Lecture Notes on

ELECTROMAGNETIC FIELDS AND WAVES

(227-0052-10L)

Prof. Dr. Lukas Novotny
ETH Zürich, Photonics Laboratory

February 9, 2013

Introduction

The properties of electromagnetic fields and waves are most commonly discussed in terms of the electric field $\mathbf{E}(\mathbf{r}, t)$ and the magnetic induction field $\mathbf{B}(\mathbf{r}, t)$. The vector $\mathbf{r}$ denotes the location in space where the fields are evaluated. Similarly, t is the time at which the fields are evaluated. Note that the choice of $\mathbf{E}$ and $\mathbf{B}$ is arbitrary and that one could also proceed with combinations of the two, for example, with the vector and scalar potentials $\mathbf{A}$ and ϕ, respectively.

The fields $\mathbf{E}$ and $\mathbf{B}$ have been originally introduced to escape the dilemma of "action-at-distance', that is, the question of how forces are transferred between two separate locations in space. To illustrate this, consider the situation depicted in Figure 1. If we shake a charge at $\mathbf{r}_1$ then a charge at location $\mathbf{r}_2$ will respond. But how did this action travel from $\mathbf{r}_1$ to $\mathbf{r}_2$? Various explanations were developed over the years, for example, by postulating an aether that fills all space and that acts as a transport medium, similar to water waves. The fields $\mathbf{E}$ and $\mathbf{B}$ are pure constructs to deal with the "action-at-distance' problem. Thus, forces generated by

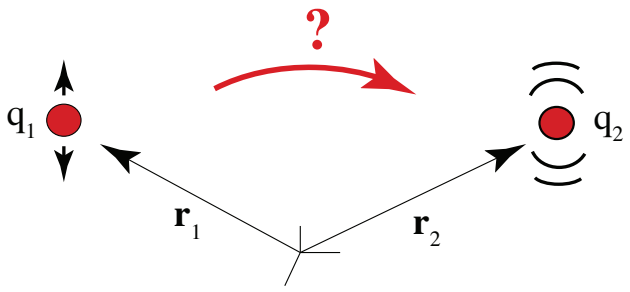

Figure 1: Illustration of "action-at-distance". Shaking a charge at $\mathbf{r}_1$ makes a second charge at $\mathbf{r}_2$ respond.

electrical charges and currents are explained in terms of $\mathbf{E}$ and $\mathbf{B}$, quantities that we cannot measure directly.

Basic Properties

As mentioned above, $\mathbf{E}$ and $\mathbf{B}$ have been introduced to explain forces acting on charges and currents. The Coulomb force (electric force) is mediated by the electric field and acts on the charge q, that is, $\mathbf{F}_\mathrm{e} = q\mathbf{E}$. It accounts for the attraction or repulsion between static charges. The interaction of static charges is referred to as *electrostatics*. On the other hand, the Lorentz force (magnetic force) accounts for the interaction between static currents (charges traveling at constant velocities $\mathbf{v} = \dot{\mathbf{r}}$) according to $\mathbf{F}_\mathrm{m} = q\mathbf{v} \times \mathbf{B}$. The interaction of static currents is referred to as *magnetostatics*. Taken the electric and magnetic forces together we arrive at

$$\mathbf{F}(\mathbf{r}, t) = q\left[\mathbf{E}(\mathbf{r}, t) \; + \; \mathbf{v}(\mathbf{r}, t) \times \mathbf{B}(\mathbf{r}, t)\right] \tag{1}$$

In the SI unit system, force is measured in Newtons (N = J / m = A V s / m) and charge in terms of Coulomb (Cb = A s). Eq. (1) therefore imposes the following units on the fields: [E] = V/m and [B] = V s / m^2. The latter is also referred to as Tesla (T).

Interestingly, the fields $\mathbf{E}$ and $\mathbf{B}$ depend on the observer's reference frame. In fact, the field $\mathbf{E}$ in one inertial frame can be equal to the field $\mathbf{B}$ in another inertial

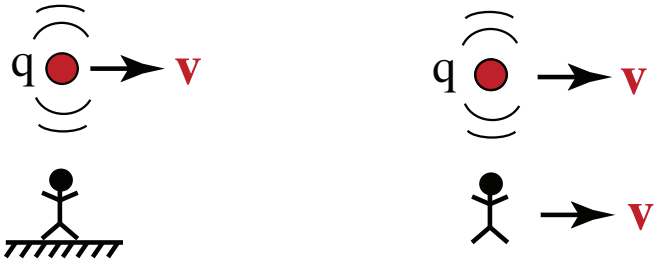

Figure 2: The fields $\mathbf{E}$ and $\mathbf{B}$ depend on the inertial frame. An observer at rest sees a $\mathbf{B}$ field when a charge at velocity $\mathbf{v}$ moves by (left). However, an observer moving at the same speed will experience only an $\mathbf{E}$ field.

frame. To illustrate this consider the two situations shown in Figure 2 where an observer is measuring the fields of a charge moving at velocity $\mathbf{v}$. An observer at rest will measure a $\mathbf{B}$ field whereas an observer moving at the same speed as the charge will only experience only an $\mathbf{E}$ field. Why? Because the charge appears to be at rest from the observer's point of view.

In general, the electric field measured by an observer at $\mathbf{r}_{\rm o}$ and at time t can be expressed as (see R. Feynman 'Lectures on Physics', Vol. II, p 21-1)

$$\mathbf{E}(\mathbf{r}_{\rm o},t) \;=\; \frac{q}{4\pi\varepsilon_0}\left[\frac{\mathbf{n}_{\rm r'}}{\rm r'^2} + \frac{\rm r'}{c}\frac{\rm d}{{\rm d}t}\left(\frac{\mathbf{n}_{\rm r'}}{\rm r'^2}\right) + \frac{1}{c^2}\frac{{\rm d}^2}{{\rm d}t^2}\mathbf{n}_{\rm r'}\right] \;, \tag{2}$$

where $c = 2.99792456\,..\,10^8$ m/s is the speed of light. As shown in Figure 3, $\rm r'$ is the distance between the charge and the observer at the earlier time $(t - {\rm r'}/c)$. Similarly, $\mathbf{n}_{\rm r'}$ is the unit vector point from the charge towards the observer at the earlier time $(t - {\rm r'}/c)$. Thus, the field at the observation point $\mathbf{r}_{\rm o}$ at the time t depends on the motion of the charge at the earlier time $(t - {\rm r'}/c)$! The reason is that it takes a time $\Delta t = {\rm r'}/c$ for the field to travel the distance $\rm r'$ to the observer.

Let us understand the different terms in Eq. (2). The first term is proportional to the *position* of the charge and describes a retarded Coulomb field. The second term is proportional to the *velocity* of the charge. Together with the first term it describes the instantaneous Coulomb field. The third term is proportional to the *acceleration* of the charge and is associated with electromagnetic radiation.

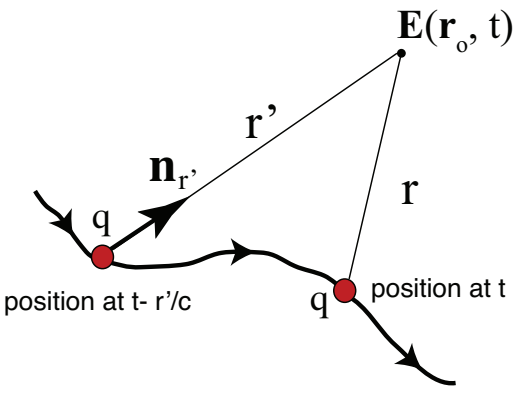

Figure 3: The field at the observation point $\mathbf{r}_{\rm o}$ at the time t depends on the motion of the charge at the earlier time $(t - {\rm r'}/c)$.

The objective of this course is to establish the theoretical foundations that lead to Eq. (2) and to develop an understanding for the generation and propagation of electromagnetic fields.

Microscopic and Macroscopic Electromagnetism

In microscopic electromagnetism one deals with discrete point charges $\mathrm{q_i}$ located at $\mathbf{r}_\mathrm{n}$ (see Figure 4). The charge density ρ and the current density $\mathbf{j}$ are then expressed as sums over Dirac delta functions

$$\rho(\mathbf{r}) \;=\; \sum_\mathrm{n} \mathrm{q_n}\,\delta[\mathbf{r}-\mathbf{r}_\mathrm{n}]\;, \tag{3}$$

$$\mathbf{j}(\mathbf{r}) \;=\; \sum_\mathrm{n} \mathrm{q_n}\,\dot{\mathrm{r}}_\mathrm{n}\,\delta[\mathbf{r}-\mathbf{r}_\mathrm{n}]\;, \tag{4}$$

where $\mathbf{r}_\mathrm{n}$ denotes the position vector of the nth charge and $\dot{\mathbf{r}}_\mathrm{n}=\mathbf{v}_\mathrm{n}$ its velocity. The total charge and current of the particle are obtained by a volume integration over ρ and $\mathbf{j}$. In terms of ρ and $\mathbf{j}$, the force law in Eq.(1) can be written as

$$\mathbf{F}(\mathbf{r},t)=\int_V [\rho(\mathbf{r},t)\mathbf{E}(\mathbf{r},t)\;+\;\mathbf{j}(\mathbf{r},t)\times\mathbf{B}(\mathbf{r},t)]\,\mathrm{d}V \tag{5}$$

where V is the volume that contains all the charges q_n.

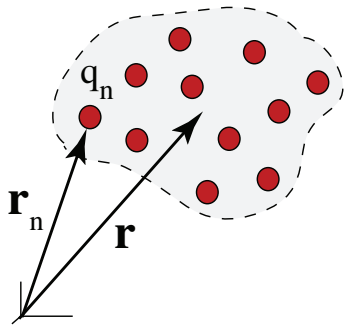

Figure 4: In the microscopic picture, optical radiation interacts with the discrete charges $\mathrm{q_n}$ of matter.

In macroscopic electromagnetism, ρ and $\mathbf{j}$ are viewed as continuous functions of position. Thus, the microscopic structure of matter is not considered and the fields become local spatial averages over microscopic fields. This is similar to the flow of water, for which the atomic scale is irrelevant.

In this course we will predominantly consider macroscopic fields for which ρ and $\mathbf{j}$ are smooth functions in space. However, the discrete nature can always be recovered by substituting Eqs. (4).

Pre-Maxwellian Electrodynamics

Let us review and summarize the laws of Ampère (Oersted), Faraday and Gauss, as introduced in the courses *Netzwerke und Schaltungen I & II*.

$$\begin{aligned}
\int_{\partial V} \mathbf{E}(\mathbf{r},t)\cdot\mathbf{n}\,da &= \frac{1}{\varepsilon_0}\int_V \rho(\mathbf{r},t)\,dV && \text{Gauss}'\text{ law (Cavendish 1772)}\\
\int_{\partial A} \mathbf{E}(\mathbf{r},t)\cdot d\mathbf{s} &= -\frac{\partial}{\partial t}\int_A \mathbf{B}(\mathbf{r},t)\cdot\mathbf{n}\,da && \text{Faraday}'\text{s law (Faraday 1825)}\\
\int_{\partial A} \mathbf{B}(\mathbf{r},t)\cdot d\mathbf{s} &= \mu_0\int_A \mathbf{j}(\mathbf{r},t)\cdot\mathbf{n}\,da && \text{Ampere}'\text{s law (Oersted 1819)}\\
\int_{\partial V} \mathbf{B}(\mathbf{r},t)\cdot\mathbf{n}\,da &= 0 && \text{No magnetic monopoles}
\end{aligned} \tag{6}$$

In our notation, V is a volume composed of infinitesimal volume elements dV, A is a surface composed of infinitesimal surface elements da, and $d\mathbf{s}$ is an infinitesimal line element. ∂V denotes the closed surface of the volume V. Similarly, ∂A is the circumference of the area A. $\mathbf{n}$ is a unit vector normal to the surface ∂V or circumference ∂A. The constants appearing in Eq. (6) are defined as follows

$$\mu_0 \;=\; 4\pi\;10^{-7}\,\frac{\mathrm{Vs}}{\mathrm{Am}} \;=\; 1.2566370\;10^{-6}\,\frac{\mathrm{Vs}}{\mathrm{Am}} \quad \text{(magnetic permeability)}$$

$$\varepsilon_0 \;=\; \frac{1}{\mu_0\,c^2} = 8.8541878\;10^{-12}\,\frac{\mathrm{As}}{\mathrm{Vm}} \quad \text{(electric permittivity)}$$

where $c = 2.99792458\,10^8$ m/s is the vacuum speed of light. Fig. 5 illustrates the meaning of the four equations (6). The first equation, Gauss' law, states that the flux of electric field through a closed surface ∂V is equal to the total charge q inside ∂V. The second equation, Faraday's law, predicts that the electric field integrated along a loop ∂A corresponds to the time rate of change of the magnetic flux through the loop. Similarly, Ampère's law, states that the magnetic field integrated along a loop ∂A is equal to the current flowing through the loop. Finally, the fourth equation states that the flux of magnetic field through a closed surface is always zero which, in a analogy to Gauss' law, indicates that there are no magnetic charges.

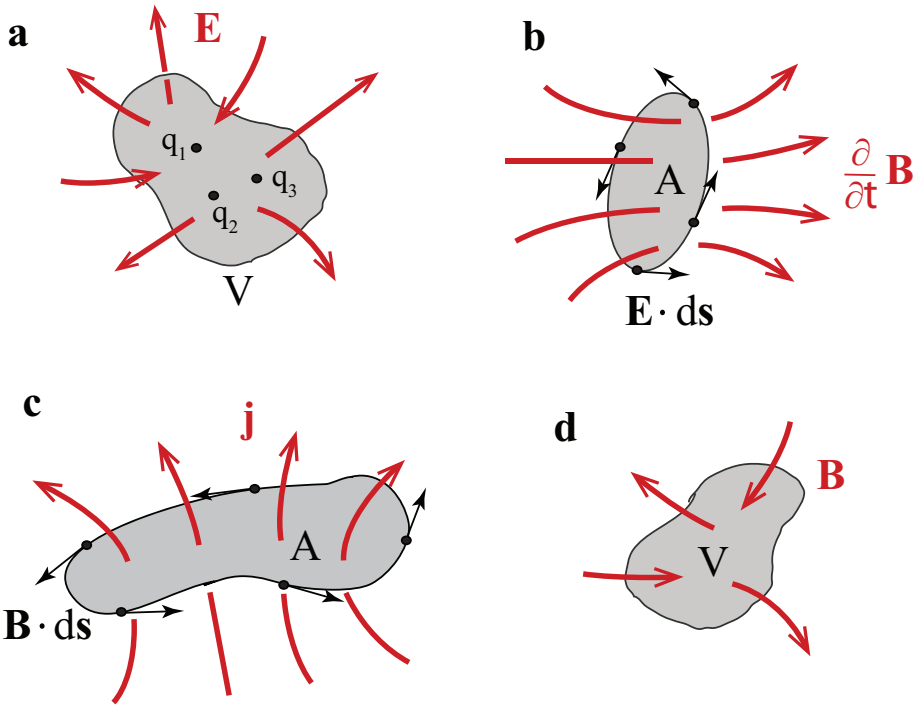

Figure 5: Illustration of (a) Gauss' law, (b) Faraday's law, (c) Ampere's law, and (d) the non-existence of magnetic charges.

