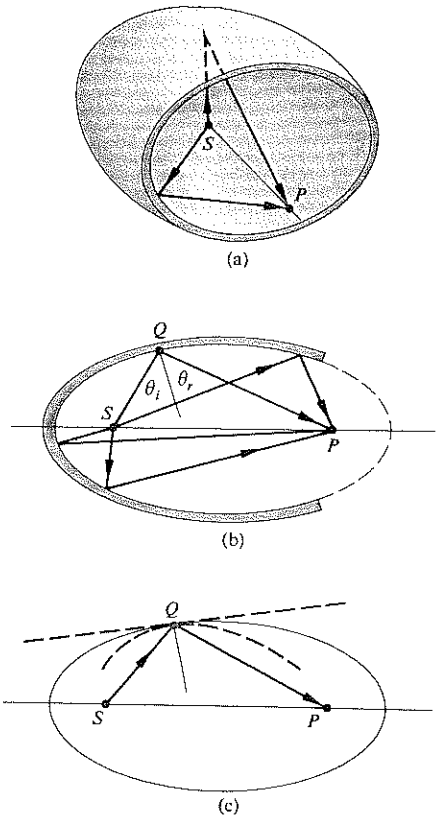
lution of variational principles brings us back to Optics via the modern formalism of Quantum Optics.



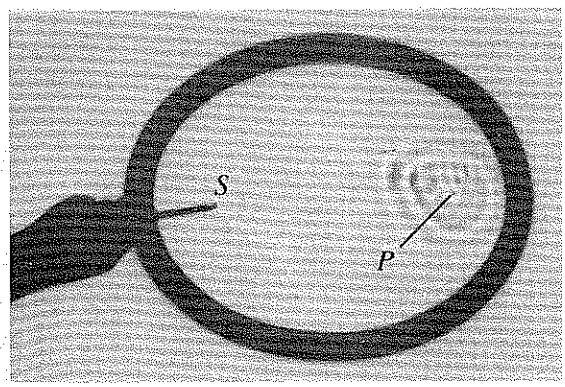


Figure 4.37 Reflection off an ellipsoidal surface. Observe the reflection of waves using a frying pan filled with water. Even though these are usually circular, it is well worth playing with. (Photo courtesy *PSSC College Physics*, D. C. Heath & Co., 1968.)

Fermat's Principle is not so much a computational device as it is a concise way of thinking about the propagation of light. It is a statement about the grand scheme of things without any concern for the contributing mechanisms, and as such it will yield insights under a myriad of circumstances.

4.6 The Electromagnetic Approach

Thus far, we have studied reflection and refraction from the perspectives of Scattering Theory, the Theorem of Malus and Dupin, and Fermat's Principle. Yet another and even more powerful approach is provided by Electromagnetic Theory. Unlike the previous techniques, which say nothing about the incident, reflected, and transmitted radiant flux densities (i.e., I_i, I_r, I_t, respectively), Electromagnetic Theory treats these within the framework of a far more complete description.

4.6.1 Waves at an Interface

Suppose that the incident monochromatic lightwave is planar, so that it has the form

$$\vec{\mathbf{E}}_i = \vec{\mathbf{E}}_{0i} \exp\left[i(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t)\right] \tag{4.11}$$

or, more simply,

$$\vec{\mathbf{E}}_i = \vec{\mathbf{E}}_{0i} \cos\left(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t\right) \tag{4.12}$$

Assume that $\vec{\mathbf{E}}_{0i}$ is constant in time; that is, the wave is linearly or plane polarized. We'll find in Chapter 8 that any form of light can be represented by two orthogonal linearly polarized waves, so that this doesn't actually represent a restriction. Note that just as the origin in time, $t = 0$, is arbitrary, so too is the origin O in space, where $\vec{\mathbf{r}} = 0$. Thus, making no assumptions about their directions, frequencies, wavelengths, phases, or amplitudes, we can write the reflected and transmitted waves as

$$\vec{\mathbf{E}}_r = \vec{\mathbf{E}}_{0r} \cos\left(\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} - \omega_r t + \varepsilon_r\right) \tag{4.13}$$

and

$$\vec{\mathbf{E}}_t = \vec{\mathbf{E}}_{0t} \cos\left(\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} - \omega_t t + \varepsilon_t\right) \tag{4.14}$$

Here ε_r and ε_t are *phase constants* relative to $\vec{\mathbf{E}}_i$ and are introduced because the position of the origin is not unique. Figure 4.38 depicts the waves in the vicinity of the planar interface between two homogeneous lossless dielectric media of indices n_i and n_t.

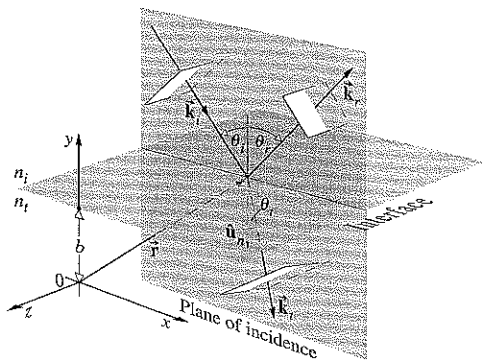


Figure 4.38 Plane waves incident on the boundary between two homogeneous, isotropic, lossless dielectric media.

The laws of Electromagnetic Theory (Section 3.1) lead to certain requirements that must be met by the fields, and they are referred to as the *boundary conditions*. Specifically, one of these is that the component of the electric field $\vec{\mathbf{E}}$ that is tangent to the interface must be continuous across it (the same is true for $\vec{\mathbf{H}}$). In other words, the total tangential component of $\vec{\mathbf{E}}$ on one side of the surface must equal that on the other (Problem 4.37). Thus, since $\hat{\mathbf{u}}_n$ is the unit vector normal to the interface, regardless of the direction of the electric field within the wavefront, the cross-product of it with $\hat{\mathbf{u}}_n$ will be perpendicular to $\hat{\mathbf{u}}_n$ and therefore tangent to the interface. Hence

$$\hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_i + \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_r = \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_t \tag{4.15}$$

or

$$\begin{aligned}\hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_{0i} \cos\,(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t) \\ + \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_{0r} \cos\,(\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} - \omega_r t + \varepsilon_r) \\ = \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_{0t} \cos\,(\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} - \omega_t t + \varepsilon_t)\end{aligned} \tag{4.16}$$

This relationship must obtain at any instant in time and at any point on the interface ($y = b$). Consequently, $\vec{\mathbf{E}}_i$, $\vec{\mathbf{E}}_r$, and $\vec{\mathbf{E}}_t$ must have precisely the same functional dependence on the variables t and r, which means that

$$\begin{aligned}(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t)\big|_{y=b} = (\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} - \omega_r t + \varepsilon_r)\big|_{y=b} \\ = (\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} - \omega_t t + \varepsilon_t)\big|_{y=b}\end{aligned} \tag{4.17}$$

With this as the case, the cosines in Eq. (4.16) cancel, leaving an expression independent of t and r, as indeed it must be. Inasmuch as this has to be true for all values of time, the coefficients of t must be equal, to wit

$$\omega_i = \omega_r = \omega_t \tag{4.18}$$

Recall that the electrons within the media are undergoing (linear) forced vibrations at the frequency of the incident wave. Whatever light is scattered has that same frequency. Furthermore,

$$(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}})\big|_{y=b} = (\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} + \varepsilon_r)\big|_{y=b} = (\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} + \varepsilon_t)\big|_{y=b} \tag{4.19}$$

wherein $\vec{\mathbf{r}}$ terminates on the interface. The values of ε_r and ε_t correspond to a given position of O, and thus they allow the relation to be valid regardless of that location. (For example, the origin might be chosen such that $\vec{\mathbf{r}}$ was perpendicular to $\vec{\mathbf{k}}_i$ but not to $\vec{\mathbf{k}}_r$ or $\vec{\mathbf{k}}_t$.) From the first two terms we obtain

$$[(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r) \cdot \vec{\mathbf{r}}]_{y=b} = \varepsilon_r \tag{4.20}$$

Recalling Eq. (2.43), this expression simply says that the endpoint of $\vec{\mathbf{r}}$ sweeps out a plane (which is of course the interface) perpendicular to the vector $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$. To phrase it slightly differently, $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$ is parallel to $\hat{\mathbf{u}}_n$. Notice, however, that since the incident and reflected waves are in the same medium, $k_i = k_r$. From the fact that $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$ has no component in the plane of the interface, that is, $\hat{\mathbf{u}}_n \times (\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r) = 0$, we conclude that

$$k_i \sin\,\theta_i = k_r \sin\,\theta_r$$

Hence we have the Law of Reflection; that is,

$$\theta_i = \theta_r$$

Furthermore, since $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$ is parallel to $\hat{\mathbf{u}}_n$ all three vectors, $\vec{\mathbf{k}}_i$, $\vec{\mathbf{k}}_r$, and $\hat{\mathbf{u}}_n$, are in the same plane, the plane-of-incidence. Again, from Eq. (4.19)

$$[(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_t) \cdot \vec{\mathbf{r}}]_{y=b} = \varepsilon_t \tag{4.21}$$

and therefore $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_t)$ is also normal to the interface. Thus $\vec{\mathbf{k}}_i$, $\vec{\mathbf{k}}_r$, $\vec{\mathbf{k}}_t$, and $\hat{\mathbf{u}}_n$ are all coplanar. As before, the tangential components of $\vec{\mathbf{k}}_i$ and $\vec{\mathbf{k}}_t$ must be equal, and consequently

$$k_i \sin\,\theta_i = k_t \sin\,\theta_t \tag{4.22}$$

But because $\omega_i = \omega_t$, we can multiply both sides by c/ω_i to get

$$n_i \sin\,\theta_i = n_t \sin\,\theta_t$$

which is Snell's Law. Finally, if we had chosen the origin O to

be in the interface, it is evident from Eqs. (4.20) and (4.21) that ε_r and ε_t would both have been zero. That arrangement, though not as instructive, is certainly simpler, and we'll use it from here on.

4.6.2 The Fresnel Equations

We have just found the relationship that exists among the phases of $\vec{\mathbf{E}}_i(\vec{r}, t)$, $\vec{\mathbf{E}}_r(\vec{r}, t)$, and $\vec{\mathbf{E}}_t(\vec{r}, t)$ at the boundary. There is still an interdependence shared by the amplitudes $\vec{\mathbf{E}}_{0i}$, $\vec{\mathbf{E}}_{0r}$, and $\vec{\mathbf{E}}_{0t}$, which can now be evaluated. To that end, suppose that a plane monochromatic wave is incident on the planar surface separating two isotropic media. Whatever the polarization of the wave, we shall resolve its $\vec{\mathbf{E}}$- and $\vec{\mathbf{B}}$-fields into components parallel and perpendicular to the plane-of-incidence and treat these constituents separately.

Case 1: $\vec{\mathbf{E}}$ perpendicular to the plane-of-incidence. Assume that $\vec{\mathbf{E}}$ is perpendicular to the plane-of-incidence and that $\vec{\mathbf{B}}$ is parallel to it (Fig. 4.39). Recall that $E = vB$, so that

$$\hat{\mathbf{k}} \times \vec{\mathbf{E}} = v\vec{\mathbf{B}} \tag{4.23}$$

and

$$\hat{\mathbf{k}} \cdot \vec{\mathbf{E}} = 0 \tag{4.24}$$

(i.e., $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and the unit propagation vector $\hat{\mathbf{k}}$ form a right-handed system). Again, making use of the continuity of the tangential components of the $\vec{\mathbf{E}}$-field, we have at the boundary at any time and any point

$$\vec{\mathbf{E}}_{0i} + \vec{\mathbf{E}}_{0r} = \vec{\mathbf{E}}_{0t} \tag{4.25}$$

where the cosines cancel. Realize that the field vectors as shown really ought to be envisioned at $y = 0$ (i.e., at the surface), from which they have been displaced for the sake of clarity. Note too that although $\vec{\mathbf{E}}_r$ and $\vec{\mathbf{E}}_t$ must be normal to the plane-of-incidence by symmetry, *we are guessing that they point outward* at the interface when $\vec{\mathbf{E}}_i$ does. The directions of the $\vec{\mathbf{B}}$-fields then follow from Eq. (4.23).

