1

Light: Electromagnetic Radiation

1.1 Introduction

In order to understand the imaging properties of the nonlinear optical microscope, we first have to have a basic understanding of light itself. In particular, a description of light in terms of propagating waves is needed to model the formation of the tightly focused volume. Fortunately, such a description is well established, and in this chapter, we review two useful forms of propagating light, namely the plane wave and the spherical wave. We also summarize helpful notations for the polarization state of light, and briefly discuss relevant expressions for reflected and transmitted light. The final aim of this chapter is to study the way in which a thin lens modifies an incident plane wave.

1.2 Electromagnetic Fields

The study of electromagnetic radiation is fascinating, and many aspects of electromagnetic radiation are worthy topics of discussion. In this book, however, we focus only on the bare essentials. Our goal is to find good descriptions of propagating light, which we can then use to model the tightly focused volume in the microscope. To arrive at such descriptions, we first have to glance at Maxwell's equations and the wave equation that follows from them.

1.2.1 Vector Fields

Light is electromagnetic radiation. In a classical description, light radiates through space as propagating electromagnetic waves. A wave is defined through its electric and magnetic fields, which oscillate in time in a synchronized manner. In vacuum, the electric and magnetic fields are indicated as $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$, respectively, which are position dependent *vector fields* that vary as a function of time. In Cartesian coordinates, defined by the axes x, y, and z, the electric field takes on the following form

$$\mathbf{E}(\mathbf{r}, t) = E_x(\mathbf{r}, t)\hat{\mathbf{e}}_x + E_y(\mathbf{r}, t)\hat{\mathbf{e}}_y + E_z(\mathbf{r}, t)\hat{\mathbf{e}}_z \tag{1.1}$$

where $\left(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z\right)$ are unit vectors that point in the (x, y, z) directions, respectively. The electric field is expressed in SI units of V/m. At a given point $\mathbf{r}$ in space, the electric field is a vector with projections of magnitude $\left(E_x, E_y, E_z\right)$ along the respective Cartesian coordinates, see Figure 1.1.

Foundations of Nonlinear Optical Microscopy, First Edition. Eric Olaf Potma.

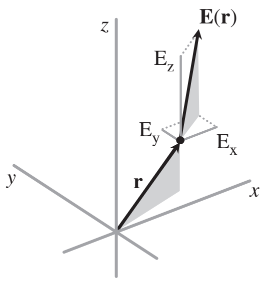

Figure 1.1 The electric field as a vector field. The vector **r** is a position vector indicating the location at which the field is considered. The field vector **E** at location **r** has projections of magnitude E_x, E_y, and E_z along the (x, y, z) coordinates, respectively.

The projections are also referred to as the orthogonal *polarization* components of the field. The corresponding expression for the magnetic field is similar, with **E** replaced by **B**, which has units of $\text{V} \cdot \text{s/m}^2$.

Electromagnetic waves in vacuum propagate at the speed of light, defined as $c = 1/\sqrt{\mu_0 \varepsilon_0}$. The quantity μ_0 is called the vacuum permeability, which in classical terms relates to the magnetic inductance of a vacuum. Similarly, ε_0, called vacuum permittivity, is a measure of the capacitance of a vacuum. Together, μ_0 and ε_0 pose a limit to how fast an electromagnetic disturbance can travel through a vacuum. The established value for the vacuum permeability is $\mu_0 = 1.257 \times 10^{-6}$ H/m. Using $c = 2.998 \times 10^8$ m/s, the value for the vacuum permittivity is $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m.

1.2.2 Wave Equation in Vacuum

Electromagnetic waves, in the form of the electric and magnetic fields, are not arbitrarily defined. Instead, the fields are described by a set of equations known as Maxwell's equations. In vacuum, the equations in differential form are written as

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B} \tag{1.2}$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial}{\partial t}\mathbf{E} \tag{1.3}$$

$$\nabla \cdot \mathbf{E} = 0 \tag{1.4}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.5}$$

where the curl operator $\nabla\times$ indicates the circulation density of the field and the divergence operator $\nabla\cdot$ denotes the flux density of the field. Here, we have written the electric field $\mathbf{E}(\mathbf{r}, t)$ and magnetic field $\mathbf{B}(\mathbf{r}, t)$ in shorthand form as **E** and **B**, respectively. Maxwell's equations show that the electric and magnetic fields are interdependent. For instance, equation (1.2), known as Maxwell–Faraday's law, states that a time-varying magnetic field induces an electric field. Similarly, a time-varying electric field gives rise to a magnetic field, as described by equation (1.3). The remaining two expressions, equations (1.4) and (1.5), indicate that in vacuum the flux density of the electric and magnetic fields is zero.

Maxwell's equations can be rewritten to bring out the wave character of the electromagnetic field. For this purpose, we take the curl of equation (1.2) and use the vector identity $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E}$. We then use the fact that $\nabla \cdot \mathbf{E}$ is zero, as per equation (1.4), and use equation (1.3) to write the curl of **B** in terms of the time derivative of **E**. These operations result in the following equation

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0 \tag{1.6}$$

This expression shows that the second-order derivative of the field in space is proportional to the field's second-order derivative in time, a characteristic of a wave equation. Equation (1.6), therefore, is known as Maxwell's wave equation in vacuum. A similar form can be derived for the magnetic field.

We are interested in time-harmonic solutions of the form $\mathcal{E}(\mathbf{r})e^{-i\omega t}$, in which the spatial part of the solution is expressed as $\mathcal{E}(\mathbf{r})$, a complex quantity, whereas the temporal part is described by $e^{-i\omega t}$.[1] More generally, we can write a monochromatic, time-harmonic field mode that oscillates at angular frequency ω as

$$\begin{aligned}\mathbf{E}(\mathbf{r},t) &= \frac{1}{2}\left\{\mathcal{E}(\mathbf{r})e^{-i\omega t} + \mathcal{E}^*(\mathbf{r})e^{i\omega t}\right\} \\ &= \mathbf{E}(\mathbf{r})e^{-i\omega t} + \mathbf{E}^*(\mathbf{r})e^{i\omega t}\end{aligned} \tag{1.7}$$

where the quantity $\mathbf{E}(\mathbf{r}) = \frac{1}{2}\mathcal{E}(\mathbf{r})$ is introduced as a matter of convenience, in order to avoid the explicit use of the $\frac{1}{2}$ prefactor. The electric field is a real quantity, but it is expressed here as a sum of complex functions. For mathematical purposes, it is often more convenient to work with the complex function $\mathbf{E}(\mathbf{r})e^{-i\omega t}$ than the full expression of the field given in (1.7). The actual (real) electric field can then be obtained by taking the real part of the complex function $2\mathbf{E}(\mathbf{r})e^{-i\omega t}$.

Example 1.1 Show that expression (1.7) represents a real quantity.

Solution The complex conjugate of $z = \mathcal{E}(\mathbf{r})e^{-i\omega t}$ is $z^* = \mathcal{E}^*(\mathbf{r})e^{i\omega t}$, which means that expression (1.7) can be written as

$$\frac{1}{2}\left\{\mathcal{E}(\mathbf{r})e^{-i\omega t} + \mathcal{E}^*(\mathbf{r})e^{i\omega t}\right\} = \frac{1}{2}(z + z^*) = \frac{1}{2}(2\mathrm{Re}\,\{z\}) = \mathrm{Re}\,\{z\}$$

The field in equation (1.7) thus equals $\mathrm{Re}\left\{\mathcal{E}(\mathbf{r})e^{-i\omega t}\right\} = 2\mathrm{Re}\left\{\mathbf{E}(\mathbf{r})e^{-i\omega t}\right\}$, which is a real quantity.

By substituting the complex time-harmonic field into equation (1.6), the wave equation can be rewritten as

$$\left(\nabla^2 + k^2\right)\mathbf{E}(\mathbf{r}) = 0 \tag{1.8}$$

where $k \equiv \omega/c$ is called the angular wave number. Equation (1.8) is known as the vector Helmholtz equation, which expresses the spatial properties of the field. If a solution for $\mathbf{E}$ can be found that complies with equation (1.8), it is also a valid solution of Maxwell's equations. Section 1.3 discusses several useful solutions of the Helmholtz equation.

1.2.3 Fields and Matter

We can measure the presence of electromagnetic waves because its electric and magnetic fields exert a force on electric charges. In general, the Lorentz force experienced by a charge q moving at a velocity $\mathbf{v}$ in the presence of an electromagnetic field is given as

$$\mathbf{F}(\mathbf{r},t) = q\left\{\mathbf{E}(\mathbf{r},t) + \mathbf{v}\times\mathbf{B}(\mathbf{r},t)\right\} \tag{1.9}$$

Because the electromagnetic field interacts with charges, it can bring about change to matter. Of particular relevance to the topic of this book is the force experienced by the electrons bound to atoms that make up materials, such as optical glasses or biological samples inspected in

1 In this book we use the $e^{-i\omega t}$ convention for the complex time-harmonic field.

microscopy experiments. Due to the action of the field, the electrons will move under the influence of the time-periodic electromagnetic force, thereby inducing a time-varying polarization in the material. Vice versa, the presence of charges can also alter the properties of the electromagnetic field. For instance, the induced polarization in the material forms the basis for the exchange of energy between fields and matter, as is the case in the process of optical absorption. In addition, the induced motion of charges in matter is also responsible for the observed propagation effects of electromagnetic waves as they encounter materials, such as the redirection of the wave's propagation direction at interfaces or the focusing of waves by lenses.

To understand these effects, we first need to consider the behavior of fields in the presence of charges and currents in a certain volume, as well as how the fields might, in turn, alter the material properties within that volume. Maxwell's equations (1.2)–(1.5) are only valid for electromagnetic fields in vacuum. To include the effects of current density $\mathbf{j}(\mathbf{r}, t)$ and charge density $\rho(\mathbf{r}, t)$ on the fields, as well as the response of the material to the presence of the fields, the equations can be expanded as

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B} \tag{1.10}$$

$$\nabla \times \mathbf{H} = \mathbf{j} + \frac{\partial}{\partial t}\mathbf{D} \tag{1.11}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.12}$$

$$\nabla \cdot \mathbf{D} = \rho \tag{1.13}$$

The two new quantities are the electric displacement field $\mathbf{D}(\mathbf{r}, t)$ and the magnetizing field $\mathbf{H}(\mathbf{r}, t)$, which are defined through the following so-called "constitutive relations"

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P} \tag{1.14}$$

$$\mathbf{B} = \mu_0\mathbf{H} + \mu_0\mathbf{M} \tag{1.15}$$

The electric displacement field describes the combined effect of the electric field and the polarization density $\mathbf{P}(\mathbf{r}, t)$ in the material (in units of $\mathrm{C/m^2}$) caused by $\mathbf{E}(\mathbf{r}, t)$. The $\mathbf{B}(\mathbf{r}, t)$ field in equation (1.15) now includes both the magnetizing field $\mathbf{H}(\mathbf{r}, t)$ as well as the material's magnetization density $\mathbf{M}(\mathbf{r}, t)$ (in units of A/m) in the presence of the magnetizing field. In vacuum $\mathbf{B} = \mu_0\mathbf{H}$, i.e. the $\mathbf{B}$ field is directly proportional to the magnetizing field. This is no longer the case in matter, and to indicate this difference, the $\mathbf{B}$ field is often referred to as *magnetic induction* or *magnetic flux density*. In this book, we refer to $\mathbf{B}$ as the magnetic field, i.e. the field that results from the magnetizing field and/or the magnetization of matter.

For nondispersive and isotropic materials, the *linear* response of the material to applied electric and magnetic fields is given as

$$\mathbf{P} = \varepsilon_0\chi_e\mathbf{E} \tag{1.16}$$

$$\mathbf{M} = \chi_m\mathbf{H} \tag{1.17}$$

Equation (1.16) shows that the polarization of the material grows as the electric field grows. The strength of the polarization is further determined by χ_e, a dimensionless quantity called the electric susceptibility of the material. Materials that are more responsive to the electric field have higher χ_e values, producing a higher polarization density. Equation (1.17) shows a similar relation for the magnetization of the material, which is directly proportional to the magnetizing field and χ_m, the magnetic susceptibility of the material. Using the definitions for **P** and **M**, equations (2.84) and (1.15) can be recast as

$$\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} \tag{1.18}$$

$$\mathbf{B} = \mu_0 \mu_r \mathbf{H} \tag{1.19}$$

where the relative permittivity ε_r and the relative permeability μ_r are defined through

$$\varepsilon_r = (1 + \chi_e) \tag{1.20}$$

$$\mu_r = (1 + \chi_m) \tag{1.21}$$

If the material is spectrally dispersive, i.e. the material properties change with ω, then the quantities above are functions of ω as well. The resulting quantities $\varepsilon_r(\omega)$ and $\mu_r(\omega)$ are important material parameters, as they describe how the electric and magnetic fields permeate the medium relative to how the fields distribute in a vacuum. Note that ε_r and μ_r are macroscopic quantities in that they describe the field properties within a volume of matter that is much larger than that occupied by a single charge or atom, i.e. they represent quantities averaged over numerous atoms that make up a volume.

