Section 3: Electromagnetic Waves in Vacuum and Simple Matter

3.1 Electromagnetic waves in vacuum

A basic feature of Maxwell's equations for the electromagnetic field is the existence of traveling wave solutions which represent the transport of energy from one point to another. The simplest and most fundamental electromagnetic waves are transverse, plane waves in vacuum.

In regions of space where there are no charges and currents, Maxwell's equations read

$$\nabla\cdot\mathbf{E}=0\,, \tag{3.1}$$

$$\nabla\cdot\mathbf{B}=0\,, \tag{3.2}$$

$$\nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}\,, \tag{3.3}$$

$$\nabla\times\mathbf{B}=\mu_0\varepsilon_0\frac{\partial\mathbf{E}}{\partial t}\,. \tag{3.4}$$

They are a set of coupled, first order, partial differential equations for **E** and **B**. They can be decoupled by applying curl to Eqs. (3.3) and (3.4):

$$\nabla\times(\nabla\times\mathbf{E})=\nabla(\nabla\cdot\mathbf{E})-\nabla^2\mathbf{E}=-\nabla\times\frac{\partial\mathbf{B}}{\partial t}=-\frac{\partial}{\partial t}(\nabla\times\mathbf{B})=-\mu_0\varepsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2}\,, \tag{3.5}$$

$$\nabla\times(\nabla\times\mathbf{B})=\nabla(\nabla\cdot\mathbf{B})-\nabla^2\mathbf{B}=\nabla\times\left(\mu_0\varepsilon_0\frac{\partial\mathbf{E}}{\partial t}\right)=\mu_0\varepsilon_0\frac{\partial}{\partial t}(\nabla\times\mathbf{E})=-\mu_0\varepsilon_0\frac{\partial^2\mathbf{B}}{\partial t^2}\,. \tag{3.6}$$

Since $\nabla\cdot\mathbf{E}=0$ and $\nabla\cdot\mathbf{B}=0$ we have

$$\nabla^2\mathbf{E}=\mu_0\varepsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2}\,, \tag{3.7}$$

$$\nabla^2\mathbf{B}=\mu_0\varepsilon_0\frac{\partial^2\mathbf{B}}{\partial t^2}\,. \tag{3.8}$$

We now have *separate* equations for **E** and **B**, but they are of *second* order. In vacuum, then, each Cartesian component of **E** and **B** satisfies the three-dimensional equation

$$\nabla^2 f=\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}\,. \tag{3.9}$$

Equation (3.9) is known as the classical wave equation, because it admits as solutions all functions in the form $f(\mathbf{r},t)=g(\mathbf{n}\cdot\mathbf{r}-ct)$, where the function f depends on the variables **r** and t in their special combination $u\equiv\mathbf{n}\cdot\mathbf{r}-ct$ with **n** being a unit vector which defines the direction of the wave propagation. To show this, we calculate $\nabla^2 f$ and $\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}$ entering Eq. (3.9):

$$\nabla^2 f=\sum_i\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right)=\sum_i\frac{\partial}{\partial x_i}\left(\frac{\partial g}{\partial u}\frac{\partial u}{\partial x_i}\right)=\sum_i\frac{\partial}{\partial x_i}\left(\frac{\partial g}{\partial u}\right)n_i=\sum_i n_i\frac{\partial}{\partial u}\left(\frac{\partial g}{\partial u}\right)\frac{\partial u}{\partial x_i}=\sum_i n_i^2\frac{\partial^2 g}{\partial u^2}=\frac{\partial^2 g}{\partial u^2}\,. \tag{3.10}$$

$$\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}=\frac{1}{c^2}\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial u}\frac{\partial u}{\partial t}\right)=\frac{1}{c^2}\frac{\partial}{\partial t}\left(\frac{\partial g}{\partial u}\right)(-c)=-\frac{1}{c}\frac{\partial}{\partial u}\left(\frac{\partial g}{\partial u}\right)\frac{\partial u}{\partial t}=\frac{\partial^2 g}{\partial u^2}\,. \tag{3.11}$$

It is seen that solutions in the form $f(\mathbf{r},t)=g(\mathbf{n}\cdot\mathbf{r}-ct)$ satisfy Eq. (3.9).

The wave nature of the solution $f(\mathbf{r},t)=g(\mathbf{n}\cdot\mathbf{r}-ct)$ can be understood as follows. Assume for simplicity that $\mathbf{n}=\hat{\mathbf{z}}$ and $f(x,y,z,t)=f(z,t)$ is independent of *x* and *y*, i.e. $f(z,t)=g(z-ct)$. Then the initial shape of the wave at $t=0$ is $f(z,0)=g(z)$. At the later time *t*, the wave shape at point *z* appears to be the same as that at a distance *ct* to the left (i.e. at $z-ct$) back at time *t* = 0, i.e. $f(z,t)=f(z-ct,0)=g(z-ct)$.

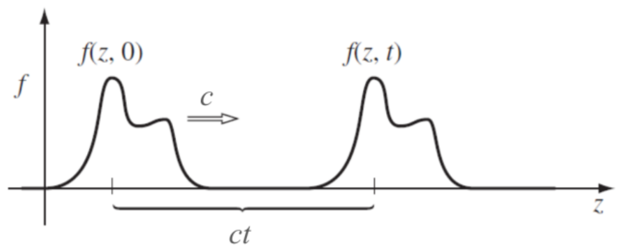

Fig. 3.1

That statement captures the essence of wave motion. It tells us that the function $f(\mathbf{r},t)$ depends on *z* and *t* in the very special combination $z-ct$. This special dependence implies that the function $f(\mathbf{r},t)$ represents a wave of fixed shape traveling in the *z* direction at speed *c* (Fig. 3.1).

Coming back to Eq. (3.9), we see that Maxwell's equations imply that support the propagation of electromagnetic waves in empty space, traveling at a speed

$$c=\frac{1}{\sqrt{\mu_0\varepsilon_0}}=3.00\cdot 10^8\,m/s \tag{3.12}$$

which happens to be precisely the velocity of light, *c*. The important implication of this result is that light represents an electromagnetic wave. While this result is not surprising today, it was a revelation in Maxwell's time. Remember that ε_0 and μ_0 came into the theory as constants in Coulomb's law and the Biot-Savart law, respectively. They can be measured in experiments involving charged balls, batteries, and wires—experiments having nothing whatever to do with light. And yet, according to Maxwell's theory one can calculate *c* from these two numbers.