We will need to invoke another of the boundary conditions in order to get one more equation. The presence of material substances that become electrically polarized by the wave has a definite effect on the field configuration. Thus, although the tangential component of $\vec{\mathbf{E}}$ is continuous across the boundary, its normal component is not. Instead, the normal component of the product $\epsilon\vec{\mathbf{E}}$ is the same on either side of the interface. Similarly, the normal component of $\vec{\mathbf{B}}$ is continuous, as is the tangential component of $\mu^{-1}\vec{\mathbf{B}}$. Here the magnetic effect of the two media appears via their permeabilities μ_i and μ_t. This boundary condition will be the simplest to use, particularly as applied to reflection from the surface of a conductor.* Thus the continuity of the tangential component of $\vec{\mathbf{B}}/\mu$ requires that

$$-\frac{B_i}{\mu_i}\cos\theta_i + \frac{B_r}{\mu_i}\cos\theta_r = -\frac{B_t}{\mu_t}\cos\theta_t \tag{4.26}$$

where the left and right sides are the total magnitudes of $\vec{\mathbf{B}}/\mu$ parallel to the interface in the incident and transmitting media, respectively. The positive direction is that of increasing x, so

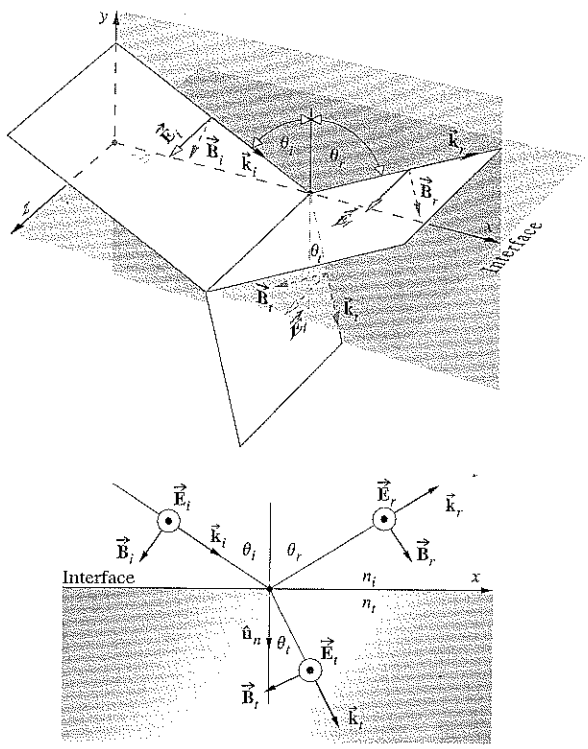


Figure 4.39 An incoming wave whose $\vec{\mathbf{E}}$-field is normal to the plane-of-incidence.

*In keeping with our intent to use only the $\vec{\mathbf{E}}$- and $\vec{\mathbf{B}}$-fields, at least in the early part of this exposition, we have avoided the usual statements in terms of $\vec{\mathbf{H}}$, where

$$\vec{\mathbf{H}} = \mu^{-1}\vec{\mathbf{B}} \qquad \text{[A1.14]}$$

that the scalar components of $\vec{\mathbf{B}}_i$ and $\vec{\mathbf{B}}_t$ appear with minus signs. From Eq. (4.23) we have

$$B_i = E_i/v_i \tag{4.27}$$

$$B_r = E_r/v_r \tag{4.28}$$

and

$$B_t = E_t/v_t \tag{4.29}$$

Since $v_i = v_r$ and $\theta_i = \theta_r$, Eq. (4.26) can be written as

$$\frac{1}{\mu_i v_i}(E_i - E_r)\cos\theta_i = \frac{1}{\mu_t v_t}E_t\cos\theta_t \tag{4.30}$$

Making use of Eqs. (4.12), (4.13), and (4.14) and remembering that the cosines therein equal one another at $y = 0$, we obtain

$$\frac{n_i}{\mu_i}(E_{0i} - E_{0r})\cos\theta_i = \frac{n_t}{\mu_t}E_{0t}\cos\theta_t \tag{4.31}$$

Combined with Eq. (4.25), this yields

$$\left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{\dfrac{n_i}{\mu_i}\cos\theta_i - \dfrac{n_t}{\mu_t}\cos\theta_t}{\dfrac{n_i}{\mu_i}\cos\theta_i + \dfrac{n_t}{\mu_t}\cos\theta_t} \tag{4.32}$$

and

$$\left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2\dfrac{n_i}{\mu_i}\cos\theta_i}{\dfrac{n_i}{\mu_i}\cos\theta_i + \dfrac{n_t}{\mu_t}\cos\theta_t} \tag{4.33}$$

The $\perp$ subscript serves as a reminder that we are dealing with the case in which $\vec{\mathbf{E}}$ is perpendicular to the plane-of-incidence. These two expressions, *which are completely general statements applying to any linear, isotropic, homogeneous media*, are two of the **Fresnel Equations**. Most often one deals with dielectrics for which $\mu_i \approx \mu_t \approx \mu_0$; consequently, the common form of these equations is simply

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{n_i\cos\theta_i - n_t\cos\theta_t}{n_i\cos\theta_i + n_t\cos\theta_t} \tag{4.34}$$

and

$$t_\perp \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2n_i\cos\theta_i}{n_i\cos\theta_i + n_t\cos\theta_t} \tag{4.35}$$

Here $r_\perp$ denotes the **amplitude reflection coefficient**, and $t_\perp$ is the **amplitude transmission coefficient**.

Case 2: $\vec{\mathbf{E}}$ parallel to the plane-of-incidence. A similar pair of equations can be derived when the incoming $\vec{\mathbf{E}}$-field lies in the plane-of-incidence, as shown in Fig. 4.40. Continuity of the tangential components of $\vec{\mathbf{E}}$ on either side of the boundary leads to

$$E_{0i}\cos\theta_i - E_{0r}\cos\theta_r = E_{0t}\cos\theta_t \tag{4.36}$$

In much the same way as before, continuity of the tangential components of $\vec{\mathbf{B}}/\mu$ yields

$$\frac{1}{\mu_i v_i}E_{0i} + \frac{1}{\mu_r v_r}E_{0r} = \frac{1}{\mu_t v_t}E_{0t} \tag{4.37}$$

Using the fact that $\mu_i = \mu_r$ and $\theta_i = \theta_r$, we can combine these formulas to obtain two more of the *Fresnel Equations*:

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\parallel = \frac{\dfrac{n_t}{\mu_t}\cos\theta_i - \dfrac{n_i}{\mu_i}\cos\theta_t}{\dfrac{n_i}{\mu_i}\cos\theta_i + \dfrac{n_t}{\mu_t}\cos\theta_t} \tag{4.38}$$

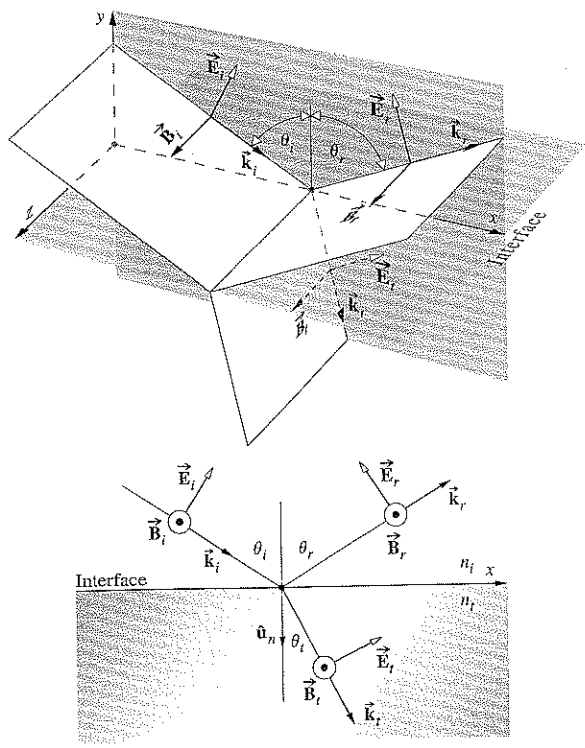


Figure 4.40 An incoming wave whose $\vec{\mathbf{E}}$-field is in the plane-of-incidence.

and

$$t_{\parallel} \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_{\parallel} = \frac{2\dfrac{n_i}{\mu_i}\cos\theta_i}{\dfrac{n_i}{\mu_i}\cos\theta_t + \dfrac{n_t}{\mu_t}\cos\theta_i} \tag{4.39}$$

When both media forming the interface are dielectrics that are essentially "nonmagnetic" (p. 66), the amplitude coefficients become

$$r_{\parallel} = \frac{n_t\cos\theta_i - n_i\cos\theta_t}{n_i\cos\theta_t + n_t\cos\theta_i} \tag{4.40}$$

and

$$t_{\parallel} = \frac{2n_i\cos\theta_i}{n_i\cos\theta_t + n_t\cos\theta_i} \tag{4.41}$$

One further notational simplification can be made using Snell's Law, whereupon the Fresnel Equations for dielectric media become (Problem 4.39)

$$r_{\perp} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \tag{4.42}$$

$$r_{\parallel} = +\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \tag{4.43}$$

$$t_{\perp} = +\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i + \theta_t)} \tag{4.44}$$

$$t_{\parallel} = +\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i + \theta_t)\cos(\theta_i - \theta_t)} \tag{4.45}$$

A note of caution must be introduced here. Bear in mind that the directions (or more precisely, the phases) of the fields in Figs. 4.39 and 4.40 were selected rather arbitrarily. For example, in Fig. 4.39 we could have assumed that $\vec{\mathbf{E}}_r$ pointed inward, whereupon $\vec{\mathbf{B}}_r$ would have had to be reversed as well. Had we done that, the sign of $r_{\perp}$ would have turned out to be positive, leaving the other amplitude coefficients unchanged. The signs appearing in Eqs. (4.42) through (4.45), which are positive except for the first, correspond to the particular set of field directions selected. The minus sign in Eq. (4.42), as we will see, just means that we didn't guess correctly concerning $\vec{\mathbf{E}}_r$ in Fig. 4.39. Nonetheless, be aware that the literature is not standardized, and all possible sign variations have been labeled the *Fresnel Equations*. To avoid confusion *they must be related to the specific field directions from which they were derived.*

4.6.3 Interpretation of the Fresnel Equations

This section examines the physical implications of the Fresnel Equations. In particular, we are interested in determining the fractional amplitudes and flux densities that are reflected and refracted. In addition we shall be concerned with any possible phase shifts that might be incurred in the process.

Amplitude Coefficients

Let's briefly examine the form of the amplitude coefficients over the entire range of θ_i values. At nearly normal incidence ($\theta_i \approx 0$) the tangents in Eq. (4.43) are essentially equal to sines, in which case

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \left[\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}\right]_{\theta_i=0}$$

We will come back to the physical significance of the minus sign presently. After expanding the sines and using Snell's Law, this expression becomes

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \left[\frac{n_t\cos\theta_i - n_i\cos\theta_t}{n_t\cos\theta_i + n_i\cos\theta_t}\right]_{\theta_i=0} \tag{4.46}$$

which follows as well from Eqs. (4.34) and (4.40). In the limit, as θ_i goes to 0, $\cos\theta_i$ and $\cos\theta_t$ both approach one, and consequently

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \frac{n_t - n_i}{n_t + n_i} \tag{4.47}$$

This equality of the reflection coefficients arises because the plane-of-incidence is no longer specified when $\theta_t = 0$. Thus, for example, at an air ($n_i = 1$) glass ($n_t = 1.5$) interface at nearly normal incidence, the amplitude reflection coefficients equal ± 0.2. (See Problem 4.45.)

When $n_t > n_i$ it follows from Snell's Law that $\theta_i > \theta_t$, and $r_{\perp}$ is negative for all values of θ_i (Fig. 4.41). In contrast, Eq. (4.43) tells us that $r_{\parallel}$ starts out positive at $\theta_i = 0$ and decreases gradually until it equals zero when $(\theta_i + \theta_t) = 90°$, since there $\tan\pi/2$ is infinite. The particular value of the incident angle for which this occurs is denoted by θ_p and referred to as the **polarization angle** (see Section 8.6.1). Notice that $r_{\parallel} \to 0$ at θ_p, just when the phase shifts 180°. That means we won't see the $\vec{\mathbf{E}}$-field do any flipping when θ_i approaches θ_p from either side. As θ_i increases beyond θ_p, $r_{\parallel}$ becomes progressively more negative, reaching -1.0 at 90°.

