

Chapter 2

Electromagnetic Theory

The treatment of absorption and scattering of light by small particles is a problem in electromagnetic theory, a subdivision of which is optics. Various aspects of this theory are found scattered throughout books on electricity and magnetism, optics, and polarization of light. To rely on such existing books as the basis for our development might, however, result in a confusing tangle of conflicting assumptions, notation, and conventions. Also, it is convenient to gather together in one place the mathematical and physical apparatus that underlies succeeding chapters. Thus, we present in this chapter our version of those concepts and equations from electromagnetic theory that are germane to the subject at hand; we shall have need of them throughout the remainder of the book.

2.1 FIELD VECTORS AND THE MAXWELL EQUATIONS

As stated in the introductory chapter, we adopt a macroscopic approach to the problem of determining absorption and scattering of electromagnetic waves by particles. Therefore, the logical point of departure is the Maxwell equations for the macroscopic electromagnetic field at interior points in matter, which in SI units may be written

$$\nabla \cdot \mathbf{D} = \rho_{\text{F}}, \quad (2.1)$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \quad (2.2)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (2.3)$$

$$\nabla \times \mathbf{H} = \mathbf{J}_{\text{F}} + \frac{\partial \mathbf{D}}{\partial t}, \quad (2.4)$$

where \mathbf{E} is the electric field and \mathbf{B} the magnetic induction. The electric displacement \mathbf{D} and magnetic field \mathbf{H} are defined by

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (2.5)$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M}, \quad (2.6)$$

where \mathbf{P} is the electric polarization (average electric dipole moment per unit volume), \mathbf{M} the magnetization (average magnetic dipole moment per unit volume), ϵ_0 the permittivity, and μ_0 the permeability of free space. Implicit in (2.5) is the assumption that quadrupole and higher moments are negligible compared with the dipole moment. In free space the polarization and magnetization vanish identically. The charge density ρ_f and current density \mathbf{J}_f are associated with so-called “free” charges. The terms “free” and “bound” are sometimes set in quotation marks, which indicates that they are slightly suspect. Indeed, Purcell (1963, pp. 342–347) has simply but convincingly shown that it is not always possible to unambiguously distinguish between free and bound charges in matter. Nevertheless, we shall assume in time-honored fashion that the ambiguity in the meanings of free and bound leads to no observable consequences in the problems with which we shall be concerned.

Although there appears to be nearly universal agreement about the *microscopic* Maxwell equations, it is not trivial to provide a derivation of the macroscopic equations that will satisfy everyone. The process of so doing was begun by Lorentz some 100 years ago, and new derivations appear regularly in the literature, each of which claims to be more general, freer from ambiguity, and more logically consistent than its predecessors. Without wishing to enter the lists of combatants, we direct the interested reader to Russakoff (1970) and Robinson (1973) for further discussion of the transition from the microscopic to the macroscopic field equations. We accept, therefore, (2.1)–(2.6) as *the* macroscopic field equations without further comment; any attempt to justify this choice rigorously and to define all terms precisely would lead to a volume that would dwarf the present work.

Equations (2.1)–(2.6) are not sufficient in themselves; they must be supplemented with *constitutive relations*, which are assumed to have the form

$$\mathbf{J}_f = \sigma \mathbf{E}, \quad (2.7)$$

$$\mathbf{B} = \mu \mathbf{H}, \quad (2.8)$$

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}, \quad (2.9)$$

where σ is the *conductivity*, μ the *permeability*, and χ the electric *susceptibility*. The *phenomenological coefficients* σ , μ , and χ depend on the **medium** under consideration, but will be assumed to be independent of the fields (**the medium is linear**), independent of position (**the medium is homogeneous**), and independent of direction (**the medium is isotropic**). There are many **classes** of materials for which these assumptions are *not* valid. Equations (2.7)–(2.9) are not universal laws of nature; they merely describe a particular **class** of materials which, fortunately, has a large number of members.

We shall not, however, assume that the *phenomenological coefficients* are independent of *frequency*. To provide some insight into this assertion, we must digress briefly. Consider, for example, the susceptibility. The polarization \mathbf{P} is

the average dipole moment per unit volume of the medium, that is, the vector sum of the dipole moments in a unit volume. An isolated sample of matter (excluding electrets) is unpolarized ($\mathbf{P} = 0$). But when placed in an external field, which we may assume to be time harmonic, it becomes polarized: the electric field induces a net dipole moment. For a linear homogeneous isotropic medium, the relation (2.9) between \mathbf{P} and \mathbf{E} indicates that χ is a measure of how easily the material can be polarized; it represents the response of the material to the field \mathbf{E} . In particular, χ may be interpreted as the amplitude of the response to a unit field. Now it is well known (to those who know it well) that the response of a mechanical system to a periodic driving force is a sensitive function of the frequency. Therefore, by analogy, it is plausible that χ is a function of the frequency. The frequency dependence of χ is discussed in much greater detail in Chapters 9 and 10.

2.2 TIME-HARMONIC FIELDS

The general time-harmonic field \mathbf{F} has the form

$$\mathbf{F} = \mathbf{A} \cos \omega t + \mathbf{B} \sin \omega t, \quad (2.10)$$

where ω is the *angular frequency*. The *real* vector fields \mathbf{A} and \mathbf{B} are independent of time but may depend on position. We note that \mathbf{F} may be written as the real part of a complex vector: $\mathbf{F} = \text{Re}(\mathbf{F}_c)$, where

$$\mathbf{F}_c = \mathbf{C} \exp(-i\omega t), \quad \mathbf{C} = \mathbf{A} + i\mathbf{B}. \quad (2.11)$$

The vector \mathbf{F}_c is a *complex representation* of the real field \mathbf{F} . If all our operations on time-harmonic fields are *linear* (e.g., addition, differentiation, integration), it is more convenient to work with the complex representation. The reason this may be done is as follows. Let \mathcal{L} be any linear operator; we can operate on the field (2.10) by operating on the complex representation (2.11) and then take the real part of the result:

$$\mathcal{L}\mathbf{F} = \mathcal{L}\text{Re}(\mathbf{F}_c) = \text{Re}(\mathcal{L}\mathbf{F}_c).$$

Note that there is a degree of arbitrariness associated with the complex representation of a real field: \mathbf{F} could just as easily have been written $\mathbf{F} = \text{Re}(\mathbf{F}_c^*)$, where $\mathbf{F}_c^* = \mathbf{C}^* \exp(i\omega t)$ and the asterisk denotes the complex conjugate. Thus, there are two possible choices for the time-dependent factor in a complex representation of a time-harmonic field: $\exp(i\omega t)$ and $\exp(-i\omega t)$. It makes no difference which choice is made: the quantities of physical interest are always real. But once a sign convention has been chosen it must be used consistently in all analysis. We shall take the time-dependent factor to be $\exp(-i\omega t)$; this is the convention found in standard books on optics (Born and Wolf, 1965) and electromagnetic theory (Stratton, 1941; Jackson, 1975) as

well as being nearly universal in solid-state physics. If comparison is made between our expressions and similar expressions in the scientific literature, the sign convention must be carefully noted.

If $\exp(-i\omega t)$ time dependence is assumed for all fields, and the constitutive relations (2.7)–(2.9) are substituted into (2.1)–(2.4), we obtain

$$\nabla \cdot (\epsilon \mathbf{E}_c) = 0, \quad (2.12)$$

$$\nabla \times \mathbf{E}_c = i\omega\mu\mathbf{H}_c, \quad (2.13)$$

$$\nabla \cdot \mathbf{H}_c = 0, \quad (2.14)$$

$$\nabla \times \mathbf{H}_c = -i\omega\epsilon\mathbf{E}_c, \quad (2.15)$$

where the complex *permittivity* is

$$\epsilon = \epsilon_0(1 + \chi) + i\frac{\sigma}{\omega}. \quad (2.16)$$

If $\epsilon \neq 0$, the electric field is divergence free; this is the general condition for a *transverse* field. Except possibly at frequencies where $\epsilon = 0$, therefore, the medium cannot support *longitudinal* fields.

Equations (2.12)–(2.16) will usually be our point of departure in scattering problems. However, to avoid a cluttered notation, we shall often omit the subscript *c* from the complex fields. In those instances where confusion might result, the subscript will be retained, although it should usually be clear from the context if we are dealing with real fields or their complex representations.

2.3 FREQUENCY-DEPENDENT PHENOMENOLOGICAL COEFFICIENTS

Although we gave some physical justification, the seemingly *ad hoc* introduction of frequency-dependent phenomenological coefficients implies that they have a well-defined meaning only if the fields are harmonic in time. But this is very unsatisfactory: it is certainly not a law of nature that all electromagnetic fields must be harmonic in time. Thus, we must answer the questions: How are \mathbf{P} and \mathbf{E} (or \mathbf{B} and \mathbf{H}) related in general, and what is the meaning of frequency-dependent phenomenological coefficients? To answer these questions we must invoke the mathematical apparatus of Fourier transforms.

2.3.1 Fourier Transforms

Consider a *real* function of time $F(t)$. The Fourier transform $\mathcal{F}(\omega)$ of $F(t)$ is defined as

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} F(t) e^{i\omega t} dt, \quad (2.17)$$

where ω is real. We assume that the integral in (2.17) converges in some sense for all functions of interest. The *inverse Fourier transform* is given by

$$F(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathfrak{F}(\omega) e^{-i\omega t} d\omega. \quad (2.18)$$

According to (2.18), an arbitrary time-dependent function can be expressed as a superposition of time-harmonic functions $\exp(-i\omega t)$, where the complex amplitude $\mathfrak{F}(\omega)$ depends on the frequency ω . The condition that $F(t)$ be real is that $\mathfrak{F}^*(\omega) = \mathfrak{F}(-\omega)$; therefore, $F(t)$ can be expressed as a superposition of time-harmonic functions with positive frequency:

$$F(t) = \text{Re}(F_c(t)),$$

$$F_c(t) = \frac{1}{\pi} \int_0^{\infty} \mathfrak{F}(\omega) e^{-i\omega t} d\omega.$$

Note that (2.17) and (2.18) are not unique: all complex functions could be replaced by their complex conjugates, and the factor $1/2\pi$ could appear either in (2.17) or (2.18). If we want our expressions to appear more symmetrical, both integrals can have the common multiplicative factor $1/\sqrt{2\pi}$. There is no universally accepted convention for Fourier transforms. However, once the form of the Fourier transform has been specified, the corresponding expression for the inverse Fourier transform is uniquely determined.

If we Fourier analyze the Maxwell equations (2.1)–(2.4), with $\rho_f = 0$, and assume that the operations of integration and differentiation may be interchanged, we obtain

$$\nabla \cdot \mathfrak{D}(\omega) = 0, \quad (2.19)$$

$$\nabla \cdot \mathfrak{B}(\omega) = 0, \quad (2.20)$$

$$\nabla \times \mathfrak{E}(\omega) = i\omega \mathfrak{B}(\omega), \quad (2.21)$$

$$\nabla \times \mathfrak{H}(\omega) = \mathfrak{J}_f(\omega) - i\omega \mathfrak{D}(\omega), \quad (2.22)$$

where

$$\mathfrak{E}(\omega) = \int_{-\infty}^{\infty} \mathbf{E}(t) e^{i\omega t} dt; \quad \text{etc.}$$

The Fourier transforms of the fields are, in general, position dependent, although this is not explicitly indicated. Because the fields are real, we need consider only positive frequencies. Let us now assume that the Fourier transforms of the fields, rather than the fields themselves, are connected by

linear constitutive relations:

$$\mathcal{P}(\omega) = \epsilon_0 \chi(\omega) \mathcal{E}(\omega), \quad (2.23)$$

$$\mathcal{J}_f(\omega) = \sigma(\omega) \mathcal{E}(\omega), \quad (2.24)$$

$$\mathcal{B}(\omega) = \mu(\omega) \mathcal{H}(\omega), \quad (2.25)$$

where $\chi^*(\omega) = \chi(-\omega)$, and so on. We also assume that $\chi(\omega)$, $\sigma(\omega)$, and $\mu(\omega)$ are scalar quantities independent of position. As we shall see, the constitutive relations (2.23)–(2.25) reduce to familiar expressions when the fields are time harmonic but, in addition, answer the question posed above: How are arbitrarily time-dependent fields related?

Before we proceed we need one more result from Fourier transform theory. If the Fourier transform $\mathcal{X}(\omega)$ of a function $X(t)$ is expressed as a product of Fourier transforms

$$\mathcal{X}(\omega) = \mathcal{Y}(\omega) \mathcal{Z}(\omega),$$

the *convolution theorem* states that

$$X(t) = \int_{-\infty}^{\infty} Y(t - t') Z(t') dt',$$

where

$$Y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{Y}(\omega) e^{-i\omega t} d\omega; \quad \text{etc.}$$

If we apply the convolution theorem to the constitutive relation (2.23), for example, we obtain

$$\mathbf{P}(t) = \int_{-\infty}^{\infty} G(t - t') \mathbf{E}(t') dt', \quad (2.26)$$

where

$$G(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \epsilon_0 \chi(\omega) e^{-i\omega t} d\omega. \quad (2.27)$$

Thus, a frequency-dependent susceptibility $\chi(\omega)$ implies that the polarization \mathbf{P} at time t depends on the electric field \mathbf{E} at all other times t' . This conclusion is consistent with simple physical reasoning. If, for example, a steady electric field is applied to a sample of matter for a sufficient period of time, a steady polarization will be induced in the sample. However, if the electric field were to be suddenly removed, the polarization would not immediately drop to zero but would decay according to characteristic times associated with microscopic processes. In this example it is clear that the polarization is not proportional to the instantaneous field.

Let us now assume that the fields are time harmonic with angular frequency ω_0 :

$$\mathbf{E}(t) = \mathbf{A} \cos \omega_0 t + \mathbf{B} \sin \omega_0 t = \text{Re}\{\mathbf{E}_c\}.$$

The Fourier transform of $\mathbf{E}(t)$ is

$$\mathcal{E}(\omega) = \pi\{(\mathbf{A} + i\mathbf{B})\delta(\omega - \omega_0) + (\mathbf{A} - i\mathbf{B})\delta(\omega + \omega_0)\},$$

where δ is the Dirac delta function. From the Fourier inversion formula (2.18) and the constitutive relation (2.23) it follows that

$$\begin{aligned} \mathbf{P}(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \epsilon_0 \chi(\omega) \mathcal{E}(\omega) e^{-i\omega t} d\omega \\ &= \frac{\epsilon_0}{2} \{(\mathbf{A} + i\mathbf{B})\chi(\omega_0) e^{-i\omega_0 t} + (\mathbf{A} - i\mathbf{B})\chi(-\omega_0) e^{i\omega_0 t}\}. \end{aligned}$$

If we use the relation $\chi(-\omega_0) = \chi^*(\omega_0)$, then

$$\mathbf{P}(t) = \text{Re}\{\mathbf{P}_c(t)\} = \text{Re}\{\epsilon_0 \chi(\omega_0) \mathbf{E}_c(t)\}.$$

Thus, the perhaps unfamiliar constitutive relations (2.23)–(2.25) yield familiar results when the fields are time harmonic; moreover, because of (2.26) and (2.27), physical meaning can now be attached to the phenomenological coefficients even for arbitrarily time-dependent fields.