3.2 Monochromatic plane waves

The simplest solution of Eq. (3.9) is the sinusoidal wave

$$f(z,t)=A\cos\left[k(z-ct)+\delta\right]. \tag{3.13}$$

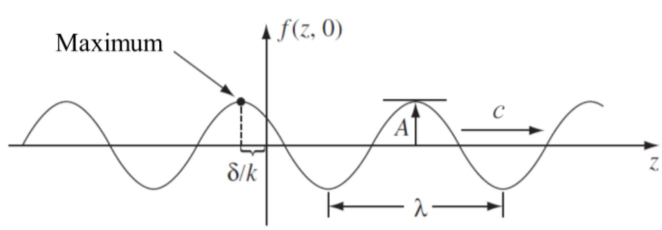

Fig. 3.2

Figure 3.2 shows this function at time *t* = 0. *A* is the *amplitude* of the wave (it is positive, and represents the maximum displacement from equilibrium). The argument of the cosine is called the *phase*, and *δ* is the *phase constant* (a value in the range $0 \le \delta < 2\pi$). Notice that at $z = ct - \delta / k$, the phase is zero corresponding to the wave maximum. If $\delta = 0$, the maximum passes the origin at time $t = 0$; more generally, δ/k is the distance by which the maximum (and therefore the entire wave) is "delayed."

Finally, k is the *wave number*, which is related to the *wavelength* λ by the equation

$$\lambda = \frac{2\pi}{k}. \tag{3.14}$$

When z advances by $2\pi/k$, the cosine executes one complete cycle. As time passes, the entire wave train proceeds to the right, at speed c. At any fixed point z, the wave vibrates up and down, undergoing one full cycle in a *period*

$$T = \frac{2\pi}{kc}. \tag{3.15}$$

The *frequency* ν (number of oscillations per unit time) is

$$\nu = \frac{1}{T} = \frac{kc}{2\pi} = \frac{c}{\lambda}. \tag{3.16}$$

A more convenient unit is the *angular frequency* ω, which represents the number of radians swept out per unit time:

$$\omega = 2\pi\nu = kc. \tag{3.17}$$

Ordinarily, Eq. (3.13) is then written in terms of ω:

$$f(z,t) = A\cos(kz - \omega t + \delta). \tag{3.18}$$

Alternatively, one cause the complex notation

$$f(z,t) = \mathrm{Re}\left[A e^{i(kz - \omega t + \delta)} \right]. \tag{3.19}$$

where Re(ξ) denotes the real part of the complex number ξ. This invites us to introduce the complex wave function

$$\tilde{f}(z,t) \equiv \tilde{A} e^{i(kz - \omega t)}, \tag{3.20}$$

with the complex amplitude $\tilde{A} \equiv A e^{i\delta}$ absorbing the phase constant. The actual wave function is the real part of $\tilde{f}$:

$$f(z,t) = \mathrm{Re}\left[\tilde{f}(z,t) \right]. \tag{3.21}$$

The advantage of the complex notation is that exponentials are much easier to manipulate than sines and cosines.

Returning to the Eqs. (3.7) and (3.8), we confine our attention to waves of frequency ω. Since different frequencies in the visible range correspond to different *colors,* such waves are called *monochromatic*.

Suppose that the waves are traveling in the z direction and have *no x or y dependence*; these are called *plane* waves, because the fields are uniform over every plane perpendicular to the direction of propagation. We are interested, then, in fields of the form

$$\mathbf{E}(z,t) = \mathbf{E}_0 e^{i(kz - \omega t)}, \tag{3.22}$$

$$\mathbf{B}(z,t) = \mathbf{B}_0 e^{i(kz - \omega t)}, \tag{3.23}$$

where $\mathbf{E}_0$ and $\mathbf{B}_0$ are the complex amplitudes (the *physical* fields, of course, are the real parts of **E** and **B**). Substituting Eqs. (3.22) and (3.23) to Eqs. (3.7) and (3.8), respectively, we find that $\omega = kc$, as expected according to Eq. (3.17).

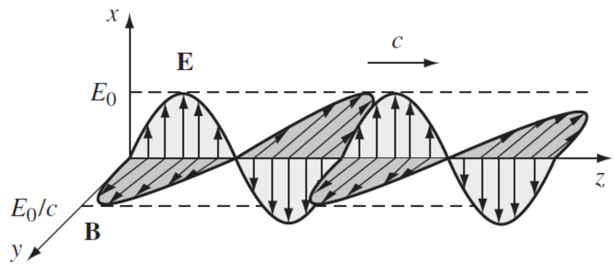

Fig. 3.3

Maxwell's equations impose extra constraints on $\mathbf{E}_0$ and $\mathbf{B}_0$. In particular, since $\nabla\cdot\mathbf{E}=0$ and $\nabla\cdot\mathbf{B}=0$, it follows that

$$\nabla\cdot\mathbf{E}=\left(\nabla\cdot\mathbf{E}_0\right)e^{i(kz-\omega t)}+\mathbf{E}_0\cdot\nabla e^{i(kz-\omega t)}=\mathbf{E}_0\cdot\hat{\mathbf{z}}e^{i(kz-\omega t)}ik=0\,, \tag{3.24}$$

and therefore

$$E_{0z}=B_{0z}=0\,. \tag{3.25}$$

That is, *electromagnetic waves are transverse:* the electric and magnetic fields are perpendicular to the direction of propagation. Moreover, Faraday's law, $\nabla\times\mathbf{E}=-\dfrac{\partial\mathbf{B}}{\partial t}$, implies a relation between the electric and magnetic amplitudes:

$$\nabla\times\mathbf{E}=-\mathbf{E}_0\times\nabla e^{i(kz-\omega t)}=-\begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ E_{0x} & E_{0y} & 0\\ 0 & 0 & ik\end{vmatrix}e^{i(kz-\omega t)}=\left(-\hat{\mathbf{x}}ikE_{0y}+\hat{\mathbf{y}}ikE_{0x}\right)e^{i(kz-\omega t)}= \\ =\left(\hat{\mathbf{x}}i\omega B_{0x}+\hat{\mathbf{y}}i\omega B_{0y}\right)e^{i(kz-\omega t)}, \tag{3.26}$$

which results in

$$-kE_{0y}=\omega B_{0x},\quad kE_{0x}=\omega B_{0y}, \tag{3.27}$$

or, more compactly:

$$\mathbf{B}_0=\frac{k}{\omega}\left(\hat{\mathbf{z}}\times\mathbf{E}_0\right)=\frac{1}{c}\left(\hat{\mathbf{z}}\times\mathbf{E}_0\right). \tag{3.28}$$

Evidently, **E** and **B** are *in phase* and *mutually perpendicular;* their (real) amplitudes are related by

$$B_0=\frac{k}{\omega}E_0=\frac{1}{c}E_0\,. \tag{3.29}$$

The fourth of Maxwell's equations, $\nabla\times\mathbf{B}=\mu_0\varepsilon_0\dfrac{\partial\mathbf{E}}{\partial t}$, does not yield an independent condition; it simply reproduces Eq. (3.26). Figure 3.3 depicts the plane wave propagating along the *z* direction.

