axis, the index of refraction is called the *extraordinary* index. The permittivity tensor for calcite is

$$\epsilon = \epsilon_0 \begin{pmatrix} 2.75 & 0 & 0 \\ 0 & 2.75 & 0 \\ 0 & 0 & 2.21 \end{pmatrix} \tag{1.11}$$

Calcite is called a negative uniaxial crystal because the extraordinary index is less than the ordinary index. We will deal with crystal optics in later chapters, especially when dealing with electrooptic modulators. For a good review of crystal optics, see Chapter 4 of Yariv and Yeh [7].

In an isotropic medium, the permittivity is independent of orientation and is described accurately by the scalar relation $D = \epsilon E$. But beware! *Isotropic* does not necessarily mean homogeneous. The permittivity can be a function of position, $\epsilon(r)$. In an inhomogeneous medium, the electric field will encounter a different permittivity, ϵ, depending upon spatial location in the material. A graded-index waveguide, discussed in Chapters 5 and 7, is a good example of an inhomogeneous medium.

For most optical dielectric materials, μ is effectively μ_0. We can ignore magnetic effects except when dealing with special magnetic optical materials, such as yttrium iron garnet (YIG), used as an optical isolator between waveguides and sources. Unless otherwise stated, it is safe to assume that the permeability, μ, is that of free space, μ_0. We will discuss the frequency dependence of μ and ϵ in Chapter 4.

1.4 THE WAVE EQUATION

The electromagnetic wave equation comes directly from Maxwell's equations. Derivation is straightforward if we assume conditions that are reasonable for optical wave propagation. These conditions are that we are operating in a *source free* ($\rho = 0$, $\mathbf{J} = 0$), *linear* (ϵ and μ are independent of $\mathbf{E}$ and $\mathbf{H}$), and *isotropic* medium. Equations 1.1–1.4 become

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.12}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \tag{1.13}$$

$$\nabla \cdot \mathbf{D} = 0 \tag{1.14}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.15}$$

These simple-looking equations completely describe the electromagnetic field in time and position. Are the assumptions reasonable? Sure. At high frequencies (e.g.,

$\nu > 10^{13}$ Hz), free charge and current are generally not the source of electromagnetic energy. The typical sources of optical energy are electric or magnetic dipoles formed by atoms and molecules undergoing transitions. Maxwell's equations account for these sources through the bulk permeability and permittivity constants.

Equations 1.12–1.15 are strongly coupled first-order differential equations. To decouple the two curl equations, we follow the usual technique of creating a single second-order differential equation. First take the curl of both sides of Equation 1.12

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \frac{-\partial \mathbf{B}}{\partial t} = \nabla \times \frac{-\partial \mu \mathbf{H}}{\partial t} \tag{1.16}$$

Assuming that $\mu(r, t)$ is independent of time and position, Equation 1.16 becomes

$$\nabla \times \nabla \times \mathbf{E} = -\mu\left(\nabla \times \frac{\partial \mathbf{H}}{\partial t}\right) \tag{1.17}$$

Since the functions are continuous, the order of the curl and time derivative operators can be reversed:

$$\nabla \times \nabla \times \mathbf{E} = -\mu\frac{\partial}{\partial t}(\nabla \times \mathbf{H}) \tag{1.18}$$

Substituting $\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t$ into Equation 1.17 and assuming ϵ is time invariant

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E} &= -\mu\frac{\partial}{\partial t}\left(\frac{\partial \mathbf{D}}{\partial t}\right) \\ &= -\mu\epsilon\frac{\partial^2 \mathbf{E}}{\partial t^2} \end{aligned} \tag{1.19}$$

Now we have a second-order differential equation with only one variable, **E**. The $(\nabla \times \nabla \times)$ operator is usually simplified using a vector identity

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \tag{1.20}$$

The ∇^2 operator should not be confused with the *scalar Laplacian* operator. The ∇^2 operator in Equation 1.18 is the *vector Laplacian* operator that acts on a vector, in this case **E**. For a rectangular coordinate system, the vector Laplacian can be written in terms of the scalar Laplacian as

$$\nabla^2 \mathbf{E} = \nabla^2 E_x \hat{x} + \nabla^2 E_y \hat{y} + \nabla^2 E_z \hat{z} \tag{1.21}$$

where $\hat{x}$, $\hat{y}$, and $\hat{z}$ represent unit vectors along the three axes. The ∇^2's on the right-hand side of Equation 1.21 are scalar, given by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.22}$$

in cartesian coordinates. Solution of the vector wave equation requires that we first break the equation into the orthogonal vector components, which is sometimes difficult but always possible, and then combine the individual vector field solutions together.

What about the term, $\nabla \cdot \mathbf{E}$? It is not necessarily equal to zero, as is often assumed. We know only that $\nabla \cdot \mathbf{D} = 0$. Simple calculus leads to an expression for $\nabla \cdot \mathbf{E}$:

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 0 \\ &= \nabla \cdot \epsilon\mathbf{E} \\ &= \nabla\epsilon \cdot \mathbf{E} + \epsilon\nabla \cdot \mathbf{E} \end{aligned} \tag{1.23}$$

Solve for $\nabla \cdot \mathbf{E}$:

$$\nabla \cdot \mathbf{E} = -\mathbf{E} \cdot \frac{\nabla\epsilon}{\epsilon} \tag{1.24}$$

Plugging this value into the linear wave equation for electromagnetic waves yields:

$$\nabla^2\mathbf{E} - \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} = -\nabla\left(\mathbf{E} \cdot \frac{\nabla\epsilon}{\epsilon}\right) \tag{1.25}$$

The right-hand side deserves special consideration. It is non-zero when there is a gradient in the permittivity of the medium. Such gradients are not uncommon in guided-wave optics. With the exception of step-index waveguides, most guided-wave structures use a graded permittivity. So how do we deal with this extra term? Well, we ignore it! For most structures, the term is negligibly small. (Problem 1.2 explores the limits of $\nabla\epsilon/\epsilon$, showing that it is almost always negligible.) Neglecting this term, the wave equation reduces to its *homogeneous* form:

$$\nabla^2\mathbf{E} - \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} = 0 \tag{1.26}$$

Had we started with Equation 1.13 instead of Equation 1.12, we could have derived a similar wave equation in terms of the magnetic field amplitude (see Problem 1.10),

$$\nabla^2\mathbf{H} - \mu\epsilon\frac{\partial^2\mathbf{H}}{\partial t^2} = 0 \tag{1.27}$$

1.5 SOLUTIONS TO THE WAVE EQUATION

Consider the units of each term in Equations 1.26 and 1.27. The ∇^2 term has units of 1/(distance)^2. The second-order time derivative clearly has units of 1/(sec)^2. In order to make physical sense, the units of $\mu\epsilon$ must be (sec/m)^2. We will show, in a later section, that $\sqrt{1/\epsilon\mu}$ is the phase velocity of light in a medium. Notice that the speed of propagation is determined by the material parameters. In free space, $\sqrt{1/\mu_0\epsilon_0} = 2.998 \times 10^8$m/sec, or c, the speed of light in vacuum. (The speed of light is now defined to be exactly 299,792,458 m/sec. The meter is defined in terms of the speed of light, being the distance light travels in 1/299,792,458 sec. This definition reflects the effort to define all fundamental constants in terms of the second.) We will discuss the speed of propagation more thoroughly in the next section.

Equations 1.26 and 1.27 are vector equations. They can be simplified by rewriting them in terms of the components of the field. In rectangular coordinates, the vector Laplacian breaks into three uncoupled components. The scalar component equations become:

$$\nabla^2 E_i - \mu\epsilon \frac{\partial^2 E_i}{\partial t^2} = 0 \tag{1.28}$$

Here the subscript indicates the i^{th} component, where i stands for x, y, or z, and ∇^2 is the scalar Laplacian given in Equation 1.22. Since the symbol for the vector and scalar Laplacian look the same, we rely on context to distinguish the operators.

The choice of coordinate system is critical to solving the wave equation. For example, choosing rectangular coordinates to describe a wave in a cylinder leads to component coupling upon reflection at the cylindrical surface. The cartesian components are inseparable in such a system. When a coordinate system can be found with no coupling between the orthogonal components, the individual equations can be solved independently. In such a case, we refer to the individual equations as the scalar wave equations. In cartesian coordinates, the scalar wave equation is written as:

$$\nabla^2 \psi - \mu\epsilon \frac{\partial^2 \psi}{\partial t^2} = 0 \tag{1.29}$$

where ψ stands for any one of the orthogonal amplitude components. To find a valid solution, we use the separation of variables technique to get:

$$\begin{aligned} \psi(\mathbf{r}, t) &= \psi(\mathbf{r})\phi(t) \\ &= \psi_0 \exp(j\mathbf{k} \cdot \mathbf{r}) \exp(j\omega t) + c.c. \end{aligned} \tag{1.30}$$

The term ψ_0 is the amplitude; the separation constant $\mathbf{k}$ is called the *wavevector* (in units of rads/meter); and ω is the *angular frequency* of the wave (in units of rads/sec). We will use the wavevector as the primary variable in most waveguide calculations. The magnitude of the wavevector is defined in terms of the angular frequency and the phase velocity:

$$|\mathbf{k}| = \omega\sqrt{\mu\epsilon} = k \tag{1.31}$$

The wavevector $\mathbf{k}$ points in the direction of travel for the plane wave. The magnitude of $|\mathbf{k}|$ describes how much phase accumulates as a plane wave travels a unit distance. Think of k as a *spatial frequency*.