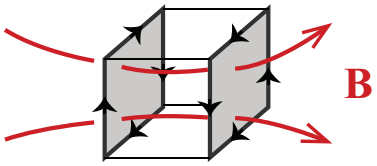

Figure 6: Ampère's law applied to a cube. Opposite faces cancel the magnetic circulation $\int_{\partial A} \mathbf{B}(\mathbf{r},t)\cdot d\mathbf{s}$, predicting that the flux of current through any closed surface is zero.

Let us consider Ampère's law for the different end faces of a small cube (c.f. Fig. 6). [1] It turns out that the magnetic field integrated along the circumference of one end face ($\int_{\partial A} \mathbf{B}(\mathbf{r},t)\cdot d\mathbf{s}$) is just the negative of the magnetic field integrated along the circumference of the opposite end face. Thus, the combined flux is zero. The same is true for the other pairs of end faces. Therefore, for any closed surface Ampère's law reduces to

$$\int_{\partial V} \mathbf{j}(\mathbf{r},t)\cdot\mathbf{n}\,da \;=\; 0 \quad \text{Kirchhoff I}\,. \tag{7}$$

In other words, the flux of current through any closed surface is zero, - what flows in has to flow out.

Eq. (7) defines the familiar current law of Kirchhoff (Knotenregel). On the other hand, Kirchhoff's voltage law (Maschenregel) follows from Faraday's law if no time-varying magnetic fields are present. In this case,

$$\int_{\partial A} \mathbf{E}(\mathbf{r},t)\cdot d\mathbf{s} \;=\; 0 \quad \text{Kirchhoff II}\,. \tag{8}$$

The two Kirchhoff laws form the basis for circuit theory and electronic design.

[1] An arbitrary volume can be viewed as a sum over infinitesimal cubes.

Chapter 1

Maxwell's Equations

Equations (6) summarize the knowledge of electromagnetism as it was understood by the mid 19th century. In 1873, however, James Clerk Maxwell introduced a critical modification that kick-started an era of wireless communication.

1.1 The Displacement Current

Eq. (7) is a statement of current conservation, that is, currents cannot be generated or destroyed, the net flux through a closed surface is zero. However, this law is flawed. For example, let's take a bunch of identical charges and hold them together (see Fig. 1.1). Once released, the charges will speed out because of Coulomb repulsion and there will be a net outward current. Evidently, the outward current is balanced by the decrease of charge inside the enclosing surface ∂V, and hence, Eq. (7) has to be corrected as follows

$$\int_{\partial V} \mathbf{j}(\mathbf{r},t)\cdot\mathbf{n}\,da \;=\; -\frac{\partial}{\partial t}\int_V \rho(\mathbf{r},t)\,dV \tag{1.1}$$

This equation describes the conservation of charge. It's general form is found in many different contexts in physics and we will encounter it again when we discuss the conservation of energy (Poynting theorem).

Because Eq. (7) has been derived from Ampère's law, we need to modify the

latter in order to end up with the correct conservation law of Eq. (1.1). This is where Maxwell comes in. He added an additional term to Ampère's law and arrived at

$$\int_{\partial A} \mathbf{B}(\mathbf{r},t)\cdot d\mathbf{s} \;=\; \mu_0 \int_A \mathbf{j}(\mathbf{r},t)\cdot \mathbf{n}\, da \;+\; \frac{1}{c^2}\,\frac{\partial}{\partial t}\int_A \mathbf{E}(\mathbf{r},t)\cdot \mathbf{n}\, da\;, \tag{1.2}$$

where $1/c^2 = \varepsilon_0\mu_0$. The last term has the form of a time-varying current. Therefore, $\varepsilon_0\,\partial\mathbf{E}/\partial t$ is referred to as the *displacement current*.

We again apply this equation to the end faces of a small cube (c.f. Fig. 6) and, as before, the left hand side vanishes. Thus,

$$\mu_0 \int_{\partial V} \mathbf{j}(\mathbf{r},t)\cdot \mathbf{n}\, da \;+\; \frac{1}{c^2}\,\frac{\partial}{\partial t}\int_{\partial V} \mathbf{E}(\mathbf{r},t)\cdot \mathbf{n}\, da \;=\; 0\;. \tag{1.3}$$

Substituting Gauss' law from Eq. (6) for the second expression yields the desired charge continuity equation (1.1).

In summary, replacing Ampère's law in (6) by Eq. (1.2) yields a set of four equations for the fields $\mathbf{E}$ and $\mathbf{B}$ that are consistent with the charge continuity equation. These four equations define what is called Maxwell's integral equations.

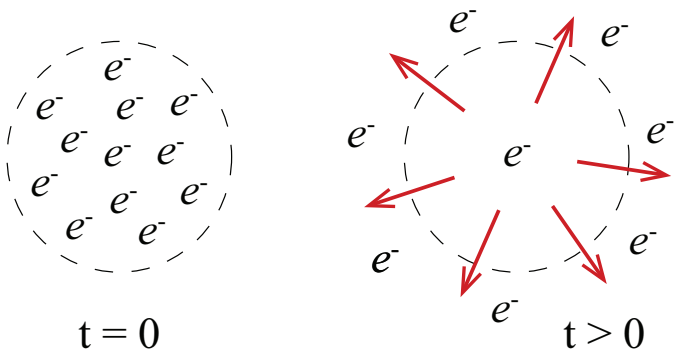

Figure 1.1: Illustration of charge conservation. A bunch of identical charges is held together at $t = 0$. Once released, the charges will spread out due to Coulomb repulsion, which gives rise to a net outward current flow.

1.2 Interaction of Fields with Matter

So far we have discussed the properties of the fields $\mathbf{E}$ and $\mathbf{B}$ in free space. The sources of these fields are charges ρ and currents $\mathbf{j}$, so-called *primary* sources. However, $\mathbf{E}$ and $\mathbf{B}$ can also interact with materials and generate *induced* charges and currents. These are then called *secondary* sources.

To account for these secondary sources we write

$$\rho_{\rm tot} \;=\; \rho \,+\, \rho_{\rm pol}\;, \tag{1.4}$$

where ρ is the charge density associated with primary sources. It is assumed that these sources are *not* affected by the fields $\mathbf{E}$ and $\mathbf{B}$. On the other hand, $\rho_{\rm pol}$ is the charge density induced in matter through the interaction with the electric field. It is referred to as the polarization charge density.[1] On a microscopic scale, the electric field slightly distorts the atomic orbitals in the material (see Fig. 1.2). On a macroscopic scale, this results in an accumulation of charges at the surface of the material (see Fig. 1.3). The net charge density inside the material remains zero.

To account for polarization charges we introduce the polarization $\mathbf{P}$ which, in analogy to Gauss' law in Eq. (6), is defined as

$$\int_{\partial V} \mathbf{P}(\mathbf{r},t)\cdot\mathbf{n}\,da \;=\; -\int_V \rho_{\rm pol}(\mathbf{r},t)\,dV\;. \tag{1.5}$$

$\mathbf{P}$ has units of Cb/m^2, which corresponds to dipole moment (Cb / m) per unit vol-

[1] The $\mathbf{B}$-field interacts only with currents and not with charges.

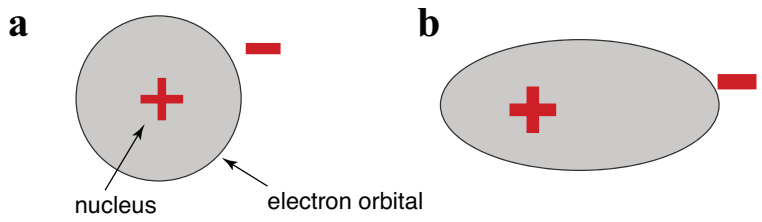

Figure 1.2: Microscopic polarization. An external electric field $\mathbf{E}$ distorts the orbital of an atom. (a) Situation with no external field. (b) Situation with external field.

ume (m^3). Inserting Eqs. (1.4) and (1.5) into Gauss' law yields

$$\int_{\partial V} [\varepsilon_0 \mathbf{E}(\mathbf{r},t) + \mathbf{P}(\mathbf{r},t)] \cdot \mathbf{n}\, da = \int_V \rho(\mathbf{r},t)\, dV \ . \tag{1.6}$$

The expression in brackets is called the *electric displacement*

$$\mathbf{D} \ = \ \varepsilon_0 \mathbf{E} + \mathbf{P} \ . \tag{1.7}$$

Time-varying polarization charges give rise to polarization currents. To see this, we take the time-derivative of Eq. (1.5) and obtain

$$\int_{\partial V} \frac{\partial}{\partial t}\mathbf{P}(\mathbf{r},t) \cdot \mathbf{n}\, da \ = \ \frac{\partial}{\partial t} \int_V \rho_{\rm pol}(\mathbf{r},t)\, dV \ , \tag{1.8}$$

which has the same appearance as the charge conservation law (1.1). Thus, we identify $\partial \mathbf{P}/\partial t$ as the *polarization current density*

$$\mathbf{j}_{\rm pol}(\mathbf{r},t) \ = \ \frac{\partial}{\partial t}\mathbf{P}(\mathbf{r},t) \ . \tag{1.9}$$

To summarize, the interaction of the $\mathbf{E}$-field with matter gives rise to polarization charges and polarization currents. The magnitude and the dynamics of these secondary sources depends on the material properties [$\mathbf{P} = f(\mathbf{E})$], which is the subject of solid-state physics.

An electric field interacting with matter not only gives rise to polarization currents but also to conduction currents. We will denote the conduction current density as $\mathbf{j}_{\rm cond}$. Furthermore, according to Ampère's law, the interaction of matter with

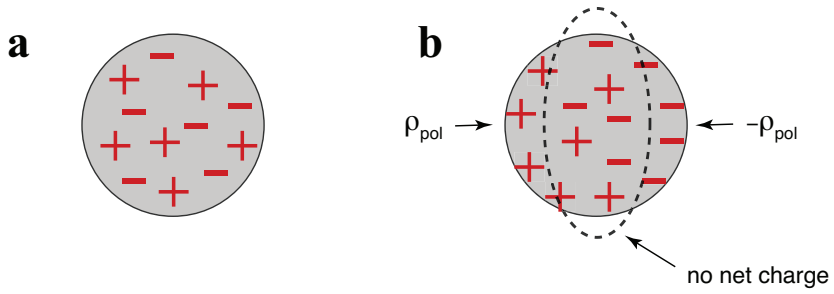

Figure 1.3: Macroscopic polarization. An external electric field $\mathbf{E}$ accumulates charges at the surface of an object. (a) Situation with no external field. (b) Situation with external field.

magnetic fields can induce magnetization currents. We will denote the magnetization current density as $\mathbf{j}_{\mathrm{mag}}$. Taken all together, the total current density can be written as

$$\mathbf{j}_{\mathrm{tot}}(\mathbf{r},t) \;=\; \mathbf{j}_0(\mathbf{r},t) + \mathbf{j}_{\mathrm{cond}}(\mathbf{r},t) + \mathbf{j}_{\mathrm{pol}}(\mathbf{r},t) + \mathbf{j}_{\mathrm{mag}}(\mathbf{r},t)\;, \tag{1.10}$$

where $\mathbf{j}_0$ is the source current density. In the following we will not distinguish between source current and conduction current and combine the two as

$$\mathbf{j}(\mathbf{r},t) \;=\; \mathbf{j}_0(\mathbf{r},t) + \mathbf{j}_{\mathrm{cond}}(\mathbf{r},t)\;. \tag{1.11}$$

$\mathbf{j}$ is simply the current density due to *free* charges, no matter whether primary or secondary. On the other hand, $\mathbf{j}_{\mathrm{pol}}$ is the current density due to *bound* charges, that is, charges that experience a restoring force to a point of origin. Finally, $\mathbf{j}_{\mathrm{mag}}$ is the current density due to *circulating* charges, associated with magnetic moments.

We now introduce (1.10) into Ampère's modified law (1.2) and obtain

$$\int_{\partial A} \mathbf{B}\cdot d\mathbf{s} \;=\; \mu_0 \int_A \left[\mathbf{j} \;+\; \left(\mathbf{j}_{\mathrm{pol}} + \varepsilon_0 \frac{\partial}{\partial t}\,\mathbf{E}\right) \;+\; \mathbf{j}_{\mathrm{mag}}\right]\cdot \mathbf{n}\, da\;, \tag{1.12}$$

where we have skipped the arguments $(\mathbf{r},t)$ for simplicity. According to Eqs. (1.9) and (1.7), the term inside the round brackets is equal to $\partial\mathbf{D}/\partial t$. To relate the induced magnetization current to the $\mathbf{B}$-field we define in analogy to Eq. (1.5)

$$\int_{\partial A} \mathbf{M}(\mathbf{r},t)\cdot d\mathbf{s} \;=\; \int_A \mathbf{j}_{\mathrm{mag}}(\mathbf{r},t)\cdot \mathbf{n}\, da\;, \tag{1.13}$$

with $\mathbf{M}$ being the magnetization. Inserting into Eq. (1.12) and rearranging terms leads to

$$\int_{\partial A} \left[\frac{1}{\mu_0}\mathbf{B} - \mathbf{M}\right]\cdot d\mathbf{s} \;=\; \int_A \left[\mathbf{j} \;+\; \frac{\partial}{\partial t}\,\mathbf{D}\right]\cdot \mathbf{n}\, da\;, \tag{1.14}$$

The expression in brackets on the left hand side is called the *magnetic field*

$$\mathbf{H} \;=\; \frac{1}{\mu_0}\mathbf{B} - \mathbf{M}\;. \tag{1.15}$$

It has units of A/m. The magnitude and the dynamics of magnetization currents depends on the specific material properties [$\mathbf{M} = f(\mathbf{B})$].

