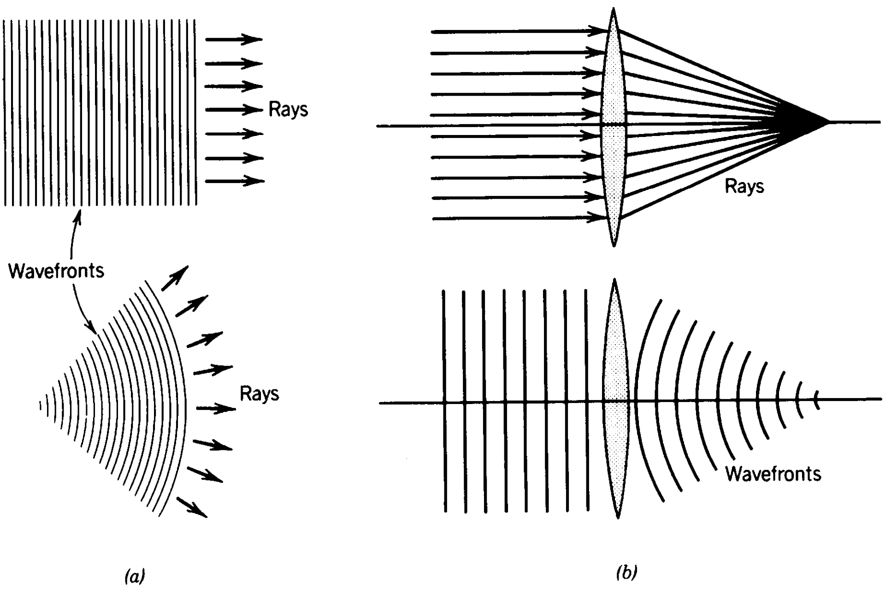


Figure 2.3-1 (*a*) The rays of ray optics are orthogonal to the wavefronts of wave optics (see also Fig. 1.3-10). (*b*) The effect of a lens on rays and wavefronts.

so that the second term of the right-hand side may be neglected in the limit $\lambda_o \to 0$ and

$$\boxed{|\nabla S|^2 \approx n^2.} \tag{2.3-4}$$

Eikonal Equation

This is the eikonal equation (1.3-18), which may be regarded as the main postulate of ray optics (Fermat's principle can be derived from the eikonal equation, and vice versa).

In conclusion: The scalar function $S(\mathbf{r})$, which is proportional to the phase in wave optics, is the eikonal of ray optics. This is also consistent with the observation that in ray optics $S(\mathbf{r}_B) - S(\mathbf{r}_A)$ equals the optical path length between the points $\mathbf{r}_A$ and $\mathbf{r}_B$.

The eikonal equation is the limit of the Helmholtz equation when $\lambda_o \to 0$. Given $n(\mathbf{r})$ we may use the eikonal equation to determine $S(r)$. By equating the imaginary part of (2.3-2) to zero, we obtain a relation between $\mathscr{a}$ and S, thereby permitting us to determine the wavefunction.

2.4 SIMPLE OPTICAL COMPONENTS

In this section we examine the effects of optical components, such as mirrors, transparent plates, prisms, and lenses, on optical waves.

A. Reflection and Refraction

Reflection from a Planar Mirror

A plane wave of wavevector $\mathbf{k}_1$ is incident onto a planar mirror located in free space in the $z = 0$ plane. A reflected plane wave of wavevector $\mathbf{k}_2$ is created. The angles of

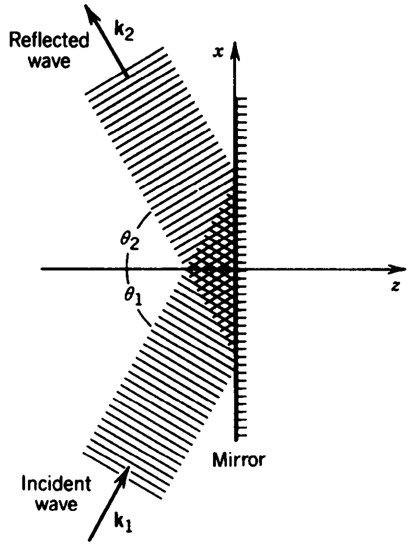


Figure 2.4-1 Reflection of a plane wave from a planar mirror. Phase matching at the surface of the mirror requires that the angles of incidence and reflection be equal.

incidence and reflection are θ_1 and θ_2, as illustrated in Fig. 2.4-1. The sum of the two waves satisfies the Helmholtz equation if $k_1 = k_2 = k_o$. Certain boundary conditions must be satisfied at the surface of the mirror. Since these conditions are the same at all points (x, y), it is necessary that the wavefronts of the two waves match, i.e., the phases must be equal,

$$\mathbf{k}_1 \cdot \mathbf{r} = \mathbf{k}_2 \cdot \mathbf{r} \qquad \text{for all } \mathbf{r} = (x, y, 0), \tag{2.4-1}$$

or differ by a constant.

Substituting $\mathbf{r} = (x, y, 0)$, $\mathbf{k}_1 = (k_o \sin\theta_1, 0, k_o \cos\theta_1)$, and $\mathbf{k}_2 = (k_o \sin\theta_2, 0, -k_o \cos\theta_2)$ into (2.4-1), we obtain $k_o \sin(\theta_1)x = k_o \sin(\theta_2)x$, from which $\theta_1 = \theta_2$, so that the angles of incidence and reflection must be equal. Thus the law of reflection of optical rays is applicable to the wavevectors of plane waves.

Reflection and Refraction at a Planar Dielectric Boundary

We now consider a plane wave of wavevector $\mathbf{k}_1$ incident on a planar boundary between two homogeneous media of refractive indices n_1 and n_2. The boundary lies in the $z = 0$ plane (Fig. 2.4-2). Refracted and reflected plane waves of wavevectors $\mathbf{k}_2$ and $\mathbf{k}_3$ emerge. The combination of the three waves satisfies the Helmholtz equation everywhere if each of the waves has the appropriate wavenumber in the medium in which it propagates ($k_1 = k_3 = n_1 k_o$ and $k_2 = n_2 k_o$).

Since the boundary conditions are invariant to x and y, it is necessary that the wavefronts of the three waves match, i.e., the phases must be equal,

$$\mathbf{k}_1 \cdot \mathbf{r} = \mathbf{k}_2 \cdot \mathbf{r} = \mathbf{k}_3 \cdot \mathbf{r} \qquad \text{for all } \mathbf{r} = (x, y, 0), \tag{2.4-2}$$

or differ by constants. Since $\mathbf{k}_1 = (n_1 k_o \sin\theta_1, 0, n_1 k_o \cos\theta_1)$, $\mathbf{k}_3 = (n_1 k_o \sin\theta_3, 0, -n_1 k_o \cos\theta_3)$, and $\mathbf{k}_2 = (n_2 k_o \sin\theta_2, 0, n_2 k_o \cos\theta_2)$, where θ_1, θ_2, and θ_3 are the angles of incidence, refraction, and reflection, respectively, it follows from (2.4-2) that $\theta_1 = \theta_3$ and $n_1 \sin\theta_1 = n_2 \sin\theta_2$. These are the laws of reflection and refraction (Snell's law) of ray optics, now applicable to the wavevectors.

It is not possible to determine the amplitudes of the reflected and refracted waves using the scalar wave theory of light since the boundary conditions are not completely specified in this theory. This will be achieved in Sec. 6.2 using the electromagnetic theory of light (Chaps. 5 and 6).

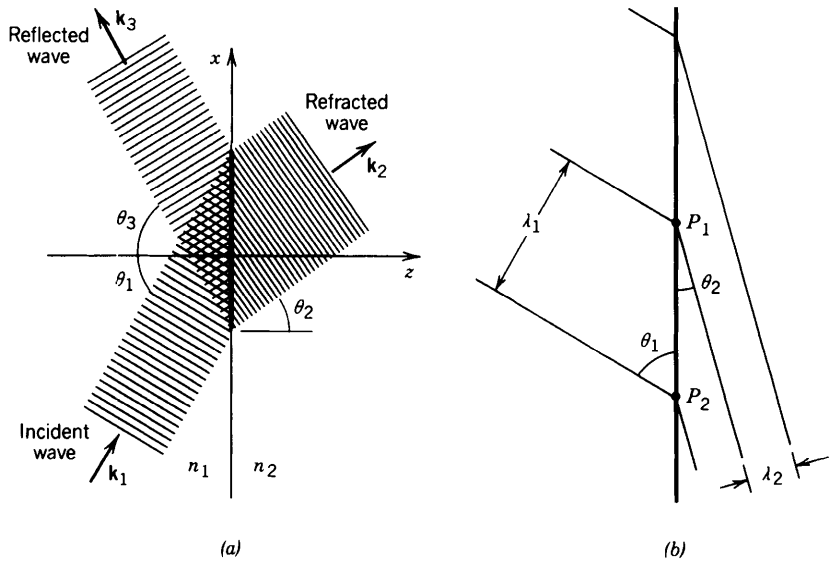


Figure 2.4-2 (a) Reflection and refraction of a plane wave at a dielectric boundary. (b) Matching the wavefronts at the boundary; the distance $\overline{P_1P_2}$ for the incident wave, $\lambda_1/\sin\theta_1 = \lambda_o/n_1 \sin\theta_1$, equals that for the refracted wave, $\lambda_2/\sin\theta_2 = \lambda_o/n_2 \sin\theta_2$, from which Snell's law follows.

B. Transmission Through Optical Components

We now proceed to examine the transmission of optical waves through transparent optical components such as plates, prisms, and lenses. The effect of reflection at the surfaces of these components will be ignored, since it cannot be properly accounted for using the scalar wave-optics model of light. The effect of absorption in the material is also ignored and relegated to Sec. 5.5. The main emphasis here is on the phase shift introduced by these components and on the associated wavefront bending.

Transmission Through a Transparent Plate

Consider first the transmission of a plane wave through a transparent plate of refractive index n and thickness d surrounded by free space. The surfaces of the plate are the planes $z = 0$ and $z = d$ and the incident wave travels in the z direction (Fig. 2.4-3). Let $U(x, y, z)$ be the complex amplitude of the wave. Since external and internal reflections are ignored, $U(x, y, z)$ is assumed to be continuous at the bound-

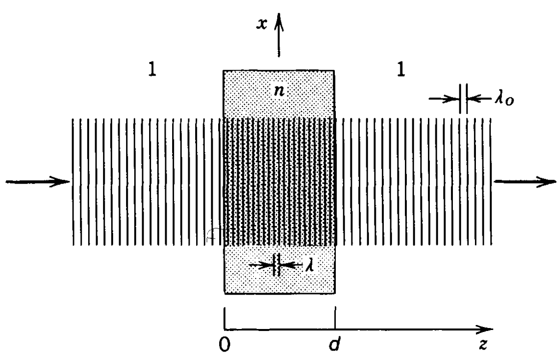


Figure 2.4-3 Transmission of a plane wave through a transparent plate.