Through proper choice of sign for each term, one can describe a wave that travels in the forward or backward direction along the axis of propagation. Figure 1.3 shows the two cases.

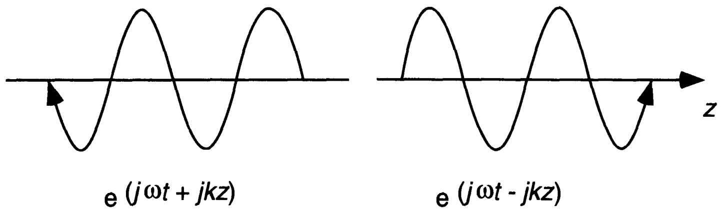

Figure 1.3 The general solution to the wave equation in a linear homogeneous medium leads to plane waves. Depending on the relative sign, the wave will travel left or right.

In optics, it is common to describe optical fields by their *wavelength.* Consider the wave in Figure 1.4.

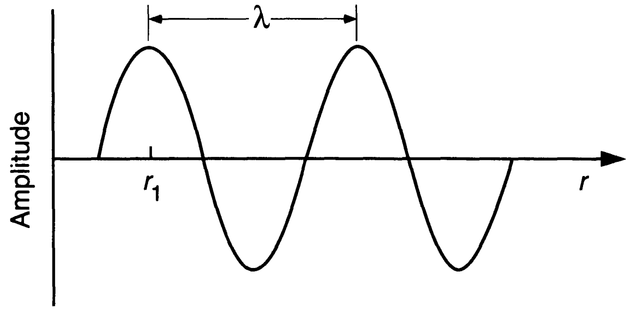

Figure 1.4 The basic description of the wavelength is that the wave accumulates 2π of phase after traveling one wavelength.

The waveform in Figure 1.4 shows the real part of the spatial component of the plane wave, $\psi(r) = \psi_0 e^{jkr}$. The distance between two adjacent peaks in amplitude is called a *wavelength,* λ. The amplitude of the wave at the first peak, e^{jkr_1}, is the same as the amplitude at the peak located one wavelength away, $e^{jk(r_1+\lambda)}$.

$$\begin{aligned} e^{jkr_1} &= e^{jk(r_1+\lambda)} \\ &= e^{jkr_1}e^{jk\lambda} \end{aligned} \tag{1.32}$$

This equality holds only if $e^{jk\lambda} = 1$, which requires that $k\lambda = 2\pi$. Solving for k:

$$k = 2\pi/\lambda \tag{1.33}$$

This is the expression for wavevector k in terms of wavelength, λ.

Example 1.2 Magnitude of the Wavevector for Visible Light

Consider a plane wave of light with $\lambda = 1\,\mu\text{m}$. The light is directed in the x–y plane at 45°. Describe the magnitude and direction of the k vector.

Solution:

The magnitude of k is:

$$\begin{aligned} k &= \frac{2\pi}{\lambda} \\ &= \frac{2\pi}{10^{-4}\ \text{cm}} \\ &= 62{,}831 \quad \text{rads/cm} \end{aligned} \tag{1.34}$$

Note the units: *radians per centimeter.* The direction is easily described with trigonometry. The k vector can be broken down into its components in the chosen coordinate system,

$$\begin{aligned} \mathbf{k} &= k(\hat{x}\cos 45^\circ + \hat{y}\sin 45^\circ) \\ &= k\,0.707(\hat{x} + \hat{y}) \\ &= 44{,}421(\hat{x} + \hat{y}) \quad \text{rads/cm} \end{aligned} \tag{1.35}$$

1.6 TRANSVERSE ELECTROMAGNETIC WAVES AND THE POYNTING VECTOR

Assume that a plane wave is propagating along the $\hat{z}$ direction and that the electric field is polarized along the $\hat{x}$ axis, $\mathbf{E}(r, t) = \hat{x}E_0 \cos(\omega t - kz)$. In complex notation, this would be described as

$$\begin{aligned} \mathbf{E}(r, t) &= \hat{x}\, E_0 \frac{1}{2}(e^{-j(kz-\omega t)} + e^{+j(kz-\omega t)}) \\ &= \hat{x}\frac{E_0}{2}e^{-j(kz-\omega t)} + c.c. \end{aligned} \tag{1.36}$$

We use complex notation because derivative and integral operations do not change the functional form. We must be careful to take the *real* part of expressions like Equation 1.36 when we want to describe a physical wave.

The magnitude of the magnetic amplitude can be derived from the electric amplitude using Maxwell's equations. Plug the electric amplitude (Equation 1.36) into Equation 1.1, and use Equations 1.7 and 1.30 to show:

$$\begin{aligned} \mathbf{H}(r, t) &= \hat{y}\frac{k}{\mu\omega}\frac{E_0}{2}e^{-jkz}e^{j\omega t} + c.c. \\ &= \hat{y}\frac{\omega\sqrt{\mu\epsilon}}{\mu\omega}\frac{E_0}{2}e^{-jkz}e^{j\omega t} + c.c. \\ &= \hat{y}\frac{1}{\eta}\frac{E_0}{2}e^{-jkz}e^{j\omega t} + c.c. \end{aligned} \tag{1.37}$$

where η is called the characteristic impedance of the medium,

$$\eta = \sqrt{\frac{\mu}{\epsilon}} \tag{1.38}$$

In vacuum, the characteristic impedance is $\eta_0 = 377\Omega$. Thus we see the magnitude of the magnetic amplitude is directly proportional to the magnitude of the electric amplitude. Note that **E** is perpendicular to **H**.

A useful concept for characterizing electromagnetic waves is the measure of power flowing through a surface. This quantity is called the *Poynting vector*, defined as

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \tag{1.39}$$

S represents the instantaneous intensity (W/m^2) of the wave. The Poynting vector points in the direction of power flow, which is perpendicular to both the **E** and **H** fields. The time average intensity for a harmonic field (i.e., sinusoidal waveform) is often given using phasor notation

$$\langle \mathbf{S} \rangle = \frac{1}{2} Re[\mathbf{E} \times \mathbf{H}^*] \tag{1.40}$$

where *Re* is the real part, and $\mathbf{H}^*$ is the complex conjugate of **H**. The total electromagnetic power moving into a volume is determined by a surface integral of the Poynting vector over the entire area of the volume. Often we are only interested in the average energy flow in one direction, e.g., the power crossing a dielectric interface. In such cases, the dot product of the Poynting vector with the unit direction vector must be evaluated, e.g.:

$$\langle S_z \rangle = \frac{1}{2} Re[\mathbf{E} \times \mathbf{H}^* \cdot \hat{z}] \tag{1.41}$$

This value of $\langle S_z \rangle$ is valid at only one point in space. To calculate the total power flow in a waveguide of finite extent, where the values of **E** and **H** vary in position, it is necessary to integrate the Poynting vector over the cross section of the guide.

1.7 PHASE VELOCITY

Two characteristic velocities describe the propagation of electromagnetic waves. These are the *phase velocity* and the *group velocity*. We will consider phase velocity first. Consider the sinusoidal electromagnetic wave plotted in Figure 1.5, traveling in the $\hat{z}$ direction. A point is attached to the top of one of the amplitude crests, as shown in Figure 1.5. How fast must this point move to stay on the crest of the wave? Since this crest represents a specific phase of the wave, the point must move at a speed such that:

$$e^{-j(kz-\omega t)} = \text{constant} \tag{1.42}$$

which is satisfied if $kz - \omega t$ = constant. It is easy to see $z(t)$ must satisfy:

$$z(t) = \frac{\omega t}{k} + \text{constant} \tag{1.43}$$

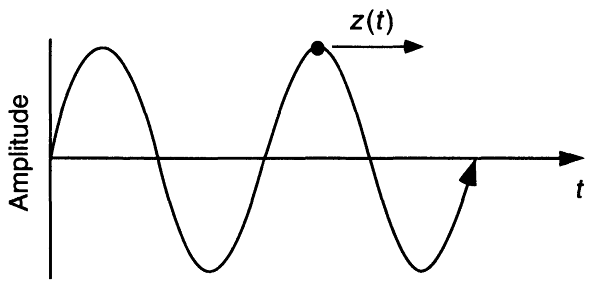

Figure 1.5 The phase velocity is determined by the speed necessary for a point to ride the crest of a wave.

We can differentiate $z(t)$ with respect to time to find the phase velocity, $v(t)$:

$$\frac{dz}{dt} = \frac{\omega}{k} = v_p \tag{1.44}$$

The phase velocity relates the angular frequency to the magnitude of the wavevector. Also, recall from Equation 1.31 that $\omega = k/\sqrt{\mu\epsilon}$, so:

$$v_p = 1/\sqrt{\mu\epsilon} \tag{1.45}$$

This is the same velocity that we derived in Equation 1.36, so the speed of light that comes out of the wave equation is the phase velocity. If permittivity $\epsilon > \epsilon_0$, then v_p is less than c, the speed of light in a vacuum. Except for unusual circumstances, such as propagation in plasmas, most materials have a permittivity, ϵ, that is greater in magnitude than ϵ_0. Do not be concerned if the phase velocity exceeds c in certain situations. Such instances are results of collective action by an oscillating medium.