Although we concentrated attention on the electric susceptibility, similar remarks hold for the conductivity and permeability. In particular, we note that μ can be frequency dependent and is not restricted to be real. Complex phenomenological coefficients imply differences in *phase* between the various time-harmonic fields. For example, if the permeability $\mu = \mu' + i\mu''$ is complex, then for a real magnetic field $\mathbf{H} = \mathbf{H}_0 \cos \omega t$, the corresponding magnetic induction \mathbf{B} is $\bar{\mu} \mathbf{H}_0 \cos(\omega t - \phi)$, where the phase angle ϕ is given by $\tan \phi = \mu''/\mu'$ and $\bar{\mu} = \{(\mu')^2 + (\mu'')^2\}^{1/2}$. In Section 2.6 we shall show that if the imaginary part of any of the phenomenological coefficients of a medium is nonzero, the amplitude of a plane wave will decrease as it propagates through such a medium because of *absorption* of electromagnetic energy. Thus, complex phenomenological coefficients or, alternatively, a phase difference between \mathbf{P} and \mathbf{E} (or \mathbf{B} and \mathbf{H}) are physically manifested by absorption, the detailed nature of which is hidden from view in a macroscopic approach.

If the Fourier-transformed Maxwell equations (2.19)–(2.22) and the constitutive relations (2.23)–(2.25) are combined, we obtain

$$\nabla \cdot \mathcal{E}(\omega) = 0, \quad (2.28)$$

$$\nabla \cdot \mathcal{H}(\omega) = 0, \quad (2.29)$$

$$\nabla \times \mathcal{E}(\omega) = i\omega\mu(\omega)\mathcal{H}(\omega), \quad (2.30)$$

$$\nabla \times \mathcal{H}(\omega) = -i\omega\epsilon(\omega)\mathcal{E}(\omega), \quad (2.31)$$

where $\epsilon(\omega) = \epsilon_0(1 + \chi(\omega)) + i\sigma(\omega)/\omega$. Note that the two sets of equations (2.12)–(2.15) and (2.28)–(2.31) are formally identical although interpreted differently. In the former set, $(\mathbf{E}_c, \mathbf{H}_c)$ is the complex representation of a time-harmonic field; whereas in the latter set, $(\mathcal{E}, \mathcal{H})$ is the Fourier transform of an arbitrarily time-dependent electromagnetic field. But the same equations have the same solutions, so we need consider only time-harmonic fields; the general time-dependent field can be constructed by Fourier synthesis.

Both the conductivity and the susceptibility contribute to the imaginary part of the permittivity: $\text{Im}\{\epsilon\} = \epsilon_0 \text{Im}\{\chi\} + \text{Re}\{\sigma/\omega\}$. A nonzero value for $\text{Im}\{\epsilon\}$ manifests itself physically by absorption of electromagnetic energy in the medium. We may associate $\text{Im}\{\chi\}$ with the “bound” charge current density and $\text{Re}\{\sigma/\omega\}$ with the “free” charge current density. Absorption is determined by the *sum* of these two quantities, however, and it is not possible to determine by absorption measurements their relative contributions. This underscores our assertion that there is no clearly defined distinction between “free” and “bound” charges.

2.3.2 Kramers–Kronig Relations

As an alternative we could have begun our discussion of constitutive relations with the *assumption* that $\mathbf{P}(t)$ and $\mathbf{E}(t)$ are related through the linear functional equation

$$\mathbf{P}(t) = \int_{-\infty}^{\infty} G(t, t') \mathbf{E}(t') dt'. \quad (2.32)$$

Suppose that the electric field is a delta function applied at time t_0 : $\mathbf{E}(t) = \delta(t - t_0)\mathbf{E}_0$; the corresponding polarization is therefore

$$\mathbf{P}(t) = G(t, t_0)\mathbf{E}_0.$$

Thus, G is the polarization resulting from a unit amplitude delta function. If the properties of the medium do not change with time, the polarization must depend only on the time elapsed between t_0 and t :

$$G(t, t_0) = G(t - t_0).$$

Therefore, we obtain (2.26), which when inverted yields the constitutive relation (2.23). We note that *causality*—the system cannot squeal before it is hurt—requires that $G(\tau) = 0$ for $\tau < 0$.

The susceptibility is the Fourier transform of $G(t)$:

$$\epsilon_0\chi(\omega) = \int_{-\infty}^{\infty} G(t) e^{i\omega t} dt = \int_0^{\infty} G(t) e^{i\omega t} dt, \quad (2.33)$$

and is a complex-valued function of the real variable ω . Let us *define* a

complex-valued function of the *complex* variable $\tilde{\omega}$ by

$$\varepsilon_0 \chi(\tilde{\omega}) = \int_0^\infty G(t) e^{i\tilde{\omega}t} dt,$$

where $\tilde{\omega} = \omega_r + i\omega_i$. The function $\chi(\tilde{\omega})$ coincides with $\chi(\omega)$ when $\tilde{\omega}$ is a point on the real axis. For any $t \geq 0$, $G(t)e^{i\tilde{\omega}t}$ is an analytic function of $\tilde{\omega}$, and $|G(t)e^{i\tilde{\omega}t}| \leq |G(t)|$ if $\omega_i > 0$. Therefore, if the integral

$$\int_0^\infty |G(t)| dt \quad (2.34)$$

converges, then

$$\int_0^\infty G(t) e^{i\tilde{\omega}t} dt$$

converges to an analytic function in the upper half of the complex $\tilde{\omega}$ plane. Convergence of (2.34) is assured if $\chi(0)$ is finite; this follows from (2.33). If $\chi(\tilde{\omega})$ is analytic, so is $\chi(\tilde{\omega})/(\tilde{\omega} - \omega)$ except at the pole $\tilde{\omega} = \omega$. Cauchy's theorem states that

$$\int_C f(\tilde{\omega}) d\tilde{\omega} = 0,$$

provided that the closed contour C encloses no poles of the analytic function $f(\tilde{\omega})$. Let us apply Cauchy's theorem to the function $\chi(\tilde{\omega})/(\tilde{\omega} - \omega)$, where ω is a point on the real axis, and the contour C , shown in Fig. 2.1, is the union of four curves with parametric representations

$$C_1: \quad \tilde{\omega} = \Omega \quad (-A \leq \Omega \leq \omega - a),$$

$$C_2: \quad \tilde{\omega} = \omega - ae^{-i\Omega} \quad (0 \leq \Omega \leq \pi),$$

$$C_3: \quad \tilde{\omega} = \Omega \quad (\omega + a \leq \Omega \leq A),$$

$$C_4: \quad \tilde{\omega} = Ae^{i\Omega} \quad (0 \leq \Omega \leq \pi).$$

Therefore, from Cauchy's theorem

$$\begin{aligned} \int_{-A}^{\omega-a} \frac{\chi(\Omega)}{\Omega - \omega} d\Omega + \int_{\omega+a}^A \frac{\chi(\Omega)}{\Omega - \omega} d\Omega + \int_0^\pi \frac{iAe^{i\Omega}\chi(Ae^{i\Omega})}{Ae^{i\Omega} - \omega} d\Omega \\ = \int_0^\pi i\chi(\omega - ae^{i\Omega}) d\Omega. \end{aligned}$$

The integral over the curve C_4 vanishes as A tends to infinity if $\lim_{|\tilde{\omega}| \rightarrow \infty} \chi(\tilde{\omega})$

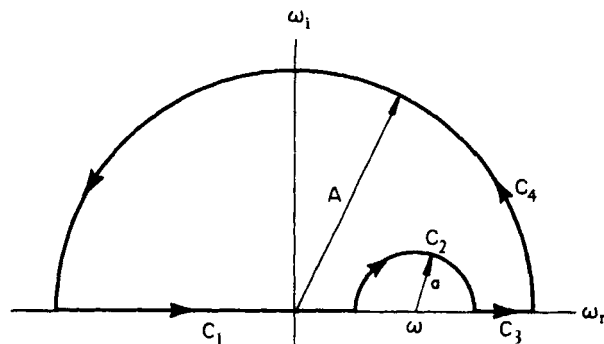


Figure 2.1 Contour of integration.

$= 0$; this implies that

$$i\pi\chi(\omega) = P \int_{-\infty}^{\infty} \frac{\chi(\Omega)}{\Omega - \omega} d\Omega, \quad (2.35)$$

where the symbol P denotes the *Cauchy principal value* of the integral, defined by

$$\lim_{a \rightarrow 0} \int_{-\infty}^{\omega-a} \frac{\chi(\Omega)}{\Omega - \omega} d\Omega + \int_{\omega+a}^{\infty} \frac{\chi(\Omega)}{\Omega - \omega} d\Omega.$$

We were slightly careless in our derivation of (2.35): $\chi(\bar{\omega})$ was required to be analytic in the upper half of the complex $\bar{\omega}$ plane, whereas part of the contour C included the real axis. But this can easily be remedied, and a more general form of (2.35) is

$$i\pi \lim_{\eta \rightarrow 0} \chi(\omega + i\eta) = \lim_{\eta \rightarrow 0} P \int_{-\infty}^{\infty} \frac{\chi(\Omega + i\eta)}{\Omega - \omega} d\Omega,$$

from which (2.35) follows if $\chi(\bar{\omega})$ is continuous.

The fundamental relation (2.35) can be written as two real integral relations:

$$\chi'(\omega) = \frac{2}{\pi} P \int_0^{\infty} \frac{\Omega \chi''(\Omega)}{\Omega^2 - \omega^2} d\Omega, \quad (2.36)$$

$$\chi''(\omega) = \frac{-2\omega}{\pi} P \int_0^{\infty} \frac{\chi'(\Omega)}{\Omega^2 - \omega^2} d\Omega, \quad (2.37)$$

where $\chi = \chi' + i\chi''$ and we have invoked the *crossing condition* $\chi^*(\omega) = \chi(-\omega)$. Equations (2.36) and (2.37) are an important example of a rather

remarkable class of mathematical relations called *Kramers–Kronig* or *dispersion relations*. Their implication, which is unexpected on physical grounds and therefore has an aura of black magic, is that the real and imaginary parts of χ are not independent but are connected by integral relations; this imposes a constraint on physically realizable susceptibilities. Moreover, if χ' is known over a sufficiently large range of frequencies around ω , $\chi''(\omega)$ can be obtained by integration, and vice versa. An interesting corollary of (2.36) is the *sum rule*:

$$\chi'(0) = \frac{2}{\pi} \int_0^\infty \frac{\chi''(\Omega)}{\Omega} d\Omega.$$

Although the Kramers–Kronig relations do not follow directly from physical reasoning, they are not devoid of physical content: underlying their derivation are the assumptions of *linearity* and *causality* and restrictions on the asymptotic behavior of χ . As we shall see in Chapter 9, the required asymptotic behavior of χ is a physical consequence of the interaction of a frequency-dependent electric field with matter.

The derivation of Kramers–Kronig relations for the susceptibility was relatively easy, perhaps misleadingly so. With a bit of extra effort, however, we can often derive similar relations for other frequency-dependent quantities that arise in physical problems. Suppose that we have two time-dependent quantities of unspecified origin, which we may call the *input* $X_i(t)$ and the *output* $X_o(t)$; the corresponding Fourier transforms are denoted by $\mathcal{X}_i(\omega)$ and $\mathcal{X}_o(\omega)$. If the relation between these transforms is linear,

$$\mathcal{X}_o(\omega) = \mathcal{R}(\omega) \mathcal{X}_i(\omega),$$

and causal (i.e., the output cannot precede the input in time), then $\mathcal{R}(\tilde{\omega})$ is an analytic function in the top half of the complex $\tilde{\omega}$ plane. It is also necessary that $\mathcal{R}(\tilde{\omega})$ vanish on the circular arc C_4 (Fig. 2.1) as A approaches infinity; if it does not, we are permitted to fiddle with it until it does. That is, we can change the asymptotic behavior of $\mathcal{R}(\tilde{\omega})$ by multiplying it by some analytic function $g(\tilde{\omega})$, or adding $g(\tilde{\omega})$ to it, without changing its analyticity. Of course, in so doing, we may also change the crossing condition, and the resultant dispersion relations may be different from (2.36) and (2.37). Techniques of fiddling with \mathcal{R} until it behaves properly are best illustrated with specific examples, which we shall encounter later in this chapter and elsewhere in the book.

2.4 SPATIAL DISPERSION

We have shown that a frequency-dependent susceptibility implies *temporal dispersion*: the polarization at time t depends on the electric field at all times previous to t . It is also possible under some circumstances to have *spatial dispersion*: the polarization at point \mathbf{x} depends on the values of the electric field at points in some neighborhood of \mathbf{x} . This nonlocal relation between \mathbf{P} and \mathbf{E}

leads to the concept of a susceptibility that depends on frequency *and* wave vector. To understand this it is helpful to introduce the three-dimensional Fourier transform of an arbitrary real function $F(\mathbf{x})$:

$$\mathcal{F}(\mathbf{k}) = \int F(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x},$$

where the integration is taken over all space. The inverse Fourier transform is

$$F(\mathbf{x}) = \frac{1}{(2\pi)^3} \int \mathcal{F}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k}.$$

Let us Fourier analyze the polarization and the electric field:

$$\mathcal{P}(\mathbf{k}, \omega) = \int \int \mathbf{P}(\mathbf{x}, t) e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega t)} d\mathbf{x} dt,$$

$$\mathcal{E}(\mathbf{k}, \omega) = \int \int \mathbf{E}(\mathbf{x}, t) e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega t)} d\mathbf{x} dt.$$

If we assume that there is a linear relation $\mathcal{P}(\mathbf{k}, \omega) = \epsilon_0 \chi(\mathbf{k}, \omega) \mathcal{E}(\mathbf{k}, \omega)$, the convolution theorem yields

$$\mathbf{P}(\mathbf{x}, t) = \int \int G(\mathbf{x} - \mathbf{x}', t - t') \mathbf{E}(\mathbf{x}', t') d\mathbf{x}' dt',$$

where

$$G(\mathbf{x}, t) = \frac{1}{(2\pi)^4} \int \int \epsilon_0 \chi(\mathbf{k}, \omega) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} d\mathbf{k} d\omega.$$

Thus, a susceptibility that depends on frequency *and* wave vector implies that the relation between $\mathbf{P}(\mathbf{x}, t)$ and $\mathbf{E}(\mathbf{x}, t)$ is nonlocal in time *and* space. Such spatially dispersive media lie outside our considerations. However, spatial dispersion can be important when the wavelength is comparable to some characteristic length in the medium (e.g., mean free path), and it is well at least to be aware of its existence: it can have an effect on absorption and scattering by small particles (Yildiz, 1963; Foley and Pattanayak, 1974; Ruppin, 1975, 1981).