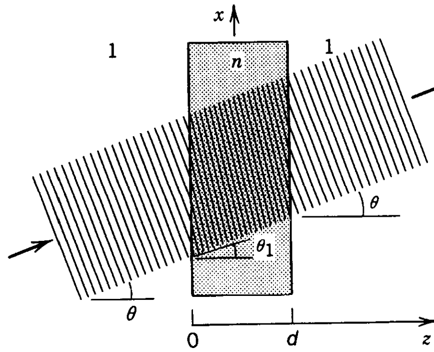


Figure 2.4-4 Transmission of an oblique plane wave through a thin transparent plate.

aries. The ratio $t(x, y) = U(x, y, d)/U(x, y, 0)$ therefore represents the **complex amplitude transmittance** of the plate. The effect of reflection is considered in Sec. 6.2 and the effect of multiple internal reflections within the plate is examined in Sec. 9.1.

The incident plane wave continues to propagate inside the plate as a plane wave with wavenumber nk_o, so that $U(x, y, z)$ is proportional to $\exp(-jnk_o z)$. Thus $U(x, y, d)/U(x, y, 0) = \exp(-jnk_o d)$, so that

$$\boxed{t(x, y) = \exp(-jnk_o d),} \tag{2.4-3}$$

Complex Amplitude Transmittance of a Transparent Plate

i.e., the plate introduces a phase shift $nk_o d = 2\pi(d/\lambda)$.

If the incident plane wave makes an angle θ with the z axis and has wavevector $\mathbf{k}$ (Fig. 2.4-4), the refracted and transmitted waves are also plane waves with wavevectors $\mathbf{k}_1$ and $\mathbf{k}$ and angles θ_1 and θ, respectively, where θ_1 and θ are related by Snell's law, $\sin\theta = n \sin\theta_1$. The complex amplitude $U(x, y, z)$ inside the plate is now proportional to $\exp(-j\mathbf{k}_1 \cdot \mathbf{r}) = \exp[-jnk_o(z\cos\theta_1 + x\sin\theta_1)]$, so that the complex amplitude transmittance of the plate $U(x, y, d)/U(x, y, 0)$ is

$$t(x, y) = \exp\left[-jnk_o(d\cos\theta_1 + x\sin\theta_1)\right].$$

If the angle of incidence θ is small (i.e., the incident wave is *paraxial*), then $\theta_1 \approx \theta/n$ is also small and the approximations $\sin\theta \approx \theta$ and $\cos\theta \approx 1 - \frac{1}{2}\theta^2$ yield $t(x, y) \approx \exp(-jnk_o d)\exp(jk_o\theta^2 d/2n - jk_o\theta x)$. If the plate is sufficiently *thin* and the angle θ is sufficiently small such that $k_o\theta^2 d/2n \ll 2\pi$ [or $(d/\lambda_o)\theta^2/2n \ll 1$] and if $(x/\lambda)\theta \ll 1$ for all values of x of interest, then the transmittance of the plate may be approximated by (2.4-3). Under these conditions the transmittance of the plate is approximately independent of the angle θ.

Thin Transparent Plate of Varying Thickness

We now determine the amplitude transmittance of a thin transparent plate whose thickness $d(x, y)$ varies smoothly as a function of x and y, assuming that the incident wave is an arbitrary *paraxial* wave. The plate lies between the planes $z = 0$ and $z = d_0$, which are regarded as the boundaries of the optical component (Fig. 2.4-5).

In the vicinity of the position $(x, y, 0)$ the incident paraxial wave may be regarded locally as a plane wave traveling along a direction making a small angle with the z axis. It crosses a thin plate of width $d(x, y)$ surrounded by thin layers of air of total width $d_0 - d(x, y)$. In accordance with the approximate relation (2.4-3), the local transmit-

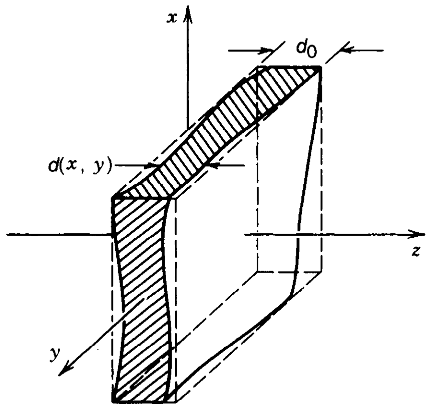


Figure 2.4-5 A transparent plate of varying thickness.

tance is the product of the transmittances of a thin layer of air of thickness $d_0 - d(x, y)$ and a thin layer of material of thickness $d(x, y)$, so that $t(x, y) \approx \exp[-jnk_o d(x, y)] \exp\{-jk_o[d_0 - d(x, y)]\}$, from which

$$\boxed{t(x, y) \approx h_0 \exp[-j(n-1)k_o d(x, y)],} \tag{2.4-4}$$

Transmittance of a Variable-Thickness Plate

where $h_0 = \exp(-jk_o d_0)$ is a constant phase factor. This relation is valid in the paraxial approximation (all angles θ are small) and when the thickness d_0 is sufficiently small so that $(d_0/\lambda)\theta^2/2n \ll 1$ at all points (x, y) for which $(x/\lambda)\theta \ll 1$ and $(y/\lambda)\theta \ll 1$.

EXERCISE 2.4-1

Transmission Through a Prism. Use (2.4-4) to show that the complex amplitude transmittance of a thin inverted prism with small angle $\alpha \ll 1$ and width d_0 (Fig. 2.4-6) is $t(x, y) = h_0 \exp[-j(n-1)k_o \alpha x]$, where $h_0 = \exp(-jk_o d_0)$. What is the effect of the

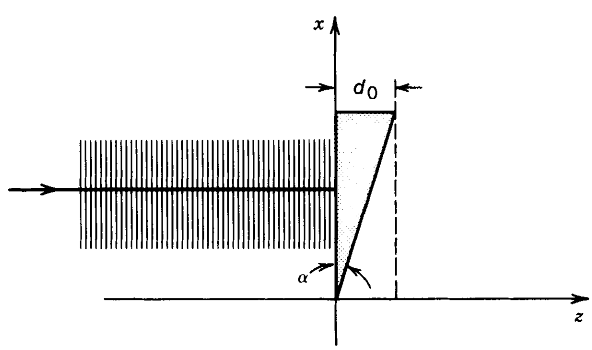


Figure 2.4-6 Transmission of a plane wave through a thin prism.

prism on an incident plane wave traveling in the z direction? Compare your results with the results obtained in the ray-optics model [see (1.2-7)].

Thin Lens

The general expression (2.4-4) for the complex amplitude transmittance of a thin transparent plate of variable thickness is now applied to the planoconvex thin lens shown in Fig. 2.4-7. Since the lens is the cap of a sphere of radius R, the thickness at the point (x, y) is $d(x, y) = d_0 - \overline{PQ} = d_0 - (R - \overline{QC})$, or

$$d(x, y) = d_0 - \left\{R - \left[R^2 - (x^2 + y^2)\right]^{1/2}\right\}. \tag{2.4-5}$$

This expression may be simplified by considering only points for which x and y are sufficiently small in comparison with R so that $x^2 + y^2 \ll R^2$. In this case

$$\left[R^2 - (x^2 + y^2)\right]^{1/2} = R\left(1 - \frac{x^2 + y^2}{R^2}\right)^{1/2} \approx R\left(1 - \frac{x^2 + y^2}{2R^2}\right),$$

and (2.4-5) gives

$$d(x, y) \approx d_0 - \frac{x^2 + y^2}{2R}.$$

Upon substitution into (2.4-4) we obtain

$$\boxed{t(x, y) \approx h_0 \exp\left[jk_o \frac{x^2 + y^2}{2f}\right],} \tag{2.4-6}$$

Transmittance of a Thin Lens

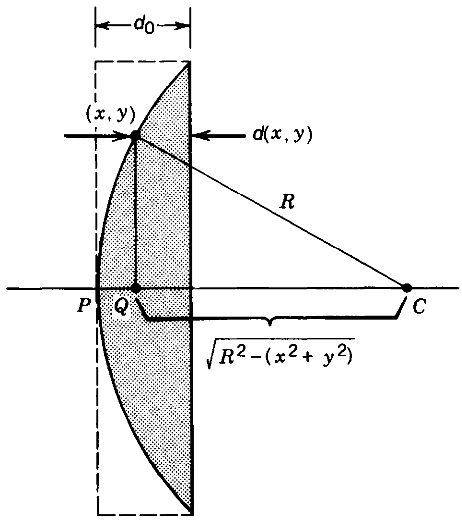


Figure 2.4-7 A planoconvex lens.

where

$$f = \frac{R}{n-1} \tag{2.4-7}$$

is the focal length of the lens (see Sec. 1.2C) and $h_0 = \exp(-jnk_o d_0)$ is a constant phase factor that is usually of no significance.

Since the lens imparts to the incident wave a phase proportional to $x^2 + y^2$, it bends the planar wavefronts of a plane wave, transforming it into a paraboloidal wave centered at a distance f from the lens, as demonstrated in Exercise 2.4-3.

EXERCISE 2.4-2

Double-Convex Lens. Show that the complex amplitude transmittance of the double-convex lens shown in Fig. 2.4-8 is given by (2.4-6) with

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{2.4-8}$$

You may prove this either by using the general formula (2.4-4) or by regarding the double-convex lens as a cascade of two planoconvex lenses. Recall that, by convention, the radius of a convex/concave surface is positive/negative, i.e., R_1 is positive and R_2 is negative for the lens in Fig. 2.4-8. The parameter f is recognized as the focal length of the lens [see (1.2-12)].

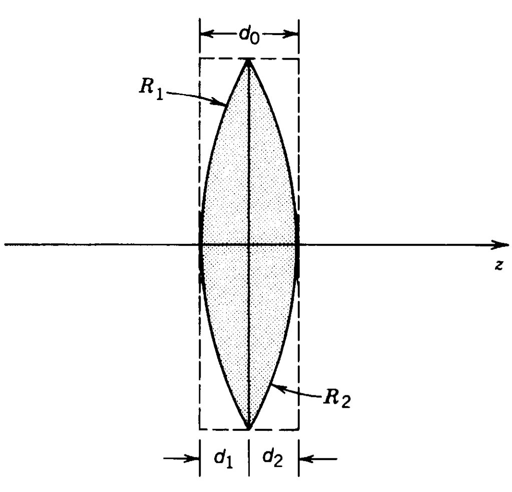


Figure 2.4-8 A double-convex lens.