1.3 Maxwell's Equations in Integral Form

Let us now summarize our knowledge electromagnetism. Accounting for Maxwell's displacement current and for secondary sources (conduction, polarization and magnetization) turns our previous set of four equations (6) into

$$\int_{\partial V} \mathbf{D}(\mathbf{r},t)\cdot\mathbf{n}\,da = \int_V \rho(\mathbf{r},t)\,dV \tag{1.16}$$

$$\int_{\partial A} \mathbf{E}(\mathbf{r},t)\cdot d\mathbf{s} = -\frac{\partial}{\partial t}\int_A \mathbf{B}(\mathbf{r},t)\cdot\mathbf{n}\,da \tag{1.17}$$

$$\int_{\partial A} \mathbf{H}(\mathbf{r},t)\cdot d\mathbf{s} = \int_A \left[\mathbf{j}(\mathbf{r},t)+\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t)\right]\cdot\mathbf{n}\,da \tag{1.18}$$

$$\int_{\partial V} \mathbf{B}(\mathbf{r},t)\cdot\mathbf{n}\,da = 0 \tag{1.19}$$

The displacement $\mathbf{D}$ and the induction $\mathbf{B}$ account for secondary sources through

$$\mathbf{D}(\mathbf{r},t) = \varepsilon_0\mathbf{E}(\mathbf{r},t)+\mathbf{P}(\mathbf{r},t)\,, \quad \mathbf{B}(\mathbf{r},t) = \mu_0\left[\mathbf{H}(\mathbf{r},t)+\mathbf{M}(\mathbf{r},t)\right] \tag{1.20}$$

These equations are always valid since they don't specify any material properties. To solve Maxwell's equations (1.16)-(1.19) we need to invoke specific material properties, i.e. $\mathbf{P} = f(\mathbf{E})$ and $\mathbf{M} = f(\mathbf{B})$, which are denoted *constitutive relations*.

1.4 Maxwell's Equations in Differential Form

For most of this course it will be more convenient to express Maxwell's equations in differential form. Using Stokes' and Gauss' theorems we can easily transform Eq. (1.16)-(1.19). However, before doing so we shall first establish the notation that we will be using.

Differential Operators

The gradient operator (grad) will be represented by the nabla symbol (∇) and is defined as a Cartesian vector

$$\nabla \equiv \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} . \tag{1.21}$$

It can be transformed to other coordinate systems in a straightforward way. Using ∇ we can express the divergence operator (div) as $\nabla\cdot$. To illustrate this, let's operate with $\nabla\cdot$ on a vector $\mathbf{F}$

$$\nabla \cdot \mathbf{F} = \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \cdot \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \frac{\partial}{\partial x}F_x + \frac{\partial}{\partial y}F_y + \frac{\partial}{\partial z}F_z \,. \tag{1.22}$$

In other words, the divergence of $\mathbf{F}$ is the scalar product of ∇ and $\mathbf{F}$.

Similarly, we express the rotation operator (rot) as $\nabla\times$ which, when applied to a vector $\mathbf{F}$ yields

$$\nabla \times \mathbf{F} = \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \times \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \begin{bmatrix} \partial F_z/\partial y - \partial F_y/\partial z \\ \partial F_x/\partial z - \partial F_z/\partial x \\ \partial F_y/\partial x - \partial F_x/\partial y \end{bmatrix} . \tag{1.23}$$

Thus, the rotation of $\mathbf{F}$ is the vector product of ∇ and $\mathbf{F}$.

Finally, the Laplacian operator (Δ) can be written as $\nabla \cdot \nabla = \nabla^2$. Applied to a scalar ψ we obtain

$$\nabla^2\psi = \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \cdot \begin{bmatrix} \partial\psi/\partial x \\ \partial\psi/\partial y \\ \partial\psi/\partial z \end{bmatrix} = \frac{\partial^2}{\partial x^2}\psi + \frac{\partial^2}{\partial y^2}\psi + \frac{\partial^2}{\partial z^2}\psi \,. \tag{1.24}$$

Very often we will encounter sequences of differential operators, such as $\nabla \times \nabla\times$. The following identities can be easily verified and are helpful to memorize

$$\nabla \times \nabla\psi = 0 \tag{1.25}$$

$$\nabla \cdot (\nabla \times \mathbf{F}) = 0 \tag{1.26}$$

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2\mathbf{F} \,. \tag{1.27}$$

The last term stands for the vector $\nabla^2\mathbf{F} = [\nabla^2 F_x,\, \nabla^2 F_y,\, \nabla^2 F_z]^T$.

The Theorems of Gauss and Stokes

The theorems of Gauss and Stokes have been derived in *Analysis II*. We won't reproduce the derivation and only state their final forms

$$\int_{\partial V} \mathbf{F}(\mathbf{r},t)\cdot\mathbf{n}\,da \;=\; \int_V \nabla\cdot\mathbf{F}(\mathbf{r},t)\,dV \qquad \text{Gauss}'\text{ theorem} \tag{1.28}$$

$$\int_{\partial A} \mathbf{F}(\mathbf{r},t)\cdot d\mathbf{s} \;=\; \int_A [\nabla\times\mathbf{F}(\mathbf{r},t)]\cdot\mathbf{n}\,da \qquad \text{Stokes}'\text{ theorem} \tag{1.29}$$

Using these theorems we can turn Maxwell's integral equations (1.16)-(1.19) into differential form.

Differential Form of Maxwell's Equations

Applying Gauss' theorem to the left hand side of Eq. (1.16) replaces the surface integral over ∂V by a volume integral over V. The same volume integration is performed on the right hand side, which allows us to write

$$\int_{\partial V} \left[\nabla\cdot\mathbf{D}(\mathbf{r},t) \;-\; \rho(\mathbf{r},t)\right] dV \;=\; 0\,. \tag{1.30}$$

This result has to hold for any volume V, which can only be guaranteed if the integrand is zero, that is,

$$\nabla\cdot\mathbf{D}(\mathbf{r},t) \;=\; \rho(\mathbf{r},t)\,. \tag{1.31}$$

This is Gauss' law in differential form. Similar steps and arguments can be applied to the other three Maxwell equations, and we end up with *Maxwell's equations in differential form*

$$\nabla\cdot\mathbf{D}(\mathbf{r},t) \;=\; \rho(\mathbf{r},t) \tag{1.32}$$

$$\nabla\times\mathbf{E}(\mathbf{r},t) \;=\; -\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r},t) \tag{1.33}$$

$$\nabla\times\mathbf{H}(\mathbf{r},t) \;=\; \frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t) \;+\; \mathbf{j}(\mathbf{r},t) \tag{1.34}$$

$$\nabla\cdot\mathbf{B}(\mathbf{r},t) \;=\; 0 \tag{1.35}$$

It has to be noted that it was Oliver Heaviside who in 1884 has first written Maxwell's equations in this compact vectorial form. Maxwell had written most of his equations in Cartesian coordinates, which yielded long and complicated expressions.

Maxwell's equations form a set of four coupled differential equations for the fields $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{H}$. The components of these vector fields constitute a set of 16 unknowns. Depending on the considered medium, the number of unknowns can be reduced considerably. For example, in linear, isotropic, homogeneous and source-free media the electromagnetic field is entirely defined by two scalar fields. Maxwell's equations combine and complete the laws formerly established by Faraday, Oersted, Ampère, Gauss, Poisson, and others. Since Maxwell's equations are differential equations they do not account for any fields that are constant in space and time. Any such field can therefore be added to the fields.

Let us remind ourselves that the concept of fields was introduced to explain the transmission of forces from a source to a receiver. The physical observables are therefore forces, whereas the fields are definitions introduced to explain the troublesome phenomenon of the "action at a distance".

The conservation of charge is implicitly contained in Maxwell's equations. Taking the divergence of Eq. (1.34), noting that $\nabla \cdot \nabla \times \mathbf{H}$ is identical zero, and substituting Eq. (1.32) for $\nabla \cdot \mathbf{D}$ one obtains the continuity equation

$$\nabla \cdot \mathbf{j}(\mathbf{r},t) + \frac{\partial}{\partial t}\rho(\mathbf{r},t) \;=\; 0 \tag{1.36}$$

consistent with the integral form (1.1) derived earlier.

Chapter 2

The Wave Equation

After substituting the fields D and B in Maxwell's *curl* equations by the expressions in (1.20) and combining the two resulting equations we obtain the inhomogeneous wave equations

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial}{\partial t}\left(\mathbf{j} + \frac{\partial \mathbf{P}}{\partial t} + \nabla \times \mathbf{M}\right) \tag{2.1}$$

$$\nabla \times \nabla \times \mathbf{H} + \frac{1}{c^2}\frac{\partial^2 \mathbf{H}}{\partial t^2} = \nabla \times \mathbf{j} + \nabla \times \frac{\partial \mathbf{P}}{\partial t} - \frac{1}{c^2}\frac{\partial^2 \mathbf{M}}{\partial t^2} \tag{2.2}$$

where we have skipped the arguments $(\mathbf{r}, t)$ for simplicity. The expression in the round brackets corresponds to the *total current density*

$$\mathbf{j} = \mathbf{j} + \frac{\partial \mathbf{P}}{\partial t} + \nabla \times \mathbf{M} \ , \tag{2.3}$$

where $\mathbf{j}$ is the source and the conduction current density, $\partial \mathbf{P}/\partial t$ the polarization current density, and $\nabla \times \mathrm{M}$ the magnetization current density. The wave equations as stated in Eqs. (2.1) and (2.2) do not impose any conditions on the media and hence are generally valid.

2.1 Homogeneous Solution in Free Space

We first consider the solution of the wave equations in free space, in absence of matter and sources. For this case the right hand sides of the wave equations are

zero. The operation $\nabla \times \nabla \times$ can be replaced by the identity (1.27), and since in free space $\nabla \cdot \mathbf{E} = 0$ the wave equation for $\mathbf{E}$ becomes

$$\nabla^2 \mathbf{E}(\mathbf{r},t) \; - \; \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{r},t) \; = \; 0 \tag{2.4}$$

with an identical equation for the $\mathbf{H}$-field. Each equation defines three independent scalar equations, namely one for E_x, one for E_y, and one for E_z.

In the one-dimensional scalar case, that is $E(x,t)$, Eq. (2.4) is readily solved by the ansatz of d'Alembert $E(x,t) = E(x-ct)$, which shows that the field propagates through space at the constant velocity c. To tackle three-dimensional vectorial fields we proceed with standard separation of variables

$$\mathbf{E}(\mathbf{r},t) \; = \; \mathbf{R}(\mathbf{r})\, T(t)\,. \tag{2.5}$$

Inserting into Eq. (2.4) equation leads to

$$c^2\,\frac{\nabla^2 \mathbf{R}(\mathbf{r})}{\mathbf{R}(\mathbf{r})} \; - \; \frac{1}{T(t)}\frac{\partial^2 T(t)}{\partial\, t^2} \; = \; 0\,. \tag{2.6}$$

The first term depends only on spatial coordinates $\mathbf{r}$ whereas the second one depends only on time t. Both terms have to add to zero, independent of the values of $\mathbf{r}$ and t. This is only possible if each term is constant. We will denote this constant as $-\omega^2$. The equations for $T(t)$ and $\mathbf{R}(\mathbf{r})$ become

$$\frac{\partial^2}{\partial\, t^2} T(t) + \omega^2\, T(t) \; = \; 0 \tag{2.7}$$

$$\nabla^2 \mathbf{R}(\mathbf{r}) + \frac{\omega^2}{c^2}\,\mathbf{R}(\mathbf{r}) \; = \; 0\,. \tag{2.8}$$

Note that both $\mathbf{R}(\mathbf{r})$ and $T(t)$ are real functions of real variables.

Eq. (2.7) is a harmonic differential equation with the solutions

$$T(t) \; = \; c'_\omega \cos[\omega t] + c''_\omega \sin[\omega t] \; = \; \mathrm{Re}\{c_\omega \exp[-i\omega t]\}\;, \tag{2.9}$$

where c'_ω and c''_ω are real constants and $c_\omega = c'_\omega + i c''_\omega$ is a complex constant. Thus, according to ansatz (2.5) we find the solutions

$$\mathbf{E}(\mathbf{r},t) \; = \; \mathbf{R}(\mathbf{r})\,\mathrm{Re}\{c_\omega \exp[-i\omega t]\} \; = \; \mathrm{Re}\{c_\omega \mathbf{R}(\mathbf{r})\, \exp[-i\omega t]\}\,. \tag{2.10}$$

In what follows, we will denote $c_\omega R(\mathbf{r})$ as the *complex field amplitude* and abbreviate it by $\mathbf{E}(\mathbf{r})$. Thus,

$$\mathbf{E}(\mathbf{r},t) \;=\; \mathrm{Re}\{\mathbf{E}(\mathbf{r})\,\mathrm{e}^{-i\omega t}\} \tag{2.11}$$

Notice that $\mathbf{E}(\mathbf{r})$ is a *complex* field whereas the true field $\mathbf{E}(\mathbf{r},t)$ is real. The symbol $\mathbf{E}$ will be used for both, the real time-dependent field and the complex spatial part of the field. The introduction of a new symbol is avoided in order to keep the notation simple. Eq. (2.11) describes the solution of a *time-harmonic* electric field, a field that oscillates in time at the fixed angular frequency ω. Such a field is also referred to as *monochromatic* field.

After inserting (2.11) into Eq. (2.4) we obtain

$$\nabla^2\mathbf{E}(\mathbf{r}) \;+\; k^2\,\mathbf{E}(\mathbf{r}) \;=\; 0 \tag{2.12}$$

with $k = |\mathbf{k}| = \omega/c$. This equation is referred to as *Helmholtz equation*.

2.1.1 Plane Waves

To solve for the solutions of the Helmholtz equation (2.12) we use the ansatz

$$\mathbf{E}(\mathbf{r}) \;=\; \mathbf{E}_0\,\mathrm{e}^{\pm i\mathbf{k}\cdot\mathbf{r}} \;=\; \mathbf{E}_0\,\mathrm{e}^{\pm i(k_x x+k_y y+k_z z)} \tag{2.13}$$

which, after inserting into (2.12), yields

$$k_x^2 \;+\; k_y^2 \;+\; k_z^2 \;=\; \frac{\omega^2}{c^2} \tag{2.14}$$

The left hand side can also be represented by $\mathbf{k}\cdot bfk = k^2$. For the following we assume that k_x, k_y, and k_z are real. After inserting Eq. (2.13) into Eq. (2.11) we find the solutions

$$\mathbf{E}(\mathbf{r},t) \;=\; \mathrm{Re}\{\mathbf{E}_0\,\mathrm{e}^{\pm i\mathbf{k}\cdot\mathbf{r}-i\omega t}\} \tag{2.15}$$

which are called *plane waves* or homogeneous waves. Solutions with the $+$ sign in the exponent are waves that propagate in direction of $\mathbf{k} = [k_x,\, k_y,\, k_z]$. They are denoted *outgoing waves*. On the other hand, solutions with the $-$ sign are incoming waves and propagate against the direction of $\mathbf{k}$.

Although the field $\mathbf{E}(\mathbf{r},t)$ fulfills the wave equation it is not yet a rigorous solution of Maxwell's equations. We still have to require that the fields are divergence free, i.e. $\nabla\cdot\mathbf{E}(\mathbf{r},t) = 0$. This condition restricts the $\mathbf{k}$-vector to directions perpendicular to the electric field vector ($\mathbf{k}\cdot\mathbf{E}_0 = 0$). Fig. 2.1 illustrates the characteristic features of plane waves.