We define the *index of refraction*, n, of a medium as the ratio of the phase velocity of light in a vacuum to the velocity in the medium. Using Equation 1.45:

$$n \equiv \frac{c}{v_p} \tag{1.46}$$

or, in terms of the material properties of the medium

$$\begin{aligned} n &= \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_0\epsilon_0}} \\ &= \sqrt{\frac{\epsilon}{\epsilon_0}} \quad \text{when } \mu = \mu_0 \end{aligned} \tag{1.47}$$

The index of refraction is an important parameter in optical design and material characterization. We will explore its dependence on wavelength in later chapters. The ratio ϵ/ϵ_0 is called the *dielectric constant.* The index of refraction, n, is the square root of the dielectric constant.

We often write the wavevector k in terms of the *vacuum wavevector,* k_0, and the index of refraction. The vacuum wavevector is the magnitude of the wavevector in a vacuum, and is given by $k = 2\pi/\lambda$. Using the relation $k = \omega\sqrt{\mu_0\epsilon}$, we can rewrite this as

$$k = \omega\sqrt{\mu_0\epsilon} = \omega\sqrt{\mu_0\epsilon_0}\sqrt{\frac{\epsilon}{\epsilon_0}} = \omega\sqrt{\mu_0\epsilon_0}n = k_0 n \tag{1.48}$$

Once we know the vacuum wavevector, we can define the magnitude of the wavevector in all media based on the index of refraction.

To summarize, the angular frequency, ω, of a plane wave is identical in all media. This follows from conservation of energy, where Planck's relation, $E = \hbar\omega$, describes the energy in the wave. The wavelength of the plane wave is modified by the local index of refraction to be $\lambda = \lambda_0/n$, where λ_0 is the vacuum wavelength of the plane wave. The k vector scales as $k = k_0 n$ in a medium with dielectric constant n.

1.8 GROUP VELOCITY

Except in regions of high attenuation, energy in an electromagnetic wave travels at the *group velocity*, v_g. Information which is carried by modulation on a light wave is also carried at the group velocity. The group velocity describes the speed of propagation of a pulse of light. A simple construction allows us to develop an expression for the group velocity through a superposition of two waves with different frequencies. With the frequencies assigned,

$$\omega_1 = \omega + \Delta\omega, \qquad \omega_2 = \omega - \Delta\omega, \tag{1.49}$$

the two associated wavevectors will have values

$$k_1 = k + \Delta k, \qquad k_2 = k - \Delta k \tag{1.50}$$

Assuming the waves have equal amplitudes, E_0, the superposition can be described as:

$$\begin{aligned} E_1 + E_2 = E_0(&\cos[(\omega + \Delta\omega)t - (k + \Delta k)z] \\ &+ \cos[(\omega - \Delta\omega)t - (k - \Delta k)z]) \end{aligned} \tag{1.51}$$

Using the trigonometric identity

$$2\cos x\cos y = \cos(x + y) + \cos(x - y) \tag{1.52}$$

the electric field superposition can be rewritten as

$$E_1 + E_2 = 2E_0\cos(\omega t - kz)\cos(\Delta\omega t - \Delta kz) \tag{1.53}$$

This superposition of two waves at different frequencies leads to a temporal beat at frequency $\Delta\omega$ and a spatial beat with period Δk. Figure 1.6 shows the superposition of the two waves. The envelope of the amplitude clearly depicts the beat frequency.

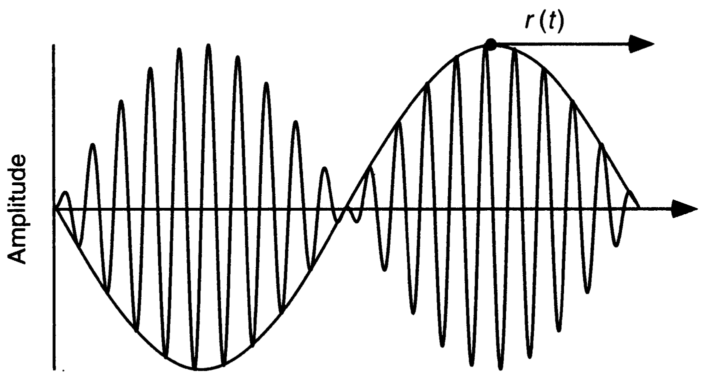

Figure 1.6 Two waves of different frequency will form a beat pattern. The envelope of the beat travels at the group velocity.

The group velocity is the speed at which a pulse, or in this case, the envelope, travels. The envelope is described by the cos $(\Delta\omega t - \Delta kz)$ term of Equation 1.53. We can again attach a point to the crest of the envelope, and ask what speed, $v(t)$, is required to stay on the crest of the envelope. Following the arguments used to derive the phase velocity, we set the phase argument $\Delta\omega t - \Delta kz =$ constant. Solving for $z(t)$,

$$z(t) = \frac{\Delta\omega t}{\Delta k} + \text{constant} \tag{1.54}$$

The group velocity is the derivative of this:

$$v_g = \frac{dz}{dt} = \frac{\Delta\omega}{\Delta k} \quad \text{becomes} \quad \lim_{\Delta\omega\to 0} \frac{\Delta\omega}{\Delta k} = \frac{d\omega}{dk} = v_g \tag{1.55}$$

The group velocity, v_g, depends on the *first derivative* of the angular frequency with respect to the wavevector. In free space, where $\omega = kc$, the relation is simple and leads to $\frac{d\omega}{dk} = c$. In a vacuum, the phase and group velocities are identical. The relation is more complicated in other media. The constitutive constants, especially ϵ, usually depend on frequency. Recall that $k = \omega n/c$. Then:

$$\begin{aligned} v_g &= \frac{d\omega}{dk} = \left[\frac{dk}{d\omega}\right]^{-1} = \left[\frac{d}{d\omega}\left(\frac{\omega n}{c}\right)\right]^{-1} = \left[\frac{n}{c} + \frac{\omega}{c}\frac{dn}{d\omega}\right]^{-1} \\ &= \frac{c}{n - \lambda\dfrac{dn}{d\lambda}} \end{aligned} \tag{1.56}$$

The last relation can be confirmed with a simple calculation. The group velocity is nearly equal to the phase velocity, but is reduced or increased by a small term proportional to the change of index of refraction with wavelength. This change in index is called *dispersion*. In regions of regular dispersion, $dn/dk > 0$, the group velocity is less than the phase velocity, c/n. Optical materials display regular dispersion throughout their transparent regions, so energy travels slower than the phase. Dispersive properties are described in more detail in Chapter 4.

1.9 BOUNDARY CONDITIONS FOR DIELECTRIC INTERFACES: REFLECTION AND REFRACTION

When two different media are adjacent to one another, the wave solutions in the two regions must be connected at the interface. The rules for connecting solutions are called *boundary conditions*. In general, if there is an index difference between two media, there will be a reflection. This is called a Fresnel reflection, after the French scientist, A. J. Fresnel (1788–1827).

Consider the interface shown in Figure 1.7. The **k** vector of an electromagnetic wave propagates from one medium into another (accompanied by a partial reflection back into the originating media). The wave has frequency ω, and is incident on the interface from region 1 at an angle of incidence, θ_i. The two regions have indices of refraction n_1 and n_2, respectively. We want to determine the amplitudes of the transmitted and reflected waves, E_t and E_r, and their respective wavevectors, $\mathbf{k}_t$ and $\mathbf{k}_r$.

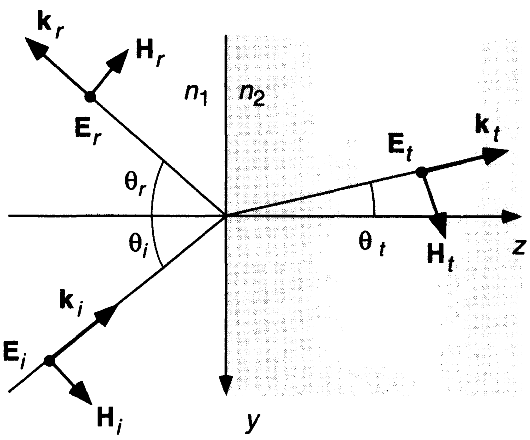

Figure 1.7 A ray incident on an interface at angle θ_i will reflect and refract into two different rays. The electric field in this figure is directed out of the page for all waves.