If you place a single sheet of glass, a microscope slide, on this page and look straight down into it ($\theta_i = 0$), the region

beneath the glass will seem decidedly grayer than the rest of the paper, because the slide will reflect at both its interfaces, and the light reaching and returning from the paper will be diminished appreciably. Now hold the slide near your eye and again view the page through it as you tilt it, increasing θ_i. The amount of light reflected will increase, and it will become more difficult to see the page through the glass. When $\theta_i \approx 90°$ the slide will look like a perfect mirror as the reflection coefficients (Fig. 4.41) go to -1.0. Even a rather poor surface (see photo), such as the cover of this book, will be mirrorlike at glancing incidence. Hold the book horizontally at the level of the middle of your eye and face a bright light; you will see the source reflected rather nicely in the cover. This suggests that even X-rays could be mirror-reflected at glancing incidence (p. 242), and modern X-ray telescopes are based on that very fact.

At normal incidence Eqs. (4.35) and (4.41) lead rather straightforwardly to

$$[t_{\parallel}]_{\theta_i=0} = [t_{\perp}]_{\theta_i=0} = \frac{2n_i}{n_i + n_t} \tag{4.48}$$

It will be shown in Problem 4.50 that the expression

$$t_{\perp} + (-r_{\perp}) = 1 \tag{4.49}$$

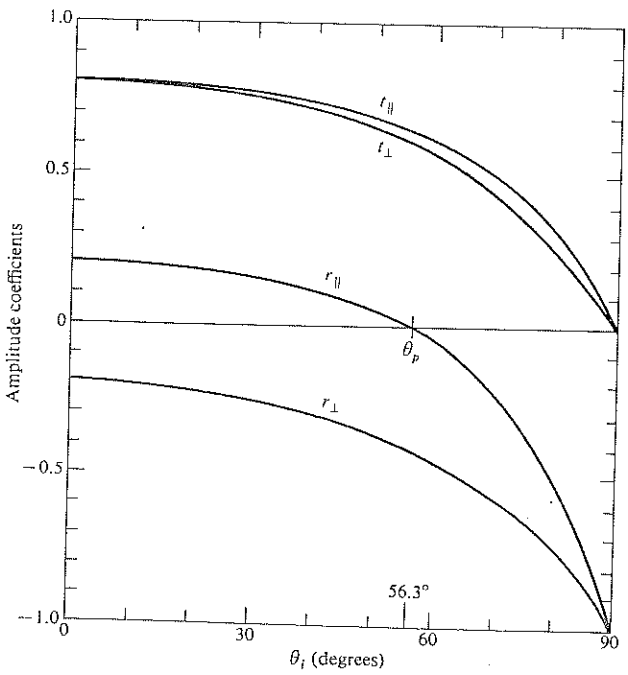
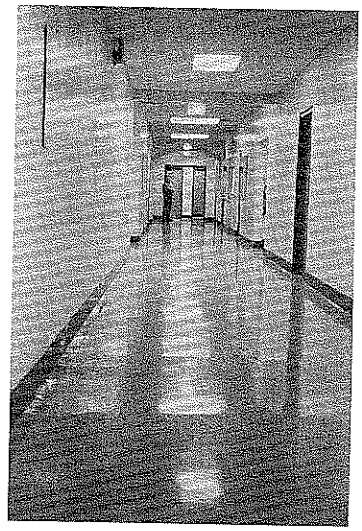


Figure 4.41 The amplitude coefficients of reflection and transmission as a function of incident angle. These correspond to external reflection $n_t > n_i$ at an air–glass interface ($n_{ti} = 1.5$).



At near-glancing incidence the walls and floor are mirrorlike—this despite the fact that the surfaces are rather poor reflectors at $\theta_i = 0°$. (Photo by E.H.)

holds for all θ_i, whereas

$$t_{\parallel} + r_{\parallel} = 1 \tag{4.50}$$

is true only at normal incidence.

The foregoing discussion, for the most part, was restricted to the case of **external reflection** (i.e., $n_t > n_i$). The opposite situation of **internal reflection**, in which the incident medium is the more dense ($n_i > n_t$), is of interest as well. In that instance $\theta_t > \theta_i$, and $r_{\perp}$, as described by Eq. (4.42), will always be positive. Figure 4.42 shows that $r_{\perp}$ increases from its initial value [Eq. (4.47)] at $\theta_i = 0$, reaching $+1$ at what is called the **critical angle**, θ_c. Specifically, θ_c is the special value of the incident angle (p. 122) for which $\theta_t = \pi/2$. Likewise, $r_{\parallel}$ starts off negatively [Eq. (4.47)] at $\theta_i = 0$ and thereafter increases, reaching $+1$ at $\theta_i = \theta_c$, as is evident from the Fresnel Equation (4.40). Again, $r_{\parallel}$ passes through zero at the *polarization angle* θ_p'. It is left for Problem 4.66 to show that the polarization angles θ_p' and θ_p for internal and external reflection at the interface between the same media are simply the complements of each other. We will return to internal reflection in Section 4.7, where it will be shown that $r_{\perp}$ and $r_{\parallel}$ are complex quantities for $\theta_i > \theta_c$.

Phase Shifts

It should be evident from Eq. (4.42) that $r_{\perp}$ is negative regardless of θ_i when $n_t > n_i$. Yet we saw earlier that had we chosen $[\vec{E}_r]_{\perp}$ in Fig. 4.37 to be in the opposite direction, the first Fres-

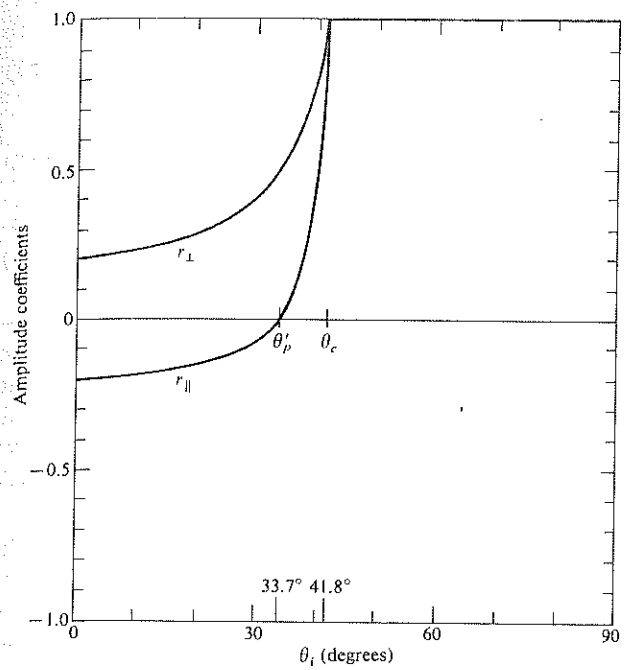


Figure 4.42 The amplitude coefficients of reflection as a function of incident angle. These correspond to internal reflection $n_t < n_i$ at an air–glass interface ($n_{ti} = 1/1.5$).

nel Equation (4.42) would have changed signs, causing $r_\perp$ to become a positive quantity. The sign of $r_\perp$ is associated with the relative directions of $[\vec{\mathbf{E}}_{0i}]_\perp$ and $[\vec{\mathbf{E}}_{0r}]_\perp$. Bear in mind that a reversal of $[\vec{\mathbf{E}}_{0r}]_\perp$ is tantamount to introducing a phase shift, $\Delta\varphi_\perp$, of π radians into $[\vec{\mathbf{E}}_r]_\perp$. Hence at the boundary $[\vec{\mathbf{E}}_i]_\perp$ and $[\vec{\mathbf{E}}_r]_\perp$ will be antiparallel and therefore π out-of-phase with each other, as indicated by the negative value of $r_\perp$. When we consider components normal to the plane-of-incidence, there is no confusion as to whether two fields are in-phase or π radians out-of-phase: if parallel, they're in-phase; if antiparallel, they're π out-of-phase. In summary, then, ***the component of the electric field normal to the plane-of-incidence undergoes a phase shift of π radians upon reflection when the incident medium has a lower index than the transmitting medium.*** Similarly, $t_\perp$ and $t_\parallel$ are always positive and $\Delta\varphi = 0$. *Furthermore, when $n_i > n_t$ no phase shift in the normal component results on reflection, that is, $\Delta\varphi_\perp = 0$ so long as $\theta_i < \theta_c$.*

Things are a bit less obvious when we deal with $[\vec{\mathbf{E}}_i]_\parallel$, $[\vec{\mathbf{E}}_r]_\parallel$, and $[\vec{\mathbf{E}}_t]_\parallel$. It now becomes necessary to define more explicitly what is meant by *in-phase*, since the field vectors are coplanar but generally not colinear. The field directions were chosen in Figs. 4.39 and 4.40 such that if you looked down any one of the propagation vectors toward the direction from which the light was coming, $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{k}}$ would appear to have the same relative orientation whether the ray was incident, reflected, or transmitted. We can use this as the required condition for two **E**-fields to be in-phase. Equivalently, but more simply, ***two fields in the incident plane are in-phase if their y-components are parallel and are out-of-phase if the components are antiparallel.*** Notice that when two $\vec{\mathbf{E}}$-fields are out-of-phase so too are their associated $\vec{\mathbf{B}}$-fields and vice versa. With this definition we need only look at the vectors normal to the plane-of-incidence, whether they be $\vec{\mathbf{E}}$ or $\vec{\mathbf{B}}$, to determine the relative phase of the accompanying fields in the incident plane. Thus in Fig. 4.43*a* $\vec{\mathbf{E}}_i$ and $\vec{\mathbf{E}}_t$ are in-phase, as are $\vec{\mathbf{B}}_i$ and $\vec{\mathbf{B}}_t$, whereas $\vec{\mathbf{E}}_i$ and $\vec{\mathbf{E}}_r$ are out-of-phase, along with $\vec{\mathbf{B}}_i$ and $\vec{\mathbf{B}}_r$. Similarly, in Fig. 4.43*b* $\vec{\mathbf{E}}_i$, $\vec{\mathbf{E}}_r$, and $\vec{\mathbf{E}}_t$ are in-phase, as are $\vec{\mathbf{B}}_i$, $\vec{\mathbf{B}}_r$, and $\vec{\mathbf{B}}_t$.

Now, the amplitude reflection coefficient for the parallel component is given by

$$r_\parallel = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t}$$

which is positive ($\Delta\varphi_\parallel = 0$) as long as

$$n_t \cos\theta_i - n_i \cos\theta_t > 0$$

that is, if

$$\sin\theta_i \cos\theta_i - \cos\theta_t \sin\theta_t > 0$$

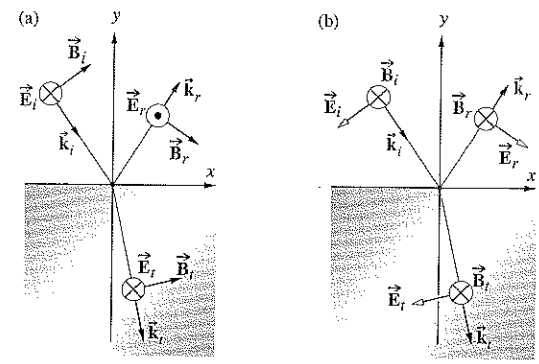


Figure 4.43 Field orientations and phase shifts.