2.5 POYNTING VECTOR

Consider an electromagnetic field (\mathbf{E}, \mathbf{H}) , which is not necessarily time harmonic. The *Poynting vector* $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ specifies the magnitude and direction of the rate of transfer of electromagnetic energy at all points of space; it is

of fundamental importance in problems of propagation, absorption, and scattering of electromagnetic waves. If the orientation of a plane surface with area A is specified by a unit normal vector $\hat{\mathbf{n}}$, the rate at which electromagnetic energy is transferred across this surface is $\mathbf{S} \cdot \hat{\mathbf{n}} A$, provided that \mathbf{S} is constant over the surface. When \mathbf{S} is a function of position and the surface has arbitrary shape, this can be generalized to

$$\int \mathbf{S} \cdot \hat{\mathbf{n}} dA.$$

The net rate W at which electromagnetic energy crosses the boundary of a closed surface A which encloses a volume V is

$$W = - \int_A \mathbf{S} \cdot \hat{\mathbf{n}} dA.$$

Why the minus sign? There are two choices for the unit normal to a closed surface: inward and outward. We have chosen the outward normal. If \mathbf{S} and $\hat{\mathbf{n}}$ are oppositely directed at a given point ($\mathbf{S} \cdot \hat{\mathbf{n}} < 0$), the minus sign ensures that the contribution to W is positive. Therefore, W is positive if there is a net transfer of electromagnetic energy *into* the volume. A positive W implies that electromagnetic energy is *absorbed* in V ; that is, electromagnetic energy is converted into other forms of energy (e.g., thermal energy) within V .

The formation of the vector product of two vectors is not a linear operation. Therefore, if the electromagnetic field is time harmonic, it is *not* true that $\mathbf{S} = \text{Re}(\mathbf{E}_c \times \mathbf{H}_c)$, although it is true that

$$\mathbf{S} = \text{Re}\{\mathbf{E}_c\} \times \text{Re}\{\mathbf{H}_c\}. \quad (2.38)$$

The instantaneous Poynting vector (2.38) is a rapidly varying function of time for frequencies that are usually of interest. Most instruments are not capable of following the rapid oscillations of the instantaneous Poynting vector, but respond to some time average $\langle \mathbf{S} \rangle$:

$$\langle \mathbf{S} \rangle = \frac{1}{\tau} \int_t^{t+\tau} \mathbf{S}(t') dt',$$

where τ is a time interval long compared with $1/\omega$. The time-averaged Poynting vector for time-harmonic fields is given by

$$\langle \mathbf{S} \rangle = \frac{1}{2} \text{Re}(\mathbf{E}_c \times \mathbf{H}_c^*).$$

When it is clear from the context that it is the time-averaged Poynting vector with which we are dealing, the brackets enclosing \mathbf{S} will be omitted.

2.6 PLANE-WAVE PROPAGATION IN UNBOUNDED MEDIA

Let us “look for” plane-wave solutions to the Maxwell equations (2.12)–(2.15). What does this statement mean? We know that the electromagnetic field (\mathbf{E}, \mathbf{H}) cannot be arbitrarily specified. Only certain electromagnetic fields, those that satisfy the Maxwell equations, are physically realizable. Therefore, because of their simple form, we should like to know under what conditions plane electromagnetic waves

$$\mathbf{E}_c = \mathbf{E}_0 \exp(i\mathbf{k} \cdot \mathbf{x} - i\omega t); \quad \mathbf{H}_c = \mathbf{H}_0 \exp(i\mathbf{k} \cdot \mathbf{x} - i\omega t), \quad (2.39)$$

where \mathbf{E}_0 and \mathbf{H}_0 are constant vectors, are compatible with the Maxwell equations. The wave vector \mathbf{k} may be complex

$$\mathbf{k} = \mathbf{k}' + i\mathbf{k}'', \quad (2.40)$$

where \mathbf{k}' and \mathbf{k}'' are real vectors. If (2.40) is substituted in (2.39), we obtain

$$\mathbf{E}_c = \mathbf{E}_0 \exp(-\mathbf{k}'' \cdot \mathbf{x}) \exp(i\mathbf{k}' \cdot \mathbf{x} - i\omega t),$$

$$\mathbf{H}_c = \mathbf{H}_0 \exp(-\mathbf{k}'' \cdot \mathbf{x}) \exp(i\mathbf{k}' \cdot \mathbf{x} - i\omega t).$$

$\mathbf{E}_0 \exp(-\mathbf{k}'' \cdot \mathbf{x})$ and $\mathbf{H}_0 \exp(-\mathbf{k}'' \cdot \mathbf{x})$ are the *amplitudes* of the electric and magnetic waves, and $\phi = \mathbf{k}' \cdot \mathbf{x} - \omega t$ is the *phase* of the waves. An equation of the form $\mathbf{K} \cdot \mathbf{x} = \text{constant}$, where \mathbf{K} is any real vector, defines a plane surface the normal to which is \mathbf{K} . Therefore, \mathbf{k}' is perpendicular to the *surfaces of constant phase*, and \mathbf{k}'' is perpendicular to the *surfaces of constant amplitude*. If \mathbf{k}' and \mathbf{k}'' are parallel, which includes the case $\mathbf{k}'' = 0$, these surfaces coincide and the waves are said to be *homogeneous*; if \mathbf{k}' and \mathbf{k}'' are not parallel, the waves are said to be *inhomogeneous*. For example, waves propagating in a vacuum are homogeneous.

Let us briefly consider propagation of surfaces of constant phase. Choose an arbitrary origin O and a plane surface over which the phase ϕ is constant (Fig. 2.2). At time t the distance from the origin O to the plane is z , where $\mathbf{k}' \cdot \mathbf{x} = k'z$ and $k'z - \omega t = \phi$. In a time interval Δt the surface of constant phase will have moved a distance Δz , where

$$k'z - \omega t = k'(z + \Delta z) - \omega(t + \Delta t) = \phi.$$

Thus, the velocity of propagation of surfaces of constant phase, the *phase velocity* v , is

$$v = \frac{\Delta z}{\Delta t} = \frac{\omega}{k'},$$

and the vector \mathbf{k}' specifies the direction of propagation.

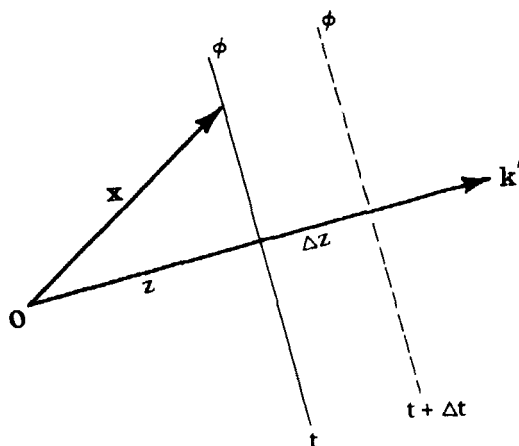


Figure 2.2 Propagation of constant phase surfaces.

The Maxwell equations for plane waves are

$$\mathbf{k} \cdot \mathbf{E}_0 = 0, \quad (2.41)$$

$$\mathbf{k} \cdot \mathbf{H}_0 = 0, \quad (2.42)$$

$$\mathbf{k} \times \mathbf{E}_0 = \omega \mu \mathbf{H}_0, \quad (2.43)$$

$$\mathbf{k} \times \mathbf{H}_0 = -\omega \epsilon \mathbf{E}_0. \quad (2.44)$$

Equations (2.41) and (2.42) are the conditions for *transversality*: \mathbf{k} is perpendicular to \mathbf{E}_0 and \mathbf{H}_0 . It is also evident from (2.43) or (2.44) that \mathbf{E}_0 and \mathbf{H}_0 are perpendicular. However, \mathbf{k} , \mathbf{E}_0 , and \mathbf{H}_0 are, in general, complex vectors, and the interpretation of the term “perpendicular” is not simple unless the waves are homogeneous; for such waves, the real fields \mathbf{E} and \mathbf{H} lie in a plane the normal to which is parallel to the direction of propagation.

If we take the vector product of both sides of (2.43) with \mathbf{k} ,

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) = \omega \mu \mathbf{k} \times \mathbf{H}_0 = -\omega^2 \epsilon \mu \mathbf{E}_0,$$

and use the **BAC–CAB** vector identity:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

together with (2.41), we obtain

$$\mathbf{k} \cdot \mathbf{k} = \omega^2 \epsilon \mu. \quad (2.45)$$

Our analysis shows that plane waves (2.39) are compatible with the Maxwell equations provided that \mathbf{k} , \mathbf{E}_0 , and \mathbf{H}_0 are perpendicular:

$$\mathbf{k} \cdot \mathbf{E}_0 = \mathbf{k} \cdot \mathbf{H}_0 = \mathbf{E}_0 \cdot \mathbf{H}_0 = 0.$$

In addition, the wave vector must satisfy (2.45); this equation may be written

$$k'^2 - k''^2 + 2i\mathbf{k}' \cdot \mathbf{k}'' = \omega^2 \epsilon \mu. \quad (2.46)$$

Note that $\epsilon \mu$ is a property of the *medium* in which the wave propagates; the vectors \mathbf{k}' and \mathbf{k}'' , however, are properties of the *wave*. If \mathbf{k}' and \mathbf{k}'' satisfy (2.46), they may otherwise be arbitrary: ϵ and μ do not uniquely specify the details of wave propagation.

The wave vector of a homogeneous wave may be written $\mathbf{k} = (\mathbf{k}' + i\mathbf{k}'')\hat{\mathbf{e}}$, where k' and k'' are nonnegative and $\hat{\mathbf{e}}$ is a real unit vector in the direction of propagation. Equation (2.45) requires that

$$k = k' + ik'' = \frac{\omega N}{c},$$

where c is the speed of light *in vacuo* and the *complex refractive index* N is

$$N = c\sqrt{\epsilon\mu} = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}}. \quad (2.47)$$

We shall write

$$N = n + ik, \quad (2.48)$$

where n and k are nonnegative. Other notation commonly encountered for the complex refractive index is $n' + in''$ and $n(1 + i\kappa)$; if we had chosen $\exp(i\omega t)$ time dependence, these would be written $n' - in''$ and $n(1 - i\kappa)$. The symbol k , where $\omega\sqrt{\epsilon\mu} = k$, is widely used to denote the wave number. On the other hand, the use of k to denote the imaginary part of the complex refractive index is equally sacrosanct, particularly among those who actually undertake the task of its measurement. We shall adhere to both of these conventions; however, to differentiate between the two quantities, we shall use italic type for the imaginary part of the refractive index and Roman type for the wave number. The use of similar symbols to denote two physically distinct but not unrelated quantities may possibly lead to confusion; but even more befuddlement would undoubtedly result to many readers if we were to meddle with well-established conventions for the sake of symbolic precision.

The free-space wave number is $\omega/c = 2\pi/\lambda$, where λ is the wavelength *in vacuo*. Therefore, a plane homogeneous wave has the form

$$\mathbf{E}_c = \mathbf{E}_0 \exp\left(-\frac{2\pi k z}{\lambda}\right) \exp\left(\frac{i2\pi n z}{\lambda} - i\omega t\right),$$

where $z = \hat{\mathbf{e}} \cdot \mathbf{x}$. Thus, the imaginary part of the complex refractive index determines the attenuation of the wave as it propagates through the medium; the real part determines the phase velocity $v = c/n$. The pair of quantities n and k are often referred to as the *optical constants*, terminology that is as widespread as it is misleading: the optical “constants” are not constant, they often strongly depend on frequency. It is not unusual for k of many common solids to range over six orders of magnitude within a relatively narrow range of frequencies.

Although refractive indices of many transparent materials at visible wavelengths have been accurately known for a long time, experimental determination of optical constants is by no means trivial in wavelength regions where a solid or liquid is appreciably absorbing. Special techniques have had to be developed to measure optical constants for various materials in different portions of the electromagnetic spectrum, and lack of reliable optical constants is still a serious impediment to small-particle scattering and absorption calculations in several areas of applied physics.

The real and imaginary parts of the complex refractive index satisfy Kramers–Kronig relations; sometimes this can be used to assess the reliability of measured optical constants. $N(\omega)$ satisfies the same crossing condition as $\chi(\omega)$: $N^*(\omega) = N(-\omega)$. However, it does not vanish in the limit of indefinitely large frequency: $\lim_{\omega \rightarrow \infty} N(\omega) = 1$. But this is a small hurdle, which can be surmounted readily enough by minor fiddling with $N(\omega)$: the quantity $N(\omega) - 1$ has the desired asymptotic behavior. If we now assume that $N(\tilde{\omega})$ is analytic in the top half of the complex $\tilde{\omega}$ plane, it follows that

$$n(\omega) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\Omega k(\Omega)}{\Omega^2 - \omega^2} d\Omega, \quad (2.49)$$

$$k(\omega) = \frac{-2\omega}{\pi} P \int_0^\infty \frac{n(\Omega)}{\Omega^2 - \omega^2} d\Omega, \quad (2.50)$$

where we used

$$P \int_0^\infty \frac{d\Omega}{\Omega^2 - \omega^2} = 0 \quad (2.51)$$

in the derivation of (2.50).