There is nothing special about the *z* direction, of course—we can easily generalize to monochromatic plane waves traveling in an arbitrary direction. The notation is facilitated by the introduction of the *wave vector*, **k**, pointing in the direction of propagation, whose magnitude is the wave number *k.* The scalar product $\mathbf{k}\cdot\mathbf{r}$ is the appropriate generalization *of kz*, so

$$\mathbf{E}(\mathbf{r},t) = E_0 \mathbf{e} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \tag{3.30}$$

$$\mathbf{B}(\mathbf{r},t) = \frac{1}{c} E_0 \left(\hat{\mathbf{k}} \times \mathbf{e}\right) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} = \frac{1}{c}\left(\hat{\mathbf{k}} \times \mathbf{E}\right). \tag{3.31}$$

Where vector $\hat{\mathbf{k}} \equiv \mathbf{k}/k$ is the unit vector in the direction of propagation of the electromagnetic wave and **e** is the polarization vector. Because **E** is transverse, $\hat{\mathbf{k}} \cdot \mathbf{e} = 0$.

3.3 Linear and circular polarizations

The plane wave (3.30) and (3.31) is a wave with its electric field vector always in the direction **e**. Such a wave is said to be *linearly polarized* with polarization vector $\mathbf{e}_1 = \mathbf{e}$. Evidently there exists another wave which is linearly polarized with polarization vector $\mathbf{e}_2 \neq \mathbf{e}_1$ and is linearly independent of the first. Thus, the two waves are

$$\begin{aligned} \mathbf{E}_1(\mathbf{r},t) &= E_1 \mathbf{e}_1 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \\ \mathbf{E}_2(\mathbf{r},t) &= E_2 \mathbf{e}_2 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \end{aligned} \tag{3.32}$$

with $\mathbf{B}_{1,2} = \frac{1}{c}\left(\hat{\mathbf{k}} \times \mathbf{E}_{1,2}\right)$. They can be combined to give the most general form of a monochromatic plane wave propagating in the direction $\mathbf{k} = k\hat{\mathbf{k}}$:

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_1(\mathbf{r},t) + \mathbf{E}_2(\mathbf{r},t) = \left(E_1 \mathbf{e}_1 + E_2 \mathbf{e}_2\right) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}. \tag{3.33}$$

The amplitudes E_1 and E_2 are complex numbers, to allow the possibility of a phase difference between waves of different linear polarization.

If E_1 and E_2 have the *same phase*, wave (3.33) represents a *linearly polarized wave,* with its polarization vector making an angle $\tan\theta = E_2 / E_1$ with respect to $\mathbf{e}_1$ and a magnitude $E = \sqrt{E_2^{\,2} + E_1^{\,2}}$ as shown in Figure 3.4 (a).

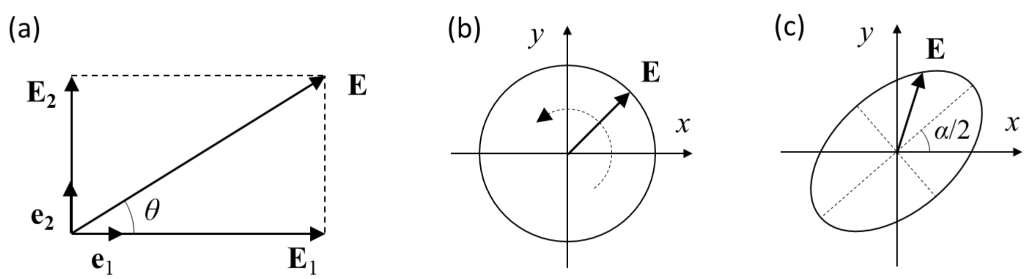

Fig. 3.4 Electric field of a linearly (a), circularly (b), and elliptically (c) polarized wave.

If E_1 and E_2 have *different phases,* the wave (3.33) is *elliptically polarized.* To understand what this means, consider the simplest case of *circular polarization.* In this case, E_1 and E_2 have the same magnitudes, but differ in phase by 90°. The wave (3.33) becomes:

$$\mathbf{E}(\mathbf{r},t) = E_0 \left(\mathbf{e}_1 \pm i\mathbf{e}_2\right) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \tag{3.34}$$

with E_0 the common real amplitude. We imagine axes chosen so that the wave is propagating in the positive *z* direction, while $\mathbf{e}_1$ and $\mathbf{e}_2$ are in the *x* and y directions, respectively. Then the components of the actual electric field, obtained by taking the real part of (3.34), are

$$E_x(\mathbf{r},t) = E_0 \cos(kz - \omega t),$$
$$E_y(\mathbf{r},t) = \mp E_0 \sin(kz - \omega t). \tag{3.35}$$

At a *fixed point in space,* the fields (3.35) are such that the electric vector is constant in magnitude, but sweeps around in a circle at a frequency ω, as shown in Figure 3.4 (b). For the upper sign $(\mathbf{e}_1 + i\mathbf{e}_2)$, the rotation is counterclockwise when the observer is facing into the oncoming wave. This wave is called *left circularly polarized* in optics. In the terminology of modern physics, however, such a wave is said to have *positive helicity.* The latter description seems more appropriate because such a wave has a positive projection of angular momentum on the z axis. For the lower sign $(\mathbf{e}_1 - i\mathbf{e}_2)$, the rotation of $\mathbf{E}$ is clock-wise when looking into the wave; the wave is *right circularly polarized* (optics): it has *negative helicity.*

The two circularly polarized waves (3.35) form an equally acceptable set of basic fields for description of a general state of polarization. If we introduce the complex orthogonal unit vectors:

$$\mathbf{e}_{\pm} = \frac{1}{\sqrt{2}}(\mathbf{e}_1 \pm i\mathbf{e}_2), \tag{3.36}$$

then a general representation, equivalent to (3.33), is

$$\mathbf{E}(\mathbf{r},t) = (E_+\mathbf{e}_+ + E_-\mathbf{e}_-)e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \tag{3.37}$$

where E_+ and E_- are complex amplitudes. If E_+ and E_- have different magnitudes, but the same phase, (3.37) represents an elliptically polarized wave with principal axes of the ellipse in the directions of $\mathbf{e}_1$ and $\mathbf{e}_2$. The ratio of semimajor to semiminor axis is $|(1+r)/(1-r)|$, where $r = E_- / E_+$. If the amplitudes have a phase difference between them, $E_- / E_+ = re^{i\alpha}$, then the ellipse traced out by the $\mathbf{E}$ vector has its axes rotated by an angle $\alpha/2$. Figure 3.4 (c) shows the general case of elliptical polarization and the ellipse traced out by $\mathbf{E}$ at a given point in space. For $r = \pm 1$ we get back a linearly polarized wave.