The corresponding magnetic field is readily found by using Maxwell's equation $\nabla\times\mathbf{E} = i\omega\mu_0\,\mathbf{H}$. We find $\mathbf{H}_0 = (\omega\mu_0)^{-1}\,(\mathbf{k}\times\mathbf{E}_0)$, that is, the magnetic field vector is perpendicular to the electric field vector and the wavevector $\mathbf{k}$.

Let us consider a plane wave with real amplitude $\mathbf{E}_0$ and propagating in direction of the z axis. This plane wave is represented by $\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0\cos[kz-\omega t]$, where $k = |\mathbf{k}| = \omega/c$. If we observe this field at a fixed position z then we'll measure an electric field $\mathbf{E}(t)$ that is oscillating with angular frequency $f = \omega\,/\,2\pi$. On the other hand, if we take a snapshot of this plane wave at $t = 0$ then we'll observe a field that spatially varies as $\mathbf{E}(\mathbf{r},t=0) = \mathbf{E}_0\cos[k\,z]$. It has a maximum at $z = 0$ and a next maximum at $k\,z = 2\pi$. The separation between maxima is $\lambda = 2\pi/k$ and is called the wavelength. After a time of $t = 2\pi/\omega$ the field reads $\mathbf{E}(\mathbf{r},t = 2\pi/\omega) = \mathbf{E}_0\cos[k\,z-2\pi] = \mathbf{E}_0\cos[k\,z]$, that is, the wave has propagated a distance of one wavelength in direction of z. Thus, the velocity of the wave is $v_0 = \lambda/(2\pi/\omega) = \omega/k = c$, the vacuum speed of light. For radio waves

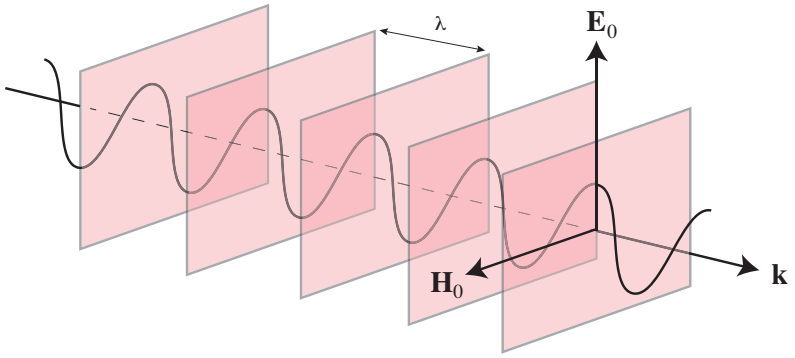

Figure 2.1: Illustration of a plane wave. In free space, the plane wave propagates with velocity c in direction of the wave vector $\mathbf{k} = (k_x,\,k_y,\,k_z)$. The electric field vector $\mathbf{E}_0$, the magnetic field vector $\mathbf{H}_0$, and $\mathbf{k}$ are perpendicular to each other.

$\lambda \sim 1$ km, for microwaves $\lambda \sim 1$ cm, for infrared radiation $\lambda \sim 10\,\mu$m, for visible light $\lambda \sim 500$ nm, and for X-rays $\lambda \sim 0.1$ nm, - the size range of atoms. Fig. 2.2 illustrates the length scales associated with the different frequency regions of the electromagnetic spectrum.

A plane wave with a fixed direction of the electric field vector $\mathbf{E}_0$ is termed *linearly polarized*. We can form other polarizations states (e.g. circularly polarized waves) by allowing $\mathbf{E}_0$ to rotate as the wave propagates. Such polarization states can be generated by superposition of linearly polarized plane waves.

Plane waves are mathematical constructs that do not exist in practice because their fields $\mathbf{E}$ and $\mathbf{H}$ are infinitely extended in space and therefore carry an infinite amount of energy. Thus, plane waves are mostly used to locally visualize or approximate more complicated fields. They are the simplest form of waves and can be used as a basis to describe other wave fields (angular spectrum representation). For an illustration of plane waves go to http://en.wikipedia.org/wiki/Plane_wave.

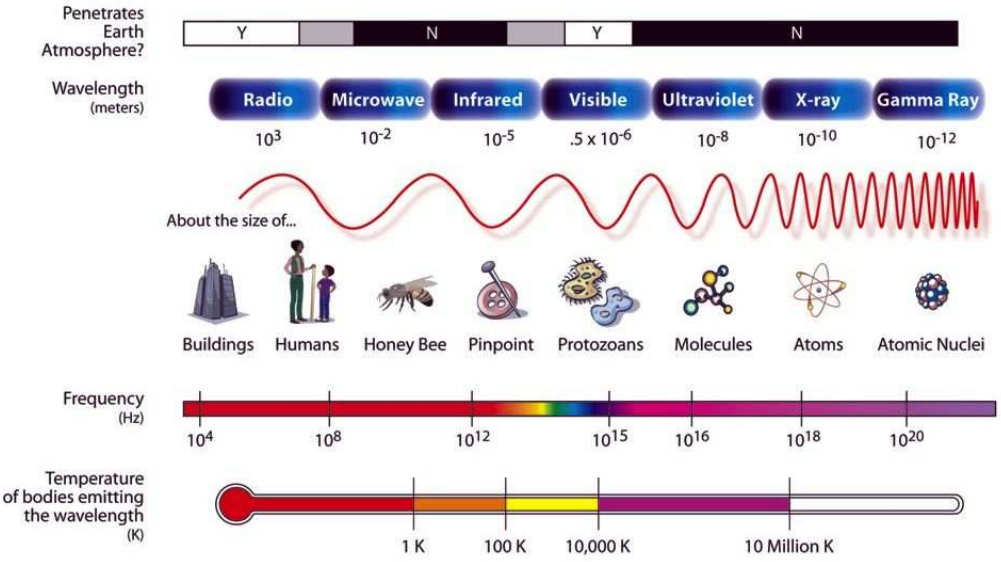

Figure 2.2: Length-scales associated with the different frequency ranges of the electromagnetic spectrum. From mynasadata.larc.nasa.gov.

2.1.2 Evanescent Waves

So far we have restricted the discussion to real k_x, k_y, and k_z. However, this restriction can be relaxed. For this purpose, let us rewrite the dispersion relation (2.14) as

$$k_z \;=\; \sqrt{(\omega^2/c^2) \;-\; (k_x^2 + k_y^2)}\;. \tag{2.16}$$

If we let $(k_x^2 + k_y^2)$ become larger than $k^2 = \omega^2/c^2$ then the square root no longer yields a real value for k_z. Instead, k_z becomes imaginary. The solution (2.15) then turns into

$$\mathbf{E}(\mathbf{r}, t) \;=\; \mathrm{Re}\{\mathbf{E}_0\, \mathrm{e}^{\pm i(k_x x + k_y y) - i\omega t}\}\; \mathrm{e}^{\mp |k_z| z}\;. \tag{2.17}$$

These waves still oscillate like plane waves in the directions of x and y, but they exponentially decay or grow in the direction of z. Typically, they have a plane of origin $z = \mathrm{const.}$ that coincides, for example, with the surface of an insulator or metal. If space is unbounded for $z > 0$ we have to reject the exponentially growing solution on grounds of energy conservation. The remaining solution is exponentially decaying and vanishes at $z \to \infty$ (evanescere = to vanish). Because of their exponential decay, evanescent waves only exist near sources (primary or secondary) of electromagnetic radiation. Evanescent waves form a source of stored energy (reactive power). In light emitting devices, for example, we want to convert

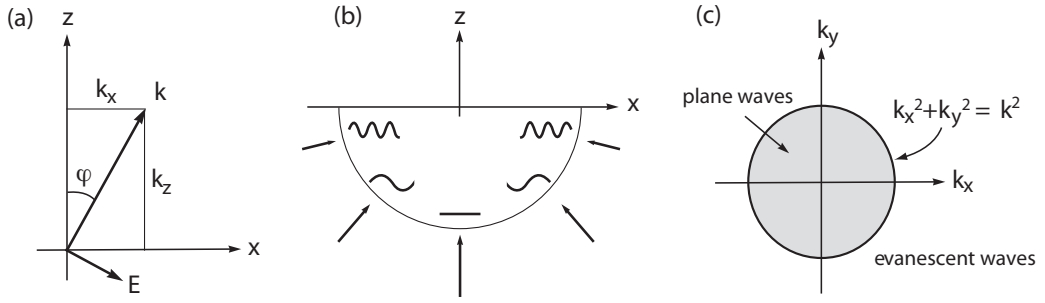

Figure 2.3: (a) Representation of a plane wave propagating at an angle φ to the z axis. (b) Illustration of the transverse spatial frequencies of plane waves incident from different angles. The transverse wavenumber $(k_x^2 + k_y^2)^{1/2}$ depends on the angle of incidence and is limited to the interval $[0 \ldots k]$. (c) The transverse wavenumbers k_x, k_y of plane waves are restricted to a circular area with radius $k = \omega/c$. Evanescent waves fill the space outside the circle.

evanescent waves into propagating waves to increase the energy efficiency.

To summarize, for a wave with a fixed $(k_x,\, k_y)$ pair we find two different characteristic solutions

$$\begin{aligned} \text{Plane waves}: &\quad \mathrm{e}^{i\,[k_x x + k_y y]}\,\mathrm{e}^{\pm i\,|k_z|\,z} &\quad (k_x^2 + k_y^2 \le k^2) \\ \text{Evanescent waves}: &\quad \mathrm{e}^{i\,[k_x x + k_y y]}\,\mathrm{e}^{-|k_z|\,|z|} &\quad (k_x^2 + k_y^2 > k^2)\,. \end{aligned} \tag{2.18}$$

Plane waves are oscillating functions in z and are restricted by the condition $k_x^2 + k_y^2 \le k^2$. On the other hand, for $k_x^2 + k_y^2 > k^2$ we encounter evanescent waves with an exponential decay along the z-axis. Figure 2.3 shows that the larger the angle between the $\mathbf{k}$-vector and the z-axis is, the larger the oscillations in the transverse plane will be. A plane wave propagating in the direction of z has no oscillations in the transverse plane ($k_x^2 + k_y^2 = 0$), whereas, in the other limit, a plane wave propagating at a right angle to z shows the highest spatial oscillations in the transverse plane ($k_x^2 + k_y^2 = k^2$). Even higher spatial frequencies are covered by evanescent waves. In principle, an infinite bandwidth of spatial frequencies $(k_x,\, k_y)$ can be achieved. However, the higher the spatial frequencies of an evanescent wave are, the faster the fields decay along the z-axis will be. Therefore, practical limitations make the bandwidth finite.

2.1.3 General Homogeneous Solution

A plane or evanescent wave characterized by the wave vector $\mathbf{k} = [k_x,\, k_y,\, k_z]$ and the angular frequency ω is only one of the many homogeneous solutions of the wave equation. To find the general solution we need to sum over all possible plane and evanescent waves, that is, we have to sum waves with all possible $\mathbf{k}$ and ω

$$\mathbf{E}(\mathbf{r}, t) \;=\; \mathrm{Re}\left\{ \sum_{n,m} \mathbf{E}_0(\mathbf{k}_n, \omega_m)\, \mathrm{e}^{\pm i \mathbf{k}_n \cdot \mathbf{r} - i\omega_m t} \right\} \quad \text{with} \quad \mathbf{k}_n \cdot \mathbf{k}_n = \omega_m^2 / c^2 \tag{2.19}$$

The condition on the right corresponds to the dispersion relation (2.14). Furthermore, the divergence condition requires that $\mathbf{E}_0(\mathbf{k}_n, \omega_m) \cdot \mathbf{k}_n = 0$. We have added the argument $(\mathbf{k}_n, \omega_m)$ to $\mathbf{E}_0$ since each plane or evanescent wave in the sum is characterized by a different complex amplitude.

The solution (2.19) assumes that there is a discrete set of frequencies and wavevectors. Such discrete sets can be generated by boundary conditions, for example, in a cavity where the fields on the cavity surface have to vanish. In free space, the sum in Eq. (2.19) becomes continuous and we obtain

$$\mathbf{E}(\mathbf{r},t) \;=\; \mathrm{Re}\left\{\int_{\mathbf{k}}\int_{\omega} \mathbf{E}_0(\mathbf{k},\omega)\,\mathrm{e}^{\pm i\mathbf{k}\cdot\mathbf{r}-i\omega t}\,d\omega\,d^3\mathbf{k}\right\} \quad \text{with} \quad \mathbf{k}\cdot\mathbf{k}=\omega^2/c^2 \tag{2.20}$$

which has the appearance of a four-dimensional Fourier transform. Notice that $\mathbf{E}_0$ is now a complex amplitude *density*, that is, amplitude per unit frequency and unit wave vector. The difference between (2.19) and (2.20) is the same as between Fourier series and Fourier transforms.

2.2 Spectral Representation

Let us consider solutions that are represented by a continuous distribution of frequencies ω, that is, solutions of finite bandwidth. For this purpose we go back to the complex notation (2.11) for monochromatic fields and sum (integrate) over all monochromatic solutions

$$\mathbf{E}(\mathbf{r},t) \;=\; \mathrm{Re}\left\{\int_{-\infty}^{\infty} \hat{\mathbf{E}}(\mathbf{r},\omega)\,\mathrm{e}^{-i\omega t}\,d\omega\right\}\,. \tag{2.21}$$

We have replaced the complex amplitude $\mathbf{E}$ by $\hat{\mathbf{E}}$ since we're now dealing with an amplitude per unit frequency, *i.e.* $\hat{\mathbf{E}} = \lim_{\Delta\omega\to 0}[\mathbf{E}/\Delta\omega]$. We have also included ω in the argument of $\hat{\mathbf{E}}$ since each solution of constant ω has its own amplitude.