We must first solve the wave equation (Equation 1.26) in each region. It is straightforward to write down general solutions to the wave equation on either side of the interface,

$$E_l(r, t) = E_l e^{-j(\mathbf{k}_l \cdot \mathbf{r} - \omega t)} \tag{1.57}$$

where E_l is the amplitude. The subscript, l, refers to the three different fields that will arise. The tough part of the problem is connecting these solutions at the interface. The boundary conditions that apply to this situation can be derived from the integral form of Maxwell's equations. For review, in a medium where there are no sources, (ρ, $\mathbf{J} = 0$), the boundary conditions are:

$$\hat{s} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \quad \text{tangential components of } \mathbf{E} \text{ are continuous;} \tag{1.58}$$

$$\hat{s} \times (\mathbf{H}_2 - \mathbf{H}_1) = 0 \quad \text{tangential components of } \mathbf{H} \text{ are continuous;} \tag{1.59}$$

$$\hat{s} \cdot (\mathbf{B}_2 - \mathbf{B}_1) = 0 \quad \text{normal component of } \mathbf{B} \text{ is continuous;} \tag{1.60}$$

$$\hat{s} \cdot (\mathbf{D}_2 - \mathbf{D}_1) = 0 \quad \text{normal component of } \mathbf{D} \text{ is continuous.} \tag{1.61}$$

Where $\hat{s}$ refers to the unit normal to the interface.

There are two possible orientations for the electric field with respect to the interface. The field can be perpendicular or parallel to the *plane of incidence.* The plane of incidence contains both the $\mathbf{k}$ vector and $\hat{s}$. When the electric field is perpendicular to the plane of incidence, it is called a *transverse electric* (TE) wave. Figure 1.7 shows the specific case of a TE wave incident on an interface at an angle θ_i.

Inspection of Figure 1.7 shows there are six field amplitudes (E_i, E_t, E_r, H_i, H_t, H_r), three wavevectors (k_i, k_t, k_r), and three angles (θ_i, θ_t, and θ_r). Some of these, like E_i and θ_i, are initial conditions of the problem, while the others are dependent variables. It is convenient to first relate the angle of incidence to the angle of reflection:

$$\theta_i = \theta_r \tag{1.62}$$

Justification is straightforward: We can apply Fermat's principle (Problem 1.4), or conservation of photon momentum (Problem 1.5).

The general description of the $\hat{x}$-polarized incident electric field is:

$$\mathbf{E}_i = E_i \hat{x} e^{-j\mathbf{k}_i \cdot \mathbf{r}} e^{j\omega t} \tag{1.63}$$

The wavevector $\mathbf{k}_i$ is described in terms of its vector components,

$$\mathbf{k}_i = (\hat{z} \cos \theta_i - \hat{y} \sin \theta_i) k_0 n_1 \tag{1.64}$$

where k_0 is the *vacuum wavevector,* defined as $k_0 = \omega/c$. Position $\mathbf{r}$ is also described in vector form:

$$\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z} \tag{1.65}$$

Substituting these terms into Equation 1.63, the complete description of the incident field is

$$\begin{aligned} \mathbf{E}_i(x, y, z, t) &= \hat{x} E_i e^{-jk_0 n_1 (\hat{z} \cos \theta_i - \hat{y} \sin \theta_i) \cdot (x\hat{x} + y\hat{y} + z\hat{z})} e^{j\omega t} \\ &= \hat{x} E_i e^{-jk_0 n_1 (z \cos \theta_i - y \sin \theta_i)} e^{j\omega t} \end{aligned} \tag{1.66}$$

The incident field is completely defined in terms of direction, frequency, and polarization. The frequency term, $e^{j\omega t}$, can be dropped from the explicit formulation because it is the same in all regions. The other electric fields in Figure 1.7 are similarly described.

$$\begin{aligned} \mathbf{E}_t(x, y, z) &= \hat{x}E_t e^{-j\mathbf{k}_t \cdot \mathbf{r}} \\ &= \hat{x}E_t e^{-jk_0 n_2(z \cos\theta_t - y \sin\theta_t)} \end{aligned} \tag{1.67}$$

$$\begin{aligned} \mathbf{E}_r(x, y, z) &= \hat{x}E_r e^{-j\mathbf{k}_r \cdot \mathbf{r}} \\ &= \hat{x}E_r e^{-jk_0 n_1(-z \cos\theta_r - y \sin\theta_r)} \end{aligned} \tag{1.68}$$

We have assumed that the electric field will continue to point out of the page for each component. This may or may not be true: in some cases the phase of the field advances by 180°, and the direction would reverse. If this happens, when we have completed our solution, one of the components will be multiplied by a negative sign. This means that we did not pick the correct orientation initially. But it has no consequence so long as we consistently apply the boundary conditions and geometric projections in our solution. So do not be too concerned about choosing the proper orientations initially, as these problems will solve themselves.

We will need to describe the magnetic fields for the three waves. The appropriate k vector for each magnetic field is the same as for the electric field. Note that for TE waves, the H fields have two vector components, a z component and a y component. The *magnitude* of the magnetic field is related to the electric field through the impedance, η, of the medium (Equation 1.37).

$$|\mathbf{H}| = |\mathbf{E}|/\eta \tag{1.69}$$

where $\eta_i = \sqrt{\mu/\epsilon_i}$. Using trigonometry, the correct expression for the H field components are

$$\mathbf{H}_i = (E_i/\eta_1)(\hat{z}\sin\theta_i + \hat{y}\cos\theta_i)e^{-jn_1k_0(z\cos\theta_i - y\sin\theta_i)} \tag{1.70}$$

$$\mathbf{H}_t = (E_t/\eta_2)(\hat{z}\sin\theta_t + \hat{y}\cos\theta_t)e^{-jn_2k_0(z\cos\theta_t - y\sin\theta_t)} \tag{1.71}$$

$$\mathbf{H}_r = (E_r/\eta_1)(\hat{z}\sin\theta_r - \hat{y}\cos\theta_r)e^{-jn_1k_0(-z\cos\theta_r - y\sin\theta_r)} \tag{1.72}$$

With a complete description of the field in all regions (Equations 1.68–1.72), we can connect the solutions at the interface, yielding formulae for transmission and reflection. First, apply the condition that the tangential component of E must be continuous across the interface:

$$\hat{z} \times (\mathbf{E}_i + \mathbf{E}_r) = \hat{z} \times \mathbf{E}_t\big|_{z=0} \tag{1.73}$$

The tangential E field at the interface is the E_x component. Expanding this at $z = 0$, and using the fact that $\hat{z} \times \hat{x} = \hat{y}$ and $\theta_i = \theta_r$, yields:

$$\hat{y}E_i e^{(jk_0n_1y\sin\theta_i)} + \hat{y}E_r e^{(jk_0n_1y\sin\theta_i)} = \hat{y}E_t e^{(jk_0n_2y\sin\theta_t)} \tag{1.74}$$

Combining terms of equal phase:

$$\hat{y}(E_i + E_r)e^{(jk_0n_1y\sin\theta_i)} = \hat{y}E_t e^{(jk_0n_2y\sin\theta_t)} \tag{1.75}$$

For this equation to hold, it must be true for all values of y. At $y = 0$, the equation becomes simply:

$$E_i + E_r = E_t \qquad \text{(continuity of magnitude)} \tag{1.76}$$

Substituting this into Equation 1.75 and canceling common terms yields:

$$e^{(jk_0 n_1 y \sin \theta_i)} = e^{(jk_0 n_2 y \sin \theta_t)} \tag{1.77}$$

which can only be true if:

$$k_0 n_1 y \sin \theta_i = k_0 n_2 y \sin \theta_t \tag{1.78}$$

Canceling common terms on both sides we arrive at Snell's Law:

$$n_1 \sin \theta_i = n_2 \sin \theta_t \tag{1.79}$$

Application of the first boundary condition provides the direction of the transmitted wave. This leaves only the amplitudes, E_t, E_r, H_t, and H_r to be determined. To determine the amplitude of E_r in terms of E_i, we resort to the magnetic boundary conditions. The continuity of tangential H requires that:

$$\hat{z} \times (\mathbf{H}_i + \mathbf{H}_r) = \hat{z} \times \mathbf{H}_t\big|_{z=0} \tag{1.80}$$

In this case, $\mathbf{H}_i$ has both z and y components, so we must be careful to carry only the y component through the cross product. Using Equations 1.70–1.72, $\hat{z} \times \hat{y} = -\hat{x}$, and $\theta_r = \theta_i$:

$$\hat{z} \times \mathbf{H}_i = \frac{-\hat{x} E_i \cos \theta_i}{\eta_1} e^{-jk_0 n_1(-y \sin \theta_i)} \tag{1.81}$$

$$\hat{z} \times \mathbf{H}_t = \frac{-\hat{x} E_t \cos \theta_t}{\eta_2} e^{-jk_0 n_2(-y \sin \theta_t)} \tag{1.82}$$

$$\hat{z} \times \mathbf{H}_r = \frac{+\hat{x} E_r \cos \theta_i}{\eta_1} e^{-jk_0 n_1(-y \sin \theta_i)} \tag{1.83}$$

where E_i, E_r, and E_t represent magnitudes, not vectors. Adding the terms according to Equation 1.80, and applying Snell's law (Equation 1.79), we get

$$(E_i - E_r) \cos \theta_i / \eta_1 = E_t \cos \theta_t / \eta_2 \tag{1.84}$$

Since $E_t = E_i + E_r$, we can replace E_t in terms of the other variables

$$(E_i - E_r) \cos \theta_i / \eta_1 = (E_i + E_r) \cos \theta_t / \eta_2 \tag{1.85}$$

and solve for the ratio of E_r/E_i:

$$\frac{E_r}{E_i} = \frac{(\eta_2 \cos \theta_i - \eta_1 \cos \theta_t)}{(\eta_2 \cos \theta_i + \eta_1 \cos \theta_t)} \tag{1.86}$$