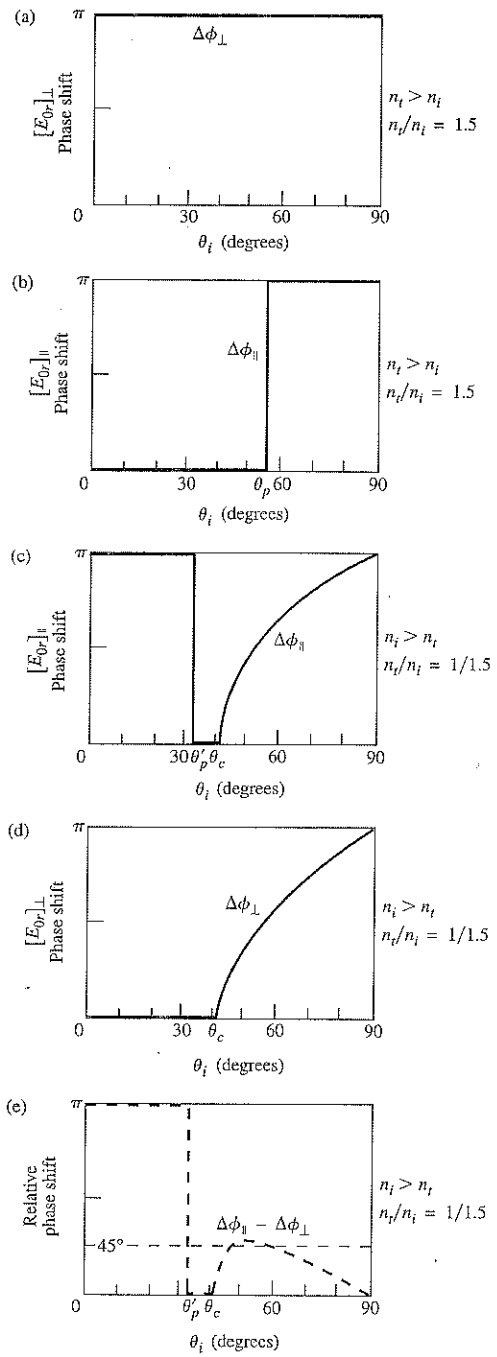


Figure 4.44 Phase shifts for the parallel and perpendicular components of the $\vec{E}$-field corresponding to internal and external reflection.

or equivalently

$$\sin(\theta_i - \theta_t)\cos(\theta_i + \theta_t) > 0 \tag{4.51}$$

This will be the case for $n_i < n_t$ if

$$(\theta_i + \theta_t) < \pi/2 \tag{4.52}$$

and for $n_i > n_t$ when

$$(\theta_i + \theta_t) > \pi/2 \tag{4.53}$$

Thus when $n_i < n_t$, $[\vec{\mathbf{E}}_{0r}]_\parallel$ and $[\vec{\mathbf{E}}_{0i}]_\parallel$ will be in-phase ($\Delta\varphi_\parallel = 0$) until $\theta_i = \theta_p$ and out-of-phase by π radians thereafter. The transition is not actually discontinuous, since $[\vec{\mathbf{E}}_{0r}]_\parallel$ goes to zero at θ_p. In contrast, for internal reflection $r_\parallel$ is negative until θ_p', which means that $\Delta\varphi_\parallel = \pi$. From θ_p' to θ_c, $r_\parallel$ is positive and $\Delta\varphi_\parallel = 0$. Beyond θ_c, $r_\parallel$ becomes complex, and $\Delta\varphi_\parallel$ gradually increases to π at $\theta_i = 90°$.

Figure 4.44, which summarizes these conclusions, will be of continued use to us. The actual functional form of $\Delta\varphi_\parallel$ and $\Delta\varphi_\perp$ for internal reflection in the region where $\theta_i > \theta_c$ can be found in the literature,* but the curves depicted here will suffice for our purposes. Figure 4.44*e* is a plot of the relative phase shift between the parallel and perpendicular components, that is, $\Delta\varphi_\parallel - \Delta\varphi_\perp$. It is included here because it will be useful later on (e.g., when we consider polarization effects). Finally, many of the essential features of this discussion are illustrated in Figs. 4.45 and 4.46. The amplitudes of the reflected vectors are in accord with those of Figs. 4.41 and 4.42 (for an air–glass interface), and the phase shifts agree with those of Fig. 4.44.

Many of these conclusions can be verified with the simplest experimental equipment, namely, two linear polarizers, a piece of glass, and a small source, such as a flashlight or high-intensity lamp. By placing one polarizer in front of the source (at 45° to the plane-of-incidence), you can easily duplicate the conditions of Fig. 4.45. For example, when $\theta_i = \theta_p$ (Fig. 4.45*b*) no light will pass through the second polarizer if its transmission axis is parallel to the plane-of-incidence. In comparison, at near-glancing incidence the reflected beam will vanish when the axes of the two polarizers are almost normal to each other.

*Born and Wolf, *Principles of Optics*, p. 49.

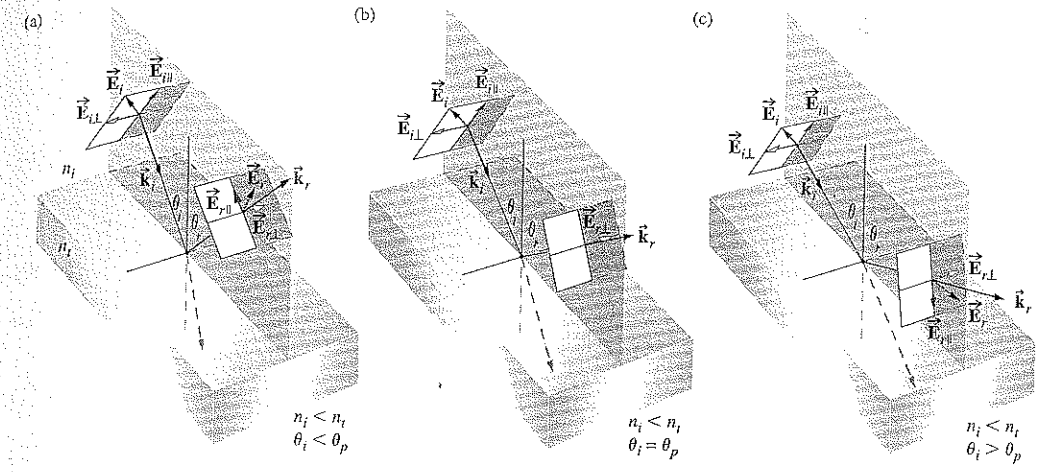


Figure 4.45 The reflected $\vec{\mathbf{E}}$-field at various angles concomitant with external reflection.

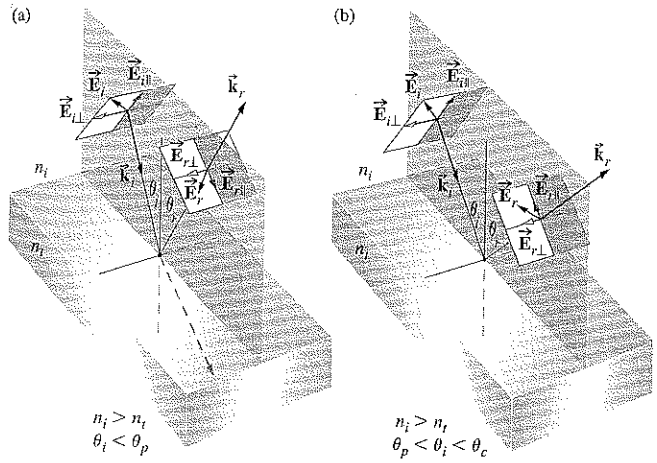


Figure 4.46 The reflected $\vec{\mathbf{E}}$-field at various angles concomitant with internal reflection.

Reflectance and Transmittance

Consider a circular beam of light incident on a surface, as shown in Fig. 4.47, such that there is an illuminated spot of area A. Recall that the power per unit area crossing a surface in vacuum whose normal is parallel to $\vec{\mathbf{S}}$, the Poynting vector, is given by

$$\vec{\mathbf{S}} = c^2\epsilon_0\vec{\mathbf{E}} \times \vec{\mathbf{B}} \qquad [3.40]$$

Furthermore, the radiant flux density (W/m^2) or irradiance is

$$I = \langle S \rangle_T = \frac{c\epsilon_0}{2} E_0^2 \qquad [3.44]$$

This is the average energy per unit time crossing a unit area normal to $\vec{\mathbf{S}}$ (in isotropic media $\vec{\mathbf{S}}$ is parallel to $\vec{\mathbf{k}}$). In the case at hand (Fig. 4.47), let I_i, I_r, and I_t be the incident, reflected, and transmitted flux densities, respectively. The cross-sectional areas of the incident, reflected, and transmitted beams are, respectively, $A\cos\theta_i$, $A\cos\theta_r$, and $A\cos\theta_t$. Accordingly, the incident power is $I_iA\cos\theta_i$; this is the energy per unit time flowing in the incident beam, and it's therefore the power

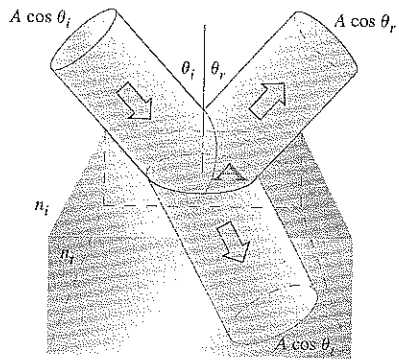


Figure 4.47 Reflection and transmission of an incident beam.

arriving on the surface over A. Similarly, $I_r A \cos \theta_r$ is the power in the reflected beam, and $I_t A \cos \theta_t$ is the power being transmitted through A. We define the **reflectance** R to be the ratio of the reflected power (or flux) to the incident power:

$$R \equiv \frac{I_r A \cos \theta_r}{I_i A \cos \theta_i} = \frac{I_r}{I_i} \tag{4.54}$$

In the same way, the **transmittance** T is defined as the ratio of the transmitted to the incident flux and is given by

$$T \equiv \frac{I_t \cos \theta_t}{I_i \cos \theta_i} \tag{4.55}$$

The quotient I_r/I_i equals $(v_r \epsilon_r E_{0r}^2/2)/(v_i \epsilon_i E_{0i}^2/2)$, and since the incident and reflected waves are in the same medium, $v_r = v_i$, $\epsilon_r = \epsilon_i$, and

$$R = \left(\frac{E_{0r}}{E_{0i}}\right)^2 = r^2 \tag{4.56}$$

In like fashion (assuming $\mu_i = \mu_t = \mu_0$),

$$T = \frac{n_t \cos \theta_t}{n_i \cos \theta_i}\left(\frac{E_{0t}}{E_{0i}}\right)^2 = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right) t^2 \tag{4.57}$$

where use was made of the fact that $\mu_0 \epsilon_t = 1/v_t^2$ and $\mu_0 v_i \epsilon_t = n_t/c$. Notice that at normal incidence, which is a situation of great practical interest, $\theta_t = \theta_i = 0$, and the transmittance [Eq. (4.55)], like the reflectance [Eq. (4.54)], is then simply the ratio of the appropriate irradiances. Since $R = r^2$, we need not worry about the sign of r in any particular formulation, and that makes reflectance a convenient notion. Observe that in Eq. (4.57) T is not simply equal to t^2, for two reasons. First, the ratio of the indices of refraction must be there, since the speeds at which energy is transported into and out of the interface are different, in other words, $I \propto v$, from Eq. (3.47). Second, the cross-sectional areas of the incident and refracted beams are different. The energy flow per unit area is affected accordingly, and that manifests itself in the presence of the ratio of the cosine terms.

Let's now write an expression representing the conservation of energy for the configuration depicted in Fig. 4.47. In other words, the total energy flowing into area A per unit time must equal the energy flowing outward from it per unit time:

$$I_i A \cos \theta_i = I_r A \cos \theta_r + I_t A \cos \theta_t \tag{4.58}$$

When both sides are multiplied by c, this expression becomes

$$n_i E_{0i}^2 \cos \theta_i = n_i E_{0r}^2 \cos \theta_i + n_i E_{0t}^2 \cos \theta_t$$

or

$$1 = \left(\frac{E_{0r}}{E_{0i}}\right)^2 + \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right)\left(\frac{E_{0t}}{E_{0i}}\right)^2 \tag{4.59}$$

But this is simply

$$R + T = 1 \tag{4.60}$$

where there was no absorption. It is convenient to use the component forms, that is,

$$R_\perp = r_\perp^2 \tag{4.61}$$

$$R_\parallel = r_\parallel^2 \tag{4.62}$$

$$T_\perp = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right) t_\perp^2 \tag{4.63}$$

and

$$T_\parallel = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right) t_\parallel^2 \tag{4.64}$$

which are illustrated in Fig. 4.48. Furthermore, it can be shown (Problem 4.71) that

$$R_\parallel + T_\parallel = 1 \tag{4.65}$$

and

$$R_\perp + T_\perp = 1 \tag{4.66}$$

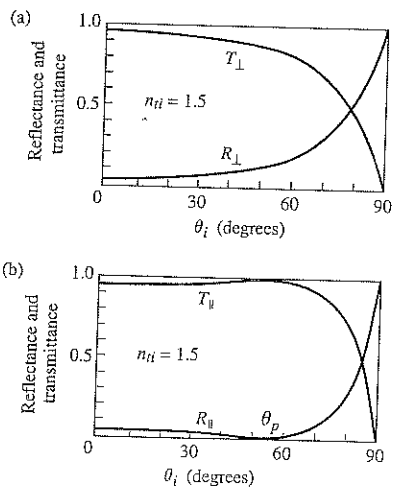
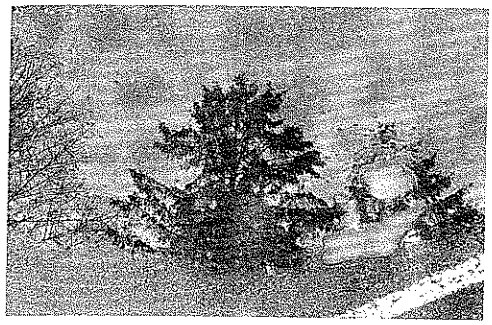


Figure 4.48 Reflectance and transmittance versus incident angle.