In order to eliminate the 'Re' sign in (2.21) we require that

$$\hat{\mathbf{E}}(\mathbf{r},-\omega) \;=\; \hat{\mathbf{E}}^{*}(\mathbf{r},\omega)\;, \tag{2.22}$$

which leads us to

$$\mathbf{E}(\mathbf{r},t) \;=\; \int_{-\infty}^{\infty} \hat{\mathbf{E}}(\mathbf{r},\omega)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega \tag{2.23}$$

This is simply the Fourier transform of $\hat{\mathbf{E}}$. In other words, $\mathbf{E}(\mathbf{r},t)$ and $\hat{\mathbf{E}}(\mathbf{r},\omega)$ form a time-frequency Fourier transform pair. $\hat{\mathbf{E}}(\mathbf{r},\omega)$ is also denoted as the *temporal*

spectrum of $\mathbf{E}(\mathbf{r},t)$. Note that $\hat{\mathbf{E}}$ is generally complex, while $\mathbf{E}$ is always real. The inverse transform reads as

$$\hat{\mathbf{E}}(\mathbf{r},\omega) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{E}(\mathbf{r},t)\,\mathrm{e}^{i\omega t}\,\mathrm{d}t\;. \tag{2.24}$$

Let us now replace each of the fields in Maxwell's equations (1.32)–(1.35) by their Fourier transforms. We then obtain

$$\nabla\cdot\hat{\mathbf{D}}(\mathbf{r},\omega) \;=\; \hat{\rho}(\mathbf{r},\omega) \tag{2.25}$$

$$\nabla\times\hat{\mathbf{E}}(\mathbf{r},\omega) \;=\; i\omega\hat{\mathbf{B}}(\mathbf{r},\omega) \tag{2.26}$$

$$\nabla\times\hat{\mathbf{H}}(\mathbf{r},\omega) \;=\; -i\omega\hat{\mathbf{D}}(\mathbf{r},\omega)\;+\;\hat{\mathbf{j}}(\mathbf{r},\omega) \tag{2.27}$$

$$\nabla\cdot\hat{\mathbf{B}}(\mathbf{r},\omega) \;=\; 0 \tag{2.28}$$

These equations are the spectral representation of Maxwell's equations. Once a solution for $\hat{\mathbf{E}}$ is found, we obtain the respective time-dependent field by the inverse transform (2.23).

2.2.1 Monochromatic Fields

A monochromatic field oscillates at a single frequency ω. According to Eq. (2.11) it can be represented by a complex amplitude $\mathbf{E}(\mathbf{r})$ as

$$\begin{aligned}\mathbf{E}(\mathbf{r},t) \;&=\; \mathrm{Re}\{\mathbf{E}(\mathbf{r})\,\mathrm{e}^{-i\omega t}\}\\ &=\; \mathrm{Re}\{\mathbf{E}(\mathbf{r})\}\cos\omega t+\mathrm{Im}\{\mathbf{E}(\mathbf{r})\}\sin\omega t\\ &=\; (1/2)\left[\mathbf{E}(\mathbf{r})\,\mathrm{e}^{-i\omega t}\;+\;\mathbf{E}^{*}(\mathbf{r})\,\mathrm{e}^{i\omega t}\right]\;.\end{aligned} \tag{2.29}$$

Inserting the last expression into Eq. (2.24) yields the temporal spectrum of a monochromatic wave

$$\hat{\mathbf{E}}(\mathbf{r},\omega') \;=\; \frac{1}{2}\left[\mathbf{E}(\mathbf{r})\,\delta(\omega'-\omega)\;+\;\mathbf{E}^{*}(\mathbf{r})\,\delta(\omega'+\omega)\right]\;. \tag{2.30}$$

Here $\delta(x)\;=\;\int\exp[ixt]\,dt/(2\pi)$ is the Dirac delta function. Iif we use Eq. (2.30) along with similar expressions for the spectra of $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{H}$, ρ_0, and $\mathbf{j}_0$ in Maxwell's

equations (2.25)–(2.28) we obtain

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r}) \tag{2.31}$$

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega \mathbf{B}(\mathbf{r}) \tag{2.32}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega \mathbf{D}(\mathbf{r}) + \mathbf{j}(\mathbf{r}) \tag{2.33}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0 \tag{2.34}$$

These equations are used whenever one deals with time-harmonic fields. They are formally identical to the spectral representation of Maxwell's equations (2.25)–(2.28). Once a solution for the complex fields is found, the real time-dependent fields are found through Eq. (2.11).

2.3 Interference of Waves

Detectors do not respond to fields, but to the intensity of fields, which is defined (in free space) as

$$I(\mathbf{r}) = \sqrt{\frac{\varepsilon_0}{\mu_0}} \left\langle \mathbf{E}(\mathbf{r},t) \cdot \mathbf{E}(\mathbf{r},t) \right\rangle , \tag{2.35}$$

with $\langle .. \rangle$ denoting the time-average. For a monochromatic wave, as defined in Eq. (2.29), this expression becomes

$$I(\mathbf{r}) = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}} \, |\mathbf{E}(\mathbf{r})|^2 . \tag{2.36}$$

with $|\mathbf{E}|^2 = \mathbf{E} \cdot \mathbf{E}^*$. The energy and intensity of electromagnetic waves will be discussed later in Chapter 5. Using Eq. (2.15), the intensity of a plane wave turns out to be $|\mathbf{E}_0|^2$ since k_x, k_y, and k_z are all real. For an evanescent wave, however, we obtain

$$I(\mathbf{r}) = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}} |\mathbf{E}_0|^2 \, \mathrm{e}^{-2\,k_z\,z} , \tag{2.37}$$

that is, the intensity decays exponentially in z-direction. The $1/e$ decay length is $L_z = 1/(2k_z)$ and characterizes the confinement of the evanescent wave.

Next, we take a look at the intensity of a pair of fields

$$\begin{aligned} I(\mathbf{r}) &= \sqrt{\varepsilon_0/\mu_0}\,\big\langle[\mathbf{E}_1(\mathbf{r},t)+\mathbf{E}_2(\mathbf{r},t)]\cdot[\mathbf{E}_1(\mathbf{r},t)+\mathbf{E}_2(\mathbf{r},t)]\big\rangle] \\ &= \sqrt{\varepsilon_0/\mu_0}\,\big[\big\langle\mathbf{E}_1(\mathbf{r},t)\cdot\mathbf{E}_1(\mathbf{r},t)\big\rangle+\big\langle\mathbf{E}_2(\mathbf{r},t)\cdot\mathbf{E}_2(\mathbf{r},t)\big\rangle+2\big\langle\mathbf{E}_1(\mathbf{r},t)\cdot\mathbf{E}_2(\mathbf{r},t)\big\rangle\big] \\ &= I_1(\mathbf{r})\,+\,I_2(\mathbf{r})\,+\,2\,I_{12}(\mathbf{r}) \end{aligned} \tag{2.38}$$

Thus, the intensity of two fields is not simply the sum of their intensities! Instead, there is a third term, a so-called *interference* term. But what about energy conservation? How can the combined power be larger than the sum of the individual power contributions? It turns out that I_{12} can be positive or negative. Furthermore, I_{12} has a directional dependence, that is, there are directions for which I_{12} is positive and other directions for which it is negative. Integrated over all directions, I_{12} cancels and energy conservation is restored.

Coherent Fields

Coherence is a term that refers to how similar two fields are, both in time and space. Coherence theory is a field on its own and we won't dig too deep here. Maximum coherence between two fields is obtained if the fields are monochromatic, their frequencies are the same, and if the two fields have a well-defined phase relationship. Let's have a look at the sum of two plane waves of the *same* frequency ω

$$\mathbf{E}(\mathbf{r},t)\;=\;\mathrm{Re}\{\left[\mathbf{E}_1\,\mathrm{e}^{i\mathbf{k}_1\cdot\mathbf{r}}+\mathbf{E}_2\,\mathrm{e}^{i\mathbf{k}_2\cdot\mathbf{r}}\right]\,\mathrm{e}^{-i\omega t}\}\;, \tag{2.39}$$

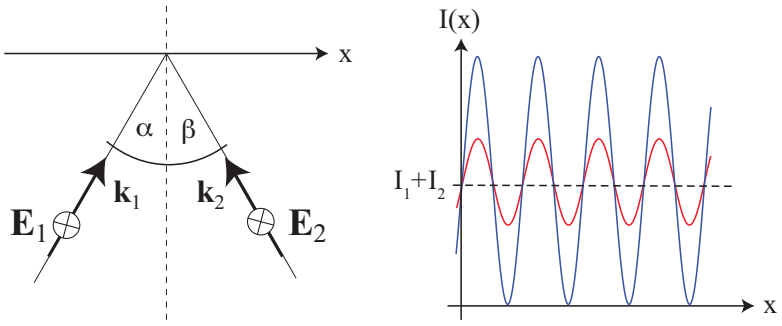

Figure 2.4: Interference of two plane waves incident at angles α and β. Left: Interference pattern along the x-axis for two different visibilities, 0.3 and 1. The average intensity is $I_1 + I_2$.

and let's denote the plane defined by the vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ as the (x,y) plane. Then, $\mathbf{k}_1 = (k_{x_1}, k_{y_1}, k_{z_1}) = k(\sin\alpha, \cos\alpha, 0)$ and $\mathbf{k}_2 = (k_{x_2}, k_{y_2}, k_{z_2}) = k(-\sin\beta, \cos\beta, 0)$ (see Figure 2.4). We now evaluate this field along the x-axis and obtain

$$\mathbf{E}(x,t) \;=\; \mathrm{Re}\{\left[\mathbf{E}_1\,\mathrm{e}^{ikx\sin\alpha} + \mathbf{E}_2\,\mathrm{e}^{-ikx\sin\beta}\right]\,\mathrm{e}^{-i\omega t}\}\,, \tag{2.40}$$

which corresponds to the intensity

$$\begin{aligned} I(x) \;&=\; \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\,\left|\left[\mathbf{E}_1\,\mathrm{e}^{ikx\sin\alpha} + \mathbf{E}_2\,\mathrm{e}^{-ikx\sin\beta}\right]\right|^2 \\ &=\; I_1 + I_2 + \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left[\mathbf{E}_1\cdot\mathbf{E}_2^*\,\mathrm{e}^{ikx(\sin\alpha+\sin\beta)} \;+\; \mathbf{E}_1^*\cdot\mathbf{E}_2\,\mathrm{e}^{-ikx(\sin\alpha+\sin\beta)}\right] \\ &=\; I_1 + I_2 + \sqrt{\varepsilon_0/\mu_0}\,\mathrm{Re}\left\{\mathbf{E}_1\cdot\mathbf{E}_2^*\,\mathrm{e}^{ikx(\sin\alpha+\sin\beta)}\right\}\,. \end{aligned} \tag{2.41}$$

This equation is valid for any complex vectors $\mathbf{E}_1$ and $\mathbf{E}_2$. We next assume that the two vectors are real and that they are polarized along the z-axis. We then obtain

$$I(x) \;=\; I_1 + I_2 + 2\,\sqrt{I_1\,I_2}\,\cos[kx(\sin\alpha + \sin\beta)]\,. \tag{2.42}$$

The cosine term oscillates between +1 and -1. Therefore, the largest and smallest signals are $I_{min} = (I_1 + I_2) \pm 2\sqrt{I_1\,I_2}$. To quantify the strength of interference one defines the *visibility*

$$\eta \;=\; \frac{I_{max} - I_{min}}{I_{max} + I_{min}}\,, \tag{2.43}$$

which has a maximum value of $\eta = 1$ for $I_1 = I_2$. The period of the interference fringes $\Delta x = \lambda/(\sin\alpha + \sin\beta)$ decreases with α and β and is shortest for $\alpha = \beta = \pi/2$, that is, when the two waves propagate head-on against each other. In this case, $\Delta x = \lambda/2$.

Incoherent Fields

Let us now consider a situation for which no interference occurs, namely for two plane waves with different frequencies ω_1 and ω_2. In this case Eq. (2.39) has to be replaced by

$$\mathbf{E}(x,t) \;=\; \mathrm{Re}\{\left[\mathbf{E}_1\,\mathrm{e}^{ik_1x\sin\alpha - i\omega_1 t} + \mathbf{E}_2\,\mathrm{e}^{-ik_2x\sin\beta - i\omega_2 t}\right]\}\,. \tag{2.44}$$

Evaluating the intensity yields

$$I(x) = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left\langle\left|\mathbf{E}_1\,\mathrm{e}^{ik_1x\sin\alpha-i\omega_1t}+\mathbf{E}_2\,\mathrm{e}^{-ik_2x\sin\beta-i\omega_2t}\right|^2\right\rangle \tag{2.45}$$

$$= I_1+I_2+2\sqrt{I_1\,I_2}\;\mathrm{Re}\left\{\mathrm{e}^{i\,[k_1\sin\alpha+k_2\sin\beta]\,x}\;\left\langle\mathrm{e}^{i\,(\omega_2-\omega_1)\,t}\right\rangle\right\}\;.$$

The expression $\langle\exp[i(\omega_2-\omega_1)t]\rangle$ is the time-average of a harmonically oscillating function and yields a result of 0. Therefore, the interference term vanishes and the total intensity becomes

$$I(x) = I_1+I_2\;, \tag{2.46}$$

that is, there is no interference.

It has to be emphasized, that the two situations that we analyzed are extreme cases. In practice there is no absolute coherence or incoherence. Any electromagnetic field has a finite line width $\Delta\omega$ that is spread around the center frequency ω. In the case of lasers, $\Delta\omega$ is a few MHz, determined by the spontaneous decay rate of the atoms in the active medium. Thus, an electromagnetic field is at best only partially coherent and its description as a monochromatic field with a single frequency ω is an approximation.

2.4 Fields as Complex Analytic Signals

The relationship (2.22) indicates that the positive frequency region contains all the information of the negative frequency region. If we restrict the integration in Eq. (2.23) to positive frequencies, we obtain what is called a *complex analytic signal*

$$\mathbf{E}^+(\mathbf{r},t) = \int_0^\infty \hat{\mathbf{E}}(\mathbf{r},\omega)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega\;, \tag{2.47}$$

with the superscript '+' denoting that only positive frequencies are included. Similarly, we can define a complex analytic signal $\mathbf{E}^-$ that accounts only for negative frequencies. The truncation of the integration range causes $\mathbf{E}^+$ and $\mathbf{E}^-$ to become complex functions of time. Because $\mathbf{E}$ is real, we have $[\mathbf{E}^+]^* = \mathbf{E}^-$. By taking the Fourier transform of $\mathbf{E}^+(\mathbf{r},t)$ and $\mathbf{E}^-(\mathbf{r},t)$ we obtain $\hat{\mathbf{E}}^+(\mathbf{r},\omega)$ and $\hat{\mathbf{E}}^-(\mathbf{r},\omega)$, respectively. It turns out that $\hat{\mathbf{E}}^+$ is identical to $\hat{\mathbf{E}}$ for $\omega > 0$ and it is zero for negative

frequencies. Similarly, $\hat{\mathbf{E}}^-$ is identical to $\hat{\mathbf{E}}$ for $\omega < 0$ and it is zero for positive frequencies. Consequently, $\hat{\mathbf{E}} = \hat{\mathbf{E}}^+ + \hat{\mathbf{E}}^-$. In quantum mechanics, $\hat{\mathbf{E}}^-$ is associated with the creation operator $\hat{a}^\dagger$ and $\hat{\mathbf{E}}^+$ with the annihilation operator $\hat{a}$.