3.4 Energy and momentum of electromagnetic waves

Energy per unit volume stored in electromagnetic fields is

$$u = \frac{1}{2}\left(\varepsilon_0 E^2 + \frac{1}{\mu_0} B^2\right). \tag{3.38}$$

Note that here E and B are real quantities (real pars of complex quantities used in the previous paragraph). In case of monochromatic plane wave

$$B_0^{\ 2} = \frac{1}{c^2} E_0^{\ 2} = \mu_0 \varepsilon_0 E_0^{\ 2}, \tag{3.39}$$

so the magnetic and electric contributions are equal resulting in

$$u = \varepsilon_0 E^2 = \varepsilon_0 E_0^{\ 2} \cos^2(kz - \omega t + \delta). \tag{3.40}$$

As the wave travels, it carriers this energy alone with it. The energy flux density (energy per unit area, per unit time) transported by the fields is given by the Poynting vector:

$$\mathbf{S} = \frac{1}{\mu_0}(\mathbf{E} \times \mathbf{B}). \tag{3.41}$$

For monochromatic plane waves propagating in the z direction,

$$\mathbf{S} = \frac{1}{\mu_0}\frac{E^2}{c}\hat{\mathbf{z}} = c\varepsilon_0 E^2\hat{\mathbf{z}} = c\varepsilon_0 {E_0}^2 \cos^2(kz - \omega t + \delta)\hat{\mathbf{z}} = cu\hat{\mathbf{z}}\,. \tag{3.42}$$

Notice that **S** is the energy density (u) times the velocity of the waves (c) – as it *should be.* In time Δt, length $c\Delta t$ passes through area A (Fig. 3.5), carrying energy $uAc\Delta t$. The energy per unit time, per unit area, transported by the wave is therefore uc.

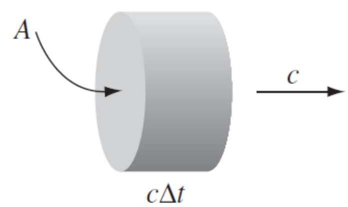

Fig. 3.5

Electromagnetic fields not only carry *energy,* they also carry *momentum.* The momentum density stored in the fields is

$$\mathbf{p}_{em} = \frac{1}{c^2}\mathbf{S}\,. \tag{3.43}$$

For monochromatic plane waves, then,

$$\mathbf{p}_{em} = \frac{1}{c}\varepsilon_0 {E_0}^2 \cos^2(kz - \omega t + \delta)\hat{\mathbf{z}} = \frac{1}{c}u\hat{\mathbf{z}}\,. \tag{3.44}$$

In the case of *light*, the wavelength is so short ($\sim 5\times 10^{-7}$ m), and the period so brief ($\sim 10^{-15}$ s), that any macroscopic measurement will encompass many cycles. Typically, therefore, we are not interested in the fluctuating cosine-squared term in the energy and momentum densities; all we want is the *average* value. Now, the average of cosine-squared over a complete cycle is ½ so

$$\langle u \rangle = \frac{1}{2}\varepsilon_0 E_0^2\,, \tag{3.45}$$

$$\langle \mathbf{S} \rangle = \frac{1}{2}\varepsilon_0 c {E_0}^2\hat{\mathbf{z}}\,, \tag{3.46}$$

$$\langle \mathbf{p}_{em} \rangle = \frac{1}{2c}\varepsilon_0 {E_0}^2\hat{\mathbf{z}}\,. \tag{3.47}$$

We use brackets, $\langle\ \rangle$, to denote the time average over a complete cycle. The average power per unit area transported by an electromagnetic wave is called the *intensity*:

$$I \equiv \langle S \rangle = \frac{1}{2}\varepsilon_0 c {E_0}^2\,. \tag{3.48}$$

When light falls (at normal incidence) on a perfect absorber, it delivers its momentum to the surface. In a time Δt, the momentum transfer is (Fig. 3.5) $\Delta\mathbf{p} = A\langle \mathbf{p}_{em}\rangle c\Delta t$, so the *radiation pressure*, defined as the average force per unit area, is

$$P = \frac{1}{A}\frac{\Delta p}{\Delta t} = \frac{1}{2}\varepsilon_0 {E_0}^2 = \frac{I}{c}\,. \tag{3.49}$$

We can account for this pressure qualitatively, as follows: The electric field (Eq. (3.22)) drives charges in the x direction, and the magnetic field then exerts on them a force $q(\mathbf{v}\times\mathbf{B})$ in the z direction. The net force on all the charges in the surface produces the pressure.

3.5 Electromagnetic waves in matter

In regions of matter where there are no *free* charges and *free* currents, Maxwell's equations are

$$\nabla\cdot\mathbf{D}=0, \tag{3.50}$$

$$\nabla\cdot\mathbf{B}=0, \tag{3.51}$$

$$\nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}, \tag{3.52}$$

$$\nabla\times\mathbf{H}=\frac{\partial\mathbf{D}}{\partial t}. \tag{3.53}$$

If the matter is *linear*

$$\mathbf{D}=\varepsilon\mathbf{E},\quad \mathbf{H}=\frac{1}{\mu}\mathbf{B} \tag{3.54}$$

and homogeneous (μ and ε are constants in space), Maxwell's equations reduce to

$$\nabla\cdot\mathbf{E}=0 \tag{3.55}$$

$$\nabla\cdot\mathbf{B}=0 \tag{3.56}$$

$$\nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t} \tag{3.57}$$

$$\nabla\times\mathbf{B}=\varepsilon\mu\frac{\partial\mathbf{E}}{\partial t}, \tag{3.58}$$

which are different from Maxwell's equations in vacuum only by the replacement of $\varepsilon_0\mu_0$ to $\varepsilon\mu$. Evidently electromagnetic waves propagate trough a linear homogeneous medium at a speed

$$v=\frac{1}{\sqrt{\varepsilon\mu}}=\frac{c}{n}, \tag{3.59}$$

where

$$n\equiv\sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}}, \tag{3.60}$$

is the *index of refraction*. For most (nonmagnetic) materials, μ is very close to μ_0 so

$$n\cong\sqrt{\varepsilon_r}, \tag{3.61}$$

where $\varepsilon_r\equiv\varepsilon/\varepsilon_0$ is the dielectric constant. Since ε_r is always greater than 1, light travels more slowly through matter than trough vacuum.