For a homogeneous material, the joint effect of the relative permittivity and permeability is captured by the refractive index, defined as

$$\tilde{n}(\omega) = \sqrt{\mu_r(\omega)\varepsilon_r(\omega)} = \sqrt{\{1 + \chi_m(\omega)\}\{1 + \chi_e(\omega)\}} \tag{1.22}$$

The refractive index is an important material-dependent quantity in the context of electromagnetic field propagation. The tilde indicates that the refractive index is a complex quantity, and this quantity can generally be written in terms of its real and imaginary parts

$$\tilde{n}(\omega) = n(\omega) + \mathrm{i}\kappa(\omega) \tag{1.23}$$

where the *index of refraction* n is related to the speed of light propagation in matter and the *extinction coefficient* κ is related to the dissipation of light energy. Recalling that $c = 1/\sqrt{\mu_0\varepsilon_0}$, we see that there is a clear connection between $\tilde{n}$ and the speed of light. Whereas c quantifies the speed by which the electromagnetic field permeates the vacuum, the refractive index provides a measure of how the field permeates a material relative to its permeation in vacuum. The refractive index, and in particular its real part n, can be understood as a correction to the speed of light when the electromagnetic field propagates in a given medium.

1.2.4 Prominence of Electric Field Interactions

It is helpful to establish the relative magnitude by which **E** and **B** are able to drive the movement of charges in matter. For this purpose, we consider the forces exerted by the electromagnetic field on a single charge q by the **E** and **B** fields. From equation (1.9), we can identify the electric force as $\mathbf{F}_\mathrm{e} = q\mathbf{E}$ and the magnetic force as $\mathbf{F}_\mathrm{m} = q\mathbf{v} \times \mathbf{B}$. We next assume that the charge is driven by a time-harmonic electromagnetic field in the form of a propagating plane wave. Section 1.3.1 discusses the properties of plane waves. Here, we are interested in the magnitudes $|\mathbf{F}_\mathrm{e}|$ and $|\mathbf{F}_\mathrm{m}|$ produced by a plane wave as it acts on the charge q. It is not difficult to show that $|\mathbf{F}_\mathrm{e}| = q|\mathbf{E}|$ and that for plane waves the magnitude of the magnetic force can be written as $|\mathbf{F}_\mathrm{m}| = qv|\mathbf{E}|/c$, where

$v = |\mathbf{v}|$. The ratio between the magnitude of the electric and magnetic forces experienced by q is then [1]

$$\frac{|\mathbf{F}_\mathrm{e}|}{|\mathbf{F}_\mathrm{m}|} = \frac{c}{v} \tag{1.24}$$

We next assume that the charge is an electron that is bound to an atom. The velocity of such an electron can be estimated as $v = \hbar/ma_0$, where $\hbar$ is Planck's constant, m is the electron mass, and a_0 is the Bohr radius. The ratio in equation (1.24) can now be written as $1/\alpha$, where α is the fine structure constant. Since $\alpha = e^2/4\pi\varepsilon_0\hbar c \approx 1/137$, where e is the electron charge, we see that the electric force exerted on the electron is more than a hundred times larger than the magnetic force. Therefore, we can quite generally state that optical effects induced by the electric field component of propagating electromagnetic fields are dominant over effects mediated by the magnetic field component. Note that this conclusion pertains to propagating fields only, and that the situation for confined optical fields in the near-zone can be quite different from that stated in equation (1.24).

For the phenomena described in this book, we can safely ignore the light–matter interactions mediated through the magnetic field component and primarily focus on the properties and interactions of the electric field. In terms of the distribution of the electromagnetic field in nonmagnetic materials, we can also assume that $\chi_m \approx 0$ and that $\mu_r = 1$. Under these conditions, the refractive index can be simplified to $\tilde{n} = \sqrt{\varepsilon_r} = \sqrt{1 + \chi_e}$. The complex relative permittivity can now be written as

$$\varepsilon_r(\omega) = \varepsilon_r'(\omega) + i\varepsilon_r''(\omega) = \tilde{n}^2(\omega) \tag{1.25}$$

which allows us to relate the real and imaginary parts of ε_r to the optical quantities n and κ

$$\varepsilon_r' = \mathrm{Re}\left\{\varepsilon_r\right\} = 1 + \mathrm{Re}\{\chi_e\} = n^2 - \kappa^2 \tag{1.26}$$

$$\varepsilon_r'' = \mathrm{Im}\left\{\varepsilon_r\right\} = \mathrm{Im}\{\chi_e\} = 2n\kappa \tag{1.27}$$

In the small dissipation limit, $\kappa \ll 1$ and thus $n^2 \gg \kappa^2$. Using equation (1.26), we then find that $n = (1 + \mathrm{Re}\{\chi_e\})^{1/2} \approx 1 + \frac{1}{2}\mathrm{Re}\{\chi_e\}$, which provides a useful relation between the index of refraction and the real part of the susceptibility. This relation states that, in the limit of a linear material response, $\mathrm{Re}\left\{\chi_e\right\}$ is responsible for the retardation of electromagnetic waves in matter relative to their propagation in vacuum.

1.2.5 Wave Equation in Matter

Maxwell's wave equation (1.6) describes the properties of the electric field $\mathbf{E}$ in vacuum. This equation is no longer valid when the field permeates matter, and corrections are needed. Starting from equations (1.10)–(1.13), we can derive a new wave equation by taking the curl of equation (1.10), using relations (1.14) and (1.15), and substituting equation (1.11). The result is the inhomogeneous wave equation for the electric field

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial}{\partial t}\left\{\mathbf{j} + \frac{\partial \mathbf{P}}{\partial t} + \nabla \times \mathbf{M}\right\} \tag{1.28}$$

The right-hand side of equation (1.28) describes how changes in current densities affect the electric field, in the form of the electrical current density ($\mathbf{j}$), the polarization current density ($\partial\mathbf{P}/\partial t$), and the circulation current density due to the material magnetization ($\nabla \times \mathbf{M}$).

With reference to Section 1.2.4, we only consider materials in which the optical response is dominated by electrons bound to atoms and molecules. Such materials lack free or very loosely bound electrons and can be classified as *dielectric* materials. This includes, for instance, virtually all biological materials, but it excludes metals and strongly conducting semiconducting materials. For dielectric materials, the absence of free charges ensures that $\rho = 0$ and the lack of electrical current densities, i.e. $\mathbf{j} = 0$. In addition, the considered materials are assumed nonmagnetic so that $\mathbf{M} = 0$. Under these conditions, the wave equation can be rewritten as

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial^2}{\partial t^2}\mathbf{P} \tag{1.29}$$

For the materials considered the condition $\nabla \cdot (\varepsilon_0 \varepsilon_r \mathbf{E}) = 0$ holds. This implies that we can use the same vector identity that led to equation (1.6). After expressing the polarization density as in equation (1.16), we obtain

$$-\nabla^2 \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \varepsilon_0 \chi_e \frac{\partial^2}{\partial t^2}\mathbf{E}$$

This result can be simplified by moving the second-order time derivative on the right-hand side to the left-hand side of the equation

$$-\nabla^2 \mathbf{E} + \frac{1}{c^2}(1 + \chi_e)\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0$$

and by using $\tilde{n}^2 = 1 + \chi_e$. This produces the following form for the wave equation in a homogeneous medium

$$\nabla^2 \mathbf{E} - \frac{\tilde{n}^2}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \tag{1.30}$$

Note that the only difference between equation (1.6), which holds for fields in vacuum, and equation (1.30) is the effect of the refractive index $\tilde{n}$ in the latter. Assuming a time-harmonic form of the electric field, equation (1.30) reduces to the vector Helmholtz equation given by (1.8), with the important difference that the angular wave number is now defined as

$$k(\omega) = \frac{\tilde{n}(\omega)\omega}{c} = \tilde{n}(\omega)k_0 \tag{1.31}$$

where k_0 is the angular wavenumber in vacuum. Because $\tilde{n}$ is complex, k is now a complex number as well. In the absence of light absorption by the material, $\tilde{n} \approx n$, and the wave number is given as $k(\omega) = n(\omega)\omega/c$. In the limit that the material responds linearly to the incident light, the propagation of light in matter is thus remarkably similar to propagation in vacuum, with the difference taken up by the new definition of k. In vacuum, $n = 1$ and the wave number is identical to its former form. In matter, the speed of light is altered as $v = c/n$, which, in the absence of light absorption, reduces the angular wave number to $k = \omega/v$. We are now ready to explore solutions to the Helmholtz equation for the electric field in isotropic, homogeneous materials.

1.3 Transverse Waves

Depending on the boundary conditions, finding solutions to the Helmholtz equation can be challenging. However, it is helpful to consider several particular solutions that are physically intuitive and that can be used as a starting point for analyzing the propagation of electromagnetic radiation in more complex situations. In this section, we study fields in an isotropic, homogeneous medium in the absence of any material boundaries. In this case, the solutions of the Helmholz equation are propagating waves in which the electric and magnetic field components oscillate in a direction perpendicular to the wave's propagation direction. Such waves are also called transverse waves. The most prominent example of a transverse wave is the plane wave, discussed in Section 1.3.1. Another example of a transverse wave is the spherical wave, discussed in Section 1.3.2. Both wave forms are relevant to the discussion of wave propagation in optical microscopes.

1.3.1 Plane Waves

Using the field notation defined in equation (1.7), the plane wave solution to the vector Helmholz equation can be written as

$$\mathbf{E}(\mathbf{r},t) = 2\mathrm{Re}\{\mathbf{E}(\mathbf{r})\mathrm{e}^{-\mathrm{i}\omega t}\} \tag{1.32}$$

where

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \tag{1.33}$$

Here, $\mathbf{E}_0$ is a real vector and $\mathbf{k}$ is the *wave vector* of the propagating wave. The scalar product $\mathbf{k}\cdot\mathbf{r}$ determines the spatial propagation phase of the wave. The wave vector is expressed in terms of its Cartesian components as

$$\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}} \tag{1.34}$$

The magnitude of the wave vector is

$$|\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2} = k \tag{1.35}$$

which equals the angular wave number defined in equation (1.31). The direction of the wave vector is $\hat{\mathbf{k}} = \mathbf{k}/k$, which defines the direction of wave propagation. The electric field $\mathbf{E}(\mathbf{r},t)$ of a plane wave is governed by the (real) field vector $2\mathbf{E}_0$, whose direction and magnitude are defined by the radiation source. Although the field vector $\mathbf{E}_0$ itself is independent of location $\mathbf{r}$ and time t, it is modulated by the function $\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, which is a harmonic function of space and time.

Example 1.2 The plane wave expression in equation (1.32) is written in terms of a complex exponential function. Reformulate the plane wave expression in terms of a real function.

Solution Using $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$ and the fact that $\mathbf{E}_0$ is real, we can write

$$\mathbf{E}(\mathbf{r},t) = 2\mathrm{Re}\left\{\mathbf{E}_0\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}-\omega t}\right\} = 2\mathbf{E}_0\cos(\mathbf{k}\cdot\mathbf{r}-\omega t)$$

Note that the factor 2 is a consequence of our definition of the complex function $\mathbf{E}(\mathbf{r})$ in equation (1.7).