Chapter 3

Constitutive Relations

Maxwell's equations define the fields that are generated by currents and charges. However, they do not describe how these currents and charges are generated. Thus, to find a self-consistent solution for the electromagnetic field, Maxwell's equations must be supplemented by relations that describe the behavior of matter under the influence of fields. These material equations are known as constitutive relations.

The constitutive relations express the secondary sources $\mathbf{P}$ and $\mathbf{M}$ in terms of the fields $\mathbf{E}$ and $\mathbf{H}$, that is $\mathbf{P} = f[\mathbf{E}]$ and $\mathbf{M} = f[\mathbf{H}]$.[1] According to Eq. (1.20) this is equivalent to $\mathbf{D} = f[\mathbf{E}]$ and $\mathbf{B} = f[\mathbf{E}]$. If we expand these relations into power series

$$D = D_0 + \left.\frac{\partial D}{\partial E}\right|_{D=0} E + \frac{1}{2}\left.\frac{\partial^2 D}{\partial E^2}\right|_{D=0} E^2 + \ldots . \tag{3.1}$$

we find that the lowest-order term depends *linearly* on $\mathbf{E}$. In most practical situations the interaction of radiation with matter is weak and it suffices to truncate the power series after the linear term. The nonlinear terms come into play when the fields acting on matter become comparable to the atomic Coulomb potential. This is the territory of strong field physics. Here we will entirely focus on the linear properties of matter.

[1] In some exotic cases we can have $\mathbf{P} = f[\mathbf{E}, \mathbf{H}]$ and $\mathbf{M} = f[\mathbf{H}, \mathbf{E}]$, which are so-called bi-isotropic or bi-anisotropic materials. These are special cases and won't be discussed here.

3.1 Linear Materials

The most general linear relationship between $\mathbf{D}$ and $\mathbf{E}$ can be written as

$$\mathbf{D}(\mathbf{r},t) \;=\; \varepsilon_0 \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{0} \tilde{\varepsilon}(\mathbf{r}-\mathbf{r}',\, t-t')\; \mathbf{E}(\mathbf{r}',t')\, \mathrm{d}^3\mathbf{r}'\, \mathrm{d}t' \;, \tag{3.2}$$

which states that the response $\mathbf{D}$ at the location $\mathbf{r}$ and at time t not only depends on the excitation $\mathbf{E}$ at $\mathbf{r}$ and t, but also on the excitation $\mathbf{E}$ in all other locations $\mathbf{r}'$ and all previous times t'. The integrals represent summations over all space and over all previous times. For reasons of causality (no response before excitation), the time integral only runs to $t' = 0$. The response function $\tilde{\varepsilon}$ is a tensor of rank two. It maps a vector $\mathbf{E}$ onto a vector $\mathbf{D}$ according to $D_i = \sum_j \tilde{\varepsilon}_{ij} E_j$, where $\{i,j\} \in \{x,y,z\}$. A material is called *temporally dispersive* if its response function at time t depends on previous times. Similarly, a material is called *spatially dispersive* if its response at $\mathbf{r}$ depends also on other locations. A spatially dispersive medium is also designated as a *nonlocal medium*.

Note that Eq. (3.2) is a convolution in space and time. Using the Fourier transform with respect to both time and space, that is,

$$\hat{\mathbf{D}}(k_x,k_y,k_z,\omega) \;=\; \frac{1}{(2\pi)^4} \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \mathbf{D}(x,y,z,t)\, \mathrm{e}^{ik_x x}\, \mathrm{e}^{ik_y y}\, \mathrm{e}^{ik_z z}\, \mathrm{e}^{i\omega t}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}t \;, \tag{3.3}$$

allows us to rewrite Eq. (3.2) as

$$\hat{\mathbf{D}}(\mathbf{k},\omega) \;=\; \varepsilon_0\, \varepsilon(\mathbf{k},\omega)\; \hat{\mathbf{E}}(\mathbf{k},\omega) \;, \tag{3.4}$$

where ε is the Fourier transform of $\tilde{\varepsilon}$. Note that the response at $(\mathbf{k},\omega)$ now only depends on the excitation at $(\mathbf{k},\omega)$ and not on neighboring $(\mathbf{k}',\omega')$. Thus, a nonlocal relationship in space and time becomes a *local relationship* in Fourier space! This is the reason why life often is simpler in Fourier space.

Spatial dispersion, i.e. a nonlocal response, is encountered near material surfaces or in objects whose size is comparable with the mean-free path of electrons. In general, nonlocal effects are very difficult to account for. In most cases of interest the effect is very weak and we can safely ignore it. Temporal dispersion, on

the other hand, is a widely encountered phenomenon and it is important to take it accurately into account. Thus, we will be mostly concerned with relationships of the sort

$$\hat{\mathbf{D}}(\mathbf{r},\omega) \;=\; \varepsilon_0\,\varepsilon(\omega)\;\hat{\mathbf{E}}(\mathbf{r},\omega)\;, \tag{3.5}$$

where $\varepsilon(\omega)$ is called the *dielectric function*, also called the relative electric permittivity. Similarly, for the magnetic field we obtain

$$\hat{\mathbf{B}}(\mathbf{r},\omega) \;=\; \mu_0\,\mu(\omega)\;\hat{\mathbf{H}}(\mathbf{r},\omega)\;, \tag{3.6}$$

with $\mu(\omega)$ being the relative magnetic permeability. Notice that the spectral representation of Maxwell's equations [(2.25)–(2.28)] is formally identical to the complex notation used for time-harmonic fields [(2.31)–(2.34)] . Therefore, Eqs. (3.5) and (3.6) also hold for the complex amplitudes of time-harmonic fields

$$\mathbf{D}(\mathbf{r}) \;=\; \varepsilon_0\varepsilon(\omega)\,\mathbf{E}(\mathbf{r})\,,\;\; \mathbf{B}(\mathbf{r}) \;=\; \mu_0\mu(\omega)\mathbf{H}(\mathbf{r}) \tag{3.7}$$

However, these equations generally do not hold for time-dependent fields $\mathbf{E}(\mathbf{r},t)$! One can use (3.7) for time-dependent fields only if dispersion can be ignored, that is $\varepsilon(\omega)=\varepsilon$ and $\mu(\omega)=\mu$. The only medium that is strictly dispersion-free is vacuum.

3.1.1 Electric and Magnetic Susceptibilities

The linear relationships (3.7) are often expressed in terms of the electric and magnetic susceptibilites $\chi_{\rm e}$ and $\chi_{\rm m}$, respectively. These are defined as

$$\mathbf{P}(\mathbf{r}) \;=\; \varepsilon_0\chi_e(\omega)\,\mathbf{E}(\mathbf{r})\,,\;\; \mathbf{M}(\mathbf{r}) \;=\; \chi_m(\omega)\mathbf{H}(\mathbf{r})\;. \tag{3.8}$$

Using the relations (1.20) we find that $\varepsilon=(1+\chi_e)$ and $\mu=(1+\chi_m)$.

3.1.2 Conductivity

The conductivity σ relates an induced conduction current $\mathbf{j}_{\rm bond}$ in a linear fashion to an exciting field $\mathbf{E}$. Similar to Eq. (3.7), this relationship can be represented as

$$\mathbf{j}_{\rm cond}(\mathbf{r}) \;=\; \sigma(\omega)\,\mathbf{E}(\mathbf{r})\;. \tag{3.9}$$

It turns out that the conduction current is accounted for by the imaginary part of $\varepsilon(\omega)$ as we shall show in the following.

Let us explicitly split the current density $\mathbf{j}$ into a source and a conduction current density according to Eq. (1.11). Maxwell's curl equation for the magnetic field (2.33) then becomes

$$\nabla \times \mathbf{H}(\mathbf{r}) \;=\; -i\omega \mathbf{D}(\mathbf{r}) \;+\; \mathbf{j}_{\mathrm{cond}}(\mathbf{r}) \;+\; \mathbf{j}_0(\mathbf{r}) \;. \tag{3.10}$$

We now introduce the linear relationships (3.7) and (3.9) and obtain

$$\begin{aligned} \nabla \times \mathbf{H}(\mathbf{r}) \;&=\; -i\omega\varepsilon_0\varepsilon(\omega)\,\mathbf{E}(\mathbf{r}) \;+\; \sigma(\omega)\,\mathbf{E}(\mathbf{r}) \;+\; \mathbf{j}_0(\mathbf{r}) \\ &=\; -i\omega\varepsilon_0 \left[\varepsilon(\omega) + i\frac{\sigma(\omega)}{\varepsilon_0\omega}\right] \mathbf{E}(\mathbf{r}) \;+\; \mathbf{j}_0(\mathbf{r}) \;. \end{aligned} \tag{3.11}$$

Thus, we see that the conductivity acts like the imaginary part of the electric permeability and that we can simply accommodate σ in ε by using a complex dielectric function

$$[\varepsilon' \;+\; i\sigma/(\omega\,\varepsilon_0)] \;\rightarrow\; \varepsilon \tag{3.12}$$

where ε' denotes the purely real polarization-induced dielectric constant. In the complex notation one does not distinguish between conduction currents and polarization currents. Energy dissipation is associated with the imaginary part of the dielectric function (ε'') whereas energy storage is associated with its real part (ε').

With the new definition of ε, the wave equations for the complex fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ in linear media are

$$\nabla \times \mu(\omega)^{-1} \nabla \times \mathbf{E}(\mathbf{r}) \;-\; k_0^2\,\varepsilon(\omega)\,\mathbf{E}(\mathbf{r}) \;=\; \mathrm{i}\omega\mu_0\,\mathbf{j}_0(\mathbf{r}) \;, \tag{3.13}$$

$$\nabla \times \varepsilon(\omega)^{-1} \nabla \times \mathbf{H}(\mathbf{r}) \;-\; k_0^2\,\mu(\omega)\,\mathbf{H}(\mathbf{r}) \;=\; \nabla \times\; \varepsilon(\omega)^{-1}\,\mathbf{j}_0(\mathbf{r}) \;, \tag{3.14}$$

where $k_0 = \omega/c$ denotes the vacuum wavenumber. Note that these equations are also valid for anisotropic media, *i.e.* if ε and μ are tensors.

If μ is isotropic then we can multiply Eq. (3.13) on both sides with μ. Furthermore, if there are no sources we can drop $\mathbf{j}_0$ and obtain the Helmholtz equation (2.12), but with the difference that now $k^2 = k_0^2\,\varepsilon\mu$, that is,

$$\nabla^2\mathbf{E}(\mathbf{r}) + k^2\,\mathbf{E}(\mathbf{r}) \;=\; \nabla^2\mathbf{E}(\mathbf{r}) + k_0^2\,n^2\mathbf{E}(\mathbf{r}) \;=\; 0 \tag{3.15}$$

where $n = \sqrt{\varepsilon\mu}$ is called *index of refraction*.

Conductors

The conductivity σ is a measure for how good a conductor is. For example, quartz has a conductivity of $\sigma_{\mathrm{SiO_2}} = 10^{-16}$ A / V m, and the conductivity of copper is $\sigma_{\mathrm{Au}} = 10^{8}$ A / V m. These values are different by 24 orders of magnitude! There are hardly any other physical parameters with a comparable dynamic range.

The net charge density ρ inside a conductor is zero, no matter wether it transports a current or not. This seems surprising, but it directly follows from the charge conservation (1.36) and Gauss' law (1.32). Combining the two equations and using $\mathbf{j} = \sigma\mathbf{E}$ and $\mathbf{D} = \varepsilon_0\varepsilon\mathbf{E}$ yields

$$\frac{\partial}{\partial t}\,\rho(t) \;=\; -\frac{\sigma}{\varepsilon_0\varepsilon}\,\rho(t)\;, \tag{3.16}$$

which has the solution

$$\rho(t) \;=\; \rho(t=0)\,\mathrm{e}^{-t\,\sigma/(\varepsilon_0\varepsilon)}\;. \tag{3.17}$$

Thus, any charge inside the conductor dissipates within a time of $T_\rho = \varepsilon_0\varepsilon/\sigma$. For a perfect conductor, $\sigma \to \infty$ and hence $\rho(t) = 0$. For realistic conductors with finite σ the characteristic time is $T_\rho \sim 10^{-19}$ s, which is so short that it can be neglected.

When a charge moves through a conductor it undergoes collisions with the lattice. After a collision event, the charge is accelerated by the external field until it is slowed down by the next collision. For good conductors (copper), the time between collisions is typically in the order of $\tau \sim 10^{-14}$ s. The sequence of acceleration and deceleration events results in a finite velocity $\mathbf{v}_d$ for the charge, called the *drift velocity*. The current density due to a charge density moving at finite speed is $\mathbf{j} = q\,\mathbf{v}_d\,n$, where n[2] is the charge density, *i.e.* the number of charges per unit volume. The drift velocity $\mathbf{v}_d$ is proportional to the driving field $\mathbf{E}$ and the proportionality constant is called *mobility* μ[3]. Thus, $\sigma = nq\mu$.

[2]Not to be confused with the index of refraction n.

[3]Not to be confused with the magnetic permittivity μ.

In a good conductor the polarization current $\partial\mathbf{P}/\partial t$ can be neglected because it is much smaller than the conduction current $\mathbf{j}$. In terms of a complex dielectric constant $\varepsilon = \varepsilon' + i\varepsilon''$ (c.f. Eq. 3.12) this implies that $\varepsilon'' \gg |\varepsilon'|$, that is, $\sigma \gg |\omega\,\varepsilon_0\varepsilon'|$. Evidently, the higher the frequency ω is, the more challenging it gets to fulfill this condition. In fact, at optical frequencies, metals are no longer good conductors and they are dominated by the polarization current. At lower frequencies, however, it is legitimate to ignore the polarization current when dealing with good conductors. Ignoring $\partial\mathbf{P}/\partial t$ is equivalent to ignoring the real part of the complex dielectric function (3.12), which implies $k^2 = (\omega/c)^2\mu\varepsilon \approx i\,(\omega/c)^2(\mu\sigma\,/\,\omega\,\varepsilon_0)$. Consequently, the Helmholtz equation (3.15) reads as

$$\nabla^2\mathbf{E}(\mathbf{r}) + i\omega\sigma\mu_0\mu\;\mathbf{E}(\mathbf{r}) \;=\; 0\;, \tag{3.18}$$

and because of $\mathbf{j} = \sigma\mathbf{E}$ the same equation holds for the current density $\mathbf{j}$. Note that we're using complex equation and that $\mathbf{j}(\mathbf{r},t) = \mathrm{Re}\{\mathbf{j}(\mathbf{r})\;\exp[-i\omega\,t]\}$.