Similarly, we could eliminate E_r from Equation 1.85 and solve for the ratio E_t/E_i:

$$\frac{E_t}{E_i} = \frac{2\eta_2 \cos \theta_i}{(\eta_2 \cos \theta_i + \eta_1 \cos \theta_t)} \tag{1.87}$$

It is more common to deal with the index of refraction, n_i, than with impedance, η_i, for a material (be careful to distinguish η from n). If $\mu = \mu_0$, then η can be rewritten as:

$$\eta_i = \sqrt{\frac{\mu_0}{\epsilon_i}} = \sqrt{\frac{\mu_0\epsilon_0}{\epsilon_i\epsilon_0}} = \frac{\eta_0}{n_i} \tag{1.88}$$

Substituting this expression into Equations 1.86 and 1.87 generates the more familiar forms of the amplitude transmission and reflection formulae for a transverse electric field. In these formulae, the field is incident from the n_1 side, entering into the n_2 side.

$$\frac{E_r}{E_i} = \frac{(n_1 \cos \theta_i - n_2 \cos \theta_t)}{(n_1 \cos \theta_i + n_2 \cos \theta_t)} \tag{1.89}$$

$$\frac{E_t}{E_i} = \frac{2n_1 \cos \theta_i}{(n_1 \cos \theta_i + n_2 \cos \theta_t)} \tag{1.90}$$

The expressions for transmission and reflection of a wave which has the magnetic field H perpendicular to the plane of incidence (the so-called *transverse magnetic* or TM wave) are significantly different. Their derivation is left as an exercise to show:

$$\frac{E_r}{E_i} = \frac{n_1 \cos \theta_t - n_2 \cos \theta_i}{n_1 \cos \theta_t + n_2 \cos \theta_i} \qquad \frac{E_t}{E_i} = \frac{2 \cos \theta_i}{(n_2/n_1) \cos \theta_i + \cos \theta_t} \tag{1.91}$$

One word of caution about the Fresnel formulae: they describe the *amplitude* of the transmitted and reflected field, not the *power* of the fields. One must be careful since, in some circumstances, the magnitude of the transmitted electric field can be *larger* than that of the incident electric field. This dilemma is resolved when total power is accounted for in the solution. In such cases, one can either rigorously solve for the z component of the Poynting vector for the transmitted and incident waves, or account for the geometric change of area between the incident and transmitted beams. Problem 1.8 explores the power issues of these formulae. We could also develop expressions for the H components, but these can be found simply and directly through the impedance relationships.

Example 1.3 Normal Reflection from a Glass Interface

The simplest example of Fresnel reflection is that which occurs when light strikes a glass-air interface. We experience this effect on a daily basis. Let's apply the reflection formula to this problem to illustrate the magnitude of the effect, and the phase shift which occurs.

A beam of light is incident normally on a glass-air interface as shown in Figure 1.8. What is the intensity of the reflected light if the glass has an index of refraction of $n = 1.5$?

Solution:

Plugging numbers into Equation 1.90, noting that $\cos\theta = 1$ in this case, we get

$$E_r/E_i = \frac{1 - 1.5}{1 + 1.5} = -0.2 \tag{1.92}$$

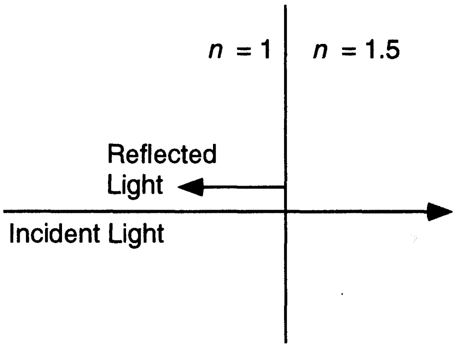

Figure 1.8 A beam of light strikes a glass interface normally, causing a small reflection.

The reflected amplitude is 20 percent of the incident amplitude. The negative sign indicates that the reflected wave is 180° out of phase with the incident wave. In general, when light strikes a surface with a higher index, the phase of the reflected wave will be reversed. Now, what is the intensity? Using the Poynting vector and the fact that $|\mathbf{H}| = |\mathbf{E}|/\eta$, we find the incident intensity is:

$$S_{inc} = \frac{1}{2}\frac{E_0^2}{\eta} \tag{1.93}$$

while the reflected intensity is only:

$$S_{ref} = \frac{1}{2}\frac{(0.2E_0)^2}{\eta} = 0.04S_{inc} \tag{1.94}$$

Thus, only 4 percent of the incident power is reflected by the glass interface. This reflection can become a significant loss in certain applications. For example, a camera lens often will consist of three or more separate lenses, representing six glass-air interfaces. The total transmission for such a system would be $T = (0.96)^6 = 0.78$ if the lenses are not modified. This represents a significant loss of power in an application where light collection efficiency is critical. Not only would the reflections require larger apertures and longer exposure times, but they also could contribute to ghost images on the film. These problems are overcome by putting an anti-reflection (AR) coating on each surface. The AR coating is basically a stack of $\lambda/4$-thick layers of dielectric material which interferometrically reduce the total reflection coefficient.

1.10 TOTAL INTERNAL REFLECTION

An important physical process in guided wave optics is *total internal reflection.* We will look at total internal reflection from two perspectives: ray tracing and the wave equation. Ray tracing is useful when the dimensions of the optical element are large compared to the wavelength of light. Ray tracing is useful for concepts such as the numerical aperture. The wave picture provides a complete description of the phase shifts and evanescent fields that accompany total internal reflection.

1.10.1 Ray Tracing

Ray tracing views light as rays traveling in straight lines between optical elements. The only action of an optical element is to redirect the ray. The angle of incidence of the ray and the properties of the optical element establish the degree to which the ray is redirected.

The important operational rules for ray tracing are *Snell's law*

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{1.95}$$

and the *Law of Reflection*

$$\theta_{incidence} = \theta_{exit} \tag{1.96}$$

illustrated in Figure 1.9. Using these two simple equations, a powerful calculus can be developed for designing and evaluating lenses and optical systems. Many excellent references [8–10] elaborate on the application of ray tracing to optical design. Powerful matrix techniques have been developed based on these simple laws which allow the engineer to design linear optical systems. The ray tracing analysis is usually of limited use for guided wave optical design, however. Its most common application is to describe graded index waveguides, and to define the numerical aperture.

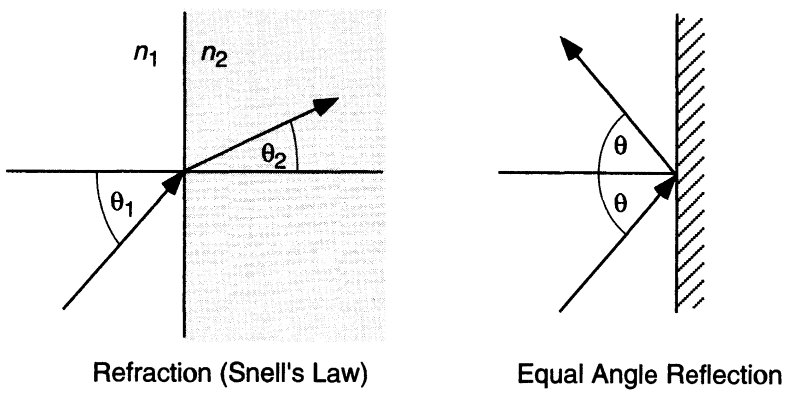

Figure 1.9 There are two principle laws of ray tracing. The left figure shows Snell's law. The right figure illustrates that the angle of incidence equals the angle of reflection.

1.10.2 Total Internal Reflection Using Ray Tracing

Total internal reflection (TIR) is the phenomenon where light is completely reflected at a dielectric interface without the help of reflective coatings. TIR is often exploited to make efficient achromatic reflectors. For example, right-angle prisms are often used to redirect light from imaging systems such as binoculars, or to serve as rugged mirrors for high-powered lasers. The application that interests us is in optical waveguides. Figure 1.10 illustrates the ray picture of a right-angle prism and of a waveguide. The key requirement for TIR is that the light must be incident on a dielectric interface from the high-index side. Thus an optical waveguide must consist of a layer of high-index dielectric surrounded by material with a lower index.

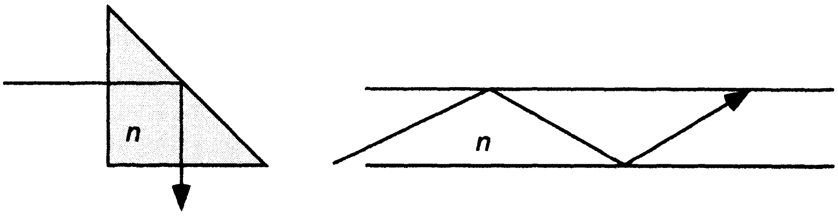

Figure 1.10 Total internal reflection can be implemented in many ways. The right-angle prism and the optical waveguide both use total internal reflection to redirect or trap light, respectively. Note that the light is incident from the high-index side of the interface in all cases of TIR.