All our previous results carry over, with the simple transcription $\varepsilon_0\to\varepsilon$, $\mu_0\to\mu$, and hence $c\to v$. The energy density is

$$u=\frac{1}{2}\left(\varepsilon E^2+\frac{1}{\mu}B^2\right), \tag{3.62}$$

and the Poynting vector is

$$\mathbf{S} = \frac{1}{\mu}(\mathbf{E}\times\mathbf{B}) = (\mathbf{E}\times\mathbf{H}). \tag{3.63}$$

For monochromatic plane waves, the frequency and wave number are related by $\omega = kv$ (Eq.(3.17)), the amplitude of **B** is $1/v$ times the amplitude of **E** (Eq. (3.29)), and the intensity is

$$I = \frac{1}{2}\varepsilon v E_0^{\,2}. \tag{3.64}$$

3.6 Reflection and transmission of EM waves at normal incidence

If a wave passes from one transparent medium into another, there is a reflected wave and a transmitted wave. The details depend on the exact nature of the electrodynamic boundary conditions which are

$$\varepsilon_1 E_1^{\perp} = \varepsilon_2 E_2^{\perp}, \tag{3.65}$$

$$\mathbf{E}_1^{\parallel} = \mathbf{E}_2^{\parallel}, \tag{3.66}$$

$$B_1^{\perp} = B_2^{\perp}, \tag{3.67}$$

$$\frac{1}{\mu_1}\mathbf{B}_1^{\parallel} = \frac{1}{\mu_2}\mathbf{B}_2^{\parallel}. \tag{3.68}$$

Here we assume that the two media which are characterized by indices 1 and 2 have different electric permittivities ($\varepsilon_1 \neq \varepsilon_2$) and magnetic permeabilities ($\mu_1 \neq \mu_2$). These equations relate the electric and magnetic fields just to the left and just to the right of the interface between two linear media. Now use them to deduce the laws governing reflection and refraction of electromagnetic waves at normal incidence.

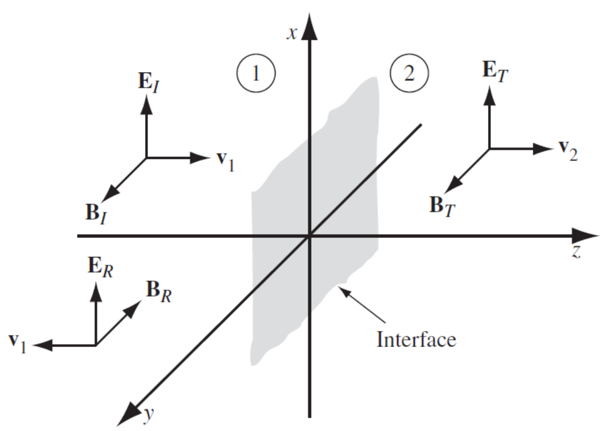

Fig. 3.6

Suppose the *xy* plane forms the boundary between two linear media. A plane wave of frequency ω, traveling in the z direction and polarized in the x direction, approaches the interface from the left as is shown in Figure 3.6:

$$\mathbf{E}_I(z,t) = E_{0I} e^{i(k_1 z - \omega t)}\hat{\mathbf{x}}, \tag{3.69}$$

$$\mathbf{B}_I(z,t) = \frac{1}{v_1} E_{0I} e^{i(k_1 z - \omega t)}\hat{\mathbf{y}}. \tag{3.70}$$

It gives rise to the reflected wave

$$\mathbf{E}_R(z,t) = E_{0R} e^{i(-k_1 z - \omega t)} \hat{\mathbf{x}} , \tag{3.71}$$

$$\mathbf{B}_R(z,t) = -\frac{1}{v_1} E_{0R} e^{i(-k_1 z - \omega t)} \hat{\mathbf{y}} , \tag{3.72}$$

which travels back to the left in medium (1), and a transmitted wave

$$\mathbf{E}_T(z,t) = E_{0T} e^{i(k_2 z - \omega t)} \hat{\mathbf{x}} \tag{3.73}$$

$$\mathbf{B}_T(z,t) = \frac{1}{v_2} E_{0I} e^{i(k_2 z - \omega t)} \hat{\mathbf{y}} , \tag{3.74}$$

which continues on the right in medium (2). Note that the minus sign in $\mathbf{B}_r$ is required by Eq. (3.31). Note that the reflected and transmitted waves are also polarized in the x direction as the incident wave, as required by the boundary conditions (3.65)−(3.68) (HW#3, Problem 1).

At $z = 0$, the combined fields on the left, $\mathbf{E}_I + \mathbf{E}_R$ and $\mathbf{B}_I + \mathbf{B}_R$, must join the fields on the right, $\mathbf{E}_T$ and $\mathbf{B}_T$, in accordance with the boundary conditions. In this case there are no components perpendicular to the surface, so Eqs. (3.65) and (3.67) are trivial. However, Eqs. (3.66) and (3.68) require that

$$E_{0I} + E_{0R} = E_{0T} . \tag{3.75}$$

$$\frac{1}{\mu_1}\left(\frac{1}{v_1} E_{0I} - \frac{1}{v_1} E_{0R}\right) = \frac{1}{\mu_1}\left(\frac{1}{v_2} E_{0T}\right). \tag{3.76}$$

Eq. (3.76) can be rewritten as follows

$$E_{0I} - E_{0R} = \beta E_{0T} , \tag{3.77}$$

where

$$\beta \equiv \frac{\mu_1 v_1}{\mu_2 v_2} = \frac{\mu_1 n_2}{\mu_2 n_1} . \tag{3.78}$$

Eqs. (3.75) and (3.77) are easily solved for the outgoing amplitudes, in terms of the incident amplitude:

$$E_{0R} = \frac{1-\beta}{1+\beta} E_{0I} , \quad E_{0T} = \frac{2}{1+\beta} E_{0I} . \tag{3.79}$$

If the permeabilities μ are close to their values in vacuum, i.e. $\mu_1 = \mu_2 = \mu_0$, (as is the case for most media), then $\beta = n_2 / n_1$, and we have

$$E_{0R} = \frac{n_1 - n_2}{n_1 + n_2} E_{0I} , \quad E_{0T} = \frac{2n_1}{n_1 + n_2} E_{0I} . \tag{3.80}$$

In order to calculate the fraction of energy which is transmitted and reflected we need to find the intensity (average power per unit area) which according to Eq. (3.48) is given by $I = \frac{1}{2}\varepsilon v E_0^{\,2}$. The ratio of the reflected intensity to the incident intensity is therefore

$$R \equiv \frac{I_R}{I_I} = \left|\frac{E_{0R}}{E_{0I}}\right|^2 = \left|\frac{n_1 - n_2}{n_1 + n_2}\right|^2 , \tag{3.81}$$

whereas the ratio of the transmitted intensity to the incident intensity is

$$T \equiv \frac{I_T}{I_I} = \frac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\left|\frac{E_{0T}}{E_{0I}}\right|^2 = \frac{4n_1 n_2}{(n_1+n_2)^2}. \tag{3.82}$$

Here, R is called the *reflection coefficient* and T the *transmission coefficient*; they measure the fraction of the incident energy that is reflected and transmitted, respectively. Notice that

$$R + T = 1, \tag{3.83}$$

as conservation of energy, of course, requires. For instance, when light passes from air ($n_1 = 1$) into glass ($n_2 = 1.5$), $R = 0.04$ and $T = 0.96$.