EXERCISE 2.4-3

Focusing of a Plane Wave by a Thin Lens. Show that when a plane wave is transmitted through a thin lens of focal length f in a direction parallel to the axis of the

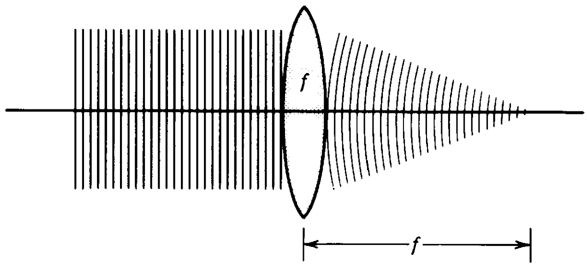


Figure 2.4-9 A thin lens transforms a plane wave into a paraboloidal wave.

lens, it is converted into a paraboloidal wave (the Fresnel approximation of a spherical wave) centered about a point at a distance f from the lens, as illustrated in Fig. 2.4-9. What is the effect of the lens on a plane wave incident at a small angle θ?

EXERCISE 2.4-4

Imaging Property of a Lens. Show that a paraboloidal wave centered at the point P_1 (Fig. 2.4-10) is converted by a lens of focal length f into a paraboloidal wave centered about P_2, where $1/z_1 + 1/z_2 = 1/f$.

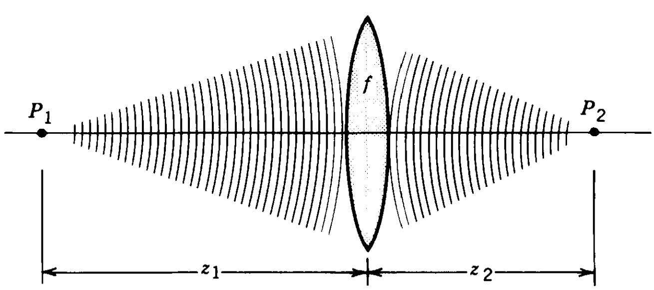


Figure 2.4-10 A lens transforms a paraboloidal wave into another paraboloidal wave. The two waves are centered at distances satisfying the imaging equation.

Diffraction Gratings

A **diffraction grating** is an optical component that serves to periodically modulate the phase or the amplitude of the incident wave. It can be made of a transparent plate with periodically varying thickness or periodically graded refractive index (see Sec. 2.4C). Repetitive arrays of diffracting elements such as apertures, obstacles, or absorbing elements can also be used (see Sec. 4.3). Reflection diffraction gratings are often fabricated by use of periodically ruled thin films of aluminum that have been evaporated onto a glass substrate.

Consider here a diffraction grating made of a thin transparent plate placed in the $z = 0$ plane whose thickness varies periodically in the x direction with period Λ (Fig. 2.4-11). As will be demonstrated in Exercise 2.4-5, this plate converts an incident plane wave of wavelength $\lambda \ll \Lambda$, traveling at a small angle θ_i with respect to the z axis, into

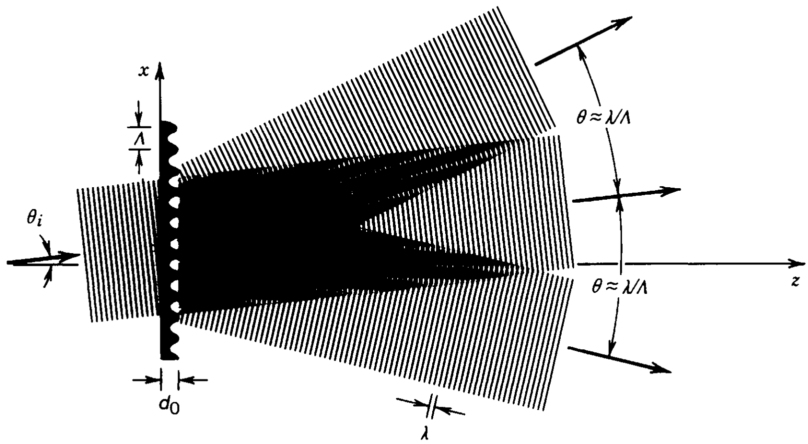


Figure 2.4-11 A thin transparent plate with periodically varying thickness serves as a diffraction grating. It splits an incident plane wave into multiple plane waves traveling in different directions.

several plane waves at small angles

$$\boxed{\theta_q = \theta_i + q\frac{\lambda}{\Lambda},} \qquad (2.4\text{-}9)$$

Grating Equation

$q = 0, \pm 1, \pm 2, \ldots,$ with the z axis, where q is called the diffraction order. The diffracted waves are separated by an angle $\theta = \lambda/\Lambda$, as shown schematically in Fig. 2.4-11.

EXERCISE 2.4-5

Transmission Through a Diffraction Grating

(a) The thickness of a thin transparent plate varies sinusoidally in the x direction, $d(x, y) = \frac{1}{2}d_0[1 + \cos(2\pi x/\Lambda)]$, as illustrated in Fig. 2.4-11. Show that the complex amplitude transmittance is $t(x, y) = h_0 \exp[-j\frac{1}{2}(n - 1)k_o d_0 \cos(2\pi x/\Lambda)]$ where $h_0 = \exp[-j\frac{1}{2}(n + 1)k_o d_0]$.

(b) Show that an incident plane wave traveling at a small angle θ_i with the z direction is transmitted in the form of a sum of plane waves traveling at angles θ_q given by (2.4-9). *Hint*: Expand the periodic function $t(x, y)$ in a Fourier series.

Equation (2.4-9) is valid only in the paraxial approximation (when all angles are small). This approximation is applicable when the period Λ is much greater than the wavelength λ. A more general analysis of thin diffraction gratings, without the use of the paraxial approximation, shows that the incident plane wave is converted into

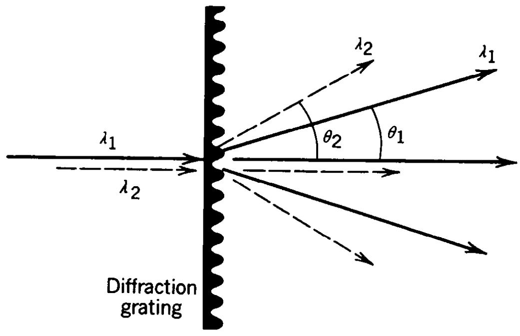


Figure 2.4-12 A diffraction grating directs two waves of wavelengths λ_1 and λ_2 into two directions θ_1 and θ_2. It therefore serves as a spectrum analyzer or a spectrometer.

several plane waves at angles θ_q satisfying[†]

$$\sin\theta_q = \sin\theta_i + q\frac{\lambda}{\Lambda}. \tag{2.4-10}$$

Diffraction gratings are used as filters and spectrum analyzers. Since the angles θ_q are dependent on the wavelength λ (and therefore on the frequency ν), an incident polychromatic wave is separated by the grating into its spectral components (Fig. 2.4-12). Diffraction gratings have found numerous applications in spectroscopy.

C. Graded-Index Optical Components

The effect of a prism, lens, or diffraction grating lies in the phase shift it imparts to the incident wave, which serves to bend the wavefront in some prescribed manner. This phase shift is controlled by the variation of the thickness of the material with the transverse distance from the optical axis (linearly, quadratically, or periodically, in the cases of the prism, lens, and diffraction grating, respectively). The same phase shift may instead be introduced by a transparent planar plate of fixed width but with varying refractive index.

The complex amplitude transmittance of a thin transparent planar plate of width d_0 and graded refractive index $n(x, y)$ is

$$\boxed{t(x, y) = \exp[-jn(x, y)k_o d_0].} \tag{2.4-11}$$

Transmittance of a Graded-Index Thin Plate

By selecting the appropriate variation of $n(x, y)$ with x and y, the action of any constant-index thin optical component can be reproduced, as demonstrated in Exercise 2.4-6.

[†]See, e.g., E. Hecht and A. Zajac, *Optics*, Addison-Wesley, Reading, MA, 1974.

EXERCISE 2.4-6

Graded-Index Lens. Show that a thin plate (Fig. 2.4-13) of uniform thickness d_0 and quadratically graded refractive index $n(x, y) = n_0[1 - \frac{1}{2}\alpha^2(x^2 + y^2)]$, where $\alpha d_0 \ll 1$, acts as a lens of focal length $f = 1/n_0\alpha^2 d_0$ (see Exercise 1.3-1).

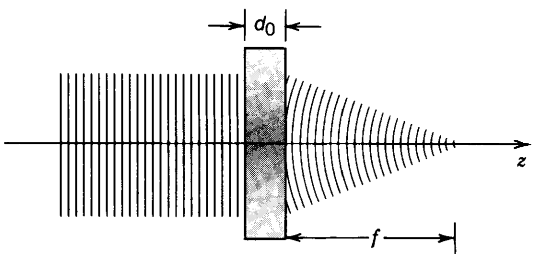


Figure 2.4-13 A graded-index plate acts as a lens.

2.5 INTERFERENCE

When two or more optical waves are present simultaneously in the same region of space, the total wavefunction is the sum of the individual wavefunctions. This basic principle of superposition follows from the linearity of the wave equation. For monochromatic waves of the same frequency, the superposition principle is also applicable to the complex amplitudes. This is consistent with the linearity of the Helmholtz equation.

The superposition principle does not apply to the optical intensity. The intensity of the superposition of two or more waves is not necessarily the sum of their intensities. The difference is attributed to the interference between these waves. Interference cannot be explained on the basis of ray optics since it is dependent on the phase relationship between the superposed waves.

In this section we examine the interference between two or more monochromatic waves of the same frequency. The interference of waves of different frequencies is discussed in Sec. 2.6.