Looking down into a puddle (that's melting snow on the right) we see a reflection of the surrounding trees. At normal incidence water reflects about 2% of the light. As the viewing angle increases here its about 40° that percentage increases. (Photo by E.H.)

When $\theta_i = 0$, the incident plane becomes undefined, and any distinction between the parallel and perpendicular components of R and T vanishes. In this case Eqs. (4.61) through (4.64), along with (4.47) and (4.48), lead to

$$R = R_{\parallel} = R_{\perp} = \left(\frac{n_t - n_i}{n_t + n_i}\right)^2 \tag{4.67}$$

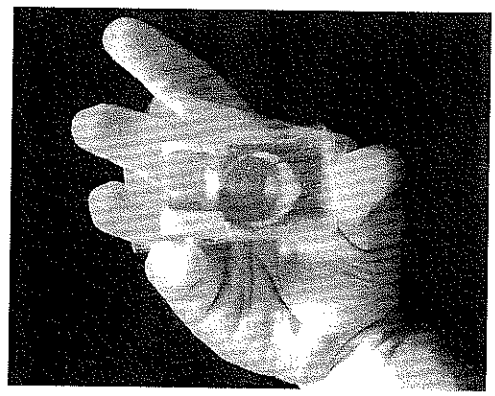
and

$$T = T_{\parallel} = T_{\perp} = \frac{4n_t n_i}{(n_t + n_i)^2} \tag{4.68}$$

Thus 4% of the light incident normally on an air–glass ($n_g = 1.5$) interface will be reflected back, whether internally, $n_i > n_t$, or externally, $n_i < n_t$ (Problem 4.72). This will be of concern to anyone who is working with a complicated lens system, which might have 10 or 20 such air–glass boundaries. Indeed, if you look perpendicularly into a stack of about 50 microscope slides (cover-glass sliders are much thinner and easier to handle in large quantities), most of the light will be reflected. The stack will look very much like a mirror (see photo). Roll up a thin sheet of clear plastic into a multiturned cylinder and it too will look like shiny metal. The many interfaces produce a large number of closely spaced *specular* reflections that send much of the light back into the incident medium, more or less, as if it had undergone a single frequency-independent reflection. A smooth gray-metal surface does pretty much the same thing —it has a large, frequency-independent specular reflectance—and looks shiny (that's what "shiny" is). If the reflection is diffuse, the surface will appear gray or even white if the reflectance is large enough.

Figure 4.49 is a plot of the reflectance at a single interface, assuming normal incidence for various transmitting media in air. Figure 4.50 depicts the corresponding dependence of the transmittance at normal incidence on the number of interfaces and the index of the medium. Of course, this is why you can't see through a roll of "clear" smooth-surfaced plastic tape, and it's also why the many elements in a periscope must be coated with antireflection films (Section 9.9.2).

At near normal incidence about 4% of the light is reflected back off each air-glass interface. Here because it's a lot brighter outside than inside the building, you have no trouble seeing the photographer. (Photo by E.H.)



Near normal reflection off a stack of microscope slides. You can see the image of the camera that took the picture. (Photo by E.H.)

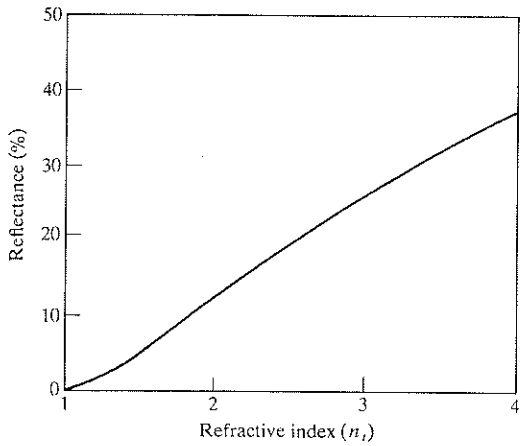


Figure 4.49 Reflectance at normal incidence in air ($n_i = 1.0$) at a single interface.

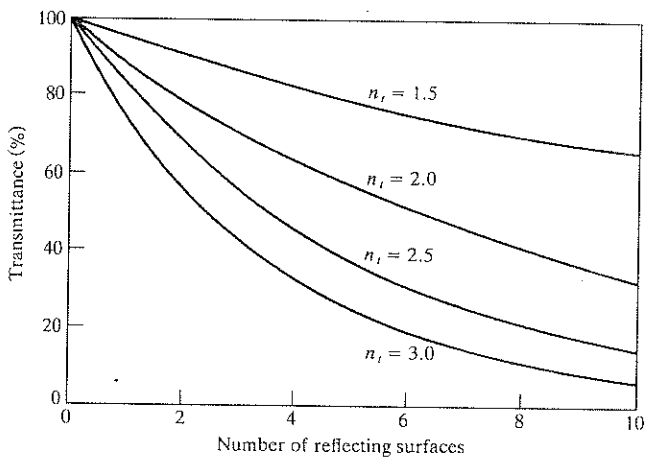


Figure 4.50 Transmittance through a number of surfaces in air ($n_i = 1.0$) at normal incidence.

4.7 Total Internal Reflection

In the previous section it was evident that something rather interesting was happening in the case of internal reflection ($n_i > n_t$) when θ_i was equal to or greater than θ_c, the so-called **critical angle**. Let's now return to that situation for a closer look. Suppose that we have a source embedded in an optically dense medium, and we allow θ_i to increase gradually, as indicated in Fig. 4.51. We know from the preceding section (Fig. 4.42) that $r_\parallel$ and $r_\perp$ increase with increasing θ_i, and therefore $t_\parallel$ and $t_\perp$ both decrease. Moreover $\theta_t > \theta_i$, since

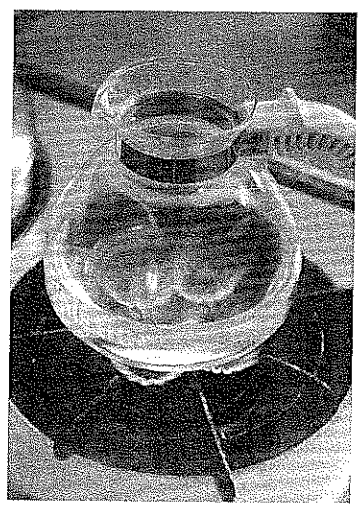
$$\sin \theta_i = \frac{n_t}{n_i} \sin \theta_t$$

and $n_i > n_t$, in which case $n_{ti} < 1$. Thus as θ_i becomes larger, the transmitted ray gradually approaches tangency with the boundary, and as it does more and more of the available energy appears in the reflected beam. Finally, when $\theta_t = 90°$, $\sin \theta_t = 1$ and

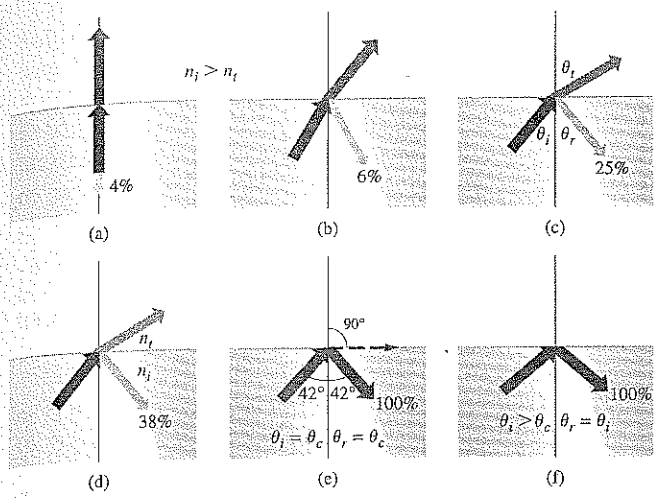
$$\sin \theta_c = n_{ti} \tag{4.69}$$

As noted earlier, *the critical angle is that special value of θ_i for which $\theta_t = 90°$*. The larger n_i is, the smaller n_{ti} is, and the smaller θ_c is. For incident angles greater than or equal to θ_c, all the incoming energy is reflected back into the incident medium in the process known as **total internal reflection** (see photo).

It should be stressed that the transition from the conditions depicted in Fig. 4.51*a* to those of 4.51*d* takes place without any discontinuities. As θ_i becomes larger, the reflected beam grows stronger and stronger while the transmitted beam grows weaker, until the latter vanishes and the former carries off all the energy at $\theta_r = \theta_c$. It's an easy matter to observe the diminution of the transmitted beam as θ_i is made larger. Just place a glass microscope slide on a printed page, this time blocking out any specularly reflected light. At $\theta_i \approx 0$, θ_t is roughly zero, and the page as seen through the glass is fairly bright and clear. But if you move your head, allowing θ_t (the angle at which you view the interface) to increase, the region



Notice that you can't see the two front flames through the water along a bright horizontal band. That's due to total internal reflection. Look at the bottom of a drinking galss through its side. Now add a few inches of water. What happens? (Photo by E.H.)



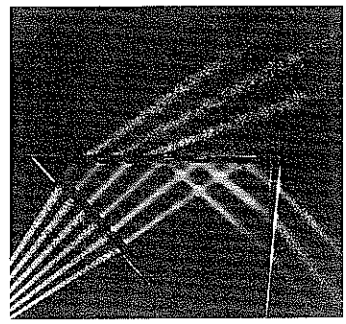


Figure 4.51 Internal reflection and the critical angle. (Photo courtesy of Educational Service, Inc.)

of the printed page covered by the glass will appear darker and darker, indicating that T has indeed been markedly reduced.

The critical angle for our air–glass interface is roughly 42° (see Table 4.2). Consequently, a ray incident normally on the left face of either of the prisms in Fig. 4.52 will have a $\theta_i > 42°$ and therefore be internally reflected. This is a convenient way to reflect nearly 100% of the incident light without having to worry about the deterioration that can occur with metallic surfaces (see photo).

Another useful way to view the situation is via Fig. 4.53, which shows a simplified representation of scattering off atomic oscillators. We know that the net effect of the presence of the homogeneous isotropic media is to alter the speed of the light from c to v_i and v_t, respectively (p. 92). The resultant wave is the superposition of these wavelets propagating at the appropriate speeds. In Fig. 4.53*a* an incident wave results in the emission of wavelets successively from scattering centers A and B. These overlap to form the transmitted wave. The reflected wave, which comes back down into the incident medium as usual ($\theta_i = \theta_r$), is not shown. In a time t the incident front travels a distance $v_i t = \overline{CB}$, while the transmitted front moves a distance $v_t t = \overline{AD} > \overline{CB}$. Since one wave moves from A to E in the same time that the other moves from

TABLE 4.2 Critical Angles

n_{it}	θ_c (degrees)	θ_c (radians)	n_{it}	θ_c (degrees)	θ_c (radians)
1.30	50.2849	0.8776	1.50	41.8103	0.7297
1.31	49.7612	0.8685	1.51	41.4718	0.7238
1.32	49.2509	0.8596	1.52	41.1395	0.7180
1.33	48.7535	0.8509	1.53	40.8132	0.7123
1.34	48.2682	0.8424	1.54	40.4927	0.7067
1.35	47.7946	0.8342	1.55	40.1778	0.7012
1.36	47.3321	0.8261	1.56	39.8683	0.6958
1.37	46.8803	0.8182	1.57	39.5642	0.6905
1.38	46.4387	0.8105	1.58	39.2652	0.6853
1.39	46.0070	0.8030	1.59	38.9713	0.6802
1.40	45.5847	0.7956	1.60	38.6822	0.6751
1.41	45.1715	0.7884	1.61	38.3978	0.6702
1.42	44.7670	0.7813	1.62	38.1181	0.6653
1.43	44.3709	0.7744	1.63	37.8428	0.6605
1.44	43.9830	0.7676	1.64	37.5719	0.6558
1.45	43.6028	0.7610	1.65	37.3052	0.6511
1.46	43.2302	0.7545	1.66	37.0427	0.6465
1.47	42.8649	0.7481	1.67	36.7842	0.6420
1.48	42.5066	0.7419	1.68	36.5296	0.6376
1.49	42.1552	0.7357	1.69	36.2789	0.6332

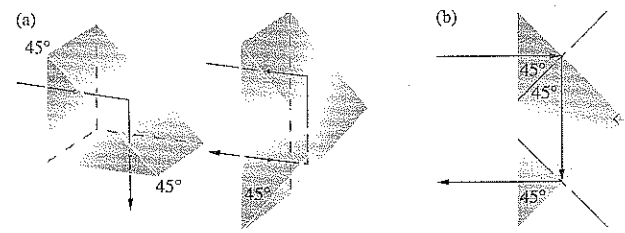


Figure 4.52 Total internal reflection.