3.7 Reflection and transmission at oblique incidence

Now we consider the more general case of *oblique* incidence, in which the incoming wave meets the boundary at an arbitrary angle θ_i (Fig. 3.7). The normal incidence considered in the previous section is really just a special case of oblique incidence, with $\theta_i = 0$.

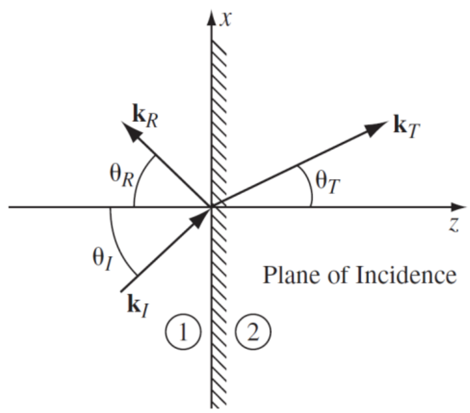

Fig. 3.7

Suppose, then, that a monochromatic plane wave

$$\mathbf{E}_I(\mathbf{r},t) = \mathbf{E}_{0I} e^{i(\mathbf{k}_I\cdot\mathbf{r}-\omega t)}; \quad \mathbf{B}_I(\mathbf{r},t) = \frac{1}{v_1}\left(\hat{\mathbf{k}}_I \times \mathbf{E}_I\right) \tag{3.84}$$

approaches from the left, giving rise to a reflected wave,

$$\mathbf{E}_R(\mathbf{r},t) = \mathbf{E}_{0R} e^{i(\mathbf{k}_R\cdot\mathbf{r}-\omega t)}; \quad \mathbf{B}_R(\mathbf{r},t) = \frac{1}{v_1}\left(\hat{\mathbf{k}}_R \times \mathbf{E}_R\right) \tag{3.85}$$

and a transmitted wave

$$\mathbf{E}_T(\mathbf{r},t) = \mathbf{E}_{0T} e^{i(\mathbf{k}_T\cdot\mathbf{r}-\omega t)}; \quad \mathbf{B}_T(\mathbf{r},t) = \frac{1}{v_2}\left(\hat{\mathbf{k}}_T \times \mathbf{E}_T\right). \tag{3.86}$$

All three waves have the same *frequency* ω that is determined once and for all at the source. The three wave numbers are related so that

$$k_I v_1 = k_R v_1 = k_T v_2 = \omega, \quad \text{or} \quad k_I = k_R = \frac{v_2}{v_1} k_T = \frac{n_1}{n_2} k_T. \tag{3.87}$$

The combined fields in medium (1), $\mathbf{E}_i + \mathbf{E}_r$ and $\mathbf{B}_i + \mathbf{B}_r$, must now be joined to the fields $\mathbf{E}_t$ and $\mathbf{B}_t$ in medium (2), using the boundary conditions (3.65) – (3.68). These all share the generic structure

$$(\ldots)e^{i(\mathbf{k}_I\cdot\mathbf{r}-\omega t)} + (\ldots)e^{i(\mathbf{k}_R\cdot\mathbf{r}-\omega t)} = (\ldots)e^{i(\mathbf{k}_T\cdot\mathbf{r}-\omega t)}. \tag{3.88}$$

The important thing to notice is that the *x, y,* and *t* dependence is confined to the exponents. Because the boundary conditions must hold at all points on the plane, and for all times, these exponential factors must be equal at $z = 0$. The time factors are already equal. As for the spatial terms, evidently at $z = 0$

$$\mathbf{k}_I \cdot \mathbf{r} = \mathbf{k}_R \cdot \mathbf{r} = \mathbf{k}_T \cdot \mathbf{r} \tag{3.89}$$

or, more explicitly,

$$(k_I)_x x + (k_I)_y y = (k_R)_x x + (k_R)_y y = (k_T)_x x + (k_T)_y y \tag{3.90}$$

for all *x* and *y*. But Eq. (3.90) can *only* hold if the components are separately equal. For $x = 0$, we obtain

$$(k_I)_y = (k_R)_y = (k_T)_y, \tag{3.91}$$

and for $y = 0$, we have

$$(k_I)_x = (k_R)_x = (k_T)_x. \tag{3.92}$$

We can always choose axes in such a way that $\mathbf{k}_I$ lies in the *xz* plane so that $(k_I)_y = 0$. Eq. (3.91) then requires $(k_R)_y = (k_T)_y = 0$. This leads to

First Law: The incident, reflected, and transmitted wave vectors form a plane (called the *plane of incidence*), which also includes the normal to the surface (here, the *z* axis).

Meanwhile, Eq. (3.92) implies that

$$k_I \sin\theta_I = k_R \sin\theta_R = k_T \sin\theta_T, \tag{3.93}$$

where θ_I is the *angle of incidence*, θ_R is the *angle of reflection*, and θ_T is the angle of transmission, more commonly known as the *angle of refraction*, all of them measured with respect to the normal (Fig. 2.7). In view of Eq. (3.93), then, we have

Second Law: The angle of incidence is equal to the angle of reflection,

$$\theta_I = \theta_R. \tag{3.94}$$

This is the *law of reflection*.

As for the transmitted angle, there is

Third Law:

$$\frac{\sin\theta_T}{\sin\theta_I} = \frac{n_1}{n_2}. \tag{3.95}$$

This is the *law of refraction* known as *Snell's law*.

These are the three fundamental laws of geometrical optics.

Now that we have taken care of the exponential factors – they cancel, given Eq. (3.89) – the boundary conditions (3.65) – (3.68) become:

$$\varepsilon_1 \left(\mathbf{E}_{0I} + \mathbf{E}_{0R}\right)_z = \varepsilon_2 \left(\mathbf{E}_{0T}\right)_z, \tag{3.96}$$

$$\left(\mathbf{B}_{0I} + \mathbf{B}_{0R}\right)_z = \left(\mathbf{B}_{0T}\right)_z, \tag{3.97}$$

$$\left(\mathbf{E}_{0I} + \mathbf{E}_{0R}\right)_{x,y} = \left(\mathbf{E}_{0T}\right)_{x,y}, \tag{3.98}$$

$$\frac{1}{\mu_1}\left(\mathbf{B}_{0I} + \mathbf{B}_{0R}\right)_{x,y} = \frac{1}{\mu_2}\left(\mathbf{B}_{0T}\right)_{x,y}, \tag{3.99}$$

where $\mathbf{B}_0(\mathbf{r},t) = \frac{1}{v}\left(\hat{\mathbf{k}} \times \mathbf{E}_0\right)$ in each case. The last two represent *pairs* of equations, one for the x-component and one for the y-component.