Total internal reflection occurs over a certain range of angles. Figure 1.11 shows a wave incident at an angle, θ_1, on a dielectric interface from the high-index side. The refracted ray in the low-index medium, n_2, exits at angle θ_2. The exit angle is:

$$\theta_2 = \sin^{-1}\left(\frac{n_1}{n_2}\sin\theta_i\right) \tag{1.97}$$

As the angle of incidence, θ_1, increases, the angle of refraction, θ_2, must also increase to satisfy the equality. But because $n_1/n_2 > 1$, the refraction angle, θ_2, will reach a value of 90° before θ_1 does. This occurs when

$$\sin\theta_1 = \frac{n_2}{n_1} \tag{1.98}$$

This value of θ_1 is known as the *critical angle*. For angles of incidence larger than the critical angle, θ_2 must be a complex number to satisfy Snell's law. A complex angle in the expressions for transmission (for example, Equation 1.90) leads directly to complex amplitudes in the low-index region. Complex amplitudes simply means that a phase shift occurs. As with all simple reflections, the angle of reflection is equal to the angle of incidence of the ray.

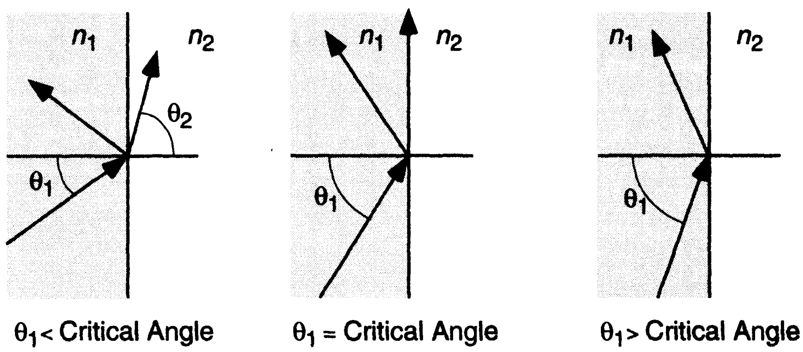

Figure 1.11 Three examples show where the angle of incidence is below, at, and above the critical angle respectively.

Total internal reflection is the key to optical waveguiding. Consider the dielectric structure shown in Figure 1.12. A dielectric slab of index n_1 is surrounded by a lower-index dielectric. A ray traveling within the high-index material will be total-internal-reflected at the upper and lower interfaces of this structure *if* the angle of incidence at the interface exceeds the critical angle. This is a simplified picture, as the actual ray picture of a waveguide is more subtle in terms of allowed directions for the rays. (This will be fully developed in the next chapter.) However, the essential idea behind the optical waveguide is that light is trapped in a high-index media through total internal reflection.

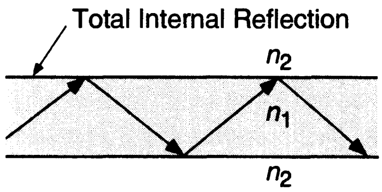

Figure 1.12 A waveguide can be formed when total internal reflection traps a wave between two surfaces.

Example 1.4 An Optical Switch

Consider the interface between a *nonlinear* dielectric of high index and a linear dielectric of lower index. The nonlinear media has an index which depends on the intensity of the electromagnetic field. This is called the optical Kerr effect. The Kerr effect modifies the index of refraction in the following way:

$$n(I) = n_0 + n_2 \cdot I \tag{1.99}$$

where n_2 is a small constant. For this example, let's assume that $n_2 = 10^{-10}$ cm^2/W. (This is unusually large. Ordinary glass has an n_2 about 6 orders of magnitude smaller than this.)

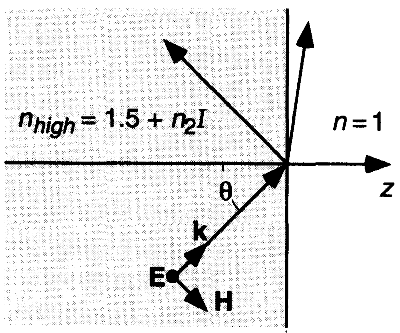

Figure 1.13 The nonlinear optical switch consists of a wave incident from the high-index side of an interface between a nonlinear medium and a linear medium.

a. Assume that $n_{high} = 1.5$ for low intensities, and $n_{low} = 1$ beyond the interface. For very low intensities (i.e., neglect the nonlinear portion of n_{high}), what is the critical angle, θ_c, for a ray incident from n_{high} to n_{low}?

b. If the angle of incidence is 0.5° smaller than the low-intensity value for θ_c, the wave will have a finite transmission into the lower index media for low intensities. What is the power of the transmitted field?

c. If the angle of incidence is 0.5° smaller than the low-intensity value for θ_c, at what intensity will the interface cease to transmit light?

Solution:

a. The critical angle is given by $\sin\theta_c = n_{low}/n_{high}$, or

$$\theta_c = \sin^{-1}\left(\frac{1}{1.5}\right) = 41.81° \tag{1.100}$$

b. If θ is set to be 41.31°, the ray will be partially transmitted at low intensity because this angle is below the critical angle. The amplitude of the transmitted field, E_t, is given by Equation 1.89. The refracted angle, θ_2, is needed to evaluate the transmitted term. From Snell's law, $\theta_2 = 81.97°$, and it follows that

$$\begin{aligned} E_t/E_i &= \frac{2n_{high}\cos\theta_1}{(n_{high}\cos\theta_1 + n_{low}\cos\theta_2)} \\ &= \frac{2\cdot 1.5\cdot\cos 41.31}{1.5\cdot\cos 41.31° + \cos 81.97°} = 1.778 \end{aligned} \tag{1.101}$$

Using this result with Equations 1.41 and 1.69, the power of the transmitted beam (recall that $\eta_1 = \eta_0/n_1$, $\eta_2 = \eta_0/n_2$) is given by

$$\begin{aligned} S_z &= \frac{n_l E_t^2}{2\eta_0}\cdot\hat{z} \\ &= \frac{(1.778E_i)^2}{2\cdot 377}\cdot\cos 81.97° = 0.00059\, E_i^2 \end{aligned} \tag{1.102}$$

The incident power is

$$\begin{aligned} S_z &= \frac{n_{high}E_i^2}{2\eta_0}\cdot\hat{z} \\ &= \frac{1.5E_i^2}{2\cdot 377}\cdot\cos 41.31° = 0.001494\, E_i^2 \end{aligned} \tag{1.103}$$

Dividing Equation 1.102 by the result of Equation 1.103, we see that approximately 39 percent of the power is transmitted through the interface.

c. To convert to total internal reflection, the index, n_{high}, must increase to

$$n_{high}(I) = \frac{1}{\sin 41.31°} \tag{1.104}$$

The critical index will be reached when $n_{high}(I) = 1.5148$. The necessary intensity is

$$n_{high}(I) = 1.5 + 10^{-10}I = 1.5148 \tag{1.105}$$

Solve for I to get $I = 1.48 \times 10^8$ W/cm^2.

Is this a practical switch? No. First, we have chosen an unrealistically large value for the nonlinear index term. Second, the amount of light actually transmitted in the "on" state is a small fraction of the incident light. A practical switch for systems application would require a more efficient transmission factor.

1.11 WAVE DESCRIPTION OF TOTAL INTERNAL REFLECTION

We claim that the ray became totally reflected for angles beyond the critical angle, yet the only evidence we offered to support this claim is that the trigonometric identity is impossible to rationalize using real angles. We can put the description on a more physical basis by examining total internal reflection using electromagnetic waves. The wave picture provides a physical explanation of the reflection, and yields information on the phase shift caused by reflection.

Consider a TE plane wave, polarized along the $\hat{x}$-axis with amplitude E_0 incident on a dielectric interface, as shown in Figure 1.14. The angle of incidence is less than θ_c. Only the spatial descriptions of the two waves are considered, since the time behavior is identical for both:

$$\begin{aligned} \mathbf{E}_1(y,z) &= \hat{x}E_0 e^{-jk_0 n_1(z\cos\theta_1 - y\sin\theta_1)} + c.c. \\ \mathbf{E}_2(y,z) &= \tau\hat{x}E_0 e^{-jk_0 n_2(z\cos\theta_2 - y\sin\theta_2)} + c.c. \end{aligned} \tag{1.106}$$

where τ is the amplitude transmission coefficient (from Equation 1.90). The angles θ_1 and θ_2 are related by Snell's law.

$$\begin{aligned} \sin\theta_2 &= \frac{n_1}{n_2}\sin\theta_1 \\ \cos\theta_2 &= \sqrt{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_1} \end{aligned} \tag{1.107}$$

Substituting these values into Equation 1.67, we get an expression for the transmitted amplitude, $\mathbf{E}_2$, that is a function of the incident angle θ_1,

$$\mathbf{E}_2 = \tau\hat{x}E_0 \exp\left\{-jk_0 n_2\left(z\sqrt{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_1} - y\frac{n_1}{n_2}\sin\theta_1\right)\right\} \tag{1.108}$$

Physically, we can understand refraction by considering what happens to the wavefronts at the interface. On the incident side, the wavefront strikes the interface and is partially reflected and partially transmitted. If $\theta_1 < \theta_c$, the wavefronts must be continuous across the interface. The node where these two wavefronts connect travels along the interface with a velocity v_{node}, as shown in Figure 1.15. The velocity of this intersection, v_{node}, is simply

$$v_{node_1} = v_{p_1}/\sin\theta_1 \tag{1.109}$$

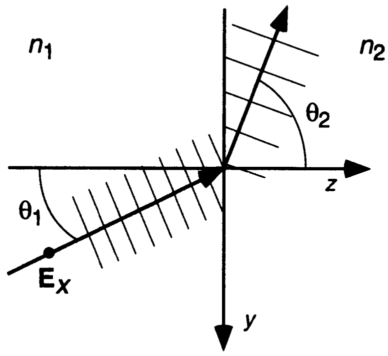

Figure 1.14 A plane wave incident on a dielectric interface at angle θ_1 will refract at an angle θ_2 in the second medium. The reflected ray is not shown for clarity.