2.6.1 Absorption of Electromagnetic Energy

The Poynting vector of a plane wave is

$$\mathbf{S} = \frac{1}{2} \text{Re} \langle \mathbf{E} \times \mathbf{H}^* \rangle = \text{Re} \left\{ \frac{\mathbf{E} \times (\mathbf{k}^* \times \mathbf{E}^*)}{2\omega\mu^*} \right\},$$

where $\mathbf{E} \times (\mathbf{k}^* \times \mathbf{E}^*) = \mathbf{k}^*(\mathbf{E} \cdot \mathbf{E}^*) - \mathbf{E}^*(\mathbf{k}^* \cdot \mathbf{E})$. If the wave is homogeneous, $\mathbf{k} \cdot \mathbf{E} = 0$ implies that $\mathbf{k}^* \cdot \mathbf{E} = 0$; for such a wave propagating in the $\hat{\mathbf{e}}$

direction, we have

$$\mathbf{S} = \frac{1}{2} \operatorname{Re} \left\{ \sqrt{\frac{\epsilon}{\mu}} \right\} |\mathbf{E}_0|^2 \exp \left(-\frac{4\pi k z}{\lambda} \right) \hat{\mathbf{e}}.$$

Not surprisingly, \mathbf{S} is in the direction of propagation. The magnitude of \mathbf{S} , which we shall denote by the symbol I , is called the *irradiance* and its dimensions are energy per unit area and time. (The term *intensity* is often used to denote irradiance; however, intensity is also used for other radiometric quantities, and we shall therefore tend to avoid this term because of possible confusion. E is now the recommended symbol for irradiance, but this hardly seems appropriate in a book where the electric field and irradiance often appear side by side.) As the wave traverses the medium, the irradiance is exponentially attenuated:

$$I = I_0 e^{-\alpha z},$$

where the *absorption coefficient* α is

$$\alpha = \frac{4\pi k}{\lambda} \quad (2.52)$$

and I_0 is the irradiance at $z = 0$. Throughout this chapter we adopt the viewpoint that bulk matter is homogeneous. However, this is only approximately true; even in media that are usually considered to be homogeneous, such as bulk samples of a pure liquid or solid, a beam of light is attenuated both by absorption and by scattering. Although absorption is usually the dominant mode of attenuation in such media, scattering is not entirely absent, and unless special techniques are used a measurement of attenuation unavoidably yields the combined effect of absorption and scattering. In later chapters we shall discuss attenuation by collections of material particles; such attenuation may or may not be dominated by absorption depending on the size and optical properties of the particles.

The rate at which electromagnetic energy is removed from the wave as it propagates through the medium is determined by the imaginary part of the complex refractive index. If the irradiances I_0 and I_t (or rather their ratio) are measured at two different positions $z = 0$ and $z = h$, then α , and hence k , can be obtained in principle from the relation

$$\alpha h = \ln \frac{I_0}{I_t}. \quad (2.53)$$

This equation is strictly valid only if the detector is optically identical with the medium for which the absorption coefficient is to be measured (Fig. 2.3a), a condition that is difficult to satisfy. The usual experimental configuration is

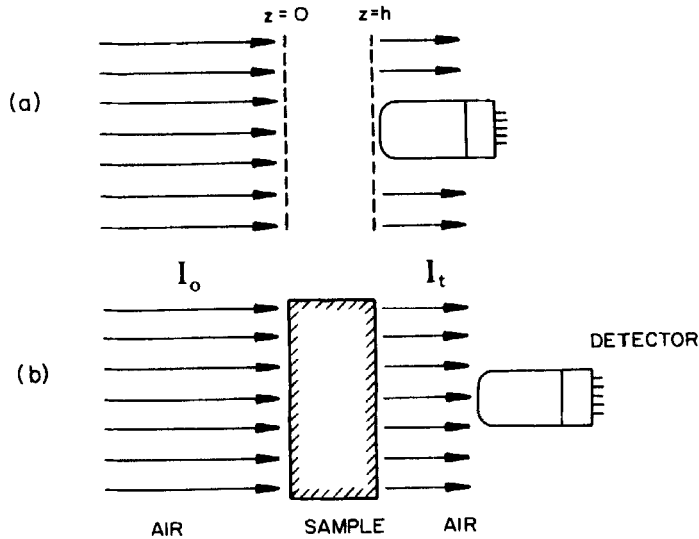


Figure 2.3 Measurement of absorption: (a) in principle and (b) in practice.

shown schematically in Fig. 2.3b. The transmission I_t/I_0 may be determined from the detector response with and without the sample interposed between source and detector. But α cannot be obtained from such a measurement unless reflections at the two interfaces are negligible.

2.7 REFLECTION AND TRANSMISSION AT A PLANE BOUNDARY

The considerations in the preceding section make it worthwhile to discuss reflection and transmission at plane boundaries: first, one plane boundary separating infinite media, then in the next section two successive plane boundaries forming a slab. In addition to providing useful results for bulk materials, these relatively simple boundary-value problems illustrate methods used in more complicated small-particle problems. Also, the optical properties of slabs often will be compared to those of small particles—both similarities and differences—to develop intuitive thinking about particles by way of the more familiar properties of bulk matter.

2.7.1 Normal Incidence

Consider a plane wave propagating in a nonabsorbing medium with refractive index $N_2 = n_2$, which is incident on a medium with refractive index $N_1 = n_1 + ik_1$ (Fig. 2.4). The amplitude of the incident electric field is E_i , and we assume that there are transmitted and reflected waves with amplitudes E_t and E_r , respectively. Therefore, plane-wave solutions to the Maxwell equations at

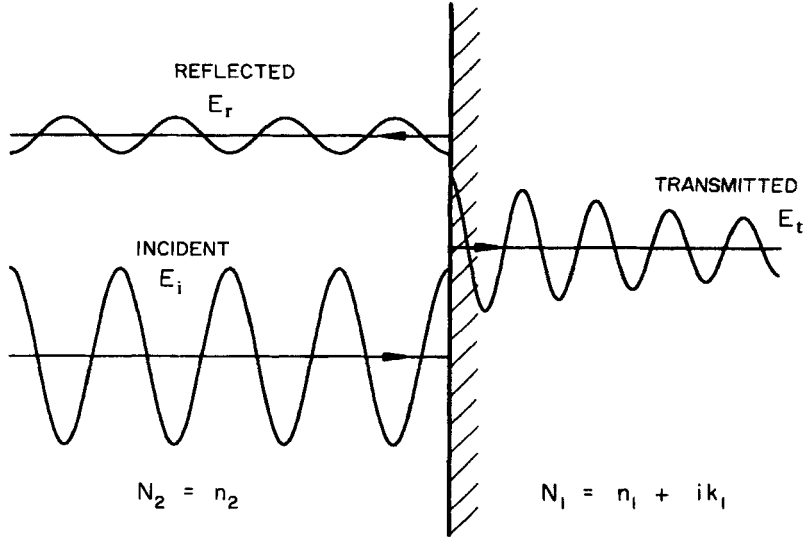


Figure 2.4 Reflection and transmission of normally incident light.

interior points on both sides of the boundary between the media are

$$E_r \exp \left[i\omega \left(\frac{N_1 z}{c} - t \right) \right] \quad (z > 0),$$

$$E_i \exp \left[i\omega \left(\frac{N_2 z}{c} - t \right) \right] + E_r \exp \left[-i\omega \left(\frac{N_2 z}{c} + t \right) \right] \quad (z < 0).$$

The tangential components of the electric field are required to be continuous across the boundary $z = 0$:

$$E_i + E_r = E_t. \quad (2.54)$$

Continuity of the tangential magnetic field yields the condition

$$E_i - E_r = \frac{N_1}{N_2} E_t, \quad (2.55)$$

where we have used (2.43) and also assumed that $\mu_1 = \mu_2$. Equations (2.54) and (2.55) are readily solved for the amplitudes E_r and E_t :

$$E_r = \tilde{r} E_i, \quad E_t = \tilde{t} E_i, \quad (2.56)$$

where the *reflection* and *transmission coefficients* are

$$\tilde{r} = \frac{1 - m}{1 + m}, \quad \tilde{t} = \frac{2}{1 + m}, \quad (2.57)$$

and $N_1/N_2 = m = n + ik$ is the refractive index of medium 1 relative to medium 2. The *reflectance* R for normally incident light, defined as the ratio of reflected to incident irradiance, is

$$R = |\tilde{r}|^2 = \left| \frac{1 - m}{1 + m} \right|^2 = \frac{(n - 1)^2 + k^2}{(n + 1)^2 + k^2}. \quad (2.58)$$

Note that $R \times 100\%$ is close to 100% if either $n \gg 1$ or $n \ll 1$ or $k \gg 1$. One might think that a material with $k \gg 1$ would be highly absorbing. But such a material is highly reflecting, and an incident wave cannot “get into” the material to be absorbed.

The amplitudes E_i and E_r in (2.56) may be interpreted as the Fourier transforms of the incident and reflected fields; therefore, from the general considerations at the end of Section 2.3, it follows that the real and imaginary parts of \tilde{r} satisfy Kramers–Kronig relations of the form (2.36) and (2.37). Although this is perhaps interesting, it is not terribly useful; but with a bit of effort we can derive Kramers–Kronig relations of great practical utility. To do so we must express \tilde{r} in terms of its amplitude (or modulus) r and phase (or argument) Θ :

$$\tilde{r} = re^{i\Theta}, \quad (2.59)$$

where $r = \sqrt{R}$. The reflection coefficient may also be written as an explicit function of n and k :

$$\tilde{r} = \frac{1 - n^2 - k^2}{(1 + n)^2 + k^2} - i \frac{2k}{(1 + n)^2 + k^2}, \quad (2.60)$$

from which it follows that

$$n = \frac{1 - R}{1 + R + 2\sqrt{R} \cos \Theta}, \quad k = \frac{-2\sqrt{R} \sin \Theta}{1 + R + 2\sqrt{R} \cos \Theta}. \quad (2.61)$$

For k to be positive, $\sin \Theta$ must be negative. The function $\log \tilde{r}$, defined by

$$\log \tilde{r} = \text{Log } r + i\Theta,$$

where $\text{Log } r = \ln r$, is analytic in the domain $r > 0$, $\pi \leq \Theta < 2\pi$. Because of causality, the function $\tilde{r}(\tilde{\omega})$ is analytic in the top half of the complex $\tilde{\omega}$ plane. If the values of $\tilde{r}(\tilde{\omega})$ lie in the domain of definition of $\log \tilde{r}$, then the function $\log \tilde{r}(\tilde{\omega})$ is also analytic in the top half of the complex $\tilde{\omega}$ plane. However, the asymptotic behavior of $\log \tilde{r}$ is unacceptable: we shall show in Chapter 9 that $\lim_{\omega \rightarrow \infty} \tilde{r}(\omega) = 1$. Once again, we are permitted to do some fiddling: the function

$$F(\omega) = \frac{\log \tilde{r}(\omega)}{\omega}$$

vanishes as ω increases without limit. Moreover, $F(\bar{\omega})$ is analytic in the top half of the complex $\bar{\omega}$ plane. Therefore, $F(\omega)$ satisfies the relation (2.35). The *phase shift dispersion relation* then follows from the crossing condition $F^*(\omega) = -F(-\omega)$:

$$\Theta(\omega) = \frac{-2\omega}{\pi} P \int_0^\infty \frac{\text{Log } r(\Omega)}{\Omega^2 - \omega^2} d\Omega. \quad (2.62)$$

Equation (2.62) is not merely a mathematical curiosity: from (2.61) it follows that at a given frequency ω the optical constants are determined by the reflectance and phase; however, if the reflectance is measured over a sufficiently large range of frequencies about ω , the phase can be obtained from (2.62).

2.7.2 Oblique Incidence

All plane waves normally incident on a plane boundary are reflected and transmitted according to (2.56) and (2.57) independently of their state of polarization (i.e., the direction of \mathbf{E}_i). This is analogous to scattering in the forward or backward directions by an isotropic sphere or a collection of randomly oriented particles, where the polarization is of no importance. However, when a plane wave is obliquely incident on a plane boundary, or when scattering angles other than forward or backward are considered, the polarization of the incident wave is indeed important. Incident unpolarized light may, upon reflection from a plane boundary or scattering by a particle, become highly polarized at certain angles. In treating reflection of a wave incident at an arbitrary angle on a plane boundary we therefore consider two polarizations: electric field vectors parallel and perpendicular to the *plane of incidence* (Fig. 2.5), defined by the direction of propagation of the incident wave and the normal to the boundary. An arbitrary wave may be written as the superposition of waves of these two kinds. Moreover, the two polarizations are independent of each other: if the incident wave is polarized parallel to the plane of incidence, for example, the reflected and transmitted waves are so polarized.

We first consider the electric vectors to be parallel to the plane of incidence. The tangential components of the electric and magnetic fields are required to be continuous across the boundary:

$$E_{\parallel i} \cos \Theta_i + E_{\parallel r} \cos \Theta_r = E_{\parallel t} \cos \Theta_t, \quad (2.63)$$

$$H_{\perp i} + H_{\perp r} = H_{\perp t}. \quad (2.64)$$

If we use (2.43) and also take the permeabilities to be equal in both media, (2.64) becomes

$$E_{\parallel i} - E_{\parallel r} = \frac{N_1}{N_2} E_{\parallel t}. \quad (2.65)$$

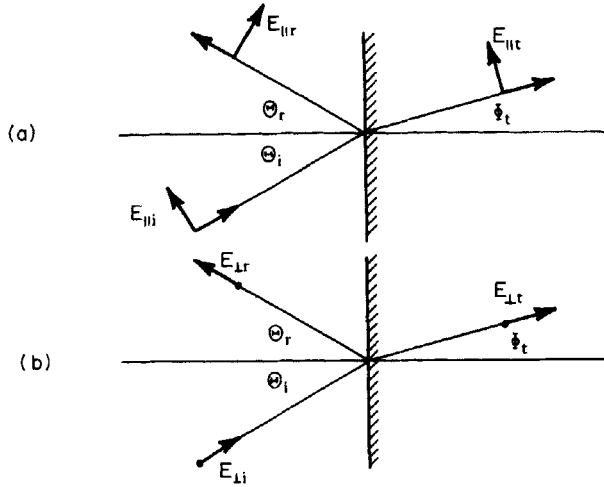


Figure 2.5 Reflection and transmission of obliquely incident light for electric vector parallel (a) and perpendicular (b) to the plane of incidence.