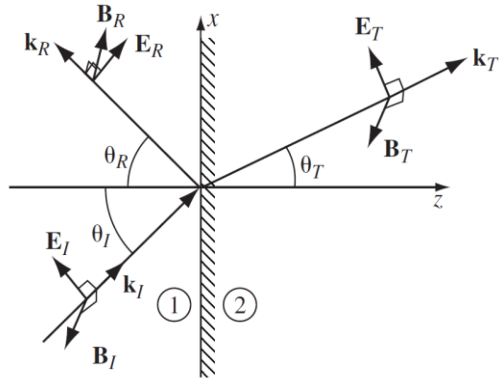

Fig. 3.8

Suppose that the polarization of the incident wave is *parallel* to the plane of incidence (the xz plane in Fig. 3.8); then the reflected and transmitted waves are also polarized in this plane. Then Eq. (3.96) reads

$$\varepsilon_1\left(-E_{0I}\sin\theta_I + E_{0R}\sin\theta_R\right) = \varepsilon_2\left(-E_{0T}\sin\theta_T\right). \tag{3.100}$$

Eq. (3.97) adds nothing (0 = 0), since the magnetic fields have no z components; Eq. (3.98) becomes

$$E_{0I}\cos\theta_I + E_{0R}\cos\theta_R = E_{0T}\cos\theta_T\,; \tag{3.101}$$

and Eq. (3.99) gives

$$\frac{1}{\mu_1 v_1}\left(E_{0I} - E_{0\backslash R}\right) = \frac{1}{\mu_2 v_2}E_{0T}\,. \tag{3.102}$$

Given the laws of reflection and refraction [(3.94) and (3.95)], Eqs. (3.100) and (3.102) reduce to

$$E_{0I} - E_{0R} = \beta E_{0T}\,, \tag{3.103}$$

where as before

$$\beta \equiv \frac{\mu_1 v_1}{\mu_2 v_2} = \frac{\mu_1 n_2}{\mu_2 n_1}. \tag{3.104}$$

Eq. (3.101) says

$$E_{0I} + E_{0R} = \alpha E_{0T}\,, \tag{3.105}$$

where

$$\alpha \equiv \frac{\cos\theta_T}{\cos\theta_I}. \tag{3.106}$$

Solving Eqs. (3.103) and (3.105) for the reflected and transmitted amplitudes, we obtain

$$E_{0R} = \frac{\alpha-\beta}{\alpha+\beta}E_{0I}; \quad E_{0T} = \frac{2}{\alpha+\beta}E_{0I}\,. \tag{3.107}$$

These are known as *Fresnel's equations*, for the case of polarization in the plane of incidence. There are two other Fresnel's equations, giving the reflected and transmitted amplitudes when the polarization is *perpendicular* to the plane of incidence. These equations you are asked to derive at home (HW#3, Problem 4).

Notice that the transmitted wave is always *in phase* with the incident one; the reflected wave is either in phase ("right side up"), if $\alpha > \beta$, or 180° out of phase ("upside down"), if $\alpha < \beta$.

The amplitudes of the transmitted and reflected waves depend on the angle of incidence, because α is a function of θ_I:

$$\alpha = \frac{\sqrt{1-\sin^2\theta_T}}{\cos\theta_I} = \frac{\sqrt{1-(n_1/n_2)^2\sin^2\theta_I}}{\cos\theta_I}. \tag{3.108}$$

In the case of normal incidence (θ_I = 0), α = 1, and we recover Eq. (3.80). At grazing incidence (θ_I = 90°), α diverges, and the wave is totally reflected. Interestingly, there is an intermediate angle, θ_B (called *Brewster's angle*), at which the reflected wave is completely extinguished. According to Eq. (3.107), this occurs when $\alpha = \beta$ or

$$\sin^2\theta_B = \frac{1-\beta^2}{(n_1/n_2)^2-\beta^2}. \tag{3.109}$$

For the typical case of $\mu_1 = \mu_2$ and $\beta = n_2/n_1$, we obtain $\sin^2\theta_B = \beta^2/(1+\beta^2)$ so that

$$\tan\theta_B = \beta = \frac{n_2}{n_1}. \tag{3.110}$$

Figure 3.9 shows a plot of the transmitted and reflected amplitudes as functions of θ_i, for light incident on glass (n_2 = 1.5) from air (n_1 = 1). On the graph, a *negative* number indicates that the wave is 180° out of phase with the incident beam – the amplitude itself is the absolute value.

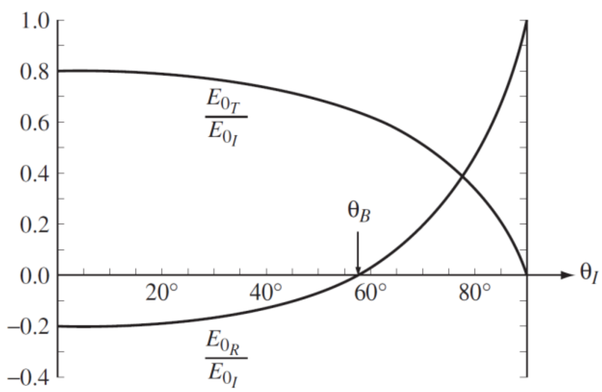

Fig. 3.9

If the wave is polarized *perpendicular* to the plane of incidence there is no Brewster's angle for *any n* (HW#3, Problem 4). Therefore, if a plane wave of mixed polarization is incident on a plane interface at the Brewster angle, the reflected radiation is *completely plane-polarized* with polarization vector *perpendicular* to the plane of incidence. This behavior can be utilized to produce beams of plane-polarized light but is not as efficient as other means employing anisotropic properties of some dielectric media. Even if the unpolarized wave is reflected at angles other than the Brewster angle, there is a tendency for the reflected wave to be predominantly polarized perpendicular to the plane of incidence.