A. Interference of Two Waves

When two monochromatic waves of complex amplitudes $U_1(\mathbf{r})$ and $U_2(\mathbf{r})$ are superposed, the result is a monochromatic wave of the same frequency and complex amplitude

$$U(\mathbf{r}) = U_1(\mathbf{r}) + U_2(\mathbf{r}). \tag{2.5-1}$$

In accordance with (2.2-10), the intensities of the constituent waves are $I_1 = |U_1|^2$ and $I_2 = |U_2|^2$ and the intensity of the total wave is

$$I = |U|^2 = |U_1 + U_2|^2 = |U_1|^2 + |U_2|^2 + U_1^* U_2 + U_1 U_2^*. \tag{2.5-2}$$

The explicit dependence on $\mathbf{r}$ has been omitted for convenience. Substituting

$$U_1 = I_1^{1/2} \exp(j\varphi_1) \qquad \text{and} \qquad U_2 = I_2^{1/2} \exp(j\varphi_2) \tag{2.5-3}$$

into (2.5-2), where φ_1 and φ_2 are the phases of the two waves, we obtain

$$\boxed{I = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos\varphi,} \tag{2.5-4}$$

Interference Equation

with

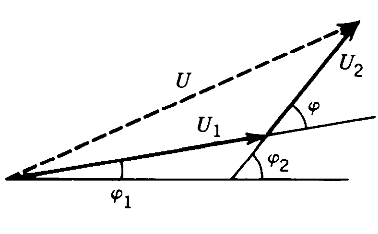
$$\varphi = \varphi_2 - \varphi_1. \tag{2.5-5}$$

This relation, called the **interference equation**, can also be seen from the geometry of the phasor diagram in Fig. 2.5-1(*a*), which demonstrates that the magnitude of the phasor U is sensitive to the phase difference φ, not only to the magnitudes of the constituent phasors.

The intensity of the sum of the two waves is *not* the sum of their intensities [Fig. 2.5-1(*b*)]; an additional term, attributed to **interference** between the two waves, is present in (2.5-4). This term may be positive or negative, corresponding to constructive or destructive interference, respectively. If $I_1 = I_2 = I_0$, for example, then $I = 2I_0(1 + \cos\varphi) = 4I_0 \cos^2(\varphi/2)$, so that for $\varphi = 0$, $I = 4I_0$ (i.e., the total intensity is four times the intensity of each of the superposed waves). For $\varphi = \pi$, the superposed waves cancel one another and the total intensity $I = 0$. When $\varphi = \pi/2$ or $3\pi/2$, the interference term vanishes and $I = 2I_0$, i.e., the total intensity is the sum of the constituent intensities. The strong dependence of the intensity I on the phase difference φ permits us to measure phase differences by detecting light intensity. This principle is used in numerous optical systems.

Interference is not observed under ordinary lighting conditions since the random fluctuations of the phases φ_1 and φ_2 cause the phase difference φ to assume random values, which are uniformly distributed between 0 and 2π, so that the average of $\cos\varphi = 0$ and the interference term is washed out. Light with such randomness is said to be *partially coherent* and Chap. 10 is devoted to its study. We limit ourselves here to the study of *coherent* light.

Interference is accompanied by a spatial redistribution of the optical intensity without violation of power conservation. For example, the two waves may have uniform intensities I_1 and I_2 in some plane, but as a result of a position-dependent phase



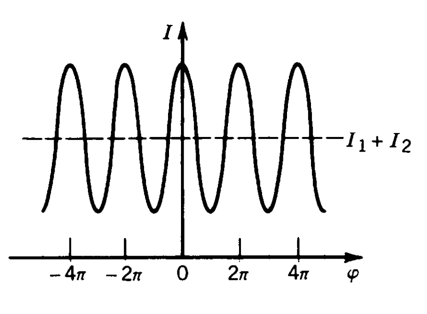


Figure 2.5-1 (*a*) Phasor diagram for the superposition of two waves of intensities I_1 and I_2 and phase difference $\varphi = \varphi_2 - \varphi_1$. (*b*) Dependence of the total intensity I on the phase difference φ.

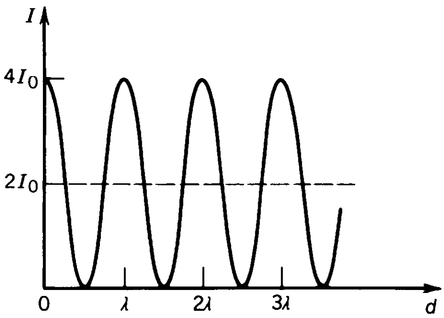


Figure 2.5-2 Dependence of the intensity I of the superposition of two waves, each of intensity I_0, on the delay distance d. When the delay distance is a multiple of λ, the interference is constructive; when it is an odd multiple of $\lambda/2$, the interference is destructive.

difference φ, the total intensity may be smaller than $I_1 + I_2$ at some positions and greater than $I_1 + I_2$ at others, with the total power (integral of the intensity) conserved.

Interferometers

Consider the superposition of two plane waves, each of intensity I_0, propagating in the z direction, and assume that one wave is delayed by a distance d with respect to the other so that $U_1 = I_0^{1/2}\exp(-jkz)$ and $U_2 = I_0^{1/2}\exp[-jk(z - d)]$. The intensity I of the sum of these two waves can be determined by substituting $I_1 = I_2 = I_0$ and $\varphi = kd = 2\pi d/\lambda$ into the interference equation (2.5-4),

$$I = 2I_0\left[1 + \cos\left(2\pi\frac{d}{\lambda}\right)\right]. \tag{2.5-6}$$

The dependence of I on the delay d is sketched in Fig. 2.5-2. If the delay is an integer multiple of λ, complete constructive interference occurs and the total intensity $I = 4I_0$. On the other hand, if d is an odd integer multiple of $\lambda/2$, complete destructive interference occurs and $I = 0$. The average intensity is the sum of the two intensities $2I_0$.

An **interferometer** is an optical instrument that splits a wave into two waves using a beamsplitter, delays them by unequal distances, redirects them using mirrors, recombines them using another (or the same) beamsplitter, and detects the intensity of their superposition. Three important examples, the **Mach–Zehnder interferometer**, the **Michelson interferometer**, and the **Sagnac interferometer**, are illustrated in Fig. 2.5-3.

Since the intensity I is sensitive to the phase $\varphi = 2\pi d/\lambda = 2\pi n d/\lambda_o = 2\pi n\nu d/c_o$, where d is the difference between the distances traveled by the two waves, the interferometer can be used to measure small variations of the distance d, the refractive index n, or the wavelength λ_o (or frequency ν). For example, if $d/\lambda_o = 10^4$, a change $\Delta n = 10^{-4}$ of the refractive index corresponds to a phase change $\Delta\varphi = 2\pi$. Also, the phase φ changes by a full 2π if d changes by a wavelength λ. An incremental change of the frequency $\Delta\nu = c/d$ has the same effect. Interferometers can serve as spectrometers, which measure the spectrum of polychromatic light (see Sec. 10.2B). In the Sagnac interferometer the optical paths are identical but opposite, so that rotation of the interferometer results in a phase shift φ proportional to the angular velocity of rotation. This system is therefore often used as a gyroscope.†

Interference of Two Oblique Plane Waves

Consider now the interference of two plane waves of equal intensities—one propagating in the z direction $U_1 = I_0^{1/2}\exp(-jkz)$, and the other at an angle θ with the z axis

†See, e.g., E. Hecht and A. Zajac, *Optics*, Addison-Wesley, Reading, MA, 1974.

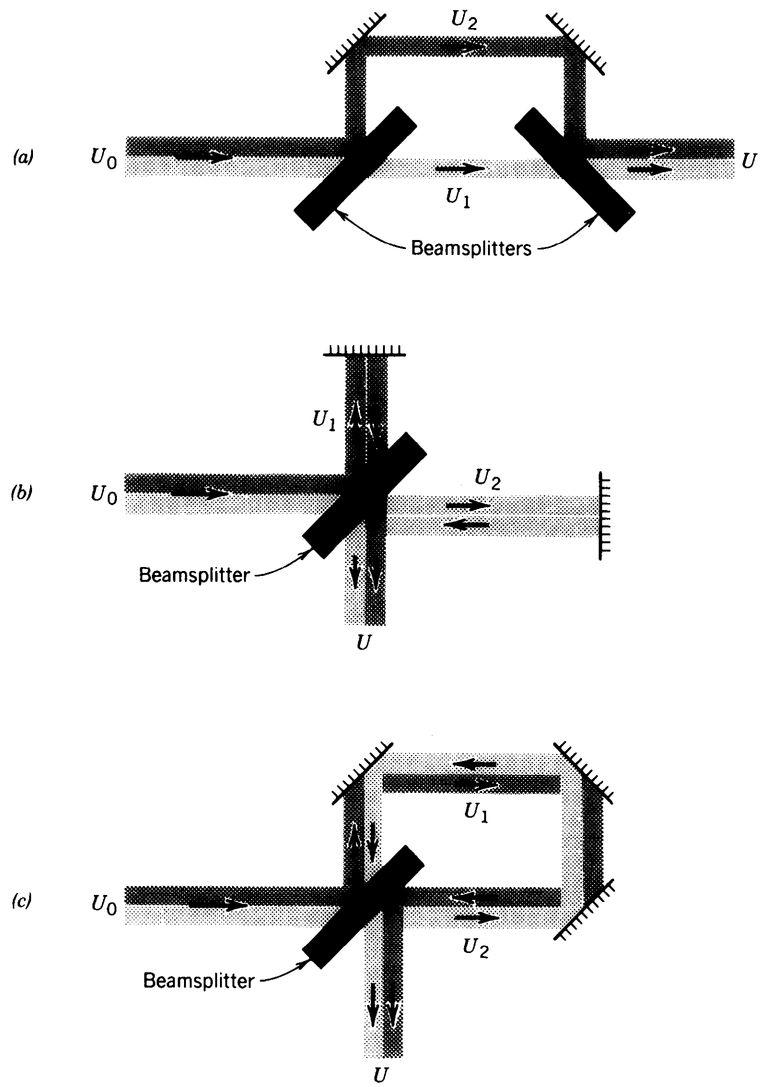


Figure 2.5-3 Interferometers: (*a*) Mach–Zehnder interferometer; (*b*) Michelson interferometer; (*c*) Sagnac interferometer. A wave U_0 is split into two waves U_1 and U_2. After traveling through different paths, the waves are recombined into a superposition wave $U = U_1 + U_2$ whose intensity is recorded. The waves are split and recombined using beamsplitters. In the Sagnac interferometer the two waves travel through the same path in opposite directions.