We also have $\Theta_r = \Theta_i$ from the law of specular reflection, and Θ_t is determined from the *generalized Snell's law*:

$$\sin \Theta_t = \frac{\sin \Theta_i}{m}. \quad (2.66)$$

Note that $\sin \Theta_t$, and hence Θ_t , is *complex* when medium 1 is absorbing. For such media the transmitted wave is inhomogeneous, although the incident and reflected waves are homogeneous (N_2 is real). In Fig. 2.5, Φ_t is the angle between the normal to the boundary and the direction of propagation of the surfaces of constant phase. The surfaces of constant amplitude are planes parallel to the boundary. If medium 1 is nonabsorbing, $\Theta_t = \Phi_t$. In general, however, the relation between Θ_t and geometrical properties of the transmitted wave is complicated. Rather than sink into the morass of trying to interpret a complex angle of refraction physically, it is much less frustrating to look upon Θ_t in (2.63) and (2.66) as merely a mathematical quantity.

Equations (2.63) and (2.65) are readily solved for the reflection and transmission coefficients

$$\tilde{r}_{\parallel} = \frac{E_{\parallel r}}{E_{\parallel i}} = \frac{\cos \Theta_t - m \cos \Theta_i}{\cos \Theta_t + m \cos \Theta_i}, \quad (2.67)$$

$$\tilde{t}_{\parallel} = \frac{E_{\parallel t}}{E_{\parallel i}} = \frac{2 \cos \Theta_i}{\cos \Theta_t + m \cos \Theta_i}. \quad (2.68)$$

When the electric vectors are perpendicular to the plane of incidence a similar

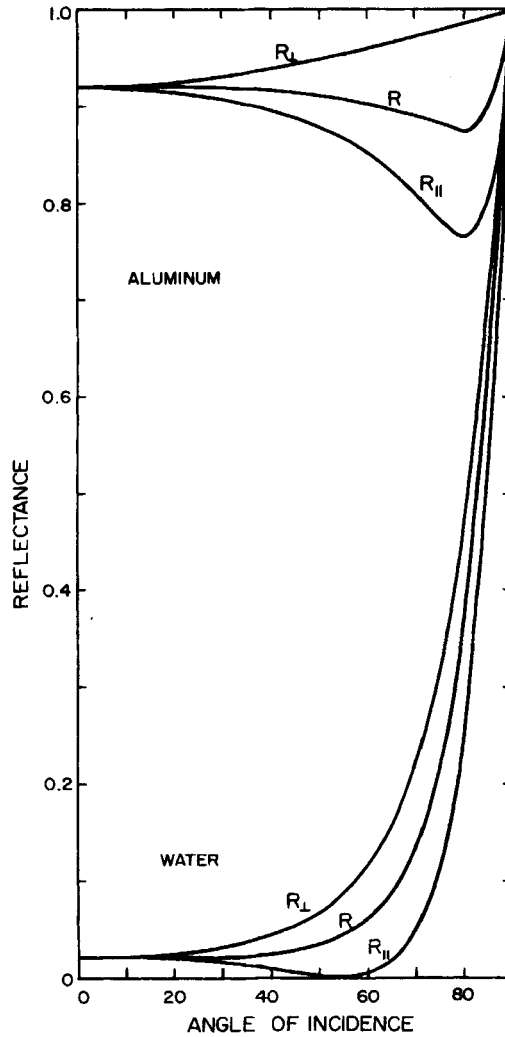


Figure 2.6 Reflectances for electric vector parallel (R_{\parallel}) and perpendicular (R_{\perp}) to the plane of incidence. R is the reflectance for unpolarized incident light.

analysis yields

$$\tilde{r}_{\perp} = \frac{E_{\perp r}}{E_{\perp i}} = \frac{\cos \Theta_i - m \cos \Theta_t}{\cos \Theta_i + m \cos \Theta_t}, \quad (2.69)$$

$$\tilde{t}_{\perp} = \frac{E_{\perp t}}{E_{\perp i}} = \frac{2 \cos \Theta_i}{\cos \Theta_i + m \cos \Theta_t}. \quad (2.70)$$

Equations (2.67)–(2.70) are the *Fresnel formulas* for reflection and transmission of light obliquely incident on a plane boundary.

The ratio of reflected to incident irradiance is more readily measured than the ratio of amplitudes; the reflectances for the two polarization states of the incident light are

$$R_{\parallel} = |\tilde{r}_{\parallel}|^2, \quad R_{\perp} = |\tilde{r}_{\perp}|^2. \quad (2.71)$$

One sometimes encounters the notation R_p and R_s for R_{\parallel} and R_{\perp} , where the p and s stand for the German *parallel* and *senkrecht* (perpendicular); R_{π} and R_{σ} are also used. The two reflectances (2.71) and the reflectance $R = \frac{1}{2}(R_{\parallel} + R_{\perp})$ for incident unpolarized light are shown as a function of the angle of incidence Θ_i in Fig. 2.6. The refractive indices are those for liquid water at visible wavelengths ($N_1 \approx 1.33$) and aluminum at a wavelength of 4958 Å ($N_1 = 0.771 + i5.91$); medium 2 is taken to be a vacuum ($N_2 = 1$) in both cases. Note that for water, an insulator, R_{\perp} is always nonzero, but there is an angle of incidence $\Theta_i = \Theta_p$ for which $R_{\parallel} = 0$. This angle is called the *polarizing* or *Brewster angle*, and it follows from (2.66) and (2.67) that

$$\tan \Theta_p = n.$$

For aluminum, a conductor, R_{\parallel} is always nonzero, although it does have a minimum at a particular angle. If the angle of incidence is other than 0 or 90°, R_{\perp} is greater than R_{\parallel} for an insulator or a conductor; therefore, if unpolarized light is incident on a plane boundary, the reflected light is partially polarized perpendicularly to the plane of incidence. At the Brewster angle (for an insulator) the reflected light is completely polarized.

2.8 REFLECTION AND TRANSMISSION BY A SLAB

We now consider reflection and transmission of a wave $E_i \exp[i\omega(N_2 z/c - t)]$ normally incident on a plane-parallel slab of arbitrary material embedded in a nonabsorbing medium (Fig. 2.7). The reflected and transmitted waves are

$$E_r \exp\left[-i\omega\left(\frac{N_2 z}{c} + t\right)\right], \quad E_t \exp\left[i\omega\left(\frac{N_2 z}{c} - t\right)\right],$$

and to satisfy all the boundary conditions we must postulate waves inside the slab that propagate in the $+z$ and $-z$ directions:

$$E_1^+ \exp\left[i\omega\left(\frac{N_1 z}{c} - t\right)\right], \quad E_1^- \exp\left[-i\omega\left(\frac{N_1 z}{c} + t\right)\right].$$

The field amplitudes are written as scalars because reflection and transmission at normal incidence are independent of polarization. At the first boundary ($z = 0$), the amplitudes satisfy the usual boundary conditions:

$$E_i + E_r = E_1^+ + E_1^-, \quad E_i - E_r = \frac{N_1}{N_2} (E_1^+ - E_1^-),$$

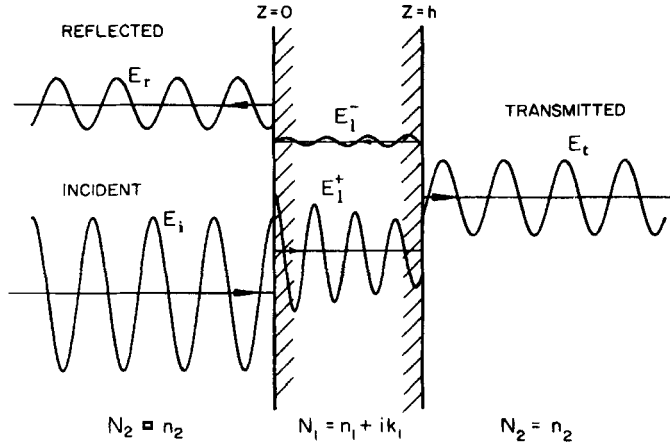


Figure 2.7 Reflection and transmission by a slab.

and at the second boundary ($z = h$) we have

$$E_1^+ \exp(ikN_1h) + E_1^- \exp(-ikN_1h) = E_t \exp(ikN_2h),$$

$$E_1^+ \exp(ikN_1h) - E_1^- \exp(-ikN_1h) = \frac{N_2}{N_1} E_t \exp(ikN_2h),$$

where $k = \omega/c = 2\pi/\lambda$. The reflection and transmission coefficients are obtained by solving these four simultaneous equations:

$$\tilde{r}_{\text{slab}} = \frac{E_r}{E_i} = \frac{\tilde{r} [1 - \exp(i2kN_1h)]}{1 - \tilde{r}^2 \exp(i2kN_1h)}, \quad (2.72)$$

$$\tilde{t}_{\text{slab}} = \frac{E_t}{E_i} = \frac{4m}{(m+1)^2} \frac{\exp(-ikN_2h)}{[\exp(-ikN_1h) - \tilde{r}^2 \exp(ikN_1h)]}, \quad (2.73)$$

where \tilde{r} is the reflection coefficient (2.57). More algebra yields the slab transmittance:

$$T_{\text{slab}} = |\tilde{t}_{\text{slab}}|^2 = \frac{(1-R)^2 + 4R \sin^2 \psi}{R^2 e^{-\alpha h} + e^{\alpha h} - 2R \cos(\zeta + 2\psi)}, \quad (2.74)$$

where

$$\psi = \tan^{-1} \left(\frac{2n_2 k_1}{n_1^2 + k_1^2 - n_2^2} \right) \quad 0 \leq \psi \leq \pi,$$

$$\zeta = \frac{4\pi n_1 h}{\lambda}, \quad \alpha = \frac{4\pi k_1}{\lambda}, \quad R = |\tilde{r}|^2.$$

Transmission experiments are usually carried out with the slab in air ($n_2 \approx 1$) and are feasible only if a measurable amount of light is transmitted. This in turn requires that for all but very thin samples (say, $h \ll \lambda$) k_1 must be small compared with unity. With this restriction ψ is small compared with ζ , and the slab transmittance is to good approximation given by

$$T_{\text{slab}} = \frac{(1 - R)^2}{R^2 e^{-\alpha h} + e^{\alpha h} - 2R \cos \zeta}. \quad (2.75)$$

Interference bands in slab transmission *may* be observed because of the oscillatory term $\cos \zeta$ in (2.75). Transmission maxima occur when $\zeta = 2\pi p$ ($p\lambda = 2n_1 h$), where $p = 1, 2, 3, \dots$. If a maximum occurs at a wavelength λ , neighboring maxima occur at $\lambda + \Delta\lambda$, where

$$\frac{\Delta\lambda}{\lambda} = \pm \frac{\lambda}{2n_1 h} \left(1 \mp \frac{\lambda}{2n_1 h} \right)^{-1},$$

provided that n_1 does not vary greatly over this interval.

We have taken the incident beam to be *perfectly monochromatic* up to this point. Such beams are produced easily enough on paper—a few strokes of the pen are sufficient! However, real beams with which experimenters must contend are composed of a finite spread, albeit narrow, of wavelengths. Thus, for practical purposes, T_{slab} in (2.74) should be looked upon as an average over a range of wavelengths. If the wavelength spread of the beam is $\delta\lambda$, interference bands are observable only if $\delta\lambda \ll |\Delta\lambda|$. Another tacit assumption is that k_1 does not vary greatly over the wavelength region of interest; if not, the absorption spectrum may obscure transmission maxima and minima that are the result of interference. This is obvious from inspection of (2.75): for sufficiently large absorption (αh) the oscillatory term is negligible compared with $\exp(\alpha h)$. Interference bands in transmission by a thin MnS crystal are shown in Fig. 2.8, where *optical density*, defined as $\log_{10}(1/T_{\text{slab}})$, is plotted as a function of wavelength. Note that in the vicinity of the absorption band at about 6000 Å the interference bands are damped somewhat.

If $\delta\lambda \gg |\Delta\lambda|$, distinct interference bands in the transmission spectrum will not be observed; in this case T_{slab} averaged over the wavelength interval $\delta\lambda$ is

$$T_{\text{slab}} = \frac{(1 - R)^2 e^{-\alpha h}}{1 - R^2 e^{-2\alpha h}}. \quad (2.76)$$

The strict validity of (2.76) also requires that $\delta\lambda/\lambda \ll 1$, where λ is the average wavelength in the interval $\delta\lambda$; this ensures that the maxima of $\cos \zeta$ are evenly spaced.

Even if $\delta\lambda \ll |\Delta\lambda|$, this is not sufficient to ensure that the interference bands predicted by (2.75) are experimentally observable. This equation also

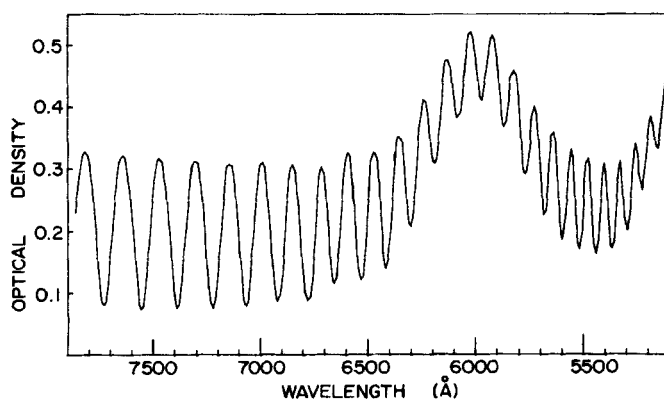


Figure 2.8 Transmission by an MnS crystal ($h \approx 6.3 \mu\text{m}$, $n_1 \approx 2.68$) measured with a Cary 14 spectrophotometer ($\delta\lambda \approx 5 \text{ Å}$).

contains the implicit requirement that $\delta h \ll \lambda/4n_1$, where δh is the average departure from parallelism of the faces of the slab over the area illuminated by the incident beam. We shall encounter a similar restriction when we deal with scattering by spheres: maxima and minima in the scattering cross section for a suspension of spheres are observable only if the dispersion of sphere radii is sufficiently narrow.

It is sometimes desirable to eliminate interference bands from transmission spectra; this is particularly true in the measurement of absorption spectra (α as a function of wavelength), where interference bands can be a nuisance. If $\delta h \gg \lambda/4n_1$, a condition that can be achieved by using a wedge-shaped sample, for example, then the observed transmittance is given by (2.76).

Yet another implicit assumption underlying both (2.75) and (2.76), and indeed all expressions in this section, is that the surfaces are *optically smooth*; that is, the surface roughness is sufficiently small compared with the wavelength of the incident light. It is difficult to state a precise criterion for smoothness, but according to the Rayleigh criterion a surface is reckoned to be smooth if $d < \lambda/(8 \cos \Theta)$, where d is the height of surface irregularities (Beckmann and Spizzichino, 1963, p. 9). If the surfaces are not smooth, incident light may be *diffusely reflected* over a distribution of angles rather than *specularly reflected* at a single angle.