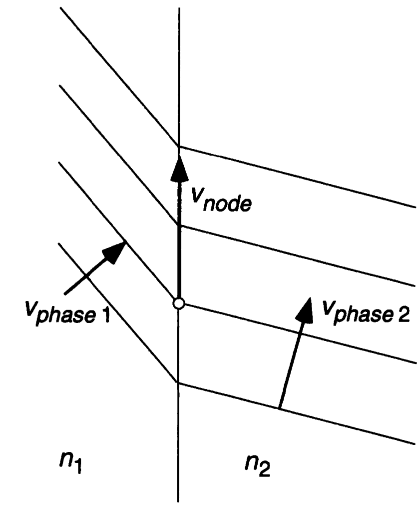

Figure 1.15 The plane waves on either side of the interface must connect as they cross the interface. These connecting nodes travel along the interface at a velocity that depends on the angle of incidence.

where v_{p_1} is the phase velocity c/n_1 in the first medium. The transmitted wave, $\mathbf{E}_2$, must travel in such a direction that the velocity of the nodes of its phase front is identical to that of the incident field. Since the phase velocity in medium n_2 is different, the only way the node velocities can be matched is if the direction of the transmitted field refracts to angle θ_2 such that

$$v_{p_1}/\sin\theta_1 = v_{p_2}/\sin\theta_2 \tag{1.110}$$

This is simply a restatement of Snell's law.

As the angle of incidence, θ_1, increases, the transmitted waves must make a larger angle, θ_2, to maintain the proper velocity of the intersection at the interface. At $\theta_1 = \theta_{cr}$, $\cos\theta_2$ goes to zero, and the transmitted field contains only one component,

$$\mathbf{E}_2 = \tau\hat{x}E_0e^{(jk_0n_2y)} + c.c. \qquad \text{at } \theta_1 = \theta_{cr} \tag{1.111}$$

This is the description of a plane wave traveling parallel to the interface in the $\hat{y}$ direction. This direction will yield a node velocity that is as slow as can be achieved in medium n_2. In the ray picture, we would say that the transmitted ray is parallel to the plane of incidence. Figure 1.16 shows this condition.

What happens as θ_1 increases *beyond* the critical angle? The radical in Equation 1.107, which describes $\cos\theta_2$, becomes imaginary, so the transmitted electric amplitude is described as

$$\mathbf{E}_2 = \tau\hat{x}E_0e^{-k_0n_2\sqrt{(n_1^2/n_2^2)\sin^2\theta_1 - 1}\,z}\,e^{jk_0n_1\sin\theta_1 y} \tag{1.112}$$

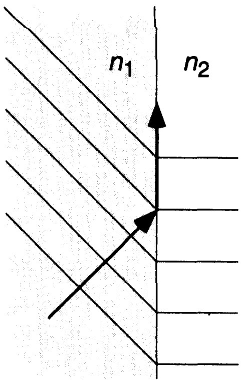

Figure 1.16 At the critical angle, the transmitted plane waves travel parallel to the interface.

where we choose the proper sign of the radical to ensure that the amplitude decays as distance from the interface increases. This cumbersome form is often written as

$$\mathbf{E}_2 = \tau \hat{x} E_0 e^{-\gamma z} e^{j\beta y} \tag{1.113}$$

where γ represents the *attenuation coefficient* (units: cm^{-1})

$$\gamma = k_0 n_2 \sqrt{\frac{n_1^2}{n_2^2} \sin^2 \theta_1 - 1} \tag{1.114}$$

and β represents the *propagation coefficient* (units: rads/cm)

$$\beta = k_0 n_1 \sin \theta_1 \tag{1.115}$$

Inspection of Equation 1.113 shows that the field amplitude decays exponentially away from the interface. This field is called the *evanescent* field. The evanescent field contains real values of **E** and **H**, but they are 90° out of phase with each other. The evanescent field contains *reactive power*, not real power. In reactive power, no work is done, but energy is stored. This evanescent field is very important to device applications. It is possible to tap some of the energy away using special structures. We will see many such devices in later chapters concerning switches, modulators, and couplers.

Returning to the physical picture, when θ_1 is increased beyond the critical angle, as in Figure 1.17, the node velocity in n_1 is slower than the minimum possible velocity of nodes in medium n_2. In medium n_2, the phase fronts advance beyond their generating counterparts in n_1. As the transmitted wave fronts travel ahead, they run up on wavefronts emitted from earlier nodes. At a certain distance, the fronts in n_2 will be 180° out of phase with the nodes of n_1, and destructive interference will occur. The larger the angle of incidence θ_1, the slower the node velocity in n_1 will be. Destructive interference will occur sooner, leading to increased attenuation. We see from Equation 1.114 that the attenuation coefficient γ increases as the angle of incidence is increased.

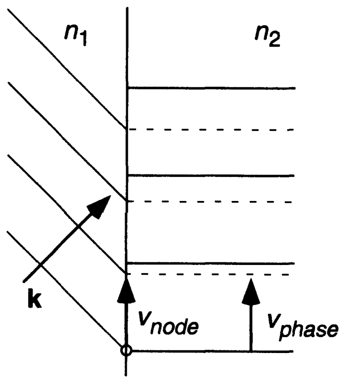

Figure 1.17 Beyond the critical angle, the plane waves on the low-index side of the interface travel faster than the nodes due to the incident field. They get ahead of their source nodes, and then react back against them.

1.12 PHASE SHIFT UPON REFLECTION

A more subtle yet critically important effect that occurs in TIR is the phase shift of the light upon reflection. These phase shifts help determine which modes propagate in a waveguide. After reflection, the optical signal slightly lags in phase compared to the incident wave. One can view this phase shift as being due to the extra distance the light travels when going into and returning from the low-index media during its evanescent phase (the Goos-Hänchen shift, described in Problem 1.22 and Appendix A), or one can view the phase shift as occurring due to the mixing of two waves that are slightly out of phase (the reflected and evanescent wave).

How big is the phase shift? For a TE wave, the phase shift can be determined directly by writing the amplitude reflection formula, Equation 1.90, in polar form:

$$\frac{E_r}{E_i} = \frac{(n_1 \cos \theta_1 - n_2 \cos \theta_2)}{(n_1 \cos \theta_1 + n_2 \cos \theta_2)} = |r|\, e^{j2\phi} \tag{1.116}$$

The reflection coefficient is described in terms of its magnitude, $|r|$, and phase shift, 2Φ. Beyond the critical angle, $\cos \theta_2$ becomes purely imaginary ($\cos \theta_2 = \sqrt{1 - n_1^2/n_2^2 \sin^2 \theta_1}$). Letting $\alpha = n_1 \cos \theta_1$, and $j\beta = n_2 \cos \theta_2$, Equation 1.116 can be rewritten as:

$$\frac{E_r}{E_i} = \frac{\alpha - j\beta}{\alpha + j\beta} \tag{1.117}$$

Substituting the value of $\cos \theta_2$ from Equation 1.107, the phase of this transfer function is:

$$
\begin{aligned}
2\Phi_{TE} &= \tan^{-1}\left(\frac{-\beta}{\alpha}\right) - \tan^{-1}\left(\frac{\beta}{\alpha}\right) \\
&= 2\tan^{-1}\left(\frac{-\beta}{\alpha}\right) \\
&= 2\tan^{-1}\frac{-\sqrt{n_1^2\sin^2\theta_1 - n_2^2}}{n_1\cos\theta_1}
\end{aligned}
\tag{1.118}
$$

This equation is only valid for $\theta_1 > \theta_{cr}$. The magnitude of r is obviously unity. We leave it as an exercise to show that the correct formula for TM waves is given by:

$$
\Phi_{TM} = \tan^{-1}\left(\frac{-n_1^2}{n_2^2}\frac{\sqrt{n_1^2\sin^2\theta_1 - n_2^2}}{n_1\cos\theta_1}\right) \tag{1.119}
$$

Figure 1.18 shows the dependence of Φ_{TE} as a function of the angle of incidence θ_1 for two ratios n_1/n_2. The ratios 0.3 and 0.7 correspond to the approximate values of a GaAs-air and glass-air interface, respectively. The phase shift for the TM case is similar.