The power per unit area striking the interface is $\mathbf{S}\cdot\hat{\mathbf{z}}$. Thus, the incident intensity is

$$I_I = \frac{1}{2}\varepsilon_1 v_1 E_{0I}^2 \cos\theta_I, \tag{3.111}$$

while the reflected and transmitted intensities are

$$I_R = \frac{1}{2}\varepsilon_1 v_1 E_{0R}^2 \cos\theta_R ,$$

$$I_T = \frac{1}{2}\varepsilon_2 v_2 E_{0T}^2 \cos\theta_T . \tag{3.112}$$

The cosines are there because the intensities represent the average power per unit area of *interface,* and the interface is at an angle to the wave front. The reflection and transmission coefficients for waves polarized parallel to the plane of incidence are

$$R \equiv \frac{I_R}{I_I} = \left|\frac{E_{0R}}{E_{0I}}\right|^2 = \left|\frac{\alpha-\beta}{\alpha+\beta}\right|^2 , \tag{3.113}$$

$$T \equiv \frac{I_T}{I_I} = \frac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\left|\frac{E_{0R}}{E_{0I}}\right|^2 \frac{\cos\theta_T}{\cos\theta_I} = \alpha\beta\left(\frac{2}{\alpha+\beta}\right)^2 . \tag{3.114}$$

They are plotted as functions of the angle of incidence in Fig. 3.10 (for the air/glass interface). R is the fraction of the incident energy that is reflected – naturally, it goes to zero at Brewster's angle; T is the fraction transmitted – it goes to 1 at θ_B. Note that $R + T = 1$, as required by conservation of energy: the energy per unit time *reaching* a particular patch of area on the surface is equal to the energy per unit time *leaving* the patch.

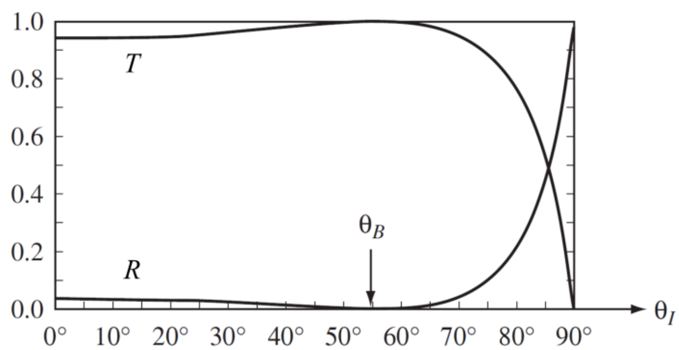

Fig. 3.10

3.8 Total internal reflection

There is another important phenomenon which is called *total internal reflection.* The word "internal" implies that the incident and reflected waves are in a medium of larger index of refraction than the refracted wave ($n_1 > n_2$). Snell's law (7.36) shows that, if $n_1 > n_2$, then $\theta_T > \theta_I$. Consequently, there is *a critical angle* when $\theta_I = \theta_c$ at which $\theta_T = \pi/2$,

$$\sin\theta_c = \frac{n_2}{n_1} , \tag{3.115}$$

i.e., the refracted wave is propagated parallel to the surface. There is no energy flow across the surface. Hence at that angle of incidence there must be total reflection (on which light pipes and fiber optics are based).

What happens if $\theta_I > \theta_c$? To answer this we first note that, for $\theta_I > \theta_c$,

$$\sin\theta_T = \frac{n_1}{n_2}\sin\theta_I = \frac{\sin\theta_I}{\sin\theta_c} > 1 . \tag{3.116}$$

This means that θ_T is a complex angle with a purely imaginary cosine:

$$\cos\theta_T = i\sqrt{\sin^2\theta_T - 1} = i\sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_I - 1}\,. \tag{3.117}$$

Obviously, θ_T can no longer be interpreted as an angle. The meaning of these complex quantities becomes clear when we consider the transmitted wave (3.86)

$$\mathbf{E}_T(\mathbf{r},t) = \mathbf{E}_{0T}e^{i(\mathbf{k}_T\cdot\mathbf{r}-\omega t)}\,. \tag{3.118}$$

The propagation factor under the exponent is

$$\begin{aligned}\mathbf{k}_T\cdot\mathbf{r} &= k_2\left(\sin\theta_T\hat{\mathbf{x}} + \cos\theta_T\hat{\mathbf{z}}\right)\cdot\left(x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}\right) = k_2\left(x\sin\theta_T + z\cos\theta_T\right) = \\ &= k_2 x\sin\theta_T + ik_2 z\sqrt{\sin^2\theta_T - 1} = k'x + ik''z,\end{aligned} \tag{3.119}$$

where

$$k' = k_2\sin\theta_T = \left(\frac{\omega n_2}{c}\right)\frac{n_1}{n_2}\sin\theta_I = \left(\frac{\omega n_1}{c}\right)\sin\theta_I\,, \tag{3.120}$$

$$k'' = \left(\frac{\omega n_2}{c}\right)\sqrt{\left(\frac{n_1}{n_2}\right)^2\sin^2\theta_I - 1} = \frac{\omega}{c}\sqrt{n_1^2\sin^2\theta_I - n_2^2}\,. \tag{3.121}$$

The transmitted wave is therefore

$$\mathbf{E}_T(\mathbf{r},t) = \mathbf{E}_{0T}e^{-k''z}e^{i(k''x-\omega t)}\,. \tag{3.122}$$

This shows that, for $\theta_I > \theta_c$, the transmitted wave is propagated only parallel to the surface and is attenuated exponentially beyond the interface. Such a wave is known as the *evanescent wave*. The attenuation occurs within a very few wavelengths of the boundary, except for $\theta_I \cong \theta_c$. Figure 3.11 shows the reflection coefficient as a function of the incident angle for || and ⊥ polarized plane waves propagating from water (*n* = 1.5) to air (*n* = 1).

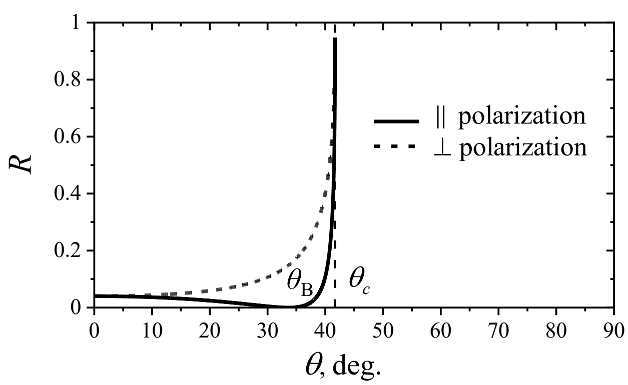

Fig. 3.11

So that for $\theta_I > \theta_c$, the reflection coefficient is 1 can be easily seen from Eq. (3.113). Taking into account that β (3.104) is real and α (3.108) is purely imaginary, i.e. we can write $\alpha = ia$ where a is real, we find for the reflection coefficient

$$R = \left|\frac{\alpha - \beta}{\alpha + \beta}\right|^2 = \left|\frac{ia - \beta}{ia + \beta}\right|^2 = \frac{(ia - \beta)}{(ia + \beta)}\frac{(-ia - \beta)}{(-ia + \beta)} = 1\,. \tag{3.123}$$