in the x-z plane, $U_2 = I_0^{1/2} \exp[-j(k \cos\theta z + k \sin\theta x)]$, as illustrated in Fig. 2.5-4. At the $z = 0$ plane the two waves have a phase difference $\varphi = kx \sin\theta$, so that the interference equation (2.5-4) yields the total intensity:

$$I = 2I_0[1 + \cos(k \sin\theta x)]. \tag{2.5-7}$$

This pattern varies sinusoidally with x, with period $2\pi/k \sin\theta = \lambda/\sin\theta$, as shown in Fig. 2.5-4. If $\theta = 30°$, for example, the period is 2λ. This suggests a method of printing a sinusoidal pattern of high resolution for use as a diffraction grating. It also suggests a method of monitoring the angle of arrival θ of a wave by mixing it with a reference

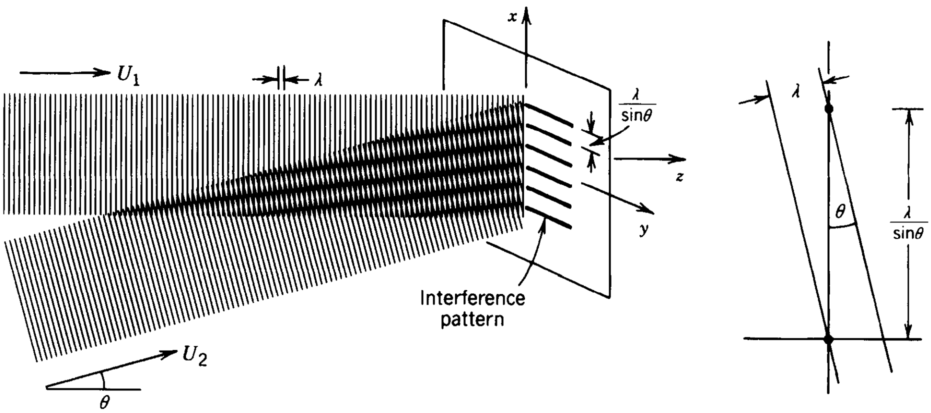


Figure 2.5-4 The interference of two plane waves at an angle θ results in a sinusoidal intensity pattern of period $\lambda/\sin\theta$.

wave and recording the resultant intensity distribution. As discussed in Sec. 4.5, this is the principle behind holography.

EXERCISE 2.5-1

Interference of a Plane Wave and a Spherical Wave. A plane wave of complex amplitude $A_1 \exp(-jkz)$ and a spherical wave approximated by the paraboloidal wave of complex amplitude $(A_2/z)\exp(-jkz)\exp[-jk(x^2+y^2)/2z]$ [see (2.2-16)], interfere in the $z = d$ plane. Derive an expression for the total intensity $I(x, y, d)$. Verify that the locus of points of zero intensity is a set of concentric rings, as illustrated in Fig. 2.5-5.

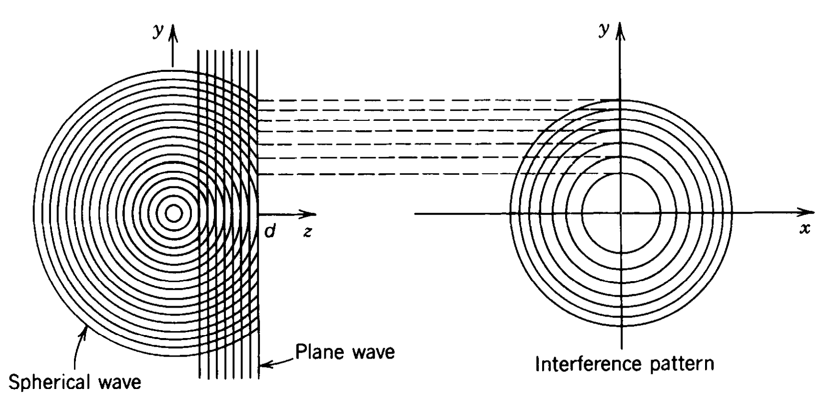


Figure 2.5-5 The interference of a plane wave and a spherical wave creates a pattern of concentric rings (illustrated at the plane $z = d$).

EXERCISE 2.5-2

Interference of Two Spherical Waves. Two spherical waves of equal intensity I_0, originating at the points $(a, 0, 0)$ and $(-a, 0, 0)$ interfere in the plane $z = d$ as illustrated in Fig. 2.5-6. The system is similar to that used by Thomas Young in his celebrated double-slit

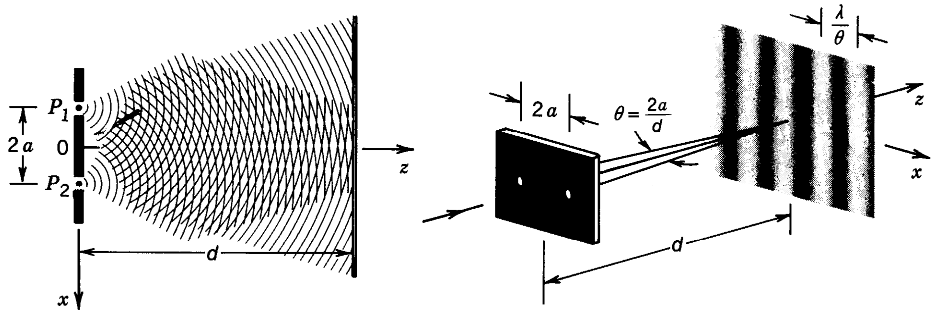


Figure 2.5-6 Interference of two spherical waves of equal intensities originating at the points P_1 and P_2. The two waves can be obtained by permitting a plane wave to impinge on two pinholes in a screen. The light intensity at an observation plane a large distance d away takes the form of a sinusoidal pattern with period $\approx \lambda/\theta$.

experiment in which he demonstrated interference. Use the paraboloidal approximation for the spherical waves to show that the detected intensity is

$$I(x, y, d) = 2I_0\left(1 + \cos\frac{2\pi x\theta}{\lambda}\right), \tag{2.5-8}$$

where $\theta = 2a/d$ is approximately the angle subtended by the centers of the two waves at the observation plane. The intensity pattern is periodic with period λ/θ.

B. Multiple-Wave Interference

When M monochromatic waves of complex amplitudes $U_1, U_2, \ldots, U_M$ and the same frequency are added, the result is a monochromatic wave with complex amplitude $U = U_1 + U_2 + \cdots + U_M$. Knowing the intensities of the individual waves, $I_1, I_2, \ldots, I_M$, is not sufficient to determine the total intensity $I = |U|^2$ since the relative phases must also be known. The role played by the phase is dramatically illustrated by the following examples.

Interference of M Waves of Equal Amplitudes and Equal Phase Differences

We first examine the interference of M waves with complex amplitudes

$$U_m = I_0^{1/2} \exp[j(m-1)\varphi], \qquad m = 1, 2, \ldots, M. \tag{2.5-9}$$

The waves have equal intensities I_0, and phase difference φ between successive waves, as illustrated in Fig. 2.5-7(a). To derive an expression for the intensity of the superposition, it is convenient to introduce $h = \exp(j\varphi)$, and write $U_m = I_0^{1/2} h^{m-1}$. The complex amplitude of the superposed wave is then

$$U = I_0^{1/2}(1 + h + h^2 + \cdots + h^{M-1}) = I_0^{1/2}\frac{1 - h^M}{1 - h}$$

$$= I_0^{1/2}\frac{1 - \exp(jM\varphi)}{1 - \exp(j\varphi)},$$

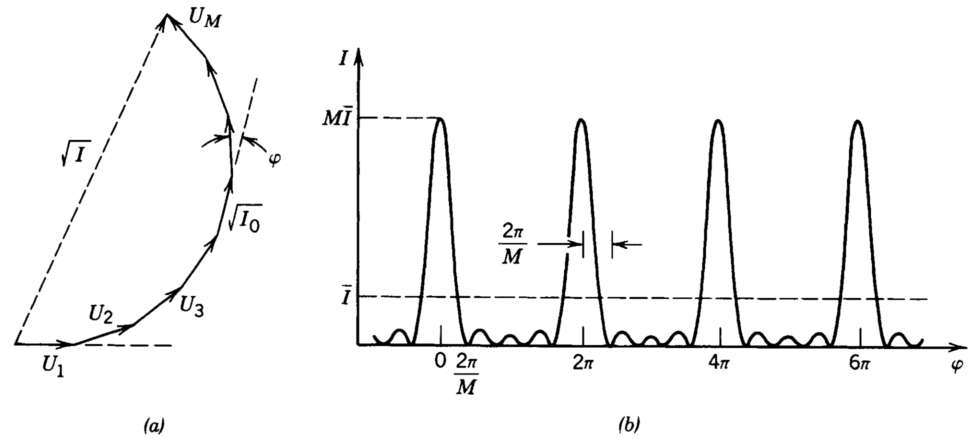


Figure 2.5-7 (*a*) The sum of M phasors of equal magnitudes and equal phase differences. (*b*) The intensity I as a function of φ. The peak intensity occurs when all the phasors are aligned; it is M times greater than the mean intensity $\bar{I} = MI_0$. In this graph $M = 5$.

and the corresponding intensity is

$$I = |U|^2 = I_0 \left| \frac{\exp(-jM\varphi/2) - \exp(jM\varphi/2)}{\exp(-j\varphi/2) - \exp(j\varphi/2)} \right|^2,$$

from which

$$\boxed{I = I_0 \frac{\sin^2(M\varphi/2)}{\sin^2(\varphi/2)}.} \qquad (2.5\text{-}10)$$

Interference of
M Waves

The intensity I is strongly dependent on the phase difference φ, as illustrated in Fig. 2.5-7(*b*) for $M = 5$. When $\varphi = 2\pi q$, where q is an integer, all the phasors are aligned so that the amplitude of the total wave is M times that of an individual component, and the intensity reaches its peak value of $M^2 I_0$. The mean intensity averaged over a uniform distribution of φ is $\bar{I} = (1/2\pi)\int_0^{2\pi} I\, d\varphi = MI_0$, which is the result obtained in the absence of interference. The peak intensity is therefore M times greater than the mean intensity. If M is large, the sensitivity to the phase can be dramatic since the peak intensity will be much greater than the mean intensity. For a phase difference φ slightly different from $2\pi q$, the intensity drops sharply. In particular, when it is $2\pi/M$ the intensity is zero. A comparison of Fig. 2.5-7(*b*) for $M = 5$ with Fig. 2.5-2 for $M = 2$ is instructive.