C to B, and since they have the same frequency and period, they must change phase by the same amount in the process. Thus the disturbance at point E must be in-phase with that at point B; both of these points must be on the same transmitted wavefront (remember Section 4.4.2).

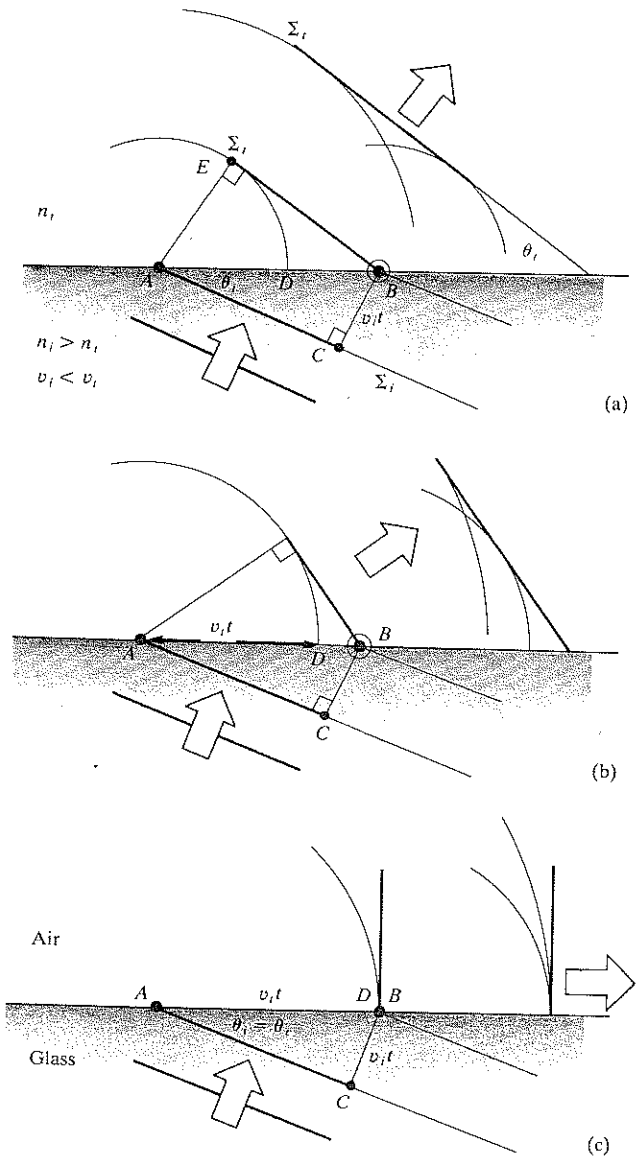
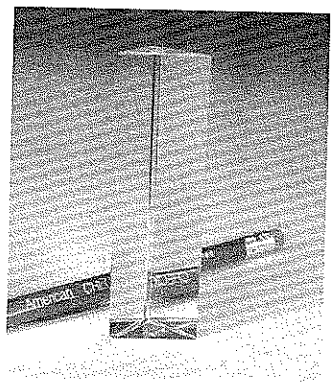


Figure 4.53 An examination of the transmitted wave in the process of total internal reflection from a scattering perspective. Here we keep θ_i and n_i constant and in successive parts of the diagram decrease n_t, thereby increasing v_t. The reflected wave ($\theta_r = \theta_i$) is not drawn.



The prism behaves like a mirror and reflects a portion of the pencil (reversing the lettering on it). The operating process is total internal reflection. (Photo by E.H.)

It can be seen that the greater v_t is in comparison to v_i, the more tilted the transmitted front will be (i.e., the larger θ_t will be). That much is depicted in Fig. 4.53*b*, where n_{ti} has been taken to be smaller by assuming n_t to be smaller. The result is a higher speed v_t, increasing $\overline{AD}$ and causing a greater transmission angle. In Fig. 4.53*c* a special case is reached: $\overline{AD} = \overline{AB} = v_i t$, and the wavelets will overlap in-phase *only along the line of the interface*, $\theta_t = 90°$. From triangle ABC, $\sin\theta_i = v_i t / v_t t = n_t / n_i$, which is Eq. (4.69). For the two given media (i.e., for the particular value of n_{ti}), the direction in which the scattered wavelets will add constructively in the transmitting medium is along the interface. The resulting disturbance ($\theta_t = 90°$) is known as a *surface wave*.

4.7.1 The Evanescent Wave

Because the frequency of X-rays is higher than the resonance frequencies of the atoms of the medium, Eq. (3.70) suggests, and experiments confirm, that the index of refraction of X-rays is less than 1.0. Thus the wave velocity of X-rays (i.e., the phase speed) in matter exceeds its value (c) in vacuum, although it usually does so by less than 1 part in 10 000, even in the densest solids. When X-rays traveling in air enter a dense material like glass, the beam bends ever so slightly *away* from the normal rather than toward it. With the above discussion of total internal reflection in mind, we should expect that X-rays will be **totally "externally" reflected** when, for example, $n_i = n_{air}$ and $n_t = n_{glass}$. This is the way it's often spoken of in the literature, but that's a misnomer; since for X-rays $n_{air} > n_{glass}$ and therefore $n_i > n_t$ (even though glass is physically more dense than air), the process is

actually still *internal* reflection. In any event, because n_t is less than, but very nearly equal to, 1 the index ratio $n_{ti} \approx 1$ and $\theta_c \approx 90°$.

In 1923 A. H. Compton reasoned that even though X-rays incident on a sample at ordinary angles are not specularly reflected, they should be totally "externally" reflected at glancing incidence. He shined 0.128 nm X-rays on a glass plate and got a critical angle of about 10 minutes of arc (0.167°) measured with respect to the surface. That yielded an index of refraction for glass that differed from 1 by -4.2×10^{-6}.

We'll come back to some important practical applications of both total internal and total "external" reflection later on (p. 200).

If we assume in the case of total internal reflection that there is no transmitted wave, it becomes impossible to satisfy the boundary conditions using only the incident and reflected waves—things are not at all as simple as they might seem. Furthermore, we can reformulate Eqs. (4.34) and (4.40) (Problem 4.75) such that

$$r_\perp = \frac{\cos\theta_i - (n_{ti}^2 - \sin^2\theta_i)^{1/2}}{\cos\theta_i + (n_{ti}^2 - \sin^2\theta_i)^{1/2}} \tag{4.70}$$

and

$$r_\parallel = \frac{n_{ti}^2\cos\theta_i - (n_{ti}^2 - \sin^2\theta_i)^{1/2}}{n_{ti}^2\cos\theta_i + (n_{ti}^2 - \sin^2\theta_i)^{1/2}} \tag{4.71}$$

Since $\sin\theta_c = n_{ti}$ when $\theta_i > \theta_c$, $\sin\theta_i > n_{ti}$, and both $r_\perp$ and $r_\parallel$ become complex quantities. Despite this (Problem 4.76), $r_\perp r_\perp^* = r_\parallel r_\parallel^* = 1$ and $R = 1$, which means that $I_r = I_i$ and $I_t = 0$. Thus, although there must be a transmitted wave, it cannot, on the average, carry energy across the boundary. We shall not perform the complete and rather lengthy computation needed to derive expressions for all the reflected and transmitted fields, but we can get an appreciation of what's happening in the following way. The wavefunction for the transmitted electric field is

$$\vec{\mathbf{E}}_t = \vec{\mathbf{E}}_{0t}\exp i(\vec{\mathbf{k}}_t\cdot\vec{\mathbf{r}} - \omega t)$$

where

$$\vec{\mathbf{k}}_t\cdot\vec{\mathbf{r}} = k_{tx}x + k_{ty}y$$

there being no z-component of $\vec{\mathbf{k}}$. But

$$k_{tx} = k_t\sin\theta_t$$

and

$$k_{ty} = k_t\cos\theta_t$$

as seen in Fig. 4.54. Once again using Snell's Law,

$$k_t\cos\theta_t = \pm k_t\left(1 - \frac{\sin^2\theta_i}{n_{ti}^2}\right)^{1/2} \tag{4.72}$$

or, since we are concerned with the case where $\sin\theta_i > n_{ti}$,

$$k_{ty} = \pm ik_t\left(\frac{\sin^2\theta_i}{n_{ti}^2} - 1\right)^{1/2} \equiv \pm i\beta$$

and

$$k_{tx} = \frac{k_t}{n_{ti}}\sin\theta_i$$

Hence

$$\vec{\mathbf{E}}_t = \vec{\mathbf{E}}_{0t}e^{\mp\beta y}e^{i(k_t x\sin\theta_i/n_{ti} - \omega t)} \tag{4.73}$$

Neglecting the positive exponential, which is physically untenable, we have a wave whose amplitude drops off exponentially as it penetrates the less dense medium. The disturbance advances in the x-direction as a surface or **evanescent wave**. Notice that the wavefronts or surfaces of constant phase (parallel to the yz-plane) are perpendicular to the surfaces of constant amplitude (parallel to the xz-plane), and as such the wave is *inhomogeneous* (p. 26). Its amplitude decays rapidly in the y-direction, becoming negligible at a distance into the second medium of only a few wavelengths.

If you are still concerned about the conservation of energy, a more extensive treatment would have shown that energy actually circulates back and forth across the interface, resulting on the average in a zero net flow through the boundary into the second medium. Yet one puzzling point remains, inas-

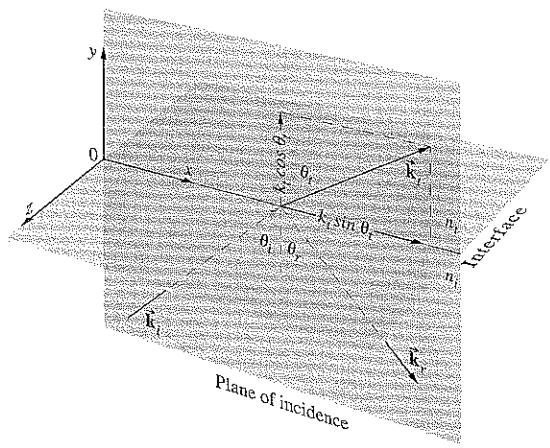


Figure 4.54 Propagation vectors for internal reflection.

much as there is still a bit of energy to be accounted for, namely, that associated with the evanescent wave that moves along the boundary in the plane-of-incidence. Since this energy could not have penetrated into the less dense medium under the present circumstances (so long as $\theta_i \geq \theta_c$), we must look elsewhere for its source. Under actual experimental conditions the incident beam would have a finite cross section and therefore would obviously differ from a true plane wave. This deviation gives rise (via diffraction) to a slight transmission of energy across the interface, which is manifested in the evanescent wave.

Incidentally, it is clear from (*c*) and (*d*) in Fig. 4.44 that the incident and reflected waves (except at $\theta_i = 90°$) do not differ in phase by π and cannot therefore cancel each other. It follows from the continuity of the tangential component of $\vec{\mathbf{E}}$ that there must be an oscillatory field in the less dense medium, with a component parallel to the interface having a frequency ω (i.e., the evanescent wave).

The exponential decay of the surface wave, or *boundary wave*, as it is also sometimes called, has been confirmed experimentally at optical frequencies.*

Imagine that a beam of light traveling within a block of glass is internally reflected at a boundary. Presumably, if you pressed another piece of glass against the first, the air–glass interface could be made to vanish, and the beam would then propagate onward undisturbed. Furthermore, you might expect this transition from total to no reflection to occur gradually as the air film thinned out. In much the same way, if you hold a drinking glass or a prism, you can see the ridges of your fingerprints in a region that, because of total internal reflection, is otherwise mirrorlike. In more general terms, when the evanescent wave extends with appreciable amplitude across the rare medium into a nearby region occupied by a higher-index material, energy may flow through the gap in what is known as **frustrated total internal reflection** (FTIR). The evanescent wave, having traversed the gap, is still strong enough to drive electrons in the "frustrating" medium; they in turn will generate a wave that significantly alters the field configuration, thereby permitting energy to flow. Figure 4.55 is a schematic representation of FTIR: the width of the lines depicting the wavefronts decreases across the gap as a reminder that the amplitude of the field behaves in the same way. The process as a whole is remarkably similar to the quantum-mechanical phenomenon of *barrier penetration* or *tunneling*, which has numerous applications in contemporary physics.