The harmonic nature of the plane wave implies that the wave repeats itself after propagating over a certain distance during the time period of an oscillation. Therefore, for the spatial part of the oscillation, there must be a distance λ along the propagation direction $\hat{\mathbf{k}}$ for which $\mathbf{E}(\mathbf{r})$ equals $\mathbf{E}(\mathbf{r}+\lambda\hat{\mathbf{k}})$. We thus require that

$$\mathbf{E}_0\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r})} = \mathbf{E}_0\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}+\lambda\hat{\mathbf{k}})}$$

which implies that $\lambda \mathbf{k} \cdot \hat{\mathbf{k}} = \lambda k = 2\pi$. This leads to the following expression of the angular wave number

$$k = \frac{2\pi}{\lambda} \tag{1.36}$$

where λ is called the *wavelength* of the electromagnetic wave. This expression underscores that the angular wave number can be understood as a spatial frequency. The quantity $1/\lambda$ represents the number of wavelengths per unit distance, or *spatial frequency*, also commonly referred to as the *wave number*. The angular wave number k is 2π radians times the wave number. Comparing equations (1.31) with (1.36), and assuming a nonabsorbing medium, gives

$$\frac{2\pi}{\lambda} = \frac{n(\omega)\omega}{c} \quad \rightarrow \quad \lambda = \frac{1}{n(\omega)} \frac{c}{\nu}$$

where $\nu = \omega/2\pi$ is the oscillation frequency of the electromagnetic field. In vacuum, where $n(\omega) = 1$ and the speed of light is c, the wavelength is $\lambda_{\text{vac}} = c/\nu$. In a homogeneous dielectric medium, the wavelength of the propagating wave is

$$\lambda = \frac{\lambda_{\text{vac}}}{n(\omega)} \tag{1.37}$$

from which we see that the effective wavelength of the field is shortened in media relative to λ_{vac}. Note that the angular frequency of the field, and thus its color, is independent of the propagation medium.

Let us next examine a couple of key properties of the plane wave.

- Transverse wave. In a medium free of unbound charges, i.e. a dielectric medium, we have $\nabla \cdot \mathbf{E} = 0$. Substituting the plane wave solution in this expression yields $\mathrm{i}\mathbf{k} \cdot \mathbf{E} = 0$, and thus

  $$\mathbf{k} \cdot \mathbf{E} = 0$$

  which means that the projection of $\mathbf{k}$ onto $\mathbf{E}$ is zero. In other words, the wave vector of a plane wave is perpendicular to the direction of the electric field vector. Since the wave advances in the direction of the wave vector, the electric field oscillations are perpendicular to the direction of propagation. Waves that have this property are called *transverse* waves. Consequently, plane waves are transverse waves. Using Maxwell's equation (1.11), and applying a similar procedure, we find that $\mathrm{i}\mathbf{k} \times \mathbf{E} = \mathrm{i}\omega\mathbf{B}$. This can be rewritten as

  $$\mathbf{k} \times \mathbf{E} = \omega\mathbf{B}$$

  which means that the $\mathbf{B}$ field is perpendicular to both $\mathbf{k}$ and $\mathbf{E}$. Hence, both the electric and magnetic fields are orthogonal to the wave vector. Furthermore, the expression above shows that the $\mathbf{E}$ and $\mathbf{B}$ fields are oscillating in phase, which is also evident in Figure 1.2.
- Planar wavefront. A plane perpendicular to $\mathbf{k}$ is formed by all the points $\mathbf{r}$ for which $\mathbf{k} \cdot \mathbf{r}$ is a constant. The electric field vector $\mathbf{E}(\mathbf{r})$ is constant throughout such a plane. The idealized plane wave, where the field is constant over the infinite extent of the plane, derives its name from this property. In Example 1.2, the plane wave is seen to propagate according to the function $\cos(\mathbf{k} \cdot \mathbf{r} - \omega t)$, an oscillatory function that evolves with a phase $\varphi = \mathbf{k} \cdot \mathbf{r} - \omega t$, a function of both space and time. The *wavefront* of the wave is defined by the surface formed by all points $(\mathbf{r}, t)$ that give rise to the same value φ, i.e. a surface of constant phase. For a given point in time, the factor ωt is a constant, which means that for a plane wave the wavefront is formed by the points $\mathbf{k} \cdot \mathbf{r} = \varphi + \text{constant}$, which is a plane perpendicular to $\mathbf{k}$. We thus recognize that plane waves have planar wavefronts.

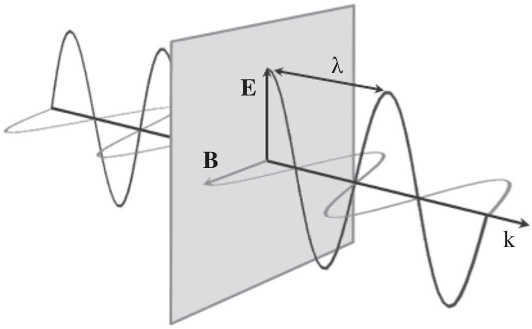

Figure 1.2 Plane wave in vacuum. The electric and magnetic field vectors are oriented perpendicular to the propagation direction $\hat{\mathbf{k}}$. The gray area indicates a plane of constant phase φ. The same phase is found for a parallel plane that is located a distance λ away.

- Phase velocity. In the direction of propagation, we can write the spatial coordinate as the scalar $r_k = \mathbf{r} \cdot \hat{\mathbf{k}}$, and the spatial propagation phase as $\mathbf{k} \cdot \mathbf{r} = kr_k$. What is the speed of propagation of the wavefront, i.e. a surface of constant phase φ, along r_k? Using the triplet product rule for partial derivatives, we may write the rate of change of location of the wavefront as [1]

$$\left(\frac{\partial r_k}{\partial t}\right)_\varphi = -\frac{(\partial\varphi/\partial t)_{r_k}}{\left(\partial\varphi/\partial r_k\right)_t}$$

Since in the direction of propagation $\varphi = kr_k - \omega t$, we find that $(\partial\varphi/\partial t)_{r_k} = -\omega$ and that $(\partial\varphi/\partial r_k)_t = k$, and thus $(\partial r_k/\partial t)_\varphi = \omega/k$. The speed of propagation of the wavefront, called the *phase velocity* v_p, is then found as

$$v_p = \frac{\omega}{k} \tag{1.38}$$

Recall that in the absence of light absorption, $k = n(\omega)\omega/c$, and thus from equation (1.38), we find $v_p = c/n(\omega)$. The phase velocity tells us the speed of propagation of a monochromatic wave of angular frequency ω in a material with an index of refraction $n(\omega)$. This analysis also provides a simple and intuitive definition of the index of refraction as $n = c/v_p$, i.e. the ratio between the speed of light propagation in vacuum over the speed of phase propagation in matter.

The plane wave solution of equation (1.32) is a harmonic wave written in vectorial form. In certain cases, it is sufficient to express the plane wave solution in scalar form. Assuming that the electric field can be described by a single polarization component E_0 in the transverse plane, the scalar plane wave is written as $\mathrm{E}(\mathbf{r}, t) = 2\mathrm{Re}\,\{E(\mathbf{r}, t)\}$, where the complex function $E(\mathbf{r}, t)$ is defined as

$$E(\mathbf{r}, t) = E_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \tag{1.39}$$

Here, we have written the complex field in italics to distinguish it from the real field $\mathrm{E}(\mathbf{r}, t)$ (nonitalic). In the case of a single plane wave, we may define a coordinate system in which the direction of propagation coincides with one of the Cartesian axes. If we choose the propagation to be along z, and the electric field to be oriented in the x direction, then equation (1.39) further simplifies to $E_x(z, t) = E_0 e^{i(kz-\omega t)}$. The latter form of the wave is a solution of the one-dimensional, homogeneous wave equation. Although the one-dimensional wave equation can be used for solving many problems in optics, it is less useful for describing the propagation and interactions of light in the tightly focusing conditions relevant to microscopy. Therefore, in general, we shall retain the three-dimensional character of wave propagation and only refer to one-dimensional problems where appropriate.

Example 1.3 A plane wave of $\lambda_{\text{vac}} = 800.0$ nm is passing a point $\mathbf{r}$ in space. What is the duration of a full oscillation of the electric field vector measured at this point?

Solution If the position remains constant, the change in phase is determined solely by the time evolution of the wave. Since a full phase cycle corresponds to 2π, we have $\omega t = 2\pi$. Using $\omega = 2\pi c/\lambda_{\text{vac}}$, we find $c \cdot t = \lambda_{\text{vac}}$, which gives

$$t = \frac{\lambda_{\text{vac}}}{c} = \frac{800.0 \times 10^{-9}\ \text{m}}{3.00 \times 10^{8}\ \text{m/s}} = 2.67 \times 10^{-15}\ \text{s} = 2.67\ \text{fs}$$

Example 1.4 A plane wave of $\lambda_{\text{vac}} = 800.0$ nm is traveling through a material with an index of refraction n. What is the time it takes for a plane of constant phase, i.e. the wavefront, to travel a distance $d = \lambda_{\text{vac}}/n$ in this material?

Solution The wavefront travels with a phase velocity $v_p = c/n$. It takes $t = d/v_p$ to cover a distance d, so that

$$t = \frac{(\lambda_{\text{vac}}/n)}{(c/n)} = \frac{\lambda_{\text{vac}}}{c} = 2.67\ \text{fs}$$

This is, of course, the same answer we found above, because it takes one full oscillation period for φ to repeat itself, during which the wavefront travels a distance $\lambda = \lambda_{\text{vac}}/n$ in the material.

1.3.2 Spherical Waves

Spherical waves form a second class of waves that is useful for describing the propagation of wavefronts. A spherical wave exhibits a curved wavefront that emanates from a source located at $\mathbf{r}_0$. From a mathematical point of view, a spherical wave is an idealized wave form that exhibits spherical symmetry, i.e. it shows no dependence on the angles θ and ϕ that define the spherical surface Ω. Instead, the spatial part of the spherical wave only depends on the radius $r = |\mathbf{r} - \mathbf{r}_0|$. The assumption of spherical symmetry is at odds with the notion of propagating (transverse) field vectors, as a transverse vector is not spherically symmetric: a spherical surface cannot be covered uniformly with tangential vectors that retain a single polarization direction. Ignoring these problems for now and proceeding regardless, the three-dimensional wave equation reduces to a one-dimensional differential equation that only depends on r. The scalar solution to this equation is given by harmonic spherical waves of the form

$$E(r,t) = A_0 \frac{e^{ikr}}{r} e^{-i\omega t} \tag{1.40}$$

where A_0 relates to the source strength and has units of V. The wavefront is defined by the condition $kr = \text{constant}$, which results in spherically symmetric wavefronts. Contrary to the plane wave, the amplitude of the spherical wave is not invariant as a function of propagation distance. Instead, the amplitude decreases as $1/r$, which implies that the energy per surface area $d\Omega$ scales as $1/r^2$. Since the area of the wavefront expands as r^2, the radial amplitude dependence complies with the notion that the total energy carried by the wave is conserved.

Given that the wave in equation (1.40) is not a proper solution of the vector Helmholtz equation and by extension of Maxwell's equations, it is perhaps surprising that we deem the spherical wave a useful representation of propagating electromagnetic radiation. There are several reasons why the spherical wave plays an important role in problems related to electromagnetic wave propagation. First and foremost, the spherical wave mimics the far-field radiation emanated from an electric dipole. The electric dipole forms an excellent model for understanding the optical response of particles that are much smaller than the wavelength of light. For instance, the electromagnetic field radiated by an excited molecule resembles that of an electric dipole emitter. Therefore, a large share of optical phenomena in materials can be understood in terms of the radiation that emanates from a collection of electric dipoles driven by the incoming electric fields.

An electric dipole moment is formed when a charge is displaced from its equilibrium position. A relevant case is the movement of a bound electron in a molecule under the influence of an incident electric field. If the equilibrium position of the electron is at point $\mathbf{r}_0$, and the tugging of the electric field will displace the electron to location $\mathbf{r}_\mathrm{d}$, then the electric dipole moment in the molecule is

$$\mathbf{p}(\mathbf{r}_0) = -e \cdot \mathbf{d} \tag{1.41}$$

where e is the electron charge and $\mathbf{d} = \mathbf{r}_\mathrm{d} - \mathbf{r}_0$. The dipole moment is a vector parallel to $\mathbf{d}$ and points from the negatively charged electron to the positive charge (hole) left behind by the electron. Because $|\mathbf{d}| \ll \lambda$, the induced dipole moment in the molecule can be considered "ideal" and thus be approximated as an Hertzian dipole. Since the driving field is time-harmonic, the time dependence of $\mathbf{p}(\mathbf{r}_0, t)$ is also time-harmonic, as long as the response is linear. Exactly how $\mathbf{p}(\mathbf{r}_0, t)$ follows from a driving field $\mathbf{E}(\mathbf{r}_0, t)$ is a topic for later discussion. Here, we are concerned with the radiation that follows from a driven dipole.