EXERCISE 2.5-3

Bragg Reflection. Light is reflected at an angle θ from M parallel reflecting planes separated by a distance d as shown in Fig. 2.5-8. Assume that only a small fraction of the light is reflected from each plane, so that the amplitudes of the M reflected waves are

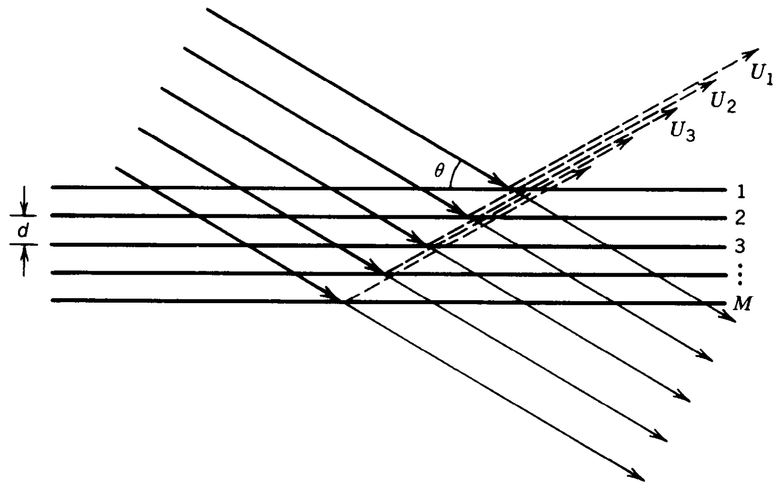


Figure 2.5-8 Reflection of a plane wave from M planes separated from each other by a distance d. The reflected waves interfere constructively and yield maximum intensity when the angle θ is the Bragg angle.

approximately equal. Show that the reflected waves have a phase difference $\varphi = k(2d \sin \theta)$ and that the angle θ at which the intensity of the total reflected light is maximum satisfies

$$\boxed{\sin \theta = \frac{\lambda}{2d}.} \qquad (2.5\text{-}11)$$

Bragg Angle

This angle is known as the Bragg angle. Such reflections are encountered when x-ray waves are reflected from atomic planes in crystalline structures. It also occurs when light is reflected from a periodic structure created by an acoustic wave (see Chap. 20).

Interference of an Infinite Number of Waves of Progressively Smaller Amplitudes and Equal Phase Differences

We now examine the superposition of an infinite number of waves with equal phase differences and with amplitudes that decrease at a geometric rate,

$$U_1 = I_0^{1/2}, \quad U_2 = hU_1, \quad U_3 = hU_2 = h^2 U_1, \quad \ldots, \qquad (2.5\text{-}12)$$

where $h = r e^{j\varphi}$, $|h| = r < 1$, and I_0 is the intensity of the initial wave. The amplitude of the mth wave is smaller than that of the $(m-1)$st wave by the factor r and the phase differs by φ. The phasor diagram is shown in Fig. 2.5-9(a).

The superposition wave has a complex amplitude

$$\begin{aligned} U &= U_1 + U_2 + U_3 + \cdots \\ &= I_0^{1/2}(1 + h + h^2 + \cdots) \\ &= \frac{I_0^{1/2}}{1-h} = \frac{I_0^{1/2}}{1 - r e^{j\varphi}}. \end{aligned} \qquad (2.5\text{-}13)$$

The intensity $I = |U|^2 = I_0/|1 - r e^{j\varphi}|^2 = I_0/[(1 - r\cos\varphi)^2 + r^2 \sin^2\varphi]$, from which

$$I = \frac{I_0}{(1 - r)^2 + 4r\sin^2(\varphi/2)}. \tag{2.5-14}$$

It is convenient to write this equation in the form

$$\boxed{I = \frac{I_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\varphi/2)},} \tag{2.5-15}$$

Intensity of an Infinite Number of Waves

where

$$I_{\max} = \frac{I_0}{(1 - r)^2} \tag{2.5-16}$$

and the quantity

$$\boxed{\mathcal{F} = \frac{\pi r^{1/2}}{1 - r}} \tag{2.5-17}$$

Finesse

is a parameter called the **finesse**.

The intensity I is a periodic function of φ with period 2π, as illustrated in Fig. 2.5-9(*b*). It reaches its maximum value $I_{\max}$ when $\varphi = 2\pi q$, where q is an integer. This occurs when the phasors align to form a straight line. (This result is not unlike that displayed in Fig. 2.5-7(*b*) for the interference of M waves of equal amplitudes and equal phase differences.) When the finesse $\mathcal{F}$ is large (i.e., the factor r is close to 1), I becomes a sharply peaked function of φ. Consider, for example, values of φ near the

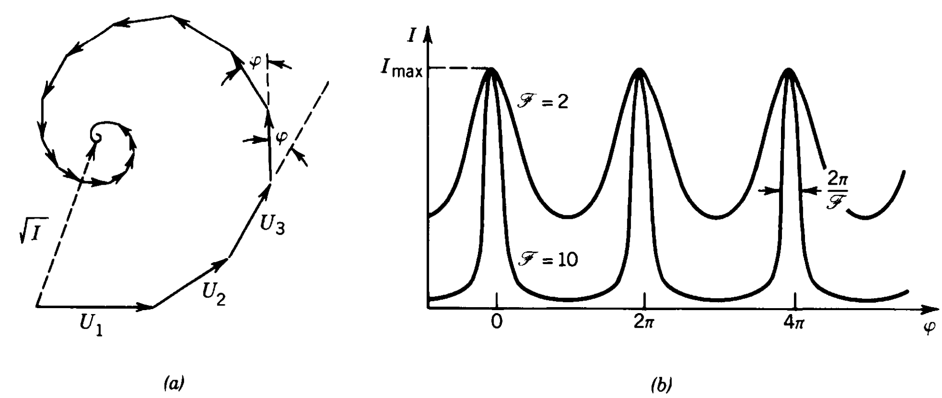


Figure 2.5-9 (*a*) The sum of an infinite number of phasors whose magnitudes are successively reduced at a geometric rate and whose phase differences φ are equal. (*b*) Dependence of the intensity I on the phase difference φ for two values of $\mathcal{F}$. Peak values occur at $\varphi = 2\pi q$. The width (FWHM) of each peak is approximately $2\pi/\mathcal{F}$ when $\mathcal{F} \gg 1$. The sharpness of the peaks increases with increasing $\mathcal{F}$.

$\varphi = 0$ peak. For $|\varphi| \ll 1$, $\sin(\varphi/2) \approx \varphi/2$ and (2.5-15) may be approximated by

$$I \approx \frac{I_{\max}}{1 + (\mathscr{F}/\pi)^2 \varphi^2}. \tag{2.5-18}$$

The intensity I decreases to half its peak valued when $\varphi = \pi/\mathscr{F}$, so that the full width at half maximum (FWHM) of the peak is

$$\boxed{\Delta\varphi = \frac{2\pi}{\mathscr{F}}.} \tag{2.5-19}$$

Width of Interference Pattern

If $\mathscr{F} \gg 1$, $\Delta\varphi \ll 2\pi$ and the assumption that $\varphi \ll 1$ is applicable. The finesse $\mathscr{F}$ is therefore the ratio between the period 2π and the FWHM of the interference pattern. It is a measure of the sharpness of the interference function, i.e., the sensitivity of the intensity to deviations of φ from the values $2\pi q$ corresponding to the peaks.

The Fabry–Perot interferometer is a useful device based on this principle. It consists of two parallel mirrors within which light undergoes multiple reflections. In the course of each round trip, the light suffers a fixed amplitude reduction $\mathscr{r}$ and a phase shift $\varphi = k2d = 4\pi\nu d/c$, where d is the mirror separation. The total light intensity depends on the phase shift φ in accordance with (2.5-15). Because the phase shift φ is proportional to the optical frequency ν, the intensity transmission of the device exhibits spectral characteristics with peaks at resonance frequencies separated by $c/2d$. The width of these resonances is $(c/2d)/\mathscr{F}$, where the finesse $\mathscr{F}$ is governed by losses (since it is related to the attenuation factor $\mathscr{r}$). The Fabry–Perot interferometer serves as a spectrum analyzer and as an optical resonator, which is one of the essential components of a laser. Optical resonators are discussed in Chap. 9.

2.6 POLYCHROMATIC LIGHT

Since the wavefunction of monochromatic light is a harmonic function of time that extends over all time (from $-\infty$ to ∞), it is an idealization that cannot be met in reality. This section is devoted to polychromatic waves of finite time duration, including optical pulses.

A. Fourier Decomposition

A polychromatic wave can be expanded as a sum of monochromatic waves by the use of Fourier methods. Since we already know how monochromatic waves are transmitted through optical components, we can determine the effect of optical systems on polychromatic light by using the principle of superposition.

An arbitrary function of time, such as the wavefunction $u(\mathbf{r}, t)$ at a fixed position $\mathbf{r}$, can be analyzed as a superposition integral of harmonic functions of different frequencies, amplitudes, and phases,

$$u(\mathbf{r}, t) = \int_{-\infty}^{\infty} U_\nu(\mathbf{r}) \exp(j2\pi\nu t)\, d\nu, \tag{2.6-1}$$

where $U_\nu(\mathbf{r})$ is determined by carrying out the **Fourier transform**

$$U_\nu(\mathbf{r}) = \int_{-\infty}^{\infty} u(\mathbf{r}, t) \exp(-j2\pi\nu t)\, dt. \tag{2.6-2}$$

A review of the Fourier transform and its properties is presented in Appendix A.

Complex Representation

Since $u(\mathbf{r}, t)$ is real, $U_\nu(\mathbf{r})$ must be a symmetric function of ν, i.e., $U_{-\nu}(\mathbf{r}) = U_\nu^*(\mathbf{r})$. The integral in (2.6-1) may therefore be simplified by use of the relation

$$\int_{-\infty}^{0} U_\nu(\mathbf{r}) \exp(j2\pi\nu t)\, d\nu = \int_{0}^{\infty} U_{-\nu}(\mathbf{r}) \exp(-j2\pi\nu t)\, d\nu$$
$$= \int_{0}^{\infty} U_\nu^*(\mathbf{r}) \exp(-j2\pi\nu t)\, d\nu,$$

so that $u(\mathbf{r}, t)$ is the sum of a complex function and its conjugate,

$$u(\mathbf{r}, t) = \int_{0}^{\infty} [U_\nu(\mathbf{r}) \exp(j2\pi\nu t) + U_\nu^*(\mathbf{r}) \exp(-j2\pi\nu t)]\, d\nu. \tag{2.6-3}$$

As in the case of monochromatic light (Sec. 2.2A), the complex wavefunction is defined as twice the first term in (2.6-3),

$$U(\mathbf{r}, t) = 2\int_{0}^{\infty} U_\nu(\mathbf{r}) \exp(j2\pi\nu t)\, d\nu, \tag{2.6-4}$$

so that its real part is the wavefunction

$$u(\mathbf{r}, t) = \mathrm{Re}\{U(\mathbf{r}, t)\} = \tfrac{1}{2}[U(\mathbf{r}, t) + U^*(\mathbf{r}, t)], \tag{2.6-5}$$

as in (2.2-3). The complex wavefunction (also called the **complex analytic signal**) is therefore obtained from the wavefunction by a process of three steps: (1) determine its Fourier transform; (2) eliminate negative frequencies and multiply by 2; and (3) determine the inverse Fourier transform. Since each of its Fourier components satisfies the wave equation, the complex wavefunction $U(\mathbf{r}, t)$ itself satisfies the wave equation.