Equations (2.75) and (2.76) represent two extremes. The former applies to a “perfectly” monochromatic beam incident on a “perfectly” parallel, smooth slab (although perfection in this sense is not absolute but lies within certain tolerance limits). On the other hand, the latter equation is appropriate to what might be called the perfectly imperfect case: a slab-beam combination that has been carefully prepared to eliminate *all* interference effects. Theory has little to say about cases intermediate between these two extremes. Thus, if quantitative data are to be extracted from transmission measurements, some care must be

taken. It takes little experimental ability to insert a sample into a spectrophotometer and press the scan button. Some kind of spectrum will dutifully emerge from the instrument. But it is an entirely different matter to extract from such spectra accurate numerical values for n and k .

Although (2.76) can be derived from (2.75) by integration, a more physically satisfying approach is to consider the multiple reflections and transmissions of an incident beam. The assumption of nonparallel slab faces or a sufficient spread of wavelengths is equivalent to assuming that we may deal with the beam without regard to its phase. Consider a beam with irradiance I_i that is incident on a slab (Fig. 2.9). A fraction R of the incident light is reflected at the first interface and the unreflected fraction traverses the slab undergoing attenuation by a factor $\exp(-\alpha h)$. At the second interface part of the light is reflected and an amount $I_i(1 - R)^2 \exp(-\alpha h)$ is transmitted. The reflected light traverses the slab, is reflected at the first interface, again traverses the slab, and an amount $I_i(1 - R)^2 R^2 \exp(-3\alpha h)$ emerges—and so on into the long hours of night. In this manner we obtain the total transmitted irradiance

$$I_t = I_i(1 - R)^2 e^{-\alpha h} (1 + R^2 e^{-2\alpha h} + R^4 e^{-4\alpha h} + \dots),$$

and the infinite series is readily summed to yield (2.76). One further approximation is useful: if $R^2 e^{-2\alpha h}$ is small compared with unity, which is true for the

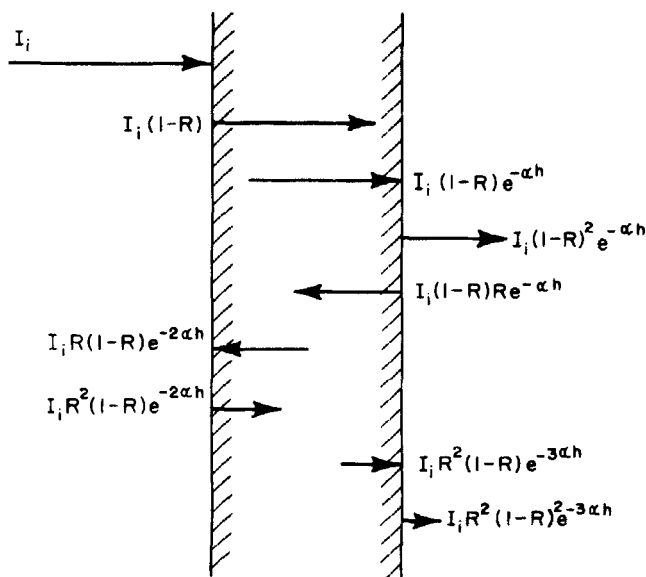


Figure 2.9 Transmission by a slab. Only the first two components that contribute to the total are shown. There is no systematic relationship among the phases of the various components.

large class of nonconductors at visible wavelengths, then (2.76) becomes

$$T_{\text{slab}} = (1 - R)^2 e^{-\alpha h}.$$

2.9 EXPERIMENTAL DETERMINATION OF OPTICAL CONSTANTS

The optical constants n and k are not directly measurable but must be derived from measurable quantities (e.g., reflection and transmission coefficients, reflectances and transmittances, and angles of refraction) by interposing a suitable theory. Most of the necessary theory has been discussed in the preceding two sections. Therefore, before considering the analogy between a particle and a slab, it is worthwhile to consider briefly how the solutions to electromagnetic boundary-value problems for plane surfaces can be used to solve the inversion problem for homogeneous media, that is, how to determine the optical constants. At a given wavelength there are, in general, two optical constants to be determined. This in turn implies that two or more measurements are required. The following methods may be used.

1. Measurement of refraction angles, such as the angle of minimum deviation of a prism; n is obtained from Snell's law. This requires samples of high transparency ($k \approx 0$).

2. Measurement of the transmittance and reflectance of a slab for light at near-normal incidence. The samples must be sufficiently transparent for measurable transmission in thin slabs, but not as transparent as required in method 1.

3. Measurement of reflectance at near-normal incidence over a wide range of frequencies. The phase shift of the reflected light is obtained from a Kramers-Kronig analysis. This technique is of great value in spectral regions where the sample is highly opaque, but requires measurements over an extended region and extrapolations into unmeasured regions.

4. Ellipsometric techniques in which amplitude ratios *and* phase shifts for reflected light are directly measured as opposed to the previous technique in which the phase shift is indirectly obtained. This is difficult to do over large wavelength regions because of requirements on optical elements such as polarizers and retarders.

5. Measurement of reflectances for incident light of various polarization states and two oblique angles of incidence; the results are analyzed with the Fresnel formulas. Large angles are required for high accuracy, and this requires large sample surfaces.

These methods and their variations are the principal means of measuring optical constants. As noted, none of them is clearly superior in all instances.

Further details may be found in the references listed at the ends of this chapter and Chapter 10.

2.10 THE ANALOGY BETWEEN A SLAB AND A PARTICLE

We have seen that the response of a plane surface to an incident beam can be rather complicated. Yet we are familiar with many of the resulting phenomena because of everyday experiences with them: reflection of sunlight from a smooth pond or a glass window; a shiny piece of metal; the colors of a soap film. Many readers probably have performed simple laboratory experiments such as measuring the dispersion of natural light by a prism or determining the Brewster angle using polarizers. All these contacts help to build intuition about bulk optical effects. By way of contrast, we have fewer opportunities to observe directly scattering and absorption by small particles the size of which may be comparable with the wavelength. Because of useful analogies between the more often experienced and inherently simpler phenomena and those less commonly experienced, we encourage the use of such analogies while recommending a good bit of caution. Many have come to grief by adhering too strongly to the analogy between small particles and plane surfaces. A classical example is the unnecessary confusion that occurred during the period in which the blue color and polarization of skylight were topics of keen scientific interest. Because of Tyndall's experiments, small particles, which we now know to be the air molecules themselves, were a strong candidate for the source of the blue sky. But one of the obstacles to the acceptance of this explanation was the polarizing angle. It was observed that the light scattered by small particles at 90° to the direction of the incident beam was highly polarized regardless of the composition of the particles. In the language of planar surfaces we would say that the polarizing angle for such particles is always 45° . This was considered by some to be a problem: the polarizing angle for planar surfaces is a function of refractive index and, hence, composition. Rayleigh's response (1871) to this objection is worth quoting:

I venture to think that the difficulty is imaginary, and is caused mainly by misuse of the word reflection. Of course there is nothing in the etymology of reflection and refraction to forbid their application in this sense; but the words have acquired technical meanings, and become associated with certain well-known laws called after them. Now a moment's consideration of the principles according to which reflection and refraction are explained in the wave theory is sufficient to show that they have no application unless the surface of the disturbing body is larger than many square wave-lengths; whereas the particles to which the sky is supposed to owe its illumination must be *smaller* than the wave-length or else the explanation of the colour breaks down. The idea of polarization by reflection is therefore out of place, and that "the law of Brewster does not apply to matter in this condition" (of extreme fineness) is only what might have been inferred from the principles of the wave theory.

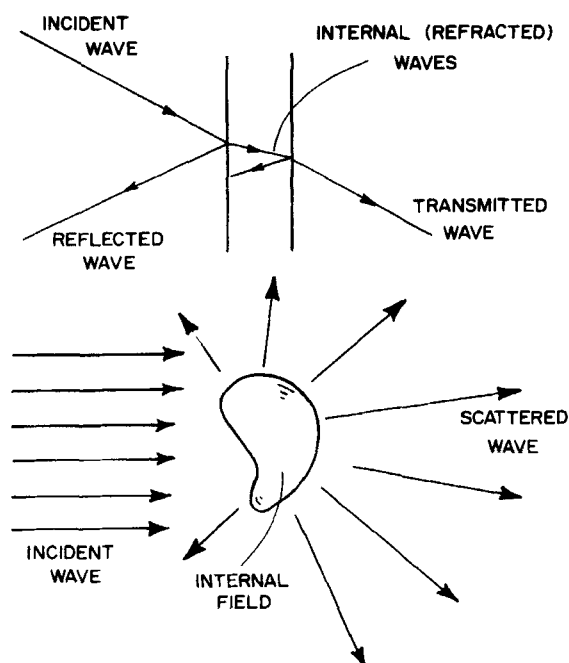


Figure 2.10 Analogy between scattering by a particle and reflection-transmission by a slab.

Rayleigh adds in a footnote that “in many departments of science a tendency may be observed to extend the field of familiar laws beyond their proper limits.” It is well to keep Rayleigh’s words in mind when comparison is made between slabs and particles.

The formal analogy between scattering by a particle and reflection-transmission by a slab, in its most general aspect, is shown schematically in Fig. 2.10 and in Table 2.1. We mentioned in Section 2.7 that unpolarized incident light may become polarized upon reflection from a plane interface *or* by scattering from a particle; throughout the rest of this book we shall be alert to further analogies, both as an aid to understanding and as a guide into the investigation of new and unfamiliar phenomena.

Table 2.1 Analogy between Slab and Particle

Slab	Particle
Incident wave	Incident wave
Reflected wave + transmitted wave	Scattered wave
Internal (refracted) waves	Internal field

2.11 POLARIZATION

In addition to irradiance and frequency, a monochromatic (i.e., time-harmonic) electromagnetic wave has a property called its *state of polarization*, a property that was briefly touched on in Section 2.7, where it was shown that the reflectance of obliquely incident light depends on the polarization of the electric field. In fact, polarization would be an uninteresting property were it not for the fact that two waves with identical frequency and irradiance, but different polarization, can behave quite differently. Before we leave the subject of plane waves it is desirable to present polarization in a systematic way, which will prove to be useful when we discuss the polarization of scattered light.

Consider a plane monochromatic wave with angular frequency ω and wave number k which is propagating in the z direction in a nonabsorbing medium. In discussions of polarization it is customary to focus attention on the electric field \mathbf{E} :

$$\begin{aligned}\mathbf{E} &= \text{Re}(\mathbf{E}_c) = \text{Re}\{(\mathbf{A} + i\mathbf{B})\exp(ikz - i\omega t)\} \\ &= \mathbf{A} \cos(kz - \omega t) - \mathbf{B} \sin(kz - \omega t),\end{aligned}\quad (2.77)$$

where the real vectors \mathbf{A} and \mathbf{B} are independent of position. The electric field vector at any point lies in a plane the normal to which is parallel to the direction of propagation. In a particular plane, say $z = 0$ for convenience, the tip of the electric vector traces out a curve:

$$\mathbf{E}(z = 0) = \mathbf{A} \cos \omega t + \mathbf{B} \sin \omega t. \quad (2.78)$$

Equation (2.78) describes an ellipse, the *vibration ellipse* (Fig. 2.11). If $\mathbf{A} = 0$ (or $\mathbf{B} = 0$), the vibration ellipse is just a straight line, and the wave is said to be *linearly polarized*; the vector \mathbf{B} then specifies the *direction of vibration*. (The term *plane polarized* is also used, but it has become less fashionable in recent years.) If $|\mathbf{A}| = |\mathbf{B}|$ and $\mathbf{A} \cdot \mathbf{B} = 0$, the vibration ellipse is a circle, and the wave is said to be *circularly polarized*. In general, a monochromatic wave of the form (2.77) is *elliptically polarized*.

A given vibration ellipse can be traced out in two opposite senses: clockwise and anticlockwise. The vibration ellipse in Fig. 2.11 is rotating clockwise as viewed from above the page. However, as viewed from the opposite direction, it is rotating in the anticlockwise sense. Thus, these terms do not have absolute meaning but depend on the direction from which the ellipse is observed. The two opposite senses of rotation lead to a classification of vibration ellipses according to their *handedness*, and herein lies a problem: there are two conventions for assigning handedness to vibration ellipses. On the one hand, the vibration ellipse may be designated as right-handed if the rotation is clockwise as viewed by an observer who is looking toward the source of light. That is, if the direction of the vector $\mathbf{A} \times \mathbf{B}$ is opposite to the direction of

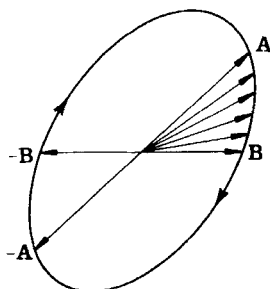


Figure 2.11 Vibration ellipse.

propagation, the vibration ellipse is said to be right-handed. Among the adherents to this convention are chemists (Djerassi, 1960) and optical physicists (Jenkins and White, 1957; Shurcliff, 1962; Stone, 1963). This convention might be called “traditional” and, as such, has been adopted by Born and Wolf (1965, p. 28), who nonetheless are slightly uncomfortable with the seeming unnaturalness of their choice—it is not compatible with the rotation behavior of a right-handed screw. Clarke (1974), however, has argued that the traditional convention need not tread on notions of what is or is not natural if we lay aside our screwdrivers and concentrate instead on the helix traced out in space by the electric field. At any instant of time (say, $t = 0$) the locus of all points described by the tip of the electric field vector is

$$\mathbf{E}(t = 0) = \mathbf{A} \cos kz - \mathbf{B} \sin kz, \quad (2.79)$$

which is the equation of a helix; (2.79) is a “snapshot” of the electric field at a particular time. With increasing time the helix moves in the direction of propagation and, in so doing, its intersection with any plane $z = \text{constant}$ describes a vibration ellipse (Fig. 2.12). If the helix is right-handed, the corresponding vibration ellipse is also right-handed according to the traditional convention. Because the handedness of a helix is independent of the direction from which it is observed, the helix associated with a given wave unambiguously assigns a handedness to that wave. We therefore adopt the traditional convention according to which an elliptically polarized wave is reckoned right-handed if the vibration ellipse is rotating in the clockwise sense as viewed by an observer looking toward the source. The opposite convention seems to be favored by astronomers (van de Hulst, 1957; Hansen and Travis, 1974; Gehrels, 1974a).