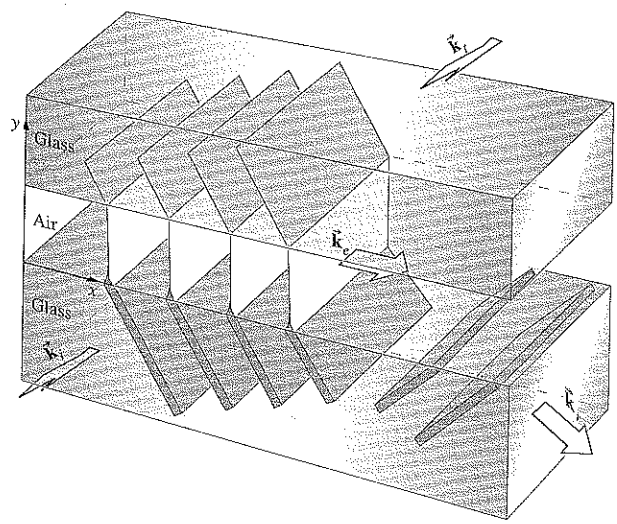
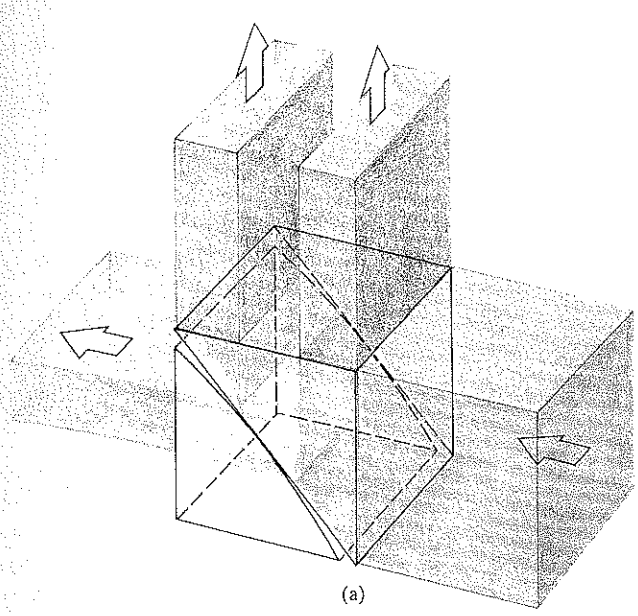


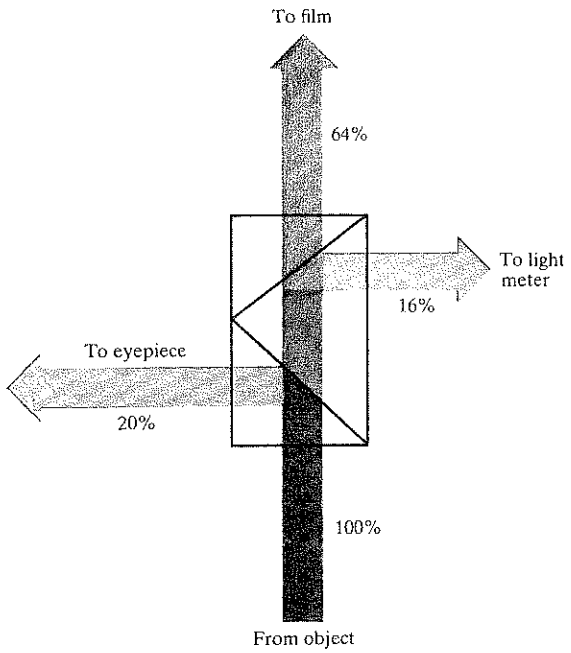
Figure 4.55 Frustrated total internal reflection.

One can demonstrate FTIR with the prism arrangement of Fig. 4.56 in a manner that is fairly self-evident. Moreover, if the hypotenuse faces of both prisms are made planar and parallel, they can be positioned so as to transmit and reflect any desired fraction of the incident flux density. Devices that perform this function are known as ***beamsplitters***. A *beamsplitter cube* can be made rather conveniently by using a thin, low-index transparent film as a precision spacer. Low-loss reflectors whose transmittance can be controlled by frustrating internal reflection are of considerable practical interest. FTIR can also be observed in other regions of the electromagnetic spectrum. Three-centimeter microwaves are particularly easy to work with, inasmuch as the evanescent wave will extend roughly 10^5 times farther than it would at optical frequencies. One can duplicate the above optical experiments with solid prisms made of paraffin or hollow ones of acrylic plastic filled with kerosene or motor oil. Any one of these would have an index of about 1.5 for 3-cm waves. It then becomes an easy matter to measure the dependence of the field amplitude on *y*.

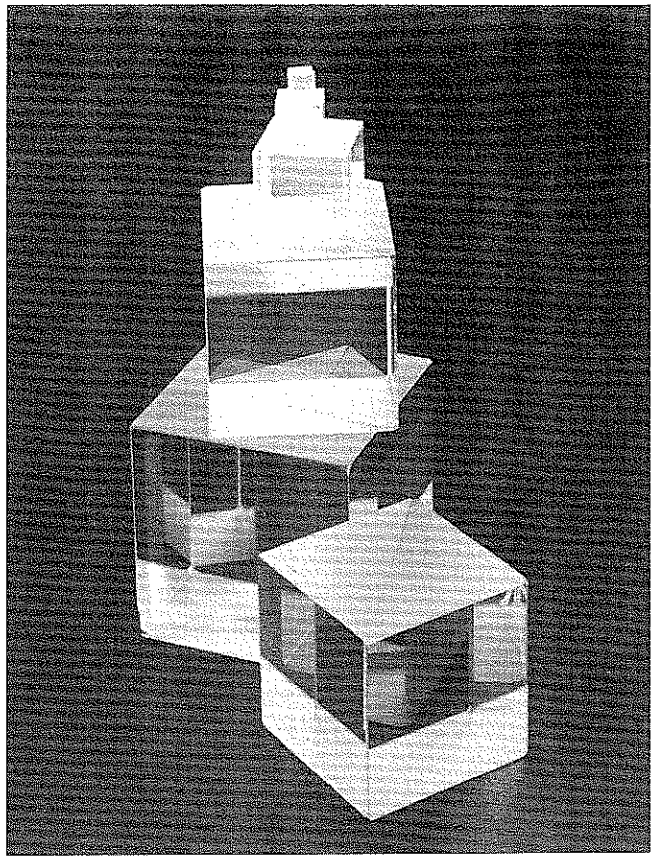
*Take a look at the fascinating article by K. H. Drexhage, "Monomolecular layers and light." *Sci. Am.* **222**, 108 (1970).



(a)



(b)



(c)

Figure 4.56 (a) A beamsplitter utilizing FTIR. (b) A typical modern application of FTIR: a conventional beamsplitter arrangement used to take photographs through a microscope. (c) Beamsplitter cubes. (Photo courtesy Melles Griot.)

4.8 Optical Properties of Metals

The characteristic feature of conducting media is the presence of a number of free electric charges (free in the sense of being unbound, i.e., able to circulate within the material). For metals these charges are of course electrons, and their motion constitutes a current. The current per unit area resulting from the application of a field $\vec{\mathbf{E}}$ is related by means of Eq. (A1.15) to the conductivity of the medium σ. For a dielectric there are no

free or conduction electrons and $\sigma = 0$, whereas for metals σ is nonzero and finite. In contrast, an idealized "perfect" conductor would have an infinite conductivity. This is equivalent to saying that the electrons, driven into oscillation by a harmonic wave, would simply follow the field's alternations. There would be no restoring force, no natural frequencies, and no absorption, only re-emission. In real metals the conduction electrons undergo collisions with the thermally agitated lattice or with imperfections and in so doing irreversibly convert electromagnetic energy into joule heat. The absorption of radiant energy by a material is a function of its conductivity.

Waves in a Metal

If we visualize the medium as continuous, Maxwell's Equations lead to

$$\frac{\partial^2 \vec{\mathbf{E}}}{\partial x^2} + \frac{\partial^2 \vec{\mathbf{E}}}{\partial y^2} + \frac{\partial^2 \vec{\mathbf{E}}}{\partial z^2} = \mu\epsilon \frac{\partial^2 \vec{\mathbf{E}}}{\partial t^2} + \mu\sigma \frac{\partial \vec{\mathbf{E}}}{\partial t} \qquad (4.74)$$

which is Eq. (A1.21) in Cartesian coordinates. The last term, $\mu\sigma\, \partial\vec{\mathbf{E}}/\partial t$, is a first-order time derivative, like the damping force in the oscillator model (p. 71). The time rate-of-change of $\vec{\mathbf{E}}$ generates a voltage, currents circulate, and since the material is resistive, light is converted to thermal energy—ergo absorption. This expression can be reduced to the unattenuated wave equation, if the permittivity is reformulated as a complex quantity. This in turn leads to a complex index of refraction, which, as we saw earlier (p. 71), is tantamount to absorption. We then need only substitute the complex index

$$\tilde{n} = n_R - in_I \qquad (4.75)$$

(where the real and imaginary indices n_R and n_I are both real numbers) into the corresponding solution for a nonconducting medium. Alternatively, we can utilize the wave equation and appropriate boundary conditions to yield a specific solution. In either event, it is possible to find a simple sinusoidal plane-wave solution applicable within the conductor. Such a wave propagating in the y-direction is ordinarily written as

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 \cos(\omega t - ky)$$

or as a function of n,

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 \cos \omega(t - \tilde{n}y/c)$$

but here the refractive index must be taken as complex. Writing the wave as an exponential and using Eq. (4.75) yields

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 e^{(-\omega n_I y/c)} e^{i\omega(t - n_R y/c)} \qquad (4.76)$$

or

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 e^{-\omega n_I y/c} \cos \omega(t - n_R y/c) \qquad (4.77)$$

The disturbance advances in the y-direction with a speed c/n_R, precisely as if n_R were the more usual index of refraction. As the wave progresses into the conductor, its amplitude, $\vec{\mathbf{E}}_0 \exp(-\omega n_I y/c)$, is exponentially attenuated. Inasmuch as irradiance is proportional to the square of the amplitude, we have

$$I(y) = I_0 e^{-\alpha y} \qquad (4.78)$$

where $I_0 = I(0)$; that is, I_0 is the irradiance at $y = 0$ (the interface), and $\alpha \equiv 2\omega n_I/c$ is called the *absorption coefficient* or (even better) the **attenuation coefficient**. The flux density will drop by a factor of $e^{-1} = 1/2.7 \approx \frac{1}{3}$ after the wave has propagated a distance $y = 1/\alpha$, known as the **skin** or **penetration depth**. For a material to be transparent, the penetration depth must be large in comparison to its thickness. The penetration depth for metals, however, is exceedingly small. For example, copper at ultraviolet wavelengths ($\lambda_0 \approx 100$ nm) has a miniscule penetration depth, about 0.6 nm, while it is still only about 6 nm in the infrared ($\lambda_0 \approx 10\,000$ nm). This accounts for the generally observed opacity of metals, which nonetheless can become partly transparent when formed into extremely thin films (e.g., in the case of partially silvered two-way mirrors). The familiar metallic sheen of conductors corresponds to a high reflectance, which exists because the incident wave cannot effectively penetrate the material. Relatively few electrons in the metal "see" the transmitted wave, and therefore, although each absorbs strongly, little total energy is dissipated by them. Instead, most of the incoming energy reappears as the reflected wave. The majority of metals, including the less common ones (e.g., sodium, potassium, cesium, vanadium, niobium, gadolinium, holmium, yttrium, scandium, and osmium) have a silvery gray appearance like that of aluminum, tin, or steel. They reflect almost all the incident light (roughly 85 to 95%) regardless of wavelengths and are therefore essentially colorless.

Equation (4.77) is certainly reminiscent of Eq. (4.73) and FTIR. In both cases there is an exponential decay of the amplitude. Moreover, a complete analysis would show that the transmitted waves are not strictly transverse, there being a component of the field in the direction of propagation in both instances.

The representation of metal as a continuous medium works fairly well in the low-frequency, long-wavelength domain of the infrared. Yet we certainly might expect that as the wavelength of the incident beam decreased the actual granular nature of matter would have to be reckoned with. Indeed, the continuum model shows large discrepancies from experimental results at optical frequencies. And so we again turn to the classical atomistic picture initially formulated by Hendrik Lorentz, Paul Karl Ludwig Drude (1863–1906), and others. This simple approach will provide qualitative agreement with the experimental data, but the ultimate treatment requires quantum theory.