The magnitudes of the Fourier transforms of the wavefunction and the complex wavefunction of a **quasi-monochromatic** wave are illustrated in Fig. 2.6-1. A quasi-monochromatic wave has Fourier components with frequencies confined within a narrow band of width $\Delta\nu$ surrounding a central frequency ν_0, such that $\Delta\nu \ll \nu_0$.

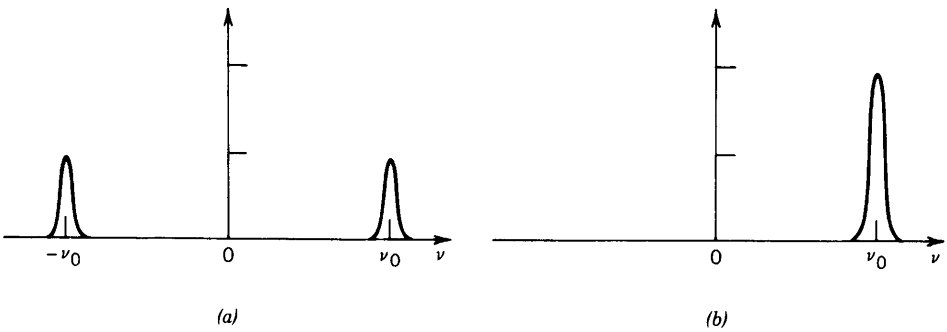


Figure 2.6-1 (*a*) The magnitude of the Fourier transform of the wavefunction. (*b*) The magnitude of the Fourier transform of the corresponding complex wavefunction.

Intensity of a Polychromatic Wave

The intensity is related to the wavefunction by

$$\begin{aligned} I(\mathbf{r}, t) &= 2\langle u^2(\mathbf{r}, t)\rangle \\ &= 2\langle\{\tfrac{1}{2}[U(\mathbf{r}, t) + U^*(\mathbf{r}, t)]\}^2\rangle \\ &= \tfrac{1}{2}\langle U^2(\mathbf{r}, t)\rangle + \tfrac{1}{2}\langle U^{*2}(\mathbf{r}, t)\rangle + \langle U(\mathbf{r}, t)U^*(\mathbf{r}, t)\rangle. \end{aligned} \tag{2.6-6}$$

If the wave is quasi-monochromatic with central frequency ν_0 and spectral width $\Delta\nu \ll \nu_0$, the average $\langle \cdot \rangle$ is taken over a time interval much longer than the time of an optical cycle $1/\nu_0$ but much shorter than $1/\Delta\nu$ (see Sec. 2.1). Since $U(\mathbf{r}, t)$ is given by (2.6-4), the term U^2 in (2.6-6) has components oscillating at frequencies $\approx 2\nu_0$. Similarly, the components of U^{*2} have frequencies $\approx -2\nu_0$. These terms are washed out by the averaging operation. The third term contains only frequency differences of the order of $\Delta\nu \ll \nu_0$. It therefore varies slowly and is unaffected by the time-averaging operation. Thus the third term survives and the light intensity is given by

$$\boxed{I(\mathbf{r}, t) = |U(\mathbf{r}, t)|^2.} \tag{2.6-7}$$

Optical Intensity of Quasi-Monochromatic Light

The intensity of a quasi-monochromatic wave is therefore given by the squared-absolute-value of its complex wavefunction. The simplicity of this result is, in fact, the rationale for introducing the concept of the complex wavefunction.

The Pulsed Plane Wave

As an example, consider a polychromatic wave each of whose monochromatic components is a plane wave traveling in the z direction with speed c. The complex wavefunction is the superposition integral

$$U(\mathbf{r}, t) = \int_0^\infty A_\nu \exp(-jkz)\exp(j2\pi\nu t)\, d\nu = \int_0^\infty A_\nu \exp\left[j2\pi\nu\left(t - \frac{z}{c}\right)\right] d\nu, \tag{2.6-8}$$

where A_ν is the complex envelope of the component of frequency ν and wavenumber $k = 2\pi\nu/c$.

Assuming that the speed $c = c_o/n$ is independent of the frequency ν, (2.6-8) may be written in the form

$$U(\mathbf{r}, t) = a\left(t - \frac{z}{c}\right), \tag{2.6-9}$$

where

$$a(t) = \int_0^\infty A_\nu \exp(j2\pi\nu t)\, d\nu. \tag{2.6-10}$$

Since A_ν may be arbitrarily chosen, (2.6-9) represents a valid wave, regardless of the function $a(\cdot)$ (provided that d^2a/dt^2 exists). Indeed, it can be easily verified that $U(\mathbf{r}, t) = a(t - z/c)$ satisfies the wave equation for an arbitrary form of $a(t)$.

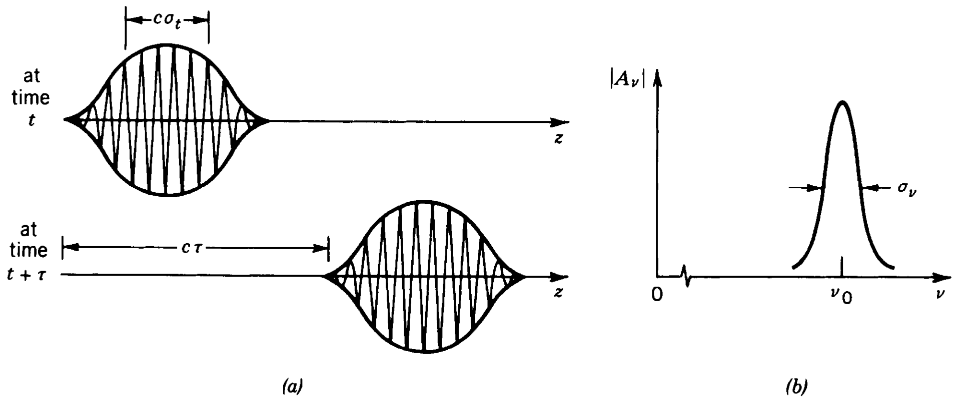


Figure 2.6-2 (a) The wavefunction $u(\mathbf{r}, t) = \mathrm{Re}\{\mathscr{a}(t - z/c)\}$ of a pulsed plane wave of time duration σ_t at times t and $t + \tau$. The pulse travels with speed c and occupies a distance $\sigma_z = c\sigma_t$. (b) The magnitude $|A_\nu|$ of the Fourier transform of the wavefunction is centered at ν_0 and has a width σ_ν.

If $\mathscr{a}(t)$ is of finite duration σ_t, for example, then the wave is a plane-wave pulse of light (a **wavepacket**) traveling in the z direction. At any time, the wavepacket extends over a distance $\sigma_z = c\sigma_t$ (Fig. 2.6-2). A pulse of duration $\sigma_t = 1$ ps, for example, extends over a distance of 0.3 mm. If the pulse intensity is Gaussian with rms width $\sigma_t = 1$ ps, its spectral bandwidth is $\sigma_\nu = 1/4\pi\sigma_t \approx 80$ GHz (see Appendix A, Sec. A.2). If the central frequency ν_0 is 5×10^{14} Hz (corresponding to $\lambda = 0.6$ μm), the condition of quasi-monochromaticity is clearly satisfied.

The propagation of optical pulses through media with frequency-dependent refractive indices (i.e., with a frequency-dependent speed of light $c = c_o/n$) is discussed in Sec. 5.6.

B. Light Beating

The dependence of the intensity of a polychromatic wave on time may be attributed to interference among the monochromatic components that constitute the wave. This concept is now demonstrated by means of two examples: interference between two monochromatic waves and interference among a finite number of monochromatic waves.

Interference Between Two Monochromatic Waves

An optical wave composed of two monochromatic waves of frequencies ν_1 and ν_2 and intensities I_1 and I_2 has a complex wavefunction at some point in space

$$U(t) = I_1^{1/2} \exp(j2\pi\nu_1 t) + I_2^{1/2} \exp(j2\pi\nu_2 t), \qquad (2.6\text{-}11)$$

where the phases are assumed to be zero. The $\mathbf{r}$ dependence has been suppressed for notational convenience. The intensity of the total wave is determined by use of the interference equation (2.5-4),

$$I(t) = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos[2\pi(\nu_2 - \nu_1)t]. \qquad (2.6\text{-}12)$$

The intensity therefore varies sinusoidally at the difference frequency $|\nu_2 - \nu_1|$, which is called the "beat frequency." The effect is called **light beating or light mixing.**

Equation (2.6-12) is analogous to (2.5-7), which describes the "spatial" interference of two monochromatic waves of the same frequency but different directions. This can be understood from the phasor diagram in Fig. 2.5-1. The two phasors U_1 and U_2 rotate at angular frequencies $\omega_1 = 2\pi\nu_1$ and $\omega_2 = 2\pi\nu_2$, so that the difference angle $\varphi = \varphi_2 - \varphi_1$ is $2\pi(\nu_2 - \nu_1)t$, in accord with (2.6-12).

Beating occurs in electronics when the sum of two sinusoidal signals drives a nonlinear (e.g., quadratic) device called a mixer and produces signals at the sum and difference frequencies. It is used in heterodyne radio receivers. In optics, the nonlinearity results from the squared-absolute-value relation between the optical intensity and the complex wavefunction. Only the difference frequency is detected in this case. The use of optical beating in optical heterodyne receivers is discussed in Sec. 22.5. Other forms of optical mixing make use of nonlinear media to generate optical frequency differences and sums, as described in Chap. 19.

EXERCISE 2.6-1

Optical Doppler Radar. As a result of the **Doppler effect** a monochromatic optical wave of frequency ν reflected from an object moving with velocity v undergoes a frequency shift $\Delta\nu = \pm(2v/c)\nu$, depending on whether the object is moving toward (+) or away (−) from the observer. Assuming that the original and reflected waves are superimposed, derive an expression for the intensity of the resultant wave. Suggest a method for measuring the velocity of a target using such an arrangement. If one of the mirrors of a Michelson interferometer moves with velocity v, use (2.5-6) to show that the beat frequency is $(2v/c)\nu$.