We shall be concerned primarily with media through which plane waves of arbitrary polarization propagate without change of polarization state. However, there are many materials that do not possess this property. For example, there are materials which, for a given direction of propagation, have different refractive indices depending on the state of linear polarization of the wave. If the real parts of the refractive indices are different, the material is said to be *linearly birefringent*; if the imaginary parts of the refractive indices are differ-

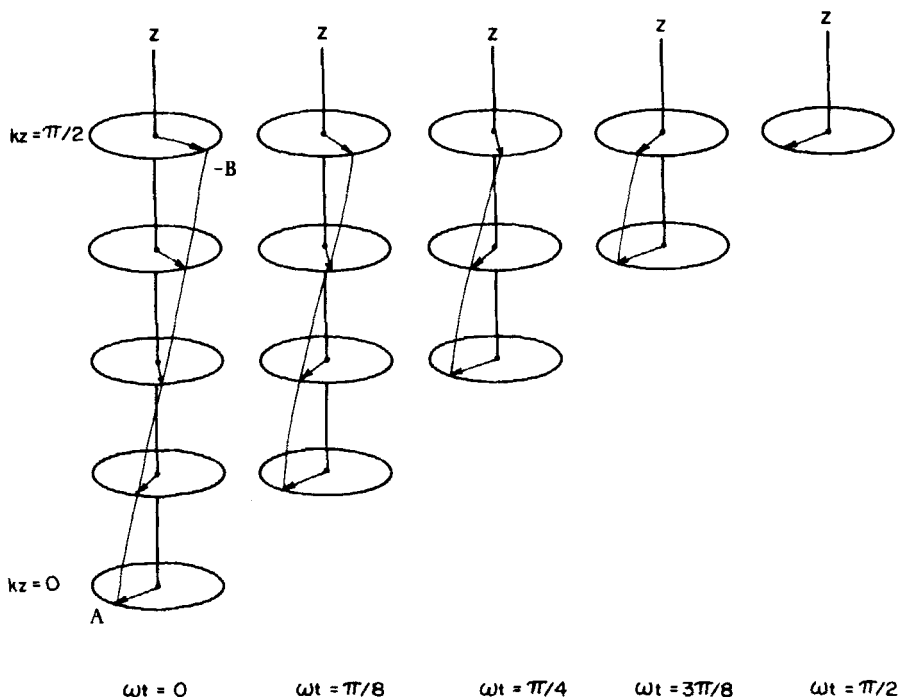


Figure 2.12 A series of snapshots of the electric field.

ent, it is said to be *linearly dichroic*. Similarly, there are *circularly birefringent* and *circularly dichroic* media, those for which the complex refractive index depends on handedness. The terms “birefringent” and “dichroic” are often used without qualification, particularly if their meaning is clear from the context. To describe such birefringent and dichroic media at the phenomenological level, the constitutive relations (2.7)–(2.9) must be modified somewhat. We shall encounter specific examples in later chapters.

In addition to its handedness, a vibration ellipse is characterized by its *ellipticity*, the ratio of the length of its semiminor axis to that of its semimajor axis, and its *azimuth*, the angle between the semimajor axis and an arbitrary reference direction (Fig. 2.13). Handedness, ellipticity, and azimuth, together with irradiance, are the *ellipsometric parameters* of a plane wave.

2.11.1 Stokes Parameters

Although the ellipsometric parameters completely specify a monochromatic wave of given frequency and are readily visualized, they are not particularly conducive to understanding the transformations of polarized light. Moreover, they are difficult to measure directly (with the exception of irradiance, which can easily be measured with a suitable detector) and are not adaptable to a

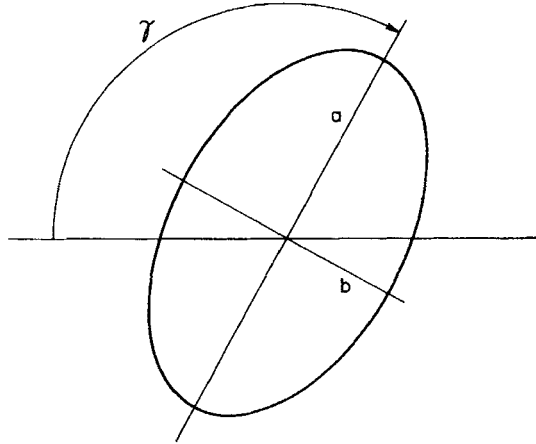


Figure 2.13 Vibration ellipse with ellipticity b/a and azimuth γ .

discussion of partially polarized light. The irradiances of two incoherently superposed beams are additive, but no such additivity exists for the other three ellipsometric parameters. As we shall see, the *Stokes parameters* are an equivalent description of polarized light, but one of greater usefulness, particularly in scattering problems.

In previous sections we stated that the polarization state of a wave may be changed by interaction with a suitable optical system (e.g., reflection at the polarizing angle or transmission through a dichroic medium). An arbitrary monochromatic wave may be expressed as a superposition of two *orthogonal* components: horizontal and vertical; right-circular and left-circular; and so on. This decomposition is more than just a mathematical device: we can construct polarizers that have the property of transmitting only one of these components. For the moment we regard such a polarizer P as a black box into which light is fed.

Let us consider a series of hypothetical experiments which can be performed with an arbitrary monochromatic beam, a detector, and various polarizers (Fig. 2.14). The detector responds to irradiance independently of the polarization state, and the polarizers are assumed to be ideal: they do not change the amplitude of the transmitted component. The electric field \mathbf{E} referred to orthogonal axes $\hat{\mathbf{e}}_{\parallel}$ and $\hat{\mathbf{e}}_{\perp}$, which we shall call “horizontal” and “vertical,” respectively, is

$$\mathbf{E} = \mathbf{E}_0 \exp(i\mathbf{k}z - i\omega t); \quad \mathbf{E}_0 = E_{\parallel} \hat{\mathbf{e}}_{\parallel} + E_{\perp} \hat{\mathbf{e}}_{\perp},$$

$$E_{\parallel} = a_{\parallel} e^{-i\delta_{\parallel}}; \quad E_{\perp} = a_{\perp} e^{-i\delta_{\perp}}.$$

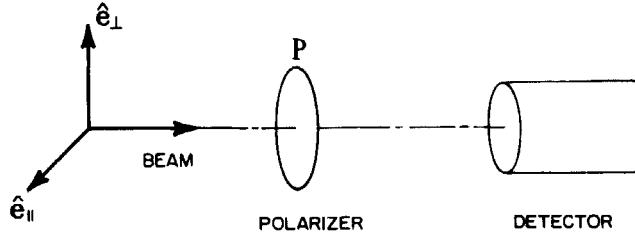


Figure 2.14 The detector measures the irradiance of the beam transmitted by the polarizer P .

Experiment I. No Polarizer. If there is no polarizer in the beam the irradiance I recorded by the detector is $E_\parallel E_\parallel^* + E_\perp E_\perp^*$, where for convenience we omit the factor $k/2\omega\mu_0$.

Experiment II. Horizontal and Vertical Polarizers. (1) Let P be a horizontal polarizer; the amplitude of the transmitted wave is E_\parallel and the irradiance I_\parallel recorded by the detector is $E_\parallel E_\parallel^*$. (2) Let P be a vertical polarizer; the amplitude of the transmitted wave is E_\perp and the irradiance I_\perp recorded by the detector is $E_\perp E_\perp^*$. The difference between these two measured irradiances is

$$I_\parallel - I_\perp = E_\parallel E_\parallel^* - E_\perp E_\perp^*.$$

Experiment III. $+45^\circ$ and -45° Polarizers. To analyze this experiment it is convenient to introduce another orthonormal set of basis vectors \hat{e}_+ and \hat{e}_- , which are obtained by rotating \hat{e}_\parallel by $+45^\circ$ and -45° (Fig. 2.15):

$$\hat{e}_+ = \frac{1}{\sqrt{2}} (\hat{e}_\parallel + \hat{e}_\perp), \quad \hat{e}_- = \frac{1}{\sqrt{2}} (\hat{e}_\parallel - \hat{e}_\perp).$$

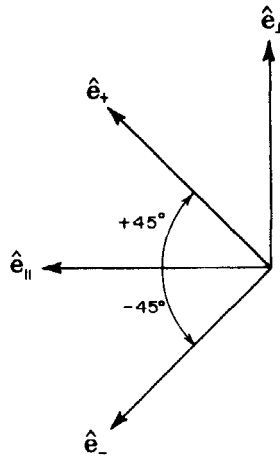


Figure 2.15 Basis vectors \hat{e}_+ and \hat{e}_- .

The electric field \mathbf{E}_0 may be written $\mathbf{E}_0 = E_+ \hat{\mathbf{e}}_+ + E_- \hat{\mathbf{e}}_-$, where

$$E_+ = \frac{1}{\sqrt{2}} (E_{\parallel} + E_{\perp}), \quad E_- = \frac{1}{\sqrt{2}} (E_{\parallel} - E_{\perp}).$$

(1) Let P be a $+45^\circ$ polarizer; the amplitude of the transmitted wave is $(E_{\parallel} + E_{\perp})/\sqrt{2}$ and its irradiance I_+ is $(E_{\parallel}E_{\parallel}^* + E_{\parallel}E_{\perp}^* + E_{\perp}E_{\parallel}^* + E_{\perp}E_{\perp}^*)/2$.
 (2) Let P be a -45° polarizer; the irradiance of the transmitted wave is $I_- = (E_{\parallel}E_{\parallel}^* - E_{\parallel}E_{\perp}^* - E_{\perp}E_{\parallel}^* + E_{\perp}E_{\perp}^*)/2$. The difference between these two irradiances is

$$I_+ - I_- = E_{\parallel}E_{\perp}^* + E_{\perp}E_{\parallel}^*.$$

Experiment IV. Circular Polarizers. We need to introduce one more set of basis vectors $\hat{\mathbf{e}}_R$ and $\hat{\mathbf{e}}_L$:

$$\hat{\mathbf{e}}_R = \frac{1}{\sqrt{2}} (\hat{\mathbf{e}}_{\parallel} + i\hat{\mathbf{e}}_{\perp}), \quad \hat{\mathbf{e}}_L = \frac{1}{\sqrt{2}} (\hat{\mathbf{e}}_{\parallel} - i\hat{\mathbf{e}}_{\perp}).$$

These basis vectors represent right-circularly and left-circularly polarized waves and are orthonormal in the sense that

$$\hat{\mathbf{e}}_R \cdot \hat{\mathbf{e}}_R^* = 1, \quad \hat{\mathbf{e}}_L \cdot \hat{\mathbf{e}}_L^* = 1, \quad \hat{\mathbf{e}}_R \cdot \hat{\mathbf{e}}_L^* = 0.$$

The incident field may be written $\mathbf{E}_0 = E_R \hat{\mathbf{e}}_R + E_L \hat{\mathbf{e}}_L$, where

$$E_R = \frac{1}{\sqrt{2}} (E_{\parallel} - iE_{\perp}), \quad E_L = \frac{1}{\sqrt{2}} (E_{\parallel} + iE_{\perp}).$$

(1) Let P be a right-handed polarizer; the transmitted irradiance I_R is $(E_{\parallel}E_{\parallel}^* - iE_{\parallel}^*E_{\perp} + iE_{\perp}^*E_{\parallel} + E_{\perp}E_{\perp}^*)/2$. (2) Let P be a left-handed polarizer; the transmitted irradiance I_L is $(E_{\parallel}E_{\parallel}^* + iE_{\perp}^*E_{\parallel} - iE_{\parallel}^*E_{\perp} + E_{\perp}E_{\perp}^*)/2$. The difference between these two irradiances is

$$I_R - I_L = i(E_{\perp}^*E_{\parallel} - E_{\parallel}^*E_{\perp}).$$

We have now done enough thought experiments to determine the *Stokes parameters* I, Q, U, V :

$$\begin{aligned} I &= E_{\parallel}E_{\parallel}^* + E_{\perp}E_{\perp}^* = a_{\parallel}^2 + a_{\perp}^2, \\ Q &= E_{\parallel}E_{\parallel}^* - E_{\perp}E_{\perp}^* = a_{\parallel}^2 - a_{\perp}^2, \\ U &= E_{\parallel}E_{\perp}^* + E_{\perp}E_{\parallel}^* = 2a_{\parallel}a_{\perp} \cos \delta, \\ V &= i(E_{\parallel}E_{\perp}^* - E_{\perp}E_{\parallel}^*) = 2a_{\parallel}a_{\perp} \sin \delta, \end{aligned} \tag{2.80}$$

where the phase difference δ is $\delta_{\parallel} - \delta_{\perp}$. Note that we have omitted the factor $k/2\omega\mu_0$ from (2.80); it is unimportant because relative, rather than absolute, irradiances are what are usually measured. Our notation for the Stokes parameters is by no means universal: there are many other symbols in use. Stokes (1852) himself used A, B, C, D ; I, Q, U, V are favored by Walker (1954), Chandrasekhar (1950), and van de Hulst (1957); Perrin (1942) and Shurcliff (1962) find I, M, C, S more to their liking; Collett (1968) prefers s_0, s_1, s_2, s_3 . Rozenberg (1960) makes no attempt to conceal his contempt for the "irrationality" of the symbols usually employed in the "foreign literature," and he offers us S_1, S_2, S_3, S_4 . To add to the confusion, the definition of the Stokes parameters can be changed without serious damage: various linear combinations of the Stokes parameters (2.80), particularly I and Q , can be, and are, used as suitable Stokes parameters. Caution is therefore in order when one leaves the pages of this book.