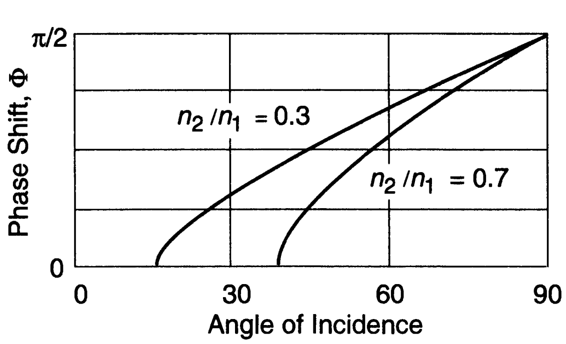

Figure 1.18 In this plot of the phase shift Φ as a function of the angle of incidence, θ, note that the phase shift below the critical angle is zero.

For angles of incidence below the critical angle, there is no phase shift upon reflection (actually, the phase shift can be 0 or π, depending on the relative indices).

Example 1.5 Fields Surrounding a Total Internal Reflection

Consider the situation shown in Figure 1.19. A TE wave is incident from GaAs with an index of 3.4 onto the GaAs-air interface, at an angle of incidence of 45°. Describe the electric fields in all regions surrounding the interface. Assume that the light has a wavelength of $\lambda = 1\mu m$.

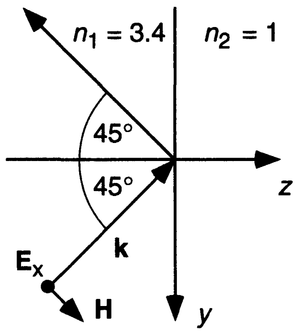

Figure 1.19 A wave incident at 45° on a dielectric interface from the high-index side is TE polarized, and has a wavelength of 1 μm.

Solution:

For this dielectric interface, the critical angle is:

$$\theta_{crit} = \sin^{-1}\frac{1}{3.4} = 17.1° \quad (1.120)$$

The incident wave is well above the critical angle, and will undergo total internal reflection. The electric field in the air region ($z > 0$) is given by:

$$\mathbf{E}(y, z) = \tau\hat{x}E_{inc}e^{-\gamma z}e^{jk_0 n_1 \sin\theta_1 y} \quad (1.121)$$

where $k_0 = 2\pi/1\mu\text{m} = 6.283 \times 10^4$ rads/cm, and

$$\gamma = k_0 n_2\sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_1 - 1}$$
$$= 6.283 \times 10^4 \cdot 1 \cdot \sqrt{3.4^2(0.707)^2 - 1} = 1.37 \times 10^5\text{cm}^{-1} \quad (1.122)$$

The value of the transmission coefficient, τ, is directly derived from Equation 1.89.

$$\tau = \frac{2n_1\cos\theta_1}{n_1\cos\theta_1 + n_2\cos\theta_2} \quad (1.123)$$

In this case, θ_2 is imaginary,

$$\cos\theta_2 = -\sqrt{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_1} = -j\,2.19 \quad (1.124)$$

where j stands for $\sqrt{-1}$. We chose the negative sign of the radical in order to assure that the field in Equation 1.108 decays with increasing z. Having chosen the sign, we must use it consistently throughout the rest of the calculation. Plugging $\cos\theta_2$ into Equation 1.123 yields

$$\tau = \frac{2\cdot(3.4)\cdot(0.707)}{(3.4\cdot 0.707) - j2.19} \quad (1.125)$$

A complex transmission coefficient simply means that there is a relative phase difference between the incident and transmitted wave. In phasor notation, the transmission coefficient, τ, is equal to

$$\tau = |\tau| \angle\left[\tan^{-1}(0) - \tan^{-1}\left(\frac{-2.19}{2.4}\right)\right]$$
$$= 1.48\angle 42.38° \tag{1.126}$$

The transmitted electric field in the air ($z > 0$) is

$$E(z, y) = 1.48E_i e^{-1.37\times 10^5 z} e^{j1.51\times 10^5 y} e^{j(\omega t+42.38°)} \tag{1.127}$$

where mixed units (radians and degrees) are used in the last term, but the meaning should be clear.

The reflected field can be found noting that $|r| = 1$, and the phase shift is given by Equation 1.116 to be:

$$\Phi_{TE} = \tan^{-1}\frac{-\sqrt{n_1^2 \sin^2\theta_1 - n_2^2}}{n_1 \cos\theta_1}$$
$$= \tan^{-1}\frac{-\sqrt{3.4^2(0.707)^2 - 1}}{3.4 \cdot 0.707} \tag{1.128}$$
$$= -42.29°$$

Thus the reflected field ($z < 0$) will be given by:

$$E(y, z) = E_i e^{-jk_0 3.4 \cdot 0.707(-y-z)} e^{j(\omega t-84.6°)} \tag{1.129}$$

Note that the amplitude of the reflected electric field is identical to the incident field, but there is a phase delay of 2Φ between the two waves.

1.13 SUMMARY

This chapter reviewed Maxwell's equations, using them to establish a set of units (MKS) and several important quantities and concepts. We derived the wave equation, and solved it in homogeneous media. From the solution, we developed expressions for phase and group velocity. The concept of the wavevector was introduced and related to the angular frequency of a wave. Using boundary conditions, we developed expressions for the reflection and refraction of electromagnetic waves from a dielectric interface.

We then explored total internal reflection. Snell's law was used to illustrate the ray picture of total internal reflection. While Snell's law, if used with complex angles, can give a total description of the evanescent fields associated with these reflections, the wave description based on Maxwell's equations provides a clearer picture. Using the wave picture, we used the Fresnel formulae for reflection and transmission at a dielectric interface to develop expressions for phase shift associated with TIR. This phase shift always accompanies TIR, and plays a unique role in establishing which rays will be allowed inside an optical waveguide.

The material parameters, μ and ϵ, play a critical role in determining the action of a wave at a dielectric interface. We alluded to the frequency dependence of these material parameters in the discussion of group velocity. This will be further developed in Chapter 3.

REFERENCES

1. F. Jenkins and H. White, *Fundamentals of Optics* (New York: McGraw-Hill, 1957).
2. E. Hecht, *Optics* (Reading, MA: Addison-Wesley Publishing Company, 1990).
3. S. G. Lipson and H. Lipson, *Optical Physics* (New York: Cambridge University Press, 1981).
4. W. Louisell, *Radiation and Noise in Quantum Electronics* (New York: McGraw-Hill, 1964).
5. R. Loudon, *Quantum Theory of Light,* 2nd ed. (Oxford, England: Clarendon Press, 1983).
6. See, for example, W. J. Tabor, "Magneto-Optic Materials," in *Laser Handbook,* vol. 2. eds. F. T. Arecchi and E. O. Schulz-Dubois (Great Britain: North-Holland Publishing Company, 1972).
7. A. Yariv and P. Yeh, *Optical Waves in Crystals* (New York: John Wiley and Sons, 1984).
8. F. Jenkins and H. White, *Fundamentals of Optics* (New York: McGraw-Hill, 1957).
9. A. Nussbaum and R. Phillips, *Contemporary Optics for Scientists and Engineers* (Englewood Cliffs, NJ: Prentice Hall, 1976).
10. Warren Smith, *Modern Optical Engineering,* 2nd ed. (New York: McGraw-Hill, 1990).

PROBLEMS

1. Derive the Fresnel amplitude reflection and transmission coefficients for an electromagnetic wave that is polarized with the electric field in the plane of incidence (TM wave).
2. We simplified Equation 1.24 by assuming that the term $-\nabla(\mathbf{E}\cdot\nabla\epsilon/\epsilon)$ is negligible. Determine how small $\nabla\epsilon$ must be for this assumption to be reasonable. Starting from the exact wave equation (with the above term included), use separation of variables to solve for the one-dimensional wave (i.e., $E = Z(z)T(t)$). Solve for $T(t)$ in terms of separation constant k and $(\mu\epsilon)^{1/2}$. From the resulting equation for $Z(z)$, find a rough value for $\nabla\epsilon$ over a characteristic distance of one wavelength of the field. How small must $\frac{\Delta\epsilon}{\epsilon}$ be to make it negligible (say less than 1 percent in magnitude) compared to the other terms in the wave equation?
3. Show that for a harmonic wave, the average value $\langle\mathbf{S}\rangle = \langle\mathbf{E}\times\mathbf{H}\rangle = \frac{\mathbf{k}}{|k|}\frac{k}{2\omega\mu}E_0^2$, for a wave with wavevector k and electric amplitude E_0.
4. Fermat's principle states that if a light ray travels between two points, it follows the path that takes the least time. Use Fermat's principle to *a)* verify that the angle of incidence equals the angle of reflection for a simple plane mirror, and *b)* derive Snell's law for a ray crossing a dielectric interface.
5. Use conservation of momentum and the fact that a photon has momentum given by $\mathbf{p} = \hbar\mathbf{k} = \hbar n\mathbf{k}_0$, where $\mathbf{k}_0$ is the vacuum wavevector of the photon, to *a)* show that the angle of incidence equals the angle of reflection for a simple plane mirror, and *b)* derive Snell's law for a ray crossing a dielectric interface.