Let us now consider a semi-infinite conductor, as illustrated in Fig. 3.1. This situation corresponds to a small section of a wire's surface. The conductor has a surface at $x = 0$ and transports a current $\mathbf{j}(\mathbf{r})$ in the z direction. Because of the

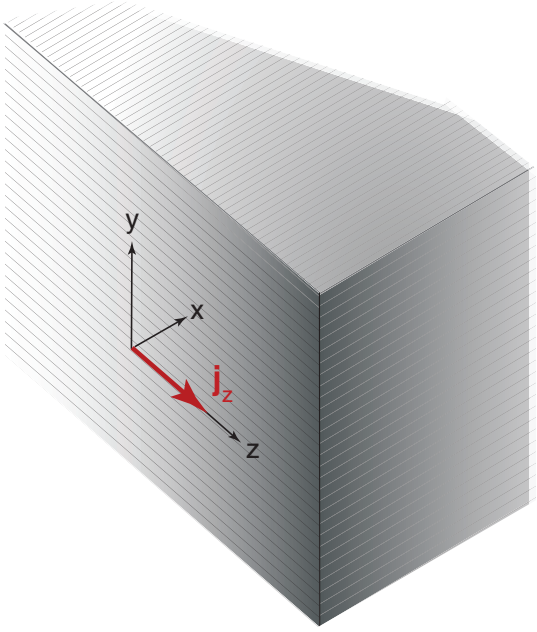

Figure 3.1: Current density $\mathbf{j} = j_z\,\mathbf{n}_z$ flowing along the surface of a conductor.

invariance in y and z we set $j_z(x) = A\exp[B\,x]$ and insert into Eq. (3.18). Noting that $\sqrt{i} = (1+i)/\sqrt{2}$ we obtain

$$j_z(x) \;=\; j_z(x=0)\,\mathrm{e}^{-(i-1)\,x/D_s} \qquad \text{with} \quad D_s = \sqrt{\frac{2}{\sigma\,\mu_0\,\mu\,\omega}}\,. \tag{3.19}$$

The length D_s is called the *skin depth*. Since $|j_z(x)/j_z(0)| = \exp[-x/D_s]$ it describes the penetration of fields and currents into the metal. Evidently, for a perfect conductor ($\sigma \to \infty$) the skin depth becomes $D_s = 0$, that is, all the current is transported on the surface of the metal. D_s also decreases with increasing frequency ω, but eventually the result (3.19) becomes inaccurate because the polarization current $\partial\mathbf{P}/\partial t$ becomes stronger than the conduction current. In wires of finite diameter the skin depth also depends on the curvature of the wire. At low frequencies, the conductance of a wire scales with the cross-section of the wire, but at high frequencies, the current is confined to the surface of the wire and hence the conductance scales with the circumference.

Chapter 4

Material Boundaries

In many practical situations, electromagnetic radiation enters from one medium into another. Examples are the refraction of light as it enters into water or the absorption of sunlight by a solar cell. The boundary between different media is not sharp. Instead, there is a transition zone defined by the length-scale of nonlocal effects, usually in the size range of a few atoms, that is $0.5 - 1$ nm. If we don't care how the fields behave on this scale we can safely describe the transition from one medium to another by a discontinuous material function. For example, at optical frequencies the dielectric function across a glass-air interface at $z = 0$ can be approximated as

$$\varepsilon(z) = \begin{cases} 2.3 & (\mathrm{z} < 0) \quad \mathrm{glass} \\ 1 & (\mathrm{z} > 0) \quad \mathrm{air} \end{cases}$$

4.1 Piecewise Homogeneous Media

Piecewise homogeneous materials consist of regions in which the material parameters are independent of position $\mathbf{r}$. In principle, a piecewise homogeneous medium is inhomogeneous and the solution can be derived from Eqs. (3.13) and (3.14). However, the inhomogeneities are entirely confined to the boundaries and it is convenient to formulate the solution for each region separately. These solutions must be connected with each other via the interfaces to form the solution for all space. Let the interface between two homogeneous regions D_i and D_j be denoted as ∂D_{ij}. If ε_i and μ_i designate the constant material parameters in region

D_i, the wave equations in that domain read as

$$(\nabla^2 + k_i^2)\,\mathbf{E}_i \;=\; -\mathrm{i}\omega\mu_0\mu_i\,\mathbf{j}_i + \frac{\nabla\rho_i}{\varepsilon_0\,\varepsilon_i}\;, \tag{4.1}$$

$$(\nabla^2 + k_i^2)\,\mathbf{H}_i \;=\; -\nabla\times\,\mathbf{j}_i\;, \tag{4.2}$$

where $k_i = (\omega/c)\sqrt{\mu_i\varepsilon_i} = k_0 n_i$ is the wavenumber and $\mathbf{j}_i$, ρ_i the sources in region D_i. To obtain these equations, the identity $\nabla\times\nabla\times = -\nabla^2 + \nabla\nabla\cdot$ was used and Maxwell's equation (1.32) was applied. Equations (4.1) and (4.2) are also denoted as the inhomogeneous vector Helmholtz equations. In most practical applications, such as scattering problems, there are no source currents or charges present and the Helmholtz equations are homogeneous, that is, the terms on the right hand side vanish.

4.2 Boundary Conditions

Eqs. (4.1) and (4.2) are only valid in the interior of the regions D_i. However, Maxwell's equations must also hold on the boundaries ∂D_{ij}. Due to the material discontinuity at the boundary between two regions it is difficult to apply the differential forms of Maxwell's equations. We will therefore use the corresponding integral forms (1.16) - (1.19). If we look at a sufficiently small section of the material boundary we can approximate the boundary as being flat and the fields as homogeneous on both sides (Fig. 4.1).

Let us first analyze Maxwell's first equation (1.16) by considering the infinitesimal rectangular pillbox illustrated in Fig. 4.1(a). Assuming that the fields are homogeneous on both sides, the surface integral of $\mathbf{D}$ becomes

$$\int_{\partial V} \mathbf{D}(\mathbf{r})\cdot\mathbf{n}\,da \;=\; A\,[\mathbf{n}\cdot\mathbf{D}_i(\mathbf{r})] - A\,[\mathbf{n}\cdot\mathbf{D}_j(\mathbf{r})]\;, \tag{4.3}$$

where A is the top surface of the pillbox. The contributions of the sides dA disappear if we shrink the thickness dL of the pillbox to zero. The right hand side of Eq. (1.16) becomes

$$\int_V \rho(\mathbf{r},t)\,dV \;=\; A\,dL\,\rho_i(\mathbf{r}) \;+\; A\,dL\,\rho_j(\mathbf{r})\;, \tag{4.4}$$

which finally yields

$$\mathbf{n} \cdot [\mathbf{D}_i(\mathbf{r}) \, - \, \mathbf{D}_j(\mathbf{r})] \; = \; \sigma(\mathbf{r}) \qquad (\mathbf{r} \in \partial D_{ij}) \; . \tag{4.5}$$

Here, $\sigma = dL\,\rho_i + dL\,\rho_j = dL\rho$ is the surface charge density. Eq. (4.5) is the boundary condition for the displacement $\mathbf{D}$. It states that if the normal component of the displacement changes across the boundary in point $\mathbf{r}$, then the change has to be compensated by a surface charge σ_0.

Let us now have a look at Faraday's law, *i.e.* Maxwell's second equation (1.17). We evaluate the path integral of the $\mathbf{E}$-field along the small rectangular path shown in Fig. 4.1(b)

$$\int_{\partial A} \mathbf{E}(\mathbf{r}) \cdot d\mathbf{s} \; = \; L\,[\mathbf{n} \times \mathbf{E}_i(\mathbf{r})] \cdot \mathbf{n}_s \, - \, L\,[\mathbf{n} \times \mathbf{E}_j(\mathbf{r})] \cdot \mathbf{n}_s \; . \tag{4.6}$$

The contributions along the normal sections vanish in the limit $dL \to 0$. We have introduced $\mathbf{n}_\mathrm{s}$ for the unit vector normal to the loop surface. It is perpendicular to $\mathbf{n}$, the unit vector normal to the boundary ∂D_{ij}. For the right hand side of Eq. (1.17) we find

$$-\frac{\partial}{\partial t} \int_A \mathbf{B}(\mathbf{r}) \cdot \mathbf{n}_\mathrm{s} \, da \; = \; 0 \; , \tag{4.7}$$

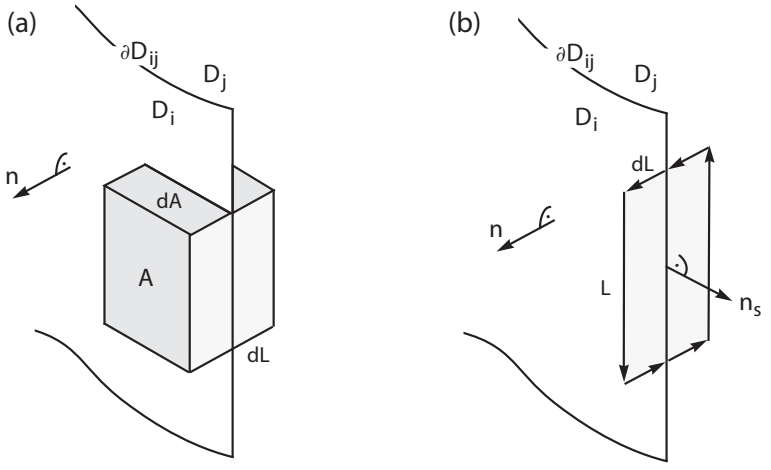

Figure 4.1: Integration paths for the derivation of the boundary conditions on the interface ∂D_{ij} between two neighboring regions D_i and D_j.

that is, the flux of the magnetic field vanishes as the area is reduced to zero. Combining the last two equations we arrive at

$$\mathbf{n} \times [\mathbf{E}_i(\mathbf{r}) - \mathbf{E}_j(\mathbf{r})] = 0 \qquad (\mathbf{r} \in \partial D_{ij}) \; . \tag{4.8}$$

We obtain a similar result for the Ampère-Maxwell law (1.18), with the exception that the flux of the current density does not vanish necessarily. If we allow for the existence of a current density that is confined to a surface layer of infinitesimal thickness then

$$\int_A \mathbf{j}(\mathbf{r}) \cdot \mathbf{n}_\mathrm{s} \, da = L \, dL \, [\mathbf{j}_i(\mathbf{r}) \cdot \mathbf{n}_s] + L \, dL \, [\mathbf{j}_j(\mathbf{r}) \cdot \mathbf{n}_s] \; . \tag{4.9}$$

Using an equation similar to (4.6) for $\mathbf{H}$ and inserting into Eq. (1.18) yields

$$\mathbf{n} \times [\mathbf{H}_i(\mathbf{r}) - \mathbf{H}_j(\mathbf{r})] = \mathbf{K}(\mathbf{r}) \qquad (\mathbf{r} \in \partial D_{ij}) \; , \tag{4.10}$$

where $\mathbf{K} = dL \, \mathbf{j}_i + dL \, \mathbf{j}_j = dL \, \mathbf{j}$ being the surface current density. The fourth Maxwell equation (1.19) yields a boundary condition similar to Eq. (4.5), but with no surface charge, that is $\mathbf{n} \cdot [\mathbf{B}_i - \mathbf{B}_j] = 0$.

Taken all equations together we obtain

$$\mathbf{n} \cdot [\mathbf{B}_i(\mathbf{r}) - \mathbf{B}_j(\mathbf{r})] = 0 \qquad (\mathbf{r} \in \partial D_{ij}) \tag{4.11}$$

$$\mathbf{n} \cdot [\mathbf{D}_i(\mathbf{r}) - \mathbf{D}_j(\mathbf{r})] = \sigma(\mathbf{r}) \qquad (\mathbf{r} \in \partial D_{ij}) \tag{4.12}$$

$$\mathbf{n} \times [\mathbf{E}_i(\mathbf{r}) - \mathbf{E}_j(\mathbf{r})] = 0 \qquad (\mathbf{r} \in \partial D_{ij}) \tag{4.13}$$

$$\mathbf{n} \times [\mathbf{H}_i(\mathbf{r}) - \mathbf{H}_j(\mathbf{r})] = \mathbf{K}(\mathbf{r}) \qquad (\mathbf{r} \in \partial D_{ij}) \tag{4.14}$$

The surface charge density σ and the surface current density $\mathbf{K}$ are confined to the very interface between the regions D_i and D_j. Such confinement is not found in practice and hence, σ and $\mathbf{K}$ are mostly of theoretical significance. The existence of σ and $\mathbf{K}$ demands materials that perfectly screen any charges, such as a metal with infinite conductivity. Such perfect conductors are often used in theoretical models, but they are approximations for real behavior. In most cases we can set $\sigma = 0$ and $\mathbf{K} = 0$. Any currents and charges confined to the boundary D_{ij} are adequately accounted for by the imaginary part of the dielectric functions ε_i and ε_j.

Note that Eq. (4.13) and (4.14) yield two equations each since the tangent of a boundary is defined by two vector components. Thus, the boundary conditions (4.11)–(4.14) constitute a total of six equations. However, these six equations

are *not* independent of each other since the fields on both sides of ∂D_{ij} are linked by Maxwell's equations. It can be easily shown, for example, that the conditions for the normal components are automatically satisfied if the boundary conditions for the tangential components hold everywhere on the boundary and Maxwell's equations are fulfilled in both regions.

If we express each field vector $\mathbf{Q} \in [\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}]$ in terms of a vector normal to the surface and vector parallel to the surface according to

$$\mathbf{Q} \;=\; \mathbf{Q}^{\parallel} + Q^{\perp}\mathbf{n}\;, \tag{4.15}$$

and we assume the we can ignore surface charges and currents ($\sigma = 0$ and $\mathbf{K} = 0$), we can represent Eqs. (4.11)-(4.14) as

$$B_i^{\perp} \;=\; B_j^{\perp} \quad \text{on} \;\; \partial D_{ij} \tag{4.16}$$

$$D_i^{\perp} \;=\; D_j^{\perp} \quad \text{on} \;\; \partial D_{ij} \tag{4.17}$$

$$\mathbf{E}_i^{\parallel} \;=\; \mathbf{E}_j^{\parallel} \quad \text{on} \;\; \partial D_{ij} \tag{4.18}$$

$$\mathbf{H}_i^{\parallel} \;=\; \mathbf{H}_j^{\parallel} \quad \text{on} \;\; \partial D_{ij} \tag{4.19}$$

4.3 Reflection and Refraction at Plane Interfaces

Applying the boundary conditions to a plane wave incident on a plane interface leads to so-called Fresnel reflection and transmission coefficients. These coefficients depend on the angle of incidence θ_i and the polarization of the incident wave, that is, the orientation of the electric field vector relative to the surface.