The solution of the dipole radiation problem can be obtained from Ampère's law given in equation (1.11), and the electric field that follows thus satisfies Maxwell's equations. The spatial dependence of the complex magnetic and electric fields radiated by the dipole into the far-field ($|\mathbf{r} - \mathbf{r}_0| \gg \lambda$) is

$$\mathbf{B}(\mathbf{r}) = \frac{k^2 p}{8\pi\varepsilon_0 c}(\hat{\mathbf{r}} \times \hat{\mathbf{p}})\frac{\mathrm{e}^{\mathrm{i}k|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r} - \mathbf{r}_0|} \tag{1.42}$$

$$\mathbf{E}(\mathbf{r}) = \frac{k^2 p}{8\pi\varepsilon_0}[(\hat{\mathbf{r}} \times \hat{\mathbf{p}}) \times \hat{\mathbf{r}}]\frac{\mathrm{e}^{\mathrm{i}k|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r} - \mathbf{r}_0|} \tag{1.43}$$

with $\hat{\mathbf{r}}$ the unit vector along $\mathbf{r} - \mathbf{r}_0$ and $\hat{\mathbf{p}} = \mathbf{p}/p$, where p is the magnitude of the dipole moment. These equations look rather cryptic but can be rewritten in terms of spherical coordinates. Assuming that the dipole is aligned along the x-axis, i.e. $\hat{\mathbf{p}} = \hat{\mathbf{x}}$, we can use the relation $\hat{\mathbf{r}} \times \hat{\mathbf{x}} = \hat{\boldsymbol{\varphi}} \sin\psi$, where ψ is the angle relative to the positive x-axis, see Figure 1.3(a). In addition, the cross product between the spherical unit vectors $\hat{\boldsymbol{\varphi}}$ and $\hat{\mathbf{r}}$ is given by $\hat{\boldsymbol{\varphi}} \times \hat{\mathbf{r}} = \hat{\boldsymbol{\psi}}$. Using $r = |\mathbf{r} - \mathbf{r}_0|$, the magnetic and electric field of a radiation dipole are found as

$$\mathbf{B}(\mathbf{r}) = -\frac{k^2 p}{8\pi\varepsilon_0 c}\sin\psi\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\hat{\boldsymbol{\varphi}} \tag{1.44}$$

$$\mathbf{E}(\mathbf{r}) = -\frac{k^2 p}{8\pi\varepsilon_0}\sin\psi\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\hat{\boldsymbol{\psi}} \tag{1.45}$$

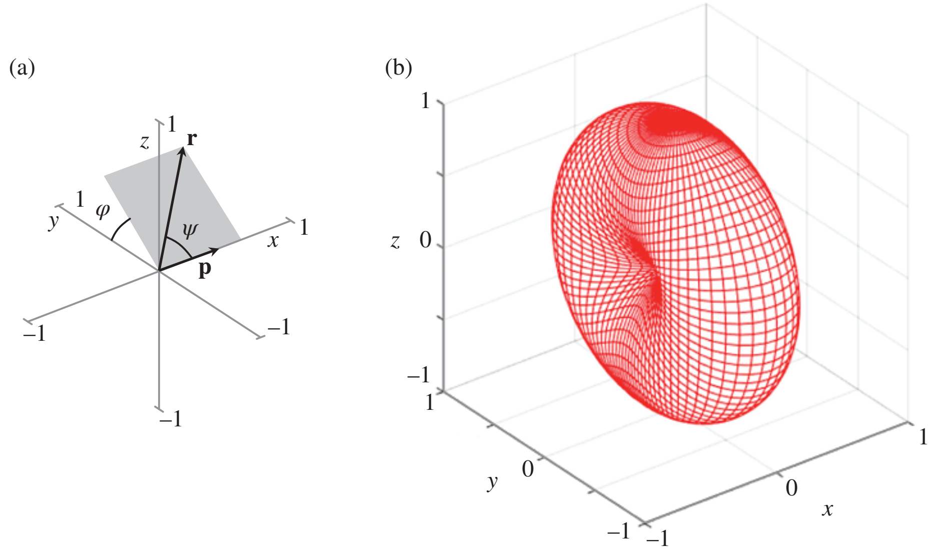

Figure 1.3 Dipole radiation. (a) Reference frame for the radiating dipole, when the dipole is aligned along the x-axis. The coordinates are defined as $x = r\cos\psi$, $y = r\sin\psi\cos\varphi$, and $z = r\sin\psi\sin\varphi$. (b) Plot of the squared amplitude $|E(\psi,\varphi)|^2$ measured on a far-field spherical surface of constant radius R, showing the characteristic Hertzian dipole radiation pattern. In this plot, the radius in the reference frame is $r = |E(\psi,\varphi)|^2/|E_{\max}|^2$.

We see that the electric field in equation (1.45) contains the function e^{ikr}/r, similar to the spatial dependence of the ideal spherical wave given in (1.40). However, the radiated electric field from a dipole is not an ideal spherical wave, and there are important differences compared to the wave described in (1.40). First, equation (1.45) represents a vectorial solution of the electric field. The electric field is directed along $\hat{\boldsymbol{\psi}}$ and complies with $\mathbf{E} = c\mathbf{B} \times \hat{\mathbf{r}}$, which tells us that the electric field is perpendicular to the magnetic field and to $\hat{\mathbf{r}}$, the direction of wave propagation. In other words, the wave described in (1.45) is a transverse wave. Second, unlike the ideal spherical wave, the amplitude of the electric field that follows from the radiating dipole is not uniform on the spherical surface. Instead, because of the $\sin\psi$ term, $\mathbf{E}(\mathbf{r})$ disappears in the directions parallel to $\hat{\mathbf{p}}$, the dipole axis, as required for a transverse vector field on a spherical surface. Figure 1.3(b) shows the angular radiation pattern $|E(\psi,\varphi)|^2$ for a dipole oriented along the x-axis.

Given the similarities between equations (1.40) and (1.45), it is tempting to generalize the harmonic spherical wave in vectorial form as

$$\mathbf{E}(\mathbf{r},t) = 2\mathrm{Re}\left\{\mathbf{A}_0 \frac{e^{ik|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r}-\mathbf{r}_0|} e^{-i\omega t}\right\} \tag{1.46}$$

where $\mathbf{A}_0$ is tangential to a spherical surface centered at $\mathbf{r}_0$, as sketched in Figure 1.4. Although expression (1.46) is not a strict solution of Maxwell's equations, it is still a good description of the field at distances far from the source and within a limited area Ω of the spherical surface. Under these conditions, the expression above represents a transverse field that propagates with the characteristics of a harmonic spherical wave.

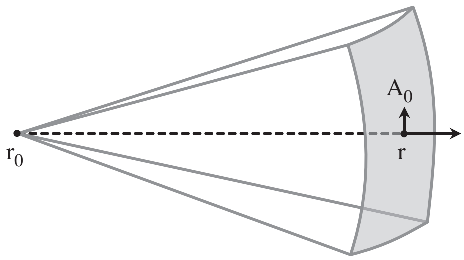

Figure 1.4 Sketch of a spherical wave. The wave originates from a source placed at $\mathbf{r}_0$. The gray area indicates a segment of a spherical surface of constant phase. At observation point $\mathbf{r}$, the radius of the spherical surface is $|\mathbf{r} - \mathbf{r}_0|$. The electric field vector is tangential to the spherical surface, whereas the propagation direction is normal to the surface.

Example 1.5 Consider a radiating dipole that has its dipole axis aligned along the x-coordinate. Determine whether, besides the regular propagation phase, there is an additional the phase-shift between the dipole oscillation and the oscillation of the radiated electric field observed at a far-field point $\mathbf{R}$ along the z-axis.

Solution We assume that the dipole is placed at the origin so that $\mathbf{r}_0 = 0$, oscillates in the $\hat{\mathbf{x}}$ direction, and that its time dependence evolves as $p(t) = p_0 e^{-i\omega t}$ in this direction, where p_0 is the oscillation amplitude. The radiated field at point $\mathbf{R}$ is given by equation (1.45). Along a far-field point on the z-axis, the unit vector $\hat{\boldsymbol{\psi}}$ is directed along the x-axis as $\hat{\boldsymbol{\psi}} = -\hat{\mathbf{x}}$. Using $\psi = \frac{1}{2}\pi$, the field can now be written as

$$\mathbf{E}(\mathbf{R}, t) = 2\mathrm{Re}\left\{ \frac{k^2 p_0}{8\pi\varepsilon_0} \frac{e^{i(kR-\omega t)}}{R} \right\} \hat{\mathbf{x}}$$

with $R = |\mathbf{R} - \mathbf{r}_0|$. Besides the regular propagation phase $\phi = kR - \omega t$, there is no additional phase shift between the dipole and the far-field observation point, and the electric field is in phase with the dipole oscillation. Along the z-axis, the polarization direction of the electric field is identical to that of the dipole itself.

1.3.3 Paraxial Propagation

In many cases, we are interested in understanding the propagation of a wave along a given linear coordinate. At larger propagation distances, the behavior of the wavefront at propagation angles directed away from the main coordinate is then less relevant. Consider a radiating source located in a plane $z_0 = 0$ at the transverse coordinates (x_0, y_0). We are interested in the electric field at point (x, y) in a transverse plane along z. The propagation distance $r = |\mathbf{r} - \mathbf{r}_0|$, used in equation (1.46), is now written as

$$r = \left\{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2\right\}^{1/2}$$

If the propagation distances along the transverse coordinates are much smaller than the distance traveled along z, then the z-coordinate is the main propagation axis. In this scenario, r is now dominated by the distance z, whereas the transverse coordinates contribute very little to r. Using $z_0 = 0$, we may thus write

$$\begin{aligned} r &= \left\{z^2 + (x - x_0)^2 + (y - y_0)^2\right\}^{1/2} \\ &= z\left\{1 + \frac{(x - x_0)^2 + (y - y_0)^2}{z}\right\}^{1/2} \\ &= z + \frac{(x - x_0)^2 + (y - y_0)^2}{2z} \end{aligned} \tag{1.47}$$

where the last expression is obtained by applying the binomial expansion of the square root and retaining only the first two terms, i.e. $(1 + a)^{1/2} \approx 1 + a/2$. This approximation is called the Fresnel approximation and the resulting distance r is sometimes referred to as the *paraxial distance*.

We can now rewrite the general spherical wave given in equation (1.46) in the limit of paraxial propagation. For this purpose, we first replace the spatial phase term with ikr, where r is now the distance as defined in (1.47). Second, we substitute the distance $|\mathbf{r} - \mathbf{r}_0|$ in the denominator for z, a good approximation if $z \gg \lambda$. Without loss of generality, we can further assume that the radiation source is located at $(x_0, y_0, z_0) = (0, 0, 0)$. The spatial part of the *paraxial spherical wave* is then found as

$$\mathbf{E}(x, y, z) = \mathbf{A}_0 \frac{e^{ikz}}{z} e^{ik\rho^2/2z} \tag{1.48}$$

where $\rho = (x^2 + y^2)^{1/2}$. Equation (1.48) describes an outgoing spherical wavefront propagating along the (positive) z direction. The last exponential factor in this expression is a measure of the wavefront curvature. The curvature is characterized by a radius of curvature that equals the propagation distance z. The paraxial spherical wave is a helpful tool in understanding wave propagation in several important situations, in particular in the discussion of scattering problems and in understanding curvature changes imposed by an optical lens.

1.4 Polarization States

The electric field $\mathbf{E}$ is a vector with polarization components (E_x, E_y, E_z) in Cartesian coordinates. For a simple monochromatic plane wave that propagates in a transparent and isotropic homogeneous medium, the polarization state of the field remains unaltered. However, the polarization state of the propagating light field may undergo changes in media that have anisotropic optical properties or at interfaces between two different media. Consideration of possible changes to the polarization state of light is relevant to optical microscopy. For instance, the polarization properties of light change upon focusing by microscope objective lenses or upon light scattering at objects in the sample. In this section, we discuss several common polarization states of light and methods for describing polarization state transformations.