The Stokes parameters are related to the ellipsometric parameters as follows:

$$\begin{aligned} I &= c^2, \\ Q &= c^2 \cos 2\eta \cos 2\gamma, \\ U &= c^2 \cos 2\eta \sin 2\gamma, \\ V &= c^2 \sin 2\eta, \end{aligned} \tag{2.81}$$

where

$$c^2 = a^2 + b^2 = (\text{semimajor axis})^2 + (\text{semiminor axis})^2,$$

$$\gamma = \text{clockwise angle between } \hat{\mathbf{e}}_{\parallel} \text{ and major axis (azimuth) } (0 \leq \gamma \leq \pi),$$

$$|\tan \eta| = \frac{b}{a} \text{ (ellipticity)} \left(-\frac{\pi}{4} \leq \eta \leq \frac{\pi}{4} \right).$$

The sign of V specifies the handedness of the vibration ellipse: positive denotes right-handed and negative denotes left-handed. We also have the relations

$$\tan 2\gamma = \frac{U}{Q}, \quad \tan 2\eta = \frac{V}{\sqrt{Q^2 + U^2}}. \tag{2.82}$$

Thus, the Stokes parameters are equivalent to the ellipsometric parameters; although less easily visualized, they are operationally defined in terms of measurable quantities (irradiances). Additional advantages of the Stokes parameters will become evident as we proceed. Note that Q and U depend on the choice of horizontal and vertical directions. If the basis vectors $\hat{\mathbf{e}}_{\parallel}$ and $\hat{\mathbf{e}}_{\perp}$

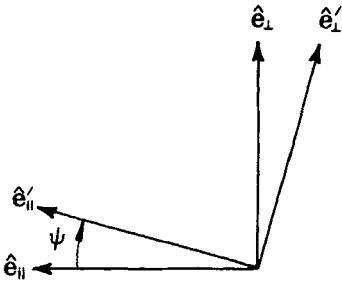


Figure 2.16 Rotation of basis vectors.

are rotated through an angle ψ (Fig. 2.16), the transformation from (I, Q, U, V) to Stokes parameters (I', Q', U', V') relative to the rotated axes \hat{e}'_{\parallel} and \hat{e}'_{\perp} is

$$\begin{pmatrix} I' \\ Q' \\ U' \\ V' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\psi & \sin 2\psi & 0 \\ 0 & -\sin 2\psi & \cos 2\psi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}. \quad (2.83)$$

It is fairly obvious from either (2.81) or (2.83) that there are three quantities associated with the Stokes parameters that are invariant under rotation of the reference directions: I , $Q^2 + U^2$, and V . In addition, the Stokes parameters are not all independent:

$$I^2 = Q^2 + U^2 + V^2.$$

A few representative sets of Stokes parameters, written as column vectors, are shown in Table 2.2; the irradiance I is normalized to unity.

Although a strictly monochromatic wave, one for which the time dependence is $\exp(-i\omega t)$, has a well-defined vibration ellipse, not all waves do. Let us consider a nearly monochromatic, or *quasi-monochromatic* beam:

$$\mathbf{E} = \mathbf{E}_0(t) \exp(ikz - i\omega t), \quad \mathbf{E}_0(t) = E_{\parallel}(t) \hat{e}_{\parallel} + E_{\perp}(t) \hat{e}_{\perp},$$

where the complex amplitudes E_{\parallel} and E_{\perp} are now functions of time but vary slowly over time intervals of the order of the period $2\pi/\omega$. However, for time intervals long compared with the period, the amplitudes fluctuate in some manner, perhaps independently of each other, or perhaps with some correlation. If $E_{\parallel}(t)$ and $E_{\perp}(t)$ are completely *uncorrelated*, the beam is said to be *unpolarized*; so-called natural light (e.g., light from the sun, incandescent and fluorescent lamps) is unpolarized. In such a beam of light the electric vector traces out a vibration ellipse the parameters of which—handedness, ellipticity, and azimuth—vary slowly in time. Moreover, there is no preferred vibration ellipse: over a sufficiently long period of time vibration ellipses of all shapes, orientations, and handedness will have been traced out. Hurwitz (1945) has

Table 2.2 Stokes Parameters for Polarized Light

<i>Linearly Polarized</i>				
0°	90°	+45°	-45°	γ
\leftrightarrow	\updownarrow	\nearrow	\nwarrow	
$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ \cos 2\gamma \\ \sin 2\gamma \\ 0 \end{pmatrix}$
<i>Circularly Polarized</i>				
	Right		Left	
	\curvearrowright		\curvearrowleft	
	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$		$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$	

discussed the statistical properties of unpolarized light in an interesting and instructive paper. If E_{\parallel} and E_{\perp} are *completely correlated*, the light is said to be *polarized*. This definition of polarization includes strictly monochromatic light but is somewhat more general: a_{\parallel} , a_{\perp} , δ_{\parallel} , δ_{\perp} may separately fluctuate provided that the ratio a_{\parallel}/a_{\perp} of the real amplitudes and the phase difference $\delta_{\parallel} - \delta_{\perp}$ are independent of time. If E_{\parallel} and E_{\perp} are *partially correlated*, the light is said to be *partially polarized*. A partially polarized beam exhibits a preference for handedness, or ellipticity, or azimuth. But this preference is not perfect: there is some statistical fluctuation.

The Stokes parameters of a quasi-monochromatic beam are given by

$$\begin{aligned}
 I &= \langle E_{\parallel} E_{\parallel}^* + E_{\perp} E_{\perp}^* \rangle, \\
 Q &= \langle E_{\parallel} E_{\parallel}^* - E_{\perp} E_{\perp}^* \rangle, \\
 U &= \langle E_{\parallel} E_{\perp}^* + E_{\perp} E_{\parallel}^* \rangle, \\
 V &= i \langle E_{\parallel} E_{\perp}^* - E_{\perp} E_{\parallel}^* \rangle,
 \end{aligned} \tag{2.84}$$

where the angular brackets indicate time averages over an interval long compared with the period. From (2.84) it follows that

$$Q^2 + U^2 + V^2 = I^2 - 4(\langle a_{\parallel}^2 \rangle \langle a_{\perp}^2 \rangle - \langle a_{\parallel} a_{\perp} e^{i\delta} \rangle \langle a_{\parallel} a_{\perp} e^{-i\delta} \rangle),$$

which implies that

$$I^2 \geq Q^2 + U^2 + V^2. \tag{2.85}$$

Equality holds if the light is polarized; if the light is unpolarized, $Q = U = V = 0$. The inequality (2.85) leads naturally to the notion of *degree of polarization* $\sqrt{Q^2 + U^2 + V^2}/I$, as well as *degree of linear polarization* $\sqrt{Q^2 + U^2}/I$, and *degree of circular polarization* V/I . For a partially polarized beam the sign of V indicates the preferential handedness of the vibration ellipses traced out by the electric vector: positive indicates right-handed and negative indicates left-handed. We may interpret the quantities U/Q and $V/\sqrt{Q^2 + U^2}$ [see (2.82)] as specifying the preferential azimuth and ellipticity of the vibration ellipses.

If two or more quasi-monochromatic beams propagating in the same direction are superposed *incoherently*, that is to say, there is no fixed relationship among the phases of the separate beams, the total irradiance is merely the sum of the individual beam irradiances. Because the definition of the Stokes parameters involves only irradiances, it follows that the Stokes parameters of a collection of incoherent sources are additive.

In the derivations above of the Stokes parameters we began with monochromatic light and then extended our results to the more general case of quasi-monochromatic light. However, the operational definition of the Stokes parameters in terms of a set of elementary experiments involving a detector and various polarizers, as opposed to the formal mathematical definitions (2.80) and (2.84), is independent of any assumed properties of the beam. Unless otherwise stated, we shall assume that all beams of interest are quasi-monochromatic, which includes as a special case monochromatic light.

2.11.2 Mueller Matrices

We may represent a beam of arbitrary polarization, including partially polarized light, by a column vector, the *Stokes vector*, the four elements of which are the Stokes parameters. In general, the state of polarization of a beam is changed on interaction with an optical element (e.g., polarizer, retarder, reflector, scatterer). Thus, it is possible to represent such optical elements by a 4×4 matrix (Mueller, 1948). The *Mueller matrix* describes the relation between "incident" and "transmitted" Stokes vectors; by "incident" is meant before interaction with the optical element, and by "transmitted" is meant after interaction. As an example, consider the Mueller matrix for an *ideal linear polarizer*. Such a polarizer transmits, without change of amplitude, only electric field components parallel to a particular axis called the *transmission axis*. Electric field components in other directions are completely removed from the transmitted beam by some means which we need not explicitly consider. The relation between incident field components ($E_{\parallel i}$, $E_{\perp i}$) and field components ($E_{\parallel t}$, $E_{\perp t}$) transmitted by the polarizer is

$$\begin{pmatrix} E_{\parallel t} \\ E_{\perp t} \end{pmatrix} = \begin{pmatrix} \cos^2 \xi & \sin \xi \cos \xi \\ \sin \xi \cos \xi & \sin^2 \xi \end{pmatrix} \begin{pmatrix} E_{\parallel i} \\ E_{\perp i} \end{pmatrix}, \quad (2.86)$$

where ξ is the (smallest) angle between $\hat{\mathbf{e}}_{\parallel}$ and the transmission axis. After a bit of algebra, we obtain from (2.86) the Mueller matrix for an ideal linear polarizer:

$$\frac{1}{2} \begin{pmatrix} 1 & \cos 2\xi & \sin 2\xi & 0 \\ \cos 2\xi & \cos^2 2\xi & \cos 2\xi \sin 2\xi & 0 \\ \sin 2\xi & \sin 2\xi \cos 2\xi & \sin^2 2\xi & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2.87)$$

The irradiance transmitted by the linear polarizer (2.87) is

$$I_t = \frac{1}{2}(I_i + Q_i \cos 2\xi + U_i \sin 2\xi).$$

Thus, as the polarizer is rotated so that ξ varies, the irradiance I_t also varies. The maximum and minimum values of I_t occur for $\xi = \gamma$ and $\xi = \gamma + \pi/2$, respectively, where $\tan 2\gamma = U_i/Q_i$:

$$\begin{aligned} I_{\max} &= \frac{1}{2}(I_i + Q_i \cos 2\gamma + U_i \sin 2\gamma), \\ I_{\min} &= \frac{1}{2}(I_i - Q_i \cos 2\gamma - U_i \sin 2\gamma). \end{aligned} \quad (2.88)$$

From (2.88) we obtain the degree of linear polarization

$$\frac{\sqrt{Q_i^2 + U_i^2}}{I_i} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}.$$

Therefore, by rotating a linear polarizer in an arbitrary beam and noting the maximum and minimum transmitted irradiance, the degree of linear polarization can be measured regardless of the value of V .

An *ideal linear retarder* divides a given incident electric vector into two linearly polarized components E_1 and E_2 , which are mutually orthogonal, and introduces a phase difference $\delta_1 - \delta_2$ between them; there is no diminution of irradiance. Thus, the relation between incident field components and field components transmitted by such a retarder is

$$\begin{pmatrix} E_{\parallel t} \\ E_{\perp t} \end{pmatrix} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} e^{i\delta_1} & 0 \\ 0 & e^{i\delta_2} \end{pmatrix} \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} E_{\parallel i} \\ E_{\perp i} \end{pmatrix}, \quad (2.89)$$

where β is the angle between $\hat{\mathbf{e}}_{\parallel}$ and $\hat{\mathbf{e}}_1$ (Fig. 2.17). It is straightforward, but laborious, to show that (2.89) yields the Mueller matrix for an ideal linear

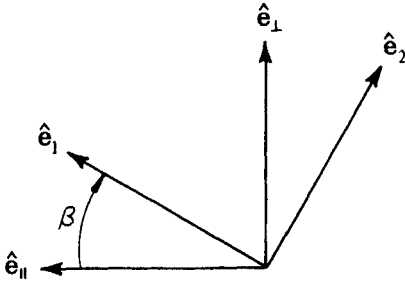


Figure 2.17 \hat{e}_1 and \hat{e}_2 specify the axes of an ideal linear retarder.

retarder:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & C^2 + S^2 \cos \delta & SC(1 - \cos \delta) & -S \sin \delta \\ 0 & SC(1 - \cos \delta) & S^2 + C^2 \cos \delta & C \sin \delta \\ 0 & S \sin \delta & -C \sin \delta & \cos \delta \end{pmatrix}, \quad (2.90)$$

where $C = \cos 2\beta$, $S = \sin 2\beta$, and the *retardance* δ is $\delta_1 - \delta_2$.

The usefulness of the Mueller formulation becomes apparent when we realize that Mueller matrices give us a simple means of determining the state of polarization of a beam transmitted by an optical element for an arbitrarily polarized incident beam. Moreover, if a series of optical elements is interposed in a beam, the combined effect of all these elements may be determined by merely multiplying their associated Mueller matrices. As an example, let us consider how a circular polarizer can be constructed by superposing a linear polarizer and a linear retarder. The beam is first incident on a linear polarizer with horizontal transmission axis ($\xi = 0^\circ$), the Mueller matrix for which is obtained from (2.87):

$$\frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2.91)$$

The beam transmitted by the polarizer (2.91) is then incident on a retarder with $\delta = 90^\circ$ and $\beta = 45^\circ$, the Mueller matrix for which is obtained from (2.90):

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The combined effect of polarizer and retarder is obtained by matrix multiplica-

tion:

$$\frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \quad (2.92)$$

Thus, if unpolarized light or, indeed, light of arbitrary polarization is incident on the optical system described by the Mueller matrix (2.92), the transmitted light will be 100% right-circularly polarized. Note that matrix multiplication is not commutative: the order of elements in a train must be properly taken into account. Further details about Mueller matrices and experimental means for realizing polarizers, retarders, and other optical elements are found in the excellent book by Shurcliff (1962).

NOTES AND COMMENTS

There are many good books on Fourier transforms. One we have found particularly useful for our purposes is by Champeney (1973).

Toll (1956) examines the logical foundations of causality and the dispersion relations. Goldberger (1960) begins his article with a good historical survey; another discussion within the context of high-energy physics is by Scadron (1979, pp. 326–329), whereas optical properties of solids form the backdrop for Stern's (1963) discussion. And an entire book by Nussenzveig (1972) is devoted to dispersion relations.

Spatial dispersion is the subject of a review article by Rukhadze and Silin (1961) and a book by Agranovich and Ginzburg (1966).

A good review article on optical constants and their measurement is that by Bell (1967). Determination of optical constants from reflectance measurements is treated by Wendlandt and Hecht (1966) and from internal reflection spectroscopy by Harrick (1967). Ellipsometric techniques are discussed at length by Azzam and Bashara (1977).

In Section 2.10 we made an analogy between slabs and particles while cautioning not to push this analogy too far. That caution is necessary is evident from calculations of volumetric absorption by slabs and spheres of the same material (Faxvog and Roessler, 1981).

McMaster (1954) takes a quantum-mechanical approach to the Stokes parameters and polarized light. Two books are devoted entirely to polarization: Shurcliff (1962) and Clarke and Grainger (1971). And a splendid collection of papers on many aspects of polarized light has been edited by Gehrels (1974a). Another collection worth consulting is that compiled by Swindell (1975); it contains several of the classical papers on polarization.