Let us consider a linearly polarized plane wave

$$\mathbf{E}_1(\mathbf{r}, t) \;=\; \mathrm{Re}\{\mathbf{E}_1\, \mathrm{e}^{i\mathbf{k}_1\cdot\mathbf{r} - i\omega t}\}\;, \tag{4.20}$$

as derived in Section 2.1.1. This plane wave is incident on a an interface at $z = 0$ (see Fig. 4.2) between two linear media characterized by $\varepsilon_1,\, \mu_1$ and $\varepsilon_2,\, \mu_2$, respectively. The magnitude of the wavevector $\mathbf{k}_1 = [k_{x_1},\, k_{y_1},\, k_{z_1}]$ is defined by

$$|\mathbf{k}_1| \;=\; k_1 \;=\; \frac{\omega}{c}\sqrt{\varepsilon_1\,\mu_1} \;=\; k_0 n_1\;, \tag{4.21}$$

The index '1' specifies the medium in which the plane wave is defined. At the interface $z = 0$ the incident wave gives rise to induced polarization and magnetization currents. These secondary sources give rise to new fields, that is, to transmitted and reflected fields. Because of the translational invariance along the interface we will look for reflected/transmitted fields that have the same form as the incident field

$$\mathbf{E}_{1r}(\mathbf{r},t) \;=\; \mathrm{Re}\{\mathbf{E}_{1r}\,\mathrm{e}^{i\mathbf{k}_{1r}\cdot\mathbf{r}-i\omega t}\} \qquad \mathbf{E}_{2}(\mathbf{r},t) \;=\; \mathrm{Re}\{\mathbf{E}_{2}\,\mathrm{e}^{i\mathbf{k}_{2}\cdot\mathbf{r}-i\omega t}\}\;, \tag{4.22}$$

where the first field is valid for $z < 0$ and the second one for $z > 0$. The two new wavevectors are defined as $\mathbf{k}_{1r} = [k_{x_{1r}},\, k_{y_{1r}},\, k_{z_{1r}}]$ and $\mathbf{k}_{1r} = [k_{x_2},\, k_{y_2},\, k_{z_2}]$, respectively. All fields have the same frequency ω because the media are assumed to be linear.

At $z = 0$, any of the boundary conditions (4.16)–(4.19) will lead to equations of the form

$$A\,\mathrm{e}^{i(\mathbf{k}_{x_1}x+\mathbf{k}_{y_1}y)} \;+\; B\,\mathrm{e}^{i(\mathbf{k}_{x_{1r}}x+\mathbf{k}_{y_{1r}}y)} \;=\; C\,\mathrm{e}^{i(\mathbf{k}_{x_2}x+\mathbf{k}_{y_2}y)}\;, \tag{4.23}$$

which has to be satisfied for all x and y along the boundary. This can only be fulfilled if the periodicities of the fields are equal on both sides of the interface, that is, $\mathbf{k}_{x_1} = \mathbf{k}_{x_{1r}} = \mathbf{k}_{x_2}$ and $\mathbf{k}_{y_1} = \mathbf{k}_{y_{1r}} = \mathbf{k}_{y_2}$. Thus, the transverse components of the wavevector $\mathbf{k}_{\parallel} = (k_x, k_y)$ are conserved. Furthermore, $\mathbf{E}_1$ and $\mathbf{E}_{1r}$ are defined in the same medium ($z < 0$) and hence $|\mathbf{k}_1|^2 = |\mathbf{k}_{1r}|^2 = k_1^2 = k_0^2\, n_1^2$. Taken these conditions together we find $k_{z_{1r}} = \pm k_{z_1}$. Here, we will have to take the minus sign

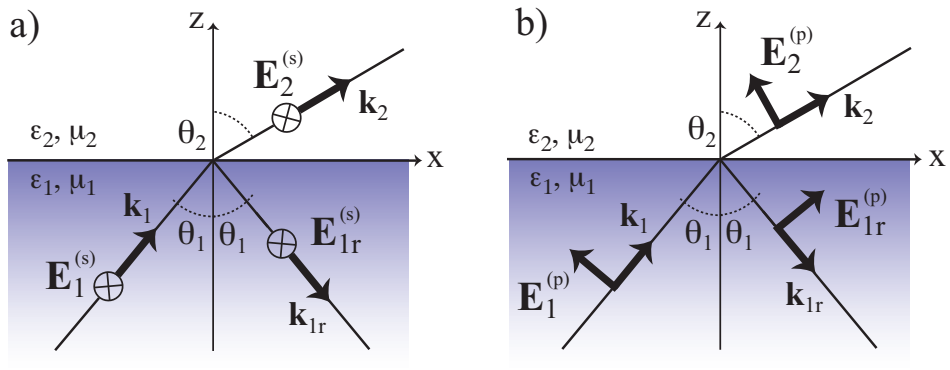

Figure 4.2: Reflection and refraction of a plane wave at a plane interface. (a) s-polarization, and (b) p-polarization.

since the reflected field is propagating away from the interface. To summarize, the three waves (incident, reflected, transmitted) can be represented as

$$\mathbf{E}_1(\mathbf{r},t) = \mathrm{Re}\{\mathbf{E}_1\, \mathrm{e}^{i(k_x x + k_y y + k_{z_1} z - \omega t)}\} \tag{4.24}$$

$$\mathbf{E}_{1r}(\mathbf{r},t) = \mathrm{Re}\{\mathbf{E}_{1r}\, \mathrm{e}^{i(k_x x + k_y y - k_{z_1} z - \omega t)}\} \tag{4.25}$$

$$\mathbf{E}_2(\mathbf{r},t) = \mathrm{Re}\{\mathbf{E}_2\, \mathrm{e}^{i(k_x x + k_y y + k_{z_2} z - \omega t)}\} \tag{4.26}$$

The magnitudes of the longitudinal wavenumbers are given by

$$k_{z_1} = \sqrt{k_1^2 - (k_x^2 + k_y^2)}, \qquad k_{z_2} = \sqrt{k_2^2 - (k_x^2 + k_y^2)}. \tag{4.27}$$

Since the transverse wavenumber $k_\parallel = [k_x^2 + k_y^2]^{1/2}$ is defined by the angle of incidence θ_1 as

$$k_\parallel = \sqrt{k_x^2 + k_y^2} = k_1 \sin\theta_1 \,, \tag{4.28}$$

it follows from (4.27) that also k_{z_1} and k_{z_2} can be expressed in terms of θ_1.

If we denote the angle of propagation of the refracted wave as θ_2 (see Fig. 4.2), then the requirement that $k_\parallel$ be continuous across the interface leads to $k_\parallel = k_1 \sin\theta_1 = k_2 \sin\theta_2$. In other words,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{4.29}$$

which is the celebrated *Snell's law*, discovered experimentally by Willebrord Snell around 1621. Furthermore, because $k_{z_{1r}} = -k_{z_1}$ we find that the angle of reflection is the same as the angle of incidence ($\theta_{1r} = \theta_1$).

4.3.1 s- and p-polarized Waves

The incident plane wave can be written as the superposition of two orthogonally polarized plane waves. It is convenient to choose these polarizations parallel and perpendicular to the plane of incidence as

$$\mathbf{E}_1 = \mathbf{E}_1^{(\mathrm{s})} + \mathbf{E}_1^{(\mathrm{p})} \,. \tag{4.30}$$

$\mathbf{E}_1^{(\mathrm{s})}$ is parallel to the interface, while $\mathbf{E}_1^{(\mathrm{p})}$ is perpendicular to the wavevector $\mathbf{k}_1$ and $\mathbf{E}_1^{(\mathrm{s})}$. The indices (s) and (p) stand for the German "senkrecht" (perpendicular) and

"parallel" (parallel), respectively, and refer to the plane of incidence. Upon reflection or transmission at the interface, the polarizations (s) and (p) are conserved, which is a consequence of the boundary condition (4.18).

s-Polarization

For s-polarization, the electric field vectors are parallel to the interface. Expressed in terms of the coordinate system shown in Fig. 4.2 they can be expressed as

$$\mathbf{E}_1 = \begin{bmatrix} 0 \\ \mathrm{E}_1^{(\mathrm{s})} \\ 0 \end{bmatrix} \qquad \mathbf{E}_{1r} = \begin{bmatrix} 0 \\ \mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})} \\ 0 \end{bmatrix} \qquad \mathbf{E}_2 = \begin{bmatrix} 0 \\ \mathrm{E}_2^{(\mathrm{s})} \\ 0 \end{bmatrix}, \tag{4.31}$$

where the superscript (s) stands for s-polarization. The magnetic field is defined through Faraday's law (2.32) as $\mathbf{H} = (\omega\mu_0\mu)^{-1}\,[\mathbf{k}\times\mathbf{E}]$, where we assumed linear and isotropic material properties ($\mathbf{B} = \mu_0\mu\mathbf{H}$). Using the electric field vectors (4.31) we find

$$\mathbf{H}_1 = \frac{1}{Z_1}\begin{bmatrix} -(k_{z_1}/k_1)\mathrm{E}_1^{(\mathrm{s})} \\ 0 \\ (k_x/k_1)\mathrm{E}_1^{(\mathrm{s})} \end{bmatrix} \qquad \mathbf{H}_{1r} = \frac{1}{Z_1}\begin{bmatrix} (k_{z_1}/k_1)\mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})} \\ 0 \\ (k_x/k_1)\mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})} \end{bmatrix}$$

$$\mathbf{H}_2 = \frac{1}{Z_2}\begin{bmatrix} -(k_{z_2}/k_2)\mathrm{E}_2^{(\mathrm{s})} \\ 0 \\ (k_x/k_2)\mathrm{E}_2^{(\mathrm{s})} \end{bmatrix}, \tag{4.32}$$

where we have introduced the wave impedance

$$Z_i = \sqrt{\frac{\mu_0\,\mu_i}{\varepsilon_0\,\varepsilon_i}} \tag{4.33}$$

In vacuum, $\mu_i = \varepsilon_i = 1$ and $Z_i \sim 377\,\Omega$. It is straightforward to show that the electric and magnetic fields are divergence free, *i.e.* $\nabla\cdot\mathbf{E} = 0$ and $\nabla\cdot\mathbf{H} = 0$. These conditions restrict the k-vector to directions perpendicular to the field vectors ($\mathbf{k}\cdot\mathbf{E} = \mathbf{k}\cdot\mathbf{H} = 0$).

Notice that we are dealing with an inhomogeneous problem, that is, the response of the system (reflection and refraction) depends on the excitation. Therefore, there are only two unknowns in the fields (4.31) and (4.32), namely $\mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})}$ and $\mathrm{E}_{2}^{(\mathrm{s})}$. $\mathrm{E}_{1}^{(\mathrm{s})}$ is the exciting field, which is assumed to be known. To solve for the two unknowns we require two boundary conditions. However, Eqs. (4.16)–(4.19) constitute a total of six boundary equations, four for the tangential field components and two for the normal components. However, only two of these six equations are independent.

The boundary conditions (4.18) and (4.19) yield

$$[\mathrm{E}_{1}^{(\mathrm{s})} + \mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})}] \;=\; \mathrm{E}_{2}^{(\mathrm{s})} \tag{4.34}$$

$$Z_1^{-1}[-(k_{z_1}/k_1)\mathrm{E}_{1}^{(\mathrm{s})} + (k_{z_1}/k_1)\mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})}] \;=\; Z_2^{-1}\left[-(k_{z_2}/k_2)\mathrm{E}_{2}^{(\mathrm{s})}\right] \;. \tag{4.35}$$

The two other tangential boundary conditions (continuity of E_x and H_y) are trivially fulfilled since $E_x = H_y = 0$. The boundary conditions for the normal components, (4.16) and (4.17), yield equations that are identical to (4.34) and (4.35) and can therefore be ignored.

Solving Eqs. (4.34) and (4.35) for $\mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})}$ yields

$$\frac{\mathrm{E}_{1\mathrm{r}}^{(\mathrm{s})}}{\mathrm{E}_{1}^{(\mathrm{s})}} \;=\; \frac{\mu_2 k_{z_1} - \mu_1 k_{z_2}}{\mu_2 k_{z_1} + \mu_1 k_{z_2}} \;\equiv\; r^{\mathrm{s}}(k_x, k_y) \;, \tag{4.36}$$

with r^{s} being the *Fresnel reflection coefficient for s-polarization*. Similarly, for $\mathrm{E}_{2}^{(\mathrm{s})}$ we obtain

$$\frac{\mathrm{E}_{2}^{(\mathrm{s})}}{\mathrm{E}_{1}^{(\mathrm{s})}} \;=\; \frac{2\mu_2 k_{z_1}}{\mu_2 k_{z_1} + \mu_1 k_{z_2}} \;\equiv\; t^{\mathrm{s}}(k_x, k_y) \;, \tag{4.37}$$

where t^{s} denotes the *Fresnel transmission coefficient for s-polarization*. Note that according to Eq. (4.27), k_{z_1} and k_{z_2} are functions of k_x and k_y. k_{z_1} can be expressed in terms of the angle of incidence as $k_{z_1} = k_1 \cos\theta_1$.

Fresnel Reflection and Transmission Coefficients

The procedure for p-polarized fields is analogous and won't be repeated here. Essentially, p-polarization corresponds to an exchange of the fields (4.31) and (4.32),

that is, the p-polarized electric field assumes the form of Eq. (4.32) and the p-polarized magnetic field the form of Eq. (4.31).The amplitudes of the reflected and transmitted waves can be represented as

$$\mathrm{E}_{1r}^{(\mathrm{p})} = \mathrm{E}_1^{(\mathrm{p})}\, r^{\mathrm{p}}(k_x, k_y) \qquad \mathrm{E}_2^{(\mathrm{p})} = \mathrm{E}_1^{(\mathrm{p})}\, t^{\mathrm{p}}(k_x, k_y)\,, \tag{4.38}$$

where r^{s} and t^{p} are the Fresnel reflection and transmission coefficients for p-polarization, respectively. In summary, for a plane wave incident on an interface between two linear and isotropic media we obtain the following reflection and transmission coefficients [1]

$$r^{\mathrm{s}}(k_x, k_y) = \frac{\mu_2 k_{z_1} - \mu_1 k_{z_2}}{\mu_2 k_{z_1} + \mu_1 k_{z_2}} \qquad r^{\mathrm{p}}(k_x, k_y) = \frac{\varepsilon_2 k_{z_1} - \varepsilon_1 k_{z_2}}{\varepsilon_2 k_{z_1} + \varepsilon_1 k_{z_2}} \tag{4.39}$$

$$t^{\mathrm{s}}(k_x, k_y) = \frac{2\mu_2 k_{z_1}}{\mu_2 k_{z_1} + \mu_1 k_{z_2}} \qquad t^{\mathrm{p}}(k_x, k_y) = \frac{2\varepsilon_2 k_{z_1}}{\varepsilon_2 k_{z_1} + \varepsilon_1 k_{z_2}} \sqrt{\frac{\mu_2 \varepsilon_1}{\mu_1 \varepsilon_2}} \tag{4.40}$$

The sign of the Fresnel coefficients depends on the definition of the electric field vectors shown in Fig. 4.2. For a plane wave at normal incidence ($\theta_1 = 