1.4.1 Linearly and Circularly Polarized Waves

To characterize the polarization state of the wave and study how the polarization states change as the wave propagates through objects and materials, it is helpful to express the field in terms of the so-called "Jones vector." For the plane wave described by equation (1.32), we can write

$$\mathbf{E}(\mathbf{r}, t) = 2\text{Re}\left\{|\mathbf{E}_0| \cdot \mathbf{JV}\ \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r} - \omega t)}\right\} \tag{1.49}$$

where $\mathbf{JV}$ is the Jones vector. We choose z as the propagation axis, which implies that the transverse polarization components lie in the xy plane. In this case, the generalized Jones vector can be written as [2]

$$\mathbf{JV} = \begin{bmatrix} e_x \\ e_y \\ 0 \end{bmatrix} \tag{1.50}$$

The parameters e_x and e_y are proportional to the complex amplitudes of the E_x (horizontal) and E_y (vertical) polarization components, respectively, such that $(e_x^2 + e_y^2)^{1/2} = 1$. If the $\mathbf{E}$ field is

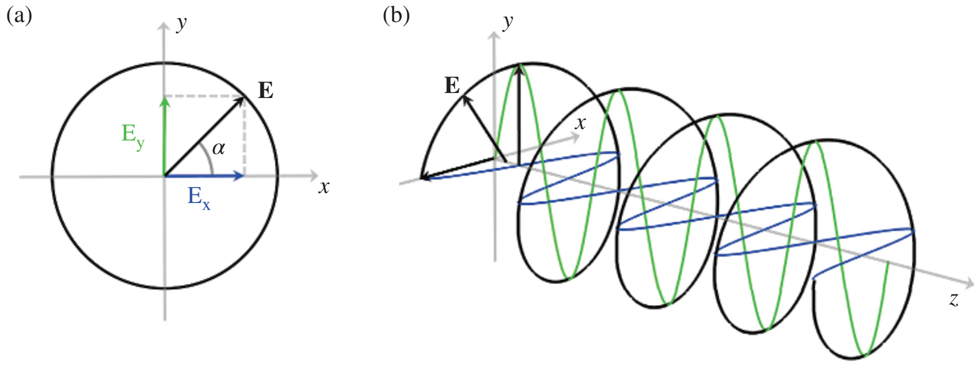

Figure 1.5 Linear (a) and circular (b) polarization states of a plane wave traveling along the z-axis, such that xy is the transverse plane. For linearly polarized light, the projections $E_x = |\mathbf{E}| \cos \alpha$ and $E_y = |\mathbf{E}| \sin \alpha$ are in phase, whereas $E_y = \pm \mathrm{i} E_x$ for circularly polarized light. The example in (b) represents the case $E_y = +\mathrm{i} E_x$, which is a left-handed circularly polarized field.

linearly polarized then the E_x and E_y components of the field are *in phase*. Denoting the angle between the field vector and the x-axis as α, the components of the Jones vector can be written as $(e_x, e_y, 0) = (\cos \alpha, \sin \alpha, 0)$, as shown in Figure 1.5(a). Some common examples of the generalized Jones vector are listed in Table 1.1.

In circularly polarized light, the E_y component is phase shifted with respect to the E_x component by an amount of $\Delta\varphi = \frac{1}{2}\pi = 90°$, i.e. the components are in *quadrature*. The E_y component thus acquires the factor $\mathrm{e}^{\pm \mathrm{i}\Delta\varphi} = \pm \mathrm{i}$. In case of a positive phase shift, the total phase becomes $\varphi = kz - (\omega t - \frac{1}{2}\pi)$, which represents a 90° *retardation* relative to E_x. If the phase shift is negative, then the E_y component is 90° *advanced* relative to E_x. These phase shift properties are summarized in Table 1.2. The corollary of the quadrature components is that the field vector no longer oscillates along a fixed direction during wave propagation. Instead, the field vector rotates in the transverse plane with an angular velocity of ω, completing a full rotation as the wave propagates a distance λ. If the wave travels toward an observer, then the viewer will perceive the positive phase shift as a left-handed (counter-clockwise) rotation of the field vector, and, vice versa, as a right-handed (clockwise) rotation if the phase shift is negative. The former is referred to as a left-handed circularly polarized (LCP) beam, while the latter is called a right-handed circularly polarized (RCP) beam. The corresponding generalized Jones vectors are listed in the bottom two rows of Table 1.1.

1.4.2 Polarization Transformation

The polarization state of the field can be altered by optical elements. Since we are interested in describing the propagation of waves in such materials, it is helpful to introduce a notation for indicating changes in the polarization state. We consider a matrix operation $\mathbb{M}$ that can change the Jones vector, such that

$$\mathbf{E}(\mathbf{r}, t) = 2\mathrm{Re}\left\{\left|\mathbf{E}_0\right| \cdot \mathbb{M} \cdot \mathbf{JV}\ \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)}\right\} \tag{1.51}$$

The linear polarizer is an example of a polarization operator. This optical element exhibits high transmission of light polarized along its main axis, and minimum transmission of field components

Table 1.1 Generalized Jones vectors of common polarization states.

Polarization state	α	Jones vector
Linear (x-axis)	0	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
Linear (y-axis)	90°	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$
Linear (+45°)	45°	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$
Linear (−45°)	−45°	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$
Circular (left-handed)	—	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \mathrm{i} \\ 0 \end{bmatrix}$
Circular (right-handed)	—	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -\mathrm{i} \\ 0 \end{bmatrix}$

Table 1.2 Phase shift properties of the propagator $\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ when multiplied by a factor $\mathrm{e}^{\mathrm{i}\Delta\varphi}$.

Phase shift	Phase advance/retardation
$\Delta\varphi > 0$	Retardation
$\Delta\varphi < 0$	Advance

polarized in the perpendicular direction. If the angle between the x-axis and the polarizer's main axis is β, then the matrix $\mathbb{M} = \mathbb{P}(\beta)$ is written as follows

$$\mathbb{P}(\beta) = \begin{bmatrix} \cos^2\beta & \sin\beta\cos\beta & 0 \\ \sin\beta\cos\beta & \sin^2\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.52}$$

For instance, if the main axis of the polarizer is aligned along the horizontal (x) direction, i.e. $\beta = 0$, then only the E_x component is transmitted, whereas the E_y component is maximally suppressed

$$\mathbb{P}(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Polarizers change the field's polarization state by attenuating the horizonal and vertical components differently, which affects the overall amplitude $2|\mathbf{E}_0|$ of the light field. Wave plates, on the other hand, manipulate the polarization state without attenuation. Wave plates are birefringent materials, which means that the index of refraction n is a function of the polarization direction. Because the phase velocity depends on n as $v_p = c/n$, the E_x and E_y polarization components can experience a different phase velocity as the wave traverses the birefringent material. The key operating principle of the wave plate is to delay one polarization component more than the other, thereby applying a relative phase shift between E_x and E_y and thus changing the polarization state of the field.

Example 1.6 Consider a linearly polarized plane wave with $\alpha = 20°$. The light field passes through a polarizer that is oriented such that $\beta = 45°$. Determine the E_x and E_y amplitudes of the field after traversing the polarizer.

Solution Since $\cos(45°) = \sin(45°) = \frac{1}{2}\sqrt{2}$, the transformation of the incident field amplitudes is

$$\begin{bmatrix} E_x \\ E_y \\ 0 \end{bmatrix} = 2|\mathbf{E}_0| \cdot \mathbb{P}(45°) \cdot \mathbf{JV}$$

$$= 2|\mathbf{E}_0| \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(20°) \\ \sin(20°) \\ 0 \end{bmatrix} = 2|\mathbf{E}_0| \begin{bmatrix} 0.641 \\ 0.641 \\ 0 \end{bmatrix}$$

We see that the field is still linearly polarized, but now at an angle $\alpha = \arctan(e_y/e_x) = 45°$, as expected. Whereas the incident amplitude of the field was $2|\mathbf{E}_0|[\cos^2(20°) + \sin^2(20°)]^{1/2} = 2|\mathbf{E}_0|$, after traversing the polarizer the amplitude is $2|\mathbf{E}_0|\sqrt{2} \cdot 0.641 = 1.81|\mathbf{E}_0|$. The field amplitude is lower because the polarizer has blocked field components that are perpendicular to the polarizer's main axis.

A wave plate is characterized by two axes, the axis along which v_p is the highest, called the fast axis, and the axis perpendicular to it, called the slow axis, along which the phase velocity is the lowest. The birefringent material of the wave plate is positioned such that the fast and slow axes are in a plane parallel to the field's transverse plane. The angle β now indicates the angle between the fast axis of the wave plate and the horizontal axis of the reference frame. The two main forms of the wave plate are the *half-wave plate* and the *quarter-wave plate*.

In a half-wave plate, the projection of the field along the fast axis is advanced by a π phase shift ($\Delta\varphi = -\pi$) relative to the field projection along the slow axis. For example, if the fast axis is aligned with the x-axis ($\beta = 0$), then the E_x component acquires a 180° phase advance relative to the E_y component. It is convention to define the phase shifts relative to the mean, i.e. for a total phase shift of $\Delta\varphi = -\pi$, E_x acquires a $e^{-i\frac{\pi}{2}}$ factor (phase advance), while E_y is modified with a $e^{i\frac{\pi}{2}}$ factor (phase retardation). For this setting, the matrix operator $\mathbb{M}$ for the half-wave plate is

$$\mathbb{W}_{\lambda/2}(0) = -i\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The common $-i$ factor affects both components and thus does not alter the polarization state. The matrix above transforms a Jones vector of $(\cos\alpha, \sin\alpha, 0)$ to $(\cos\alpha, -\sin\alpha, 0)$, which corresponds to a rotation of the field vector by an angle of -2α. The half-wave plate thus rotates the polarization direction of a linearly polarized field. For a general angle β, the matrix for the half-wave plate is given as

$$\mathbb{W}_{\lambda/2}(\beta) = -i\begin{bmatrix} \cos 2\beta & \sin 2\beta & 0 \\ \sin 2\beta & -\cos 2\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.53}$$

In a quarter-wave plate, the phase advance of the field component along the fast axis is $\frac{1}{2}\pi$ ($\Delta\varphi = -\frac{1}{2}\pi$) relative to the field component along the slow axis. The resulting matrix for the quarter-wave plate is

$$\mathbb{W}_{\lambda/4}(\beta) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 - i\cos 2\beta & -i\sin 2\beta & 0 \\ -i\sin 2\beta & 1 + i\cos 2\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.54}$$

Aligning the fast axis in the horizontal direction ($\beta = 0$) and operating on a light field that is initially linearly polarized at an angle $\alpha = +45°$, the quarter-wave plate transforms the field as

$$\mathbb{W}_{\lambda/4}(0) \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2}(1 - i)\begin{bmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = e^{-i\frac{\pi}{4}} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}$$

The operation above indicates that a field that is linearly polarized at an angle +45° relative to the fast axis of the wave plate is transformed to a LCP polarized field. The extra factor $e^{-i\frac{\pi}{4}}$ results from the convention that the phase shifts along E_x and E_y are measured relative to the mean of $\Delta\varphi$. This factor represents a phase shift that is applied to all components, has amplitude 1, and does not affect the polarization state of the field.

1.5 Reflection and Transmission at Interfaces

Our description of a plane wave works well for describing wave propagation in isotropic, homogeneous materials. In the last section, we also found a good way for describing changes to the electric field vector as the wave encounters polarizing elements such as polarizers and waveplates at normal incidence. In considering such optical elements, we assumed that they only have an effect on the polarization state of the wave, leaving the direction of propagation unaffected. In reality, a wave incident upon an optical element traverses the interface formed between the element's material and its surroundings, and light fields generally undergo changes at interfaces. The discontinuity of material properties across the interface gives rise to a reflected wave off the interface and a transmitted wave that penetrates the material. In this section, we study the changes in propagation direction of the incident wave due to a planar interface and summarize the corresponding amplitudes of reflected and transmitted waves. To simplify the discussion, we assume that the media are isotropic and homogeneous, as well as nonabsorbing (transparent). The latter restriction means that the refractive index of each material is real.

1.5.1 Waves at Interfaces

We are interested in understanding the changes to the electric field of a plane wave that is incident upon a planar interface between two materials with refractive indices n_1 and n_2, respectively. In particular, we seek to find the propagation directions of the reflected and transmitted waves, as well as the amplitudes of their polarization components. To do this, we first define a frame of reference to describe the field vectors **E** and their respective wave factors **k**. We choose the z-axis to be normal to the planar interface, with the xy plane parallel to the interface, as shown in Figure 1.6. The interface is located at the plane $z = 0$. The plane spanned by the z axis and the incident wave vector $\mathbf{k}_\mathrm{i}$ is referred to as the *plane of incidence*, and the angle between $\mathbf{k}_\mathrm{i}$ and the z-axis is called the *angle of incidence* θ_i. The projection of the incident wave vector along z is $k_{z,\mathrm{i}} = k_\mathrm{i} \cos\theta_\mathrm{i}$ and the projection in the xy plane is $k_{||,\mathrm{i}} = (k_{x,\mathrm{i}}^2 + k_{y,\mathrm{i}}^2)^{1/2} = k_\mathrm{i} \sin\theta_i$. The wave vectors of the reflected ($\mathbf{k}_\mathrm{r}$) and transmitted ($\mathbf{k}_\mathrm{t}$) waves are defined in a similar manner.

It is helpful to define the components of the incoming field vector relative to the plane of incidence. The component parallel to the plane of incidence is denoted as $E_\mathrm{i}^{(\mathrm{p})}$, known as the plane

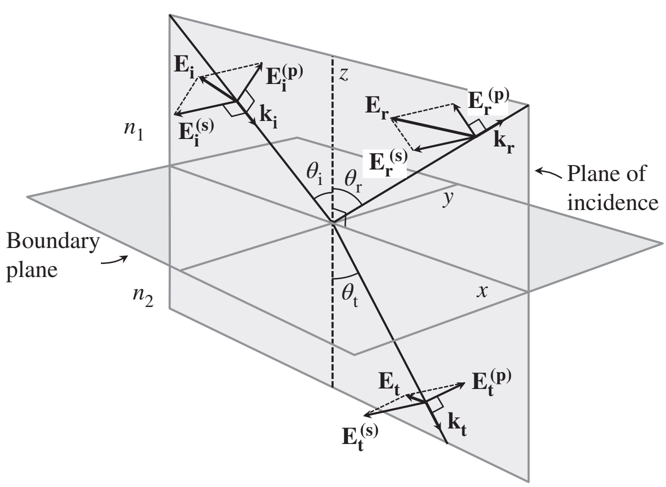

Figure 1.6 Reflection and transmission of an incident plane wave with an electric field $\mathbf{E}_\mathrm{i}$ and wave vector $\mathbf{k}_\mathrm{i}$ at an interface between a dielectric medium with an index of refraction n_1 and a dielectric medium with an index of refraction n_2. The angle of incidence θ_i, angle of reflection θ_r, and angle of transmission θ_t are defined relative to the surface normal (dotted line). Here, the boundary plane is placed in the xy plane, and the surface normal is aligned along z.

polarized component, while the component perpendicular to the plane of incidence is written as $E_i^{(s)}$, the surface polarized component. We thus have

$$\mathbf{E}_i = E_i^{(s)}\hat{\mathbf{e}}_s + E_i^{(p)}\hat{\mathbf{e}}_p \tag{1.55}$$

where $\hat{\mathbf{e}}_s$ is the unit vector perpendicular to the plane of incidence and $\hat{\mathbf{e}}_p$ is the unit vector in the direction of the component of $\mathbf{E}_i$ that is parallel to the plane of incidence. The amplitudes of the reflected ($\mathbf{E}_r$) and transmitted ($\mathbf{E}_t$) fields are commonly expressed in terms of their s and p-components.

1.5.2 Angles of Reflection and Transmission

Our first focus is the relation between the propagation direction of the incoming field and the directions of propagation of the reflected and transmitted waves. The direction of the incident field is characterized by θ_i in the plane of incidence, and we are interested in finding the corresponding angle θ_r of the reflected wave and the angle θ_t of the transmitted field. To find the angles θ_r and θ_t, we make use of the notion that at the interface the momentum of the wave is conserved. This also implies that at $z = 0$, the propagation phase of the incident, reflected, and transmitted waves is the same so that

$$\mathbf{k}_i \cdot \mathbf{r} = \mathbf{k}_r \cdot \mathbf{r} = \mathbf{k}_t \cdot \mathbf{r} \;\; (\text{at z} = 0) \tag{1.56}$$

The above condition states that for a given point $\mathbf{r} = (x, y, 0)$ in the plane of incidence, the amplitudes of the wave vector components in the xy plane are related as

$$k_{||,i} = k_{||,r} = k_{||,t} \tag{1.57}$$

Since $k_{||,i} = k_i \sin\theta_i$ and from Figure 1.6, we have $k_{||,r} = k_r \sin\theta_r$, we find the condition $k_r \sin\theta_r = k_i \sin\theta_i$. Because the reflected wave propagates in the same medium n_1 as the incident wave, we also have $k_i = n_1 k_0 = k_r$, so that the relation $\sin\theta_r = \sin\theta_i$ must hold. We recognize that the angle of reflection equals the angle of incidence, i.e. $\theta_r = \theta_i$, which is the well-known *law of reflection*. However, the reason why there should be a reflected wave in the first place is not immediately clear from this analysis. A discussion on the microscopic origin of reflection is found in Chapter 4.

We can also use equation (1.57) for analyzing the propagation angle of the transmitted field. The relation $k_{||,i} = k_{||,t}$ requires that the phase components in the xy plane must be continuous across the interface. We thus obtain $k_i \sin\theta_i = k_t \sin\theta_t$, which implies that

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{k_t}{k_i} = \frac{n_2}{n_1} = \frac{v_1}{v_2} \tag{1.58}$$

where v_1 and v_2 are the phase velocities of the wave in the respective media. Relation (1.58) is known as *Snell's law*. The law states that θ_t is different from θ_i when $n_2 \neq n_1$. Because the "bending" of the wave relies on the differences between the refractive index properties of the media, the effect is called a *refractive* effect. The physical interpretation of the difference in propagation angles across the interface is that the phase velocities in the two media are different so that the direction of propagation of the wave must change in order to maintain a continuous phase across the interface, as illustrated in Figure 1.7.

1.5.3 Amplitudes of Reflected and Transmitted Fields

The boundary conditions for the electric and magnetic field at a nonmagnetic dielectric interface ($\mathbf{j} = 0$) state that the components of $\mathbf{E}$ and $\mathbf{B}$ tangential to the boundary plane are continuous

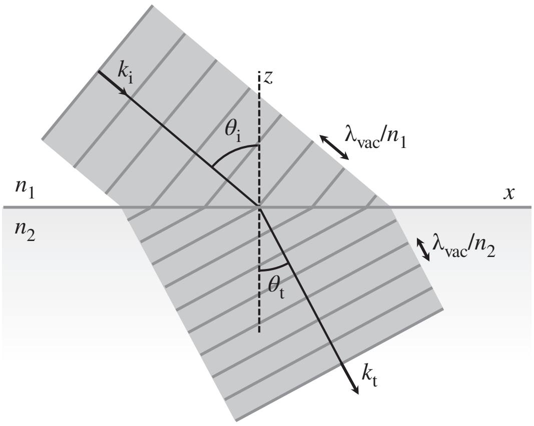

Figure 1.7 Transmission of a plane wave at an interface for the condition $n_2 > n_1$. Planes of equal phase are indicated by gray lines, underlining the continuity of the propagation phase at $z = 0$ and the necessity of bending when the phase velocity and the effective wavelength of the wave decrease.

across the interface. This requirement results in independent conditions for s-polarized waves and for p-polarized waves. We can write the reflected ($\mathbf{E}_r$) and transmitted ($\mathbf{E}_t$) fields for the s- and p-polarized components as follows

$$\begin{aligned} E_r^{(s)} &= r_s(n_1, n_2)E_i^{(s)} \\ E_t^{(s)} &= t_s(n_1, n_2)E_i^{(s)} \\ E_r^{(p)} &= r_p(n_1, n_2)E_i^{(p)} \\ E_t^{(p)} &= t_p(n_1, n_2)E_i^{(p)} \end{aligned} \tag{1.59}$$

where r_s and r_p are known as the Fresnel reflection coefficients, whereas t_s and t_p are the Fresnel transmission coefficients. From the boundary conditions for s-waves, the following expressions for the r_s and t_s coefficients can be found [3]

$$r_s = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} \tag{1.60}$$

$$t_s = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_t} \tag{1.61}$$

And, similarly, for the boundary conditions of the p-polarized components, the r_p and t_p coefficients are derived as

$$r_p = \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} \tag{1.62}$$

$$t_p = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t} \tag{1.63}$$

With the help of the Fresnel coefficients defined above, we are in a position to determine the amplitudes of the reflected and transmitted waves, with the stipulation that we have to find the projections of the incoming field vector $\mathbf{E}_i$ in the s- and p-directions first.

1.6 Transformation of the Field by a Lens

The plane waves and spherical waves introduced earlier in this chapter specify the magnitude and orientation of the electric field vector, as well as its direction and propagation phase. In the absence of material objects, these descriptions are sufficient to predict the free space propagation of the wave. In the presence of objects, such as optical elements or microscopy samples, the fields are modified and the description of the wave can grow more complex. Our goal is to know the magnitude and orientation of the vectors **E** and **k**, as well as to keep track of possible shifts in the propagation phase, at each relevant point in space.

In Section 1.4, we made modifications to a plane wave by altering the relative phase shift between the two orthogonal components of **E**, which changed the polarization state of the wave. In Section 1.5, we tracked the redirection of **E** and **k** of a plane wave, and determined the amplitude changes of its polarization components, at an interface. This allowed a description of the reflected and transmitted fields. In all these cases, the wavefront of the wave retained its original character. For a plane wave, the modifications to **E** and **k** discussed thus far did not alter the planar character of the wavefront. The situation is different for a lens, as the main purpose of a lens is to bring about changes to the wavefront of a propagating field.

Because optical lenses play a central role in microscopy, knowledge of the changes they apply to the propagating field vector is important for appreciating several key properties of the optical microscope. The lens affects the wavefront in two principal ways: (i) it applies a position-dependent phase shift to the entering wavefront, and (ii) it applies a position-dependent rotation to the field vector and its propagation direction. In this section, we examine these changes for the case of an incident plane wave to find a description of the wavefront at the exiting surface of the lens.

1.6.1 Phase Shift by a Lens

Consider the lens sketched in Figure 1.8. This is a plano convex lens, a nonabsorbing material with an index of refraction n_l shaped into an object of cylindrical symmetry with a flat surface on one side and a curved surface on the opposing side. The cylinder axis of the lens is aligned with the z-axis, which coincides with the propagation axis of an incident plane wave, entering the lens from the left onto its curved surface. On-axis, the curved surface is placed at the point $z = 0$. We see that the thickness of the lens is a function of the transverse displacement $\rho = \sqrt{x^2 + y^2}$: the

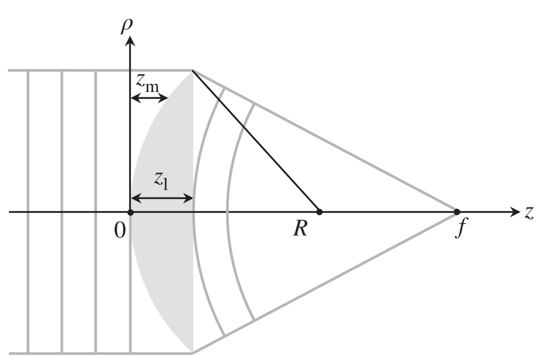

Figure 1.8 A plane wave incident on a plano-convex lens with radius of curvature R and index of refraction n_l. The lens converts a planar wavefront into a spherical wavefront with a radius of curvature f.

thickness $z_l(\rho)$ progressively decreases for larger ρ. Because of the variation in lens thickness as a function of ρ, a plane wave incident on the lens will experience a ρ-dependent phase variation. Along the main propagation direction z, the accumulated phase at a given transverse displacement depends on the distance z_l traveled through the lens as well as the remaining travel distance z_m through the surrounding medium. Assuming the index of refraction of the surrounding medium as $n_m = 1$, the accumulated spatial phase is

$$\varphi(\rho) = k_0[z_m(\rho) + n_l z_l(\rho)] \tag{1.64}$$

In the *thin lens approximation*, the lens thickness $z_l(0)$ is much smaller than the radius of curvature R. Under these conditions, we may also ignore propagation differences due to refraction at the incident surface of the lens and assume that the ray inside the lens propagates parallel to z. Within this approximation, the travel distance inside the lens can be written as $z_l(\rho) \approx z_l(0) - z_m(\rho)$. On axis, the accumulated phase is $\varphi(0) = n_l k_0 z_l(0)$. The phase difference between $\varphi(\rho)$ and $\varphi(0)$ is then

$$\Delta\varphi = \varphi(\rho) - \varphi(0) = -k(n_l - 1)z_m(\rho) \tag{1.65}$$

Note that the relative phase shift is negative, implying a relative phase advance. Portions of the wave at larger transverse displacements ρ are thus less phase delayed than portions closer to the optical axis. If the curved surface is a spherical surface, then the distance z_m can be computed from R as

$$z_m(\rho) = R - \sqrt{R^2 - \rho^2} \tag{1.66}$$

In the paraxial approximation, $\rho \ll R$, and we can make the binomial approximation for the square root term in equation (1.66), yielding

$$z_m(\rho) = R - R\left(1 - \frac{\rho^2}{2R^2}\right) = \frac{\rho^2}{2R} \tag{1.67}$$

We next use the lens maker's formula for an ideal plano convex lens, which is given as $f = R/(n_l - 1)$, where f is the focal length of the lens. We can now write the relative phase shift caused by propagation through the lens as [4]

$$\Delta\varphi = -k\frac{\rho^2}{2f} \tag{1.68}$$

At the exit surface of the lens, located at $\Delta z = z_m(\rho) + z_l(\rho)$, an entering (scalar) plane wave $E(\rho, 0)$ is altered by the phase function of the lens as

$$E(\rho, \Delta z) = E(\rho, 0)\mathrm{e}^{\mathrm{i}n_l k_0 \Delta z}\mathrm{e}^{-\mathrm{i}k\rho^2/2f} \tag{1.69}$$

Besides an overall propagation phase $n_l k_0 \Delta z$, which is independent of ρ, we see that the planar wavefront of the entering wave at $z = 0$ is modified to a wavefront that changes quadratically with ρ. This is a characteristic of a spherical wavefront. Comparing the exiting wave with the outgoing spherical wave given in equation (1.48), we recognize that the wavefronts are similar except for a difference in sign. The lens thus transforms the planar wavefront into a spherical wavefront that is converging toward a point at a distance f from the lens. Consequently, at the exit surface of the lens, the wavefront can be modeled as a spherical surface with a radius of curvature f. Note that this surface of constant phase is not the same as the physical surface of the lens. We refer to the surface of constant phase at the exit pupil of the lens as the *spherical reference surface*. We can understand the propagation of the wave after passing through the lens in terms of its propagation from this reference surface toward the focal region.

Example 1.7 The index of refraction of fused silica at an incident wavelength of $\lambda = 800$ nm is $n = 1.4533$. Determine the radius of curvature of a fused silica plano convex lens that has a focal length of $f = 100.0$ mm.

Solution The radius of curvature R of the lens is found from the lens maker's equation for a thin lens as

$$R = \{n(\lambda) - 1\}f = 45.33 \text{ mm}$$

Note that the radius of curvature of the lens surface is different from the radius of curvature of the reference surface, as the latter is given by f.

Example 1.8 Equation (1.69) states that the spatial phase properties of an incoming plane wave are modified by a plano convex lens, forming a converging spherical wave with a radius of curvature f. Consider an incident wave with a diverging spherical wavefront with radius of curvature s_1 at the entrance aperture of a lossless lens, where $s_1 > f$. At the exit aperture of the lens, a converging spherical wave is observed. Express the point of convergence s_2 of the exiting wave in terms of s_1 and f.

Solution Positioning the lens as $z = 0$, the incident wave at the lens surface is

$$E_{\text{in}}(\rho, 0) = E_0(\rho, 0)\mathrm{e}^{ik\rho^2/2s_1}$$

After passing through the lens, the wavefront has been modified, producing the following wave at the exit aperture

$$E_{\text{out}}(\rho, \Delta z) = E_0(\rho, 0)\mathrm{e}^{ik\rho^2/2s_1}\mathrm{e}^{-ik\rho^2/2f}$$

where the ρ-independent phase factor has been ignored. Since the exit wavefront resembles a converging wave with radius of curvature s_2, we may also write

$$E_{\text{out}}(\rho, \Delta z) = E_0(\rho, 0)\mathrm{e}^{-ik\rho^2/2s_2}$$

Equating the last two expressions for E_{out} yields

$$\frac{ik\rho^2}{2}\left(\frac{1}{s_1} - \frac{1}{f}\right) = -\frac{ik\rho^2}{2}\frac{1}{s_2}$$

which gives

$$\frac{1}{f} = \frac{1}{s_1} + \frac{1}{s_2}$$

This is the well-known *Gaussian thin lens equation*, which relates the distance s_1 between the object and the thin lens, to the focal length f and the distance s_2 between the thin lens and the image.

1.6.2 Vector Rotation by a Lens

In addition to a phase modification, transforming a planar wavefront into a converging spherical wavefront naturally involves a change in the orientation of $\mathbf{E}_0$ and the wave vector $\mathbf{k}$. Whereas for a planar wavefront these vectors are constant over the extent of the transverse plane, on a spherical surface the orientation of the field vector and its propagation direction are spatially variant. The lens applies a position-dependent rotation of the vector field because of refraction at its entering and exiting surfaces. This vector rotation, however, is not explicitly described by the phase transformation discussed in Section 1.6.1. To arrive at a complete description of the vector field on the spherical reference surface, we have to apply a proper transformation to the incident vector field.

An elegant way to accomplish the vector rotation of an incoming field $\mathbf{E}(\mathbf{r})$ is by use of transformation matrices [2], in analogy with the discussion in Section 1.4. By operating on the field at the entering surface with a properly chosen matrix $\mathbb{M}$ we can transform it into the field at the exiting surface. We choose a reference spherical surface with a radius of f as the exiting wavefront, defined in terms of the polar angle θ and azimuthal angle ϕ, and which intersects the optical axis at $z = -f$. We also define the so-called "meridional plane," which is aligned with the optical axis and intersects the focal (xy) plane at an angle ϕ relative to the positive x-axis, see Figure 1.9. The θ angle represents a rotation relative to the negative z-axis around an axis normal to the meridional plane. Our task is to find $\mathbf{E}_0(\theta, \phi)$ on the spherical reference surface, given the input field $\mathbf{E}_0^{\mathrm{inc}} = (E_x, E_y, 0)$ at location $z = -f$

$$\mathbf{E}_0(\theta, \phi) = \left(\frac{n_1}{n_2}\right)^{1/2} \sqrt{\cos\theta}\ \mathbb{M} \cdot \mathbf{E}_0^{\mathrm{inc}} \tag{1.70}$$

The factor $(n_1/n_2)^{1/2}$ accounts for the change in refractive index before and after the lens. The $(\cos\theta)^{1/2}$ factor is included to ensure that the energy flux remains constant upon projecting

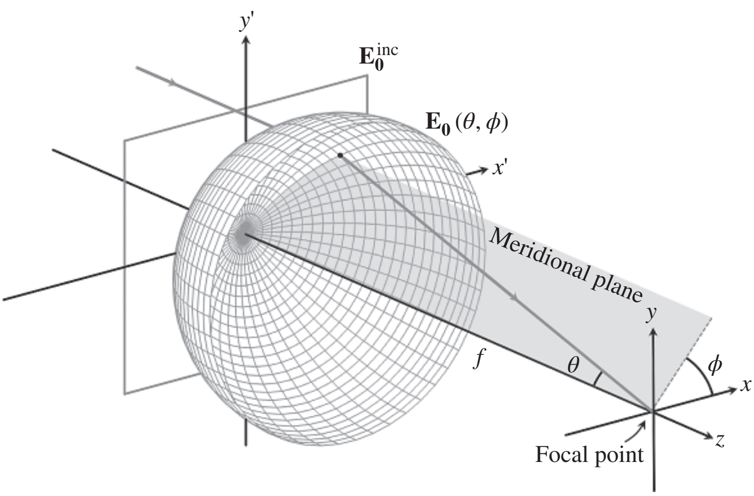

Figure 1.9 The incident planar wavefront of the field $\mathbf{E}_0^{\mathrm{inc}}$ is transformed by a lens, forming the field $\mathbf{E}_0(\theta, \phi)$ on the spherical reference surface. The plane spanned by the optical (z) axis and the ray that travels from a point (θ, ϕ) on the spherical reference surface to the focal point is called the meridional plane. Spherical reference surface is not drawn to scale.

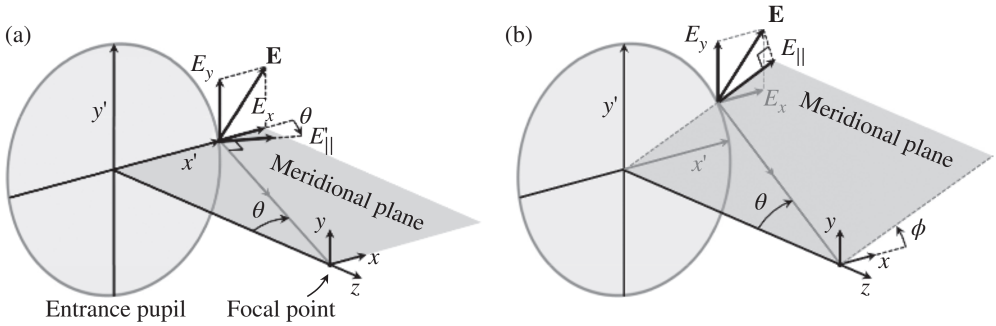

Figure 1.10 Rotation of the field vector at the reference surface. (a) Incoming field **E** at a location $(\theta, 0)$ on the reference surface, in which case E_x corresponds to $E_{\|}$ and E_y to $E_{\perp}$. Application of the $\mathbb{L}$ transformation rotates the $E_{\|}$ component over the angle θ around the y-axis within the meridional plane. (b) Incoming field at a location for which $E_{\|}$ does not coincide with E_x. Rotating the reference frame by $\mathbb{R}$ ensures that a subsequent $\mathbb{L}$ rotation around y only affects the $E_{\|}$ component.

a planar surface element onto a spherical surface element. Note that this transformation takes care of the vector rotation only and does not explicitly include propagation effects associated with converting the planar surface into a spherical surface. In the following, we ignore reflection effects at the surfaces of the lens, and thus the Fresnel reflection coefficients are set to $r_s = r_p = 0$, whereas $t_s = t_p = 1$.

Consider a vector $\mathbf{E}_0^{\text{inc}}$ incident on the reference surface at location $(\theta, 0)$, as shown in Figure 1.10(a). We can see that the E_x component of the vector is parallel to the meridional plane, which we refer to as $E_{\|}$. The E_y component is perpendicular to this plane and thus corresponds to $E_{\perp}$. Bending of the ray toward the focal point requires that the $E_{\|}$ component be rotated by θ, so that the resulting electric field vector remains tangential to the spherical wavefront. This rotation can be accomplished with the matrix $\mathbb{L}$, which performs a clockwise rotation around the y-axis, and is given as

$$\mathbb{L} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{1.71}$$

The $E_{\perp}$ component is unaffected by this θ-rotation, as it is already tangential with the spherical surface. In general, field components perpendicular to the meridional plane are unaltered by the lossless focusing lens. For points on the reference surface for which $\phi \neq 0$, such as shown in Figure 1.10(b), $E_{\|}$ does not coincide with E_x. In this case, the $\mathbb{L}$ operation no longer describes the proper rotation of the field vector, as both $E_{\|}$ and $E_{\perp}$ would be affected. We can solve this problem by first rotating the frame of reference such that the θ-rotation axis coincides with the y-axis. The latter can be achieved by rotating the frame around z by $-\phi$, as follows

$$\mathbb{R} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.72}$$

After the $\mathbb{R}$ operation, the $\mathbb{L}$ operation can be applied to rotate the $E_{\|}$ component. After this step, we perform an $\mathbb{R}^{-1}$ operation to rotate the resulting vector back to the original Cartesian frame of reference. The completed transformation matrix is then $\mathbb{M} = \mathbb{R}^{-1}\mathbb{L}\mathbb{R}$, and its operation on the incident field yields

$$\mathbb{M} \cdot \mathbf{E}_0^{\text{inc}} = \mathbb{R}^{-1}\mathbb{L}\mathbb{R} \cdot \begin{bmatrix} E_x \\ E_y \\ 0 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} E_x\{(1+\cos\theta) - (1-\cos\theta)\cos 2\phi\} - E_y(1-\cos\theta)\sin 2\phi \\ -E_x(1-\cos\theta)\sin 2\phi + E_y\{(1+\cos\theta) + (1-\cos\theta)\cos 2\phi\} \\ 2E_x\sin\theta\cos\phi + 2E_y\sin\theta\sin\phi \end{bmatrix}$$

Substituting the resulting vector for $\mathbb{M} \cdot \mathbf{E}_0^{\text{inc}}$ in equation (1.70) then gives the transformed field components at the exit surface of the lens. A useful case relevant to many microscopy applications is when the incident plane wave is linearly polarized. Aligning the field vector of the input wave along the x-axis, i.e. $\mathbf{E}_0^{\text{inc}} = (E_x, 0, 0)$, where E_x is a constant amplitude, we find the following field at the exit surface of the lens

$$\mathbf{E}_0(\theta, \phi) = \left(\frac{n_1}{n_2}\right)^{1/2} \sqrt{\cos\theta}\, \frac{E_x}{2} \begin{bmatrix} (1+\cos\theta) - (1-\cos\theta)\cos 2\phi \\ -(1-\cos\theta)\sin 2\phi \\ 2\sin\theta\cos\phi \end{bmatrix} \tag{1.73}$$

Whereas the incident field only contained E_x components, when the field is mapped onto the spherical surface the rotation of the field vector causes projections of the vector in the y- and z-directions. As can be gleaned from equation (1.73), the amplitude of the E_y and E_z components grows as the polar angle θ increases. This means that lenses that support larger θ angles can have significant field contributions of y-polarized and z-polarized light at their exit surface, as shown in Figure 1.11. We now have a good description of the vector field on the reference sphere. The spherical wave thus obtained propagates to the focal plane to form a diffraction-limited focal field distribution. To find the field near focus, we first have to discuss the principle of diffraction, which is the topic of Chapter 2.

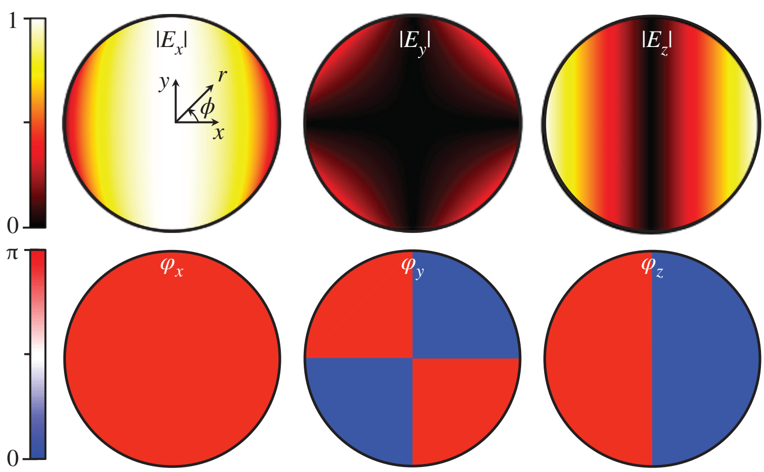

Figure 1.11 Electric field distribution on the spherical surface for an x-polarized incident plane wave of uniform amplitude. The spherical surface is represented as a two-dimensional projection, with $r = a(1 - \cos\theta)$, where a is the radius of the lens aperture. Top row shows the field amplitude of the field components, whereas the bottom row shows their relative phase.

Example 1.9 Equation (1.73) describes the field on the reference surface for an incoming plane wave that is *x*-polarized. Find a similar expression for the case that the incoming wave is *y*-polarized.

Solution We can use expression (1.73) by setting $E_x = 0$ and retaining only the components derived from the E_y component of the incident wave. We thus obtain

$$\mathbf{E}_0(\theta, \phi) = \left(\frac{n_1}{n_2}\right)^{1/2} \sqrt{\cos\theta}\, \frac{E_y}{2} \begin{bmatrix} -(1-\cos\theta)\sin 2\phi \\ (1+\cos\theta) + (1-\cos\theta)\cos 2\phi \\ 2\sin\theta\sin\phi \end{bmatrix}$$

The resulting field has similar symmetry properties as the field shown in Figure 1.11, but with the E_x and E_y patterns switching roles, and the E_z pattern rotated by $\phi = -90°$.

1.7 Intensity and Energy

We have characterized light in terms of electromagnetic fields. The field description is useful, as it tells us how an electromagnetic disturbance propagates through space. Interactions of light with charges in materials are also best understood through the concept of fields. However, to quantify the amount of light, a typical photo-detector registers the amount of energy carried by the fields rather than their amplitude and phase. We thus need a relation between the fields and the energy they carry.

1.7.1 Poynting Vector

Energy is commonly expressed in units of Joule, where $\mathrm{J} = \mathrm{kg} \cdot \mathrm{m}^2/\mathrm{s}^2$. Since electromagnetic fields transfer energy through space as a function of time, we are interested in the amount of energy transferred per unit time through a given surface area, i.e. $\mathrm{J}/(\mathrm{m}^2 \cdot \mathrm{s}) = \mathrm{W}/\mathrm{m}^2$. This quantity, electromagnetic power delivered per unit area, is also known as *irradiance* in radiometry.

The connection between electromagnetic fields and the flux of energy they represent is provided by Poynting's theorem. Starting from Maxwell's equations, Poynting's theorem establishes a relation between the rate of change of electromagnetic energy stored in a given volume V and the flux of energy across the surface S that encloses this volume. The energy density u of the electromagnetic field in V is given as

$$u = \tfrac{1}{2}\left\{\varepsilon_0 \mathbf{E}\cdot\mathbf{E} + \frac{1}{\mu_0}\mathbf{B}\cdot\mathbf{B}\right\} \tag{1.74}$$

which has units of J/m^3. We are interested in the energy flux density that traverses S, a quantity known as the instantaneous *Poynting vector*. It is defined as

$$\mathbf{S} \equiv \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} \tag{1.75}$$

and it has the desired units of power per unit area. The Poynting vector represents the flow of electromagnetic energy through space. The direction of the Poynting vector, $\hat{\mathbf{n}}_S$, corresponds to the direction of energy flow. Since the Poynting vector is a real quantity that involves a multiplication of two fields, the expressions for the fields in equation (1.75) must be real as well.

Photodetectors for visible and near-infrared light are typically not fast enough to follow the time-harmonic variations of the electromagnetic field. Instead, the energy density flux is measured over many cycles of the wave that impinges on the detector surface. The quantity of interest is, therefore, the time-averaged Poynting vector $\langle \mathbf{S} \rangle$, which is the instantaneous Poynting vector integrated in time over a full harmonic cycle of the wave. The following relation is particularly useful for calculating the time-averaged Poynting vector of time-harmonic fields

$$\langle \mathbf{S} \rangle = \frac{1}{2\mu_0} \mathrm{Re}\left\{ \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{B}}^* \right\} \tag{1.76}$$

where $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{B}}$ are the complex spatial parts of the electric and magnetic fields, respectively.

1.7.2 Intensity of Plane Waves

The Poynting vector of a plane wave can be determined by substituting the expressions for the plane wave's electric and magnetic fields into equation (1.76). For a plane wave propagating in an isotropic nonabsorbing medium with index of refraction n, the time-averaged Poynting vector is found as

$$\langle \mathbf{S} \rangle = \frac{n\varepsilon_0 c}{2} \left(\boldsymbol{\mathcal{E}} \cdot \boldsymbol{\mathcal{E}}^* \right) \cdot \hat{\mathbf{n}}_S \tag{1.77}$$

In the case of a plane wave, the direction $\hat{\mathbf{n}}_S$ is aligned with the wave vector $\mathbf{k}$, and the energy carried by a plane wave flows in the direction of wave propagation. A photodetector generally does not provide a good reading of the direction of energy flow, but registers the magnitude of electromagnetic power on its detector area instead. The measured quantity is usually referred to as the light *intensity*, which for a plane wave is written as

$$I = |\langle \mathbf{S} \rangle| = \frac{n\varepsilon_0 c}{2} |\boldsymbol{\mathcal{E}}|^2 = 2n\varepsilon_0 c \left|\mathbf{E}_0\right|^2 \tag{1.78}$$

where $|\mathbf{E}_0|^2 = |E_{0,x}|^2 + |E_{0,y}|^2 + |E_{0,z}|^2$. Similar to the Poynting vector, the units of intensity are $\mathrm{W/m^2}$.

The time-averaged energy density of the wave in a material with index of refraction n is given as

$$\langle u \rangle = \tfrac{1}{2} n^2 \varepsilon_0 |\boldsymbol{\mathcal{E}}|^2 = 2n^2 \varepsilon_0 \left|\mathbf{E}_0\right|^2 \tag{1.79}$$

If the refractive index can be approximated as $n \to 1$, then equation (1.79) simplifies to $\langle u \rangle = 2\varepsilon_0 \left|\mathbf{E}_0\right|^2$, and the intensity can be written as $I = c\langle u \rangle = 2\varepsilon_0 c \left|\mathbf{E}_0\right|^2$.

Example 1.10 A laser beam of 1.0 mW average power is focused to a spot size of 1.0 $\mu\mathrm{m}^2$ in a medium with index of refraction $n = 1.33$. Determine the amplitude of the electric field in this spot.

Solution The intensity in the illuminated spot is

$$I = \frac{1.0 \times 10^{-3}\ \mathrm{W}}{1.0 \times 10^{-12}\ \mathrm{m}^2} = 1.0 \times 10^{9}\ \mathrm{W/m^2}$$

The electric field strength $|\boldsymbol{\mathcal{E}}|$ can then be computed from equation (1.78) as

$$|\boldsymbol{\mathcal{E}}| = \frac{2I}{n\varepsilon_0 c} = 5.7 \times 10^{11}\ \mathrm{V/m}$$

Note that the field strength depends on the index of refraction of the material.

1.7.3 Power Radiated from an Electric Dipole

Since the electric dipole is a good approximation for modeling the radiative properties of a molecule, knowledge of the amount of power radiated by a dipole will come in handy when quantitative estimates of the optical signal are needed. The electric field of a radiating dipole was discussed in Section 1.3.2. Using equations (1.44) and (1.45) and substituting them in the expression for the time-averaged Poynting vector given in (1.76), we can determine $\langle \mathbf{S} \rangle$ for the radiating dipole

$$\langle \mathbf{S} \rangle = \frac{\omega^4 p^2}{32\pi^2 \varepsilon_0 c^3} \frac{\sin^2\theta}{r^2} \hat{\mathbf{r}} \tag{1.80}$$

where we have used the relation $\hat{\theta} \times \hat{\phi} = \hat{\mathbf{r}}$. We see that the energy flux is in the $\hat{\mathbf{r}}$ direction, which coincides with the direction of the wave vector $\mathbf{k}$ of the spherical wave, and that no energy flows into the direction parallel to the dipole axis ($\theta = 0$). In addition, the power density decreases quadratically with increasing distance r from the source, as we previously concluded in Section 1.3.2. In this context, it is useful to consider the optical power $d\mathcal{P}$ radiated into an element of the solid angle $d\Omega = \sin\theta d\theta d\phi$ in the outward facing radial direction, defined through

$$d\mathcal{P} = \langle \mathbf{S} \rangle \cdot \hat{\mathbf{r}}\, r^2 d\Omega \tag{1.81}$$

Substituting equation (1.80) into (1.81) and rearranging gives

$$\frac{d\mathcal{P}}{d\Omega} = \frac{\omega^4 p^2}{32\pi^2 \varepsilon_0 c^3} \sin^2\theta \tag{1.82}$$

which is known as the *differential power*, a quantity independent of the distance from the dipole. Integration of equation (1.81) over the spherical surface Ω that encloses the dipole source, gives the total radiated power $\mathcal{P}$

$$\mathcal{P} = \frac{\omega^4 p^2}{12\pi \varepsilon_0 c^3} \tag{1.83}$$

which has units of W. Once the magnitude p of the optically induced dipole is known, equation (1.83) can be used to compute the total power produced by an individual dipole radiator. Note that the power scales as ω^4, which conveys that dipole radiation becomes more efficient at shorter wavelengths.

Bibliography

1 E. Hecht *Optics*, 5th edition, Pearson, Essex, 2017.

2 J. Braat and P. Török. *Imaging Optics*. Cambridge University Press, Cambridge, 2019.

3 M. Born and E. Wolf. *Principles of Optics*, 6th edition, Cambridge University Press, Cambridge, 1998.

4 C. S. Adams and I. G. Hughes. *Optics f2f: From Fourier to Fresnel*. Oxford University Press, Oxford, 2019.