Interference of M Monochromatic Waves

The interference of a large number of monochromatic waves with equal intensities, equal phases, and equally spaced frequencies can result in the generation of narrow pulses of light. Consider an odd number $M = 2L + 1$ waves, each with intensity I_0 and zero phase, and with frequencies

$$\nu_q = \nu_0 + q\nu_F, \qquad q = -L, \ldots, 0, \ldots, L,$$

centered about the frequency ν_0 and spaced by the frequency $\nu_F \ll \nu_0$. At a given position, the total wave has a complex wavefunction

$$U(t) = I_0^{1/2} \sum_{q=-L}^{L} \exp[j2\pi(\nu_0 + q\nu_F)t]. \tag{2.6-13}$$

This is the sum of M phasors of equal magnitudes and phases differing by $\varphi = 2\pi\nu_F t$. Using the result of the analysis for an identical situation provided in (2.5-10) and Fig. 2.5-7, the intensity becomes

$$I(t) = |U(t)|^2 = I_0 \frac{\sin^2(M\pi t/T_F)}{\sin^2(\pi t/T_F)}. \tag{2.6-14}$$

As illustrated in Fig. 2.6-3 the intensity $I(t)$ is a periodic sequence of pulses with period $T_F = 1/\nu_F$, peak intensity $M^2 I_0$, and mean intensity MI_0. The peak intensity is M times greater than the mean intensity. The width of each pulse is approximately

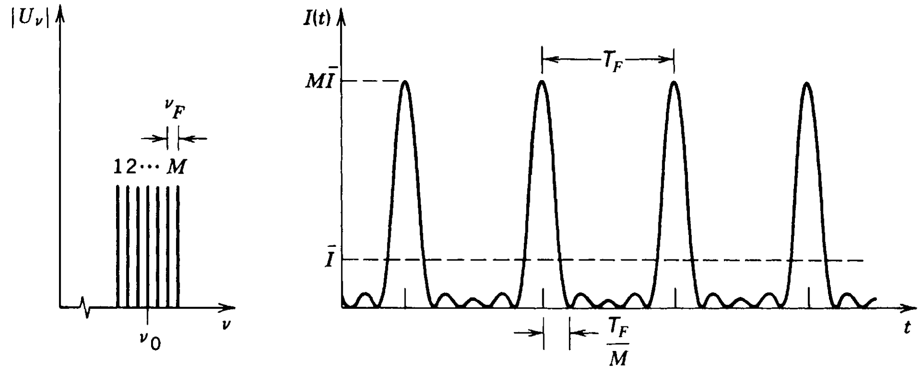


Figure 2.6-3 Time dependence of the intensity of a polychromatic wave composed of a sum of M monochromatic waves, of equal intensities, equal phases, and frequencies differing by ν_F. The intensity is a periodic train of pulses of period $T_F = 1/\nu_F$ with a peak M times greater than the mean. The duration of each pulse is M times smaller than the period. This should be compared with Fig. 2.5-7.

T_F/M. For large M, these pulses can be very narrow. If $\nu_F = 1$ GHz, for example, then $T_F = 1$ ns. If $M = 1000$, pulses of 1-ps width are generated.

This example provides a dramatic demonstration of how M monochromatic waves may cooperate to produce a train of very narrow pulses. In Chap. 14 we shall see that the modes of a laser can be "phase locked" in the fashion described above to produce narrow laser pulses.

READING LIST

General

See the general list in Chapter 1.

S. G. Lipson and H. Lipson, *Optical Physics*, Cambridge University Press, London, 1969, 2nd ed. 1981.

H. D. Young, *Fundamentals of Waves, Optics, and Modern Physics*, McGraw-Hill, New York, 2nd ed. 1976.

J. R. Pierce, *Almost All About Waves*, MIT Press, Cambridge, MA, 1974.

R. H. Webb, *Elementary Wave Optics*, Academic Press, New York, 1969.

D. H. Towne, *Wave Phenomena*, Addison-Wesley, Reading, MA, 1967.

C. Curry, *Wave Optics*, Edward Arnold, London, 1957.

Interferometry

J. M. Vaughan, *The Fabry-Perot Interferometer*, Adam Hilger, Bristol, England, 1989.

S. Tolansky, *An Introduction to Interferometry*, Wiley, New York, 1973.

A. H. Cook, *Interference of Electromagnetic Waves*, Clarendon Press, Oxford, 1971.

J. Dyson, *Interferometry as a Measuring Tool*, Machinery Publishing, Brighton, 1970.

W. H. Steel, *Interferometry*, Cambridge University Press, London, 1967.

M. Françon, *Optical Interferometry*, Academic Press, New York, 1966.

Spectroscopy

J. E. Chamberlain, *The Principles of Interferometric Spectroscopy*, Wiley, New York, 1979.

R. J. Bell, *Introductory Fourier Transform Spectroscopy*, Academic Press, New York, 1972.

J. F. James and R. S. Sternberg, *Design of Optical Spectrometers*, Chapman & Hall, London, 1969.

Diffraction Gratings

M. C. Hutley, *Diffraction Gratings*, Academic Press, New York, 1982.

R. Petit, ed., *Electromagnetic Theory of Gratings*, Springer-Verlag, New York, 1980.

S. P. Davis, *Diffraction Gratings and Spectrographs*, Holt, Rinehart and Winston, New York, 1970.

E. G. Loewen, *Diffraction Grating Handbook*, Bausch & Lomb, Rochester, NY, 1970.

Popular and Historical

J. Z. Buchwald, *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century*, University of Chicago Press, Chicago, 1989.

W. E. Kock, *Sound Waves and Light Waves*, Doubleday (Anchor Books), Garden City, NY, 1965.

C. Huygens, *Treatise on Light*, Dover, New York, 1962 (originally published in 1690).

PROBLEMS

2.2-1 **Spherical Waves.** Use a spherical coordinate system to verify that the complex amplitude of the spherical wave (2.2-15) satisfies the Helmholtz equation (2.2-7).

2.2-2 **Intensity of a Spherical Wave.** Derive an expression for the intensity I of a spherical wave at a distance r from its center in terms of the optical power P. What is the intensity at $r = 1$ m for $P = 100$ W?

2.2-3 **Cylindrical Waves.** Derive expressions for the complex amplitude and intensity of a monochromatic wave whose wavefronts are cylinders centered about the y axis.

2.2-4 **Paraxial Helmholtz Equation.** Derive the paraxial Helmholtz equation (2.2-22) using the approximations in (2.2-20) and (2.2-21).

2.2-5 **Conjugate Waves.** Compare a monochromatic wave with complex amplitude $U(\mathbf{r})$ to a monochromatic wave of the same frequency but with complex amplitude $U^*(\mathbf{r})$, with respect to intensity, wavefronts, and wavefront normals. Use the plane wave $U(\mathbf{r}) = A\exp[-jk(x+y)/\sqrt{2}]$ and the spherical wave $U(\mathbf{r}) = (A/r)\exp(-jkr)$ as examples.

2.3-1 **Wave in a GRIN Slab.** Sketch the wavefronts of a wave traveling in the graded-index SELFOC slab described in Example 1.3-1.

2.4-1 **Reflection of a Spherical Wave from a Planar Mirror.** A spherical wave is reflected from a planar mirror sufficiently far from the wave origin so that the Fresnel approximation is satisfied. By regarding the spherical wave locally as a plane wave with slowly varying direction, use the law of reflection of plane waves to determine the nature of the reflected wave.

2.4-2 **Optical Path Length.** A plane wave travels in a direction normal to a thin plate made of N thin parallel layers of thicknesses d_q and refractive indices n_q, $q = 1, 2, \ldots, N$. If all reflections are ignored, determine the complex amplitude transmittance of the plate. If the plate is replaced with a distance d of free space, what should d be so that the same complex amplitude transmittance is obtained? Show that this distance is the optical path length defined in Sec. 1.1.

2.4-3 **Diffraction Grating.** Repeat Exercise 2.4-5 for a thin transparent plate whose thickness $d(x, y)$ is a square (instead of sinusoidal) periodic function of x of period

$\Lambda \gg \lambda$. Show that the angle θ between the diffracted waves is still given by $\theta \approx \lambda/\Lambda$. If a plane wave is incident in a direction normal to the grating, determine the amplitudes of the different diffracted plane waves.

2.4-4 **Reflectance of a Spherical Mirror.** Show that the complex amplitude reflectance $\mathscr{r}(x, y)$ (the ratio of the complex amplitudes of the reflected and incident waves) of a thin spherical mirror of radius R is given by $\mathscr{r}(x, y) = h_0 \exp[-jk_o(x^2 + y^2)/R]$, where h_0 is a constant. Compare this to the complex amplitude transmittance of a lens of focal length $f = -R/2$.

2.5-1 **Standing Waves.** Derive an expression for the intensity I of the superposition of two plane waves of wavelength λ traveling in opposite directions along the z axis. Sketch I versus z.

2.5-2 **Fringe Visibility.** The visibility of an interference pattern such as that described by (2.5-4) and plotted in Fig. 2.5-1 is defined as the ratio $\mathscr{V} = (I_{\text{max}} - I_{\text{min}})/(I_{\text{max}} + I_{\text{min}})$, where I_{max} and I_{min} are the maximum and minimum values of I. Derive an expression for $\mathscr{V}$ as a function of the ratio I_1/I_2 of the two interfering waves and determine the ratio I_1/I_2 for which the visibility is maximum.

2.5-3 **Michelson Interferometer.** If one of the mirrors of the Michelson interferometer (Fig. 2.5-3(*b*)) is misaligned by a small angle $\Delta\theta$, describe the shape of the interference pattern in the detector plane. What happens to this pattern as the other mirror moves?

2.6-1 **Pulsed Spherical Wave.** (a) Show that a pulsed spherical wave has a complex wavefunction of the form $U(\mathbf{r}, t) = (1/r)\mathscr{a}(t - r/c)$, where $\mathscr{a}(t)$ is an arbitrary function. (b) An ultrashort optical pulse has a complex wavefunction with central frequency corresponding to a wavelength $\lambda_o = 585$ nm and a Gaussian envelope of rms width $\sigma_t = 6$ fs (1 fs $= 10^{-15}$ s). How many optical cycles are contained within the pulse width? If the pulse propagates in free space as a spherical wave initiated at the origin at $t = 0$, describe the spatial distribution of the intensity as a function of the radial distance at time $t = 1$ ps.