The Dispersion Equation

Envision the conductor as an assemblage of driven, damped oscillators. Some correspond to free electrons and will therefore have zero restoring force, whereas others are bound to the atom, much like those in the dielectric media of Section 3.5.1. The conduction electrons are, however, the predominant contributors to the optical properties of metals. Recall that the displacement of a vibrating electron was given by

$$x(t) = \frac{q_e/m_e}{(\omega_0^2 - \omega^2)} E(t) \qquad [3.66]$$

With no restoring force, $\omega_0 = 0$, the displacement is opposite in sign to the driving force $q_e E(t)$ and therefore 180° out-of-phase with it. This is unlike the situation for transparent dielectrics, where the resonance frequencies are above the visible and the electrons oscillate in-phase with the driving force (Fig. 4.57). Free electrons oscillating out-of-phase with the incident light will reradiate wavelets that tend to cancel the incoming disturbance. The effect, as we have already seen, is a rapidly decaying refracted wave.

Assuming that the average field experienced by an electron moving about within a conductor is just the applied field $\vec{\mathbf{E}}(t)$, we can extend the dispersion equation of a rare medium [Eq. (3.72)] to read

$$n^2(\omega) = 1 + \frac{Nq^2{}_e}{\epsilon_0 m_e}\left[\frac{f_e}{-\omega^2 + i\gamma_e\omega} + \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 + i\gamma_j\omega}\right] \qquad (4.79)$$

The first bracketed term is the contribution from the free electrons, wherein N is the number of atoms per unit volume. Each of these has f_e conduction electrons, which have no natural frequencies. The second term arises from the bound electrons and is identical to Eq. (3.72). It should be noted that if a metal has a particular color, it indicates that the atoms are partaking of selective absorption by way of the bound electrons, in addition to the general absorption characteristic of the free electrons. Recall that a medium that is very strongly absorbing at a given frequency doesn't actually absorb much of the incident light at that frequency but rather *selectively reflects* it. Gold and copper are reddish yellow because n_l increases with wavelength, and the larger values of λ are reflected more strongly. Thus, for example, gold should be fairly opaque to the longer visible wavelengths. Consequently, under white light, a gold foil less than roughly 10^{-6} m thick will indeed transmit predominantly greenish blue light.

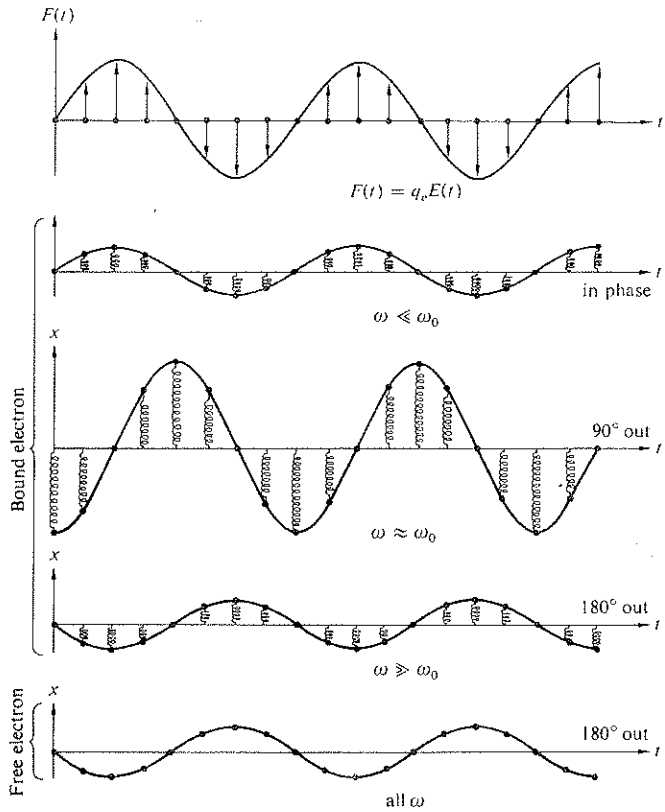


Figure 4.57 Oscillations of bound and free electrons.

We can get a rough idea of the response of metals to light by making a few simplifying assumptions. Accordingly, neglect the bound electron contribution and assume that γ_e is

also negligible for very large ω, whereupon

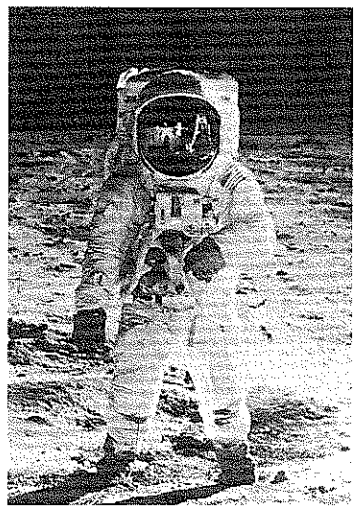
$$n^2(\omega) = 1 - \frac{Nq^2{}_e}{\epsilon_0 m_e \omega^2} \tag{4.80}$$

The latter assumption is based on the fact that at high frequencies the electrons will undergo a great many oscillations between each collision. Free electrons and positive ions within a metal may be thought of as a plasma whose density oscillates at a natural frequency ω_p, the **plasma frequency**. This in turn can be shown to equal $(Nq_e^2/\epsilon_0 m_e)^{1/2}$, and so

$$n^2(\omega) = 1 - (\omega_p/\omega)^2 \tag{4.81}$$

The plasma frequency serves as a critical value below which the index is complex and the penetrating wave drops off exponentially [Eq. (4.77)] from the boundary; at frequencies above ω_p, n is real, absorption is small, and the conductor is transparent. In the latter circumstance n is less than 1, as it was for dielectrics at very high frequencies (v can be greater than c—see p. 72). Hence we can expect metals in general to be fairly transparent to X-rays. Table 4.3 lists the plasma frequencies for some of the alkali metals that are transparent even to ultraviolet.

The index of refraction for a metal will usually be complex, and the impinging wave will suffer absorption in an amount that is frequency dependent. For example, the outer visors on the Apollo space suits were overlaid with a very thin film of gold (see photo). The coating reflected about 70% of the incident light and was used under bright conditions, such as low and forward Sun angles. It was designed to decrease the thermal load on the cooling system by strongly reflecting radiant energy in the infrared while still transmitting adequately in the visible. Inexpensive metal-coated sunglasses which are quite similar in principle are also available commercially, and they're well worth having just to experiment with.



Edwin Aldrin Jr. at Tranquility Base on the Moon. The photographer, Neil Armstrong, is reflected in the gold-coated visor. (Photo courtesy NASA.)

The ionized upper atmosphere of the Earth contains a distribution of free electrons that behave very much like those confined within a metal. The index of refraction of such a medium will be real and less than 1 for frequencies above ω_p. In July of 1965 the *Mariner IV* spacecraft made use of this effect to examine the ionosphere of the planet Mars, 216 million kilometers from Earth.*

If we wish to communicate between two distant terrestrial points, we might bounce low-frequency waves off the Earth's ionosphere. To speak to someone on the Moon, however, we should use high-frequency signals, to which the ionosphere would be transparent.

TABLE 4.3 Critical Wavelengths and Frequencies for Some Alkali Metals

Metal	λ_p (observed) nm	λ_p (calculated) nm	$\nu_p = c/\lambda_p$ (observed) Hz
Lithium (Li)	155	155	1.94×10^{15}
Sodium (Na)	210	209	1.43×10^{15}
Potassium (K)	315	287	0.95×10^{15}
Rubidium (Rb)	340	322	0.88×10^{15}

Reflection from a Metal

Imagine that a plane wave initially in air impinges on a conducting surface. The transmitted wave advancing at some angle to the normal will be inhomogeneous. But if the conductivity of the medium is increased, the wavefronts will become aligned with the surfaces of constant amplitude, whereupon $\vec{\mathbf{k}}_t$ and $\hat{\mathbf{u}}_n$ will approach parallelism. In other

*R. Von Eshelman, *Sci. Am.* **220**, 78 (1969).

words, in a good conductor the transmitted wave propagates in a direction normal to the interface regardless of θ_i.

Let's now compute the reflectance, $R = I_r/I_i$, for the simplest case of normal incidence on a metal. Taking $n_i = 1$ and $n_t = \tilde{n}$ (i.e., the complex index), we have from Eq. (4.47) that

$$R = \left(\frac{\tilde{n} - 1}{\tilde{n} + 1}\right)\left(\frac{\tilde{n} - 1}{\tilde{n} + 1}\right)^* \tag{4.82}$$

and therefore, since $\tilde{n} = n_R - in_I$,

$$R = \frac{(n_R - 1)^2 + n_I^2}{(n_R + 1)^2 + n_I^2} \tag{4.83}$$

If the conductivity of the material goes to zero, we have the case of a dielectric, whereupon in principle the index is real ($n_I = 0$), and the attenuation coefficient, α, is zero. Under those circumstances, the index of the transmitting medium n_t is n_R, and the reflectance [Eq. (4.83)] becomes identical with that of Eq. (4.67). If instead n_I is large while n_R is comparatively small, R in turn becomes large (Problem 4.81). In the unattainable limit where $\tilde{n}$ is purely imaginary, 100% of the incident flux density would be reflected ($R = 1$). Notice that it is possible for the reflectance of one metal to be greater than that of another even though its n_I is smaller. For example, at $\lambda_0 = 589.3$ nm the parameters associated with solid sodium are roughly $n_R = 0.04$, $n_I = 2.4$, and $R = 0.9$; and those for bulk tin are $n_R = 1.5$, $n_I = 5.3$, and $R = 0.8$; whereas for a gallium single crystal $n_R = 3.7$, $n_I = 5.4$, and $R = 0.7$.

The curves of $R_\parallel$ and $R_\perp$ for oblique incidence shown in Fig. 4.58 are somewhat typical of absorbing media. Thus, although R at $\theta_i = 0$ is about 0.5 for gold, as opposed to nearly 0.9 for silver in white light, the two metals have reflectances

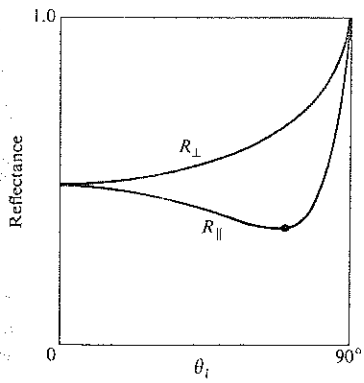


Figure 4.58 Typical reflectance for a linearly polarized beam of white light incident on an absorbing medium.

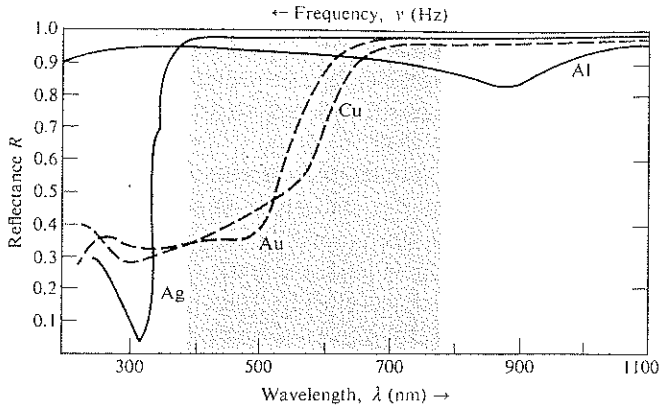


Figure 4.59 Reflectance versus wavelength for silver, gold, copper, and aluminum.

that are quite similar in shape, approaching 1.0 at $\theta_i = 90°$. Just as with dielectrics (Fig. 4.48), $R_\parallel$ drops to a minimum at what is now called the *principal angle-of-incidence*, but here that minimum is nonzero. Figure 4.59 illustrates the spectral reflectance at normal incidence for a number of evaporated metal films under ideal conditions. Observe that although gold transmits fairly well in and below the green region of the spectrum, silver, which is highly reflective across the visible, becomes transparent in the ultraviolet at about 316 nm.

Phase shifts arising from reflection off a metal occur in both components of the field (i.e., parallel and perpendicular to the plane-of-incidence). These are generally neither 0 nor π, with a notable exception at $\theta_i = 90°$, where, just as with a dielectric, both components shift phase by 180° on reflection.

4.9 Familiar Aspects of the Interaction of Light and Matter

Let's now examine some of the phenomena that paint the everyday world in a marvel of myriad colors.

As we saw earlier (p. 77), light that contains a roughly equal amount of every frequency in the visible region of the spectrum is perceived as white. A broad source of white light (whether natural or artificial) is one for which every point on its surface can be imagined as sending out a stream of light of every visible frequency. Given that we evolved on this planet, it's not surprising that a source appears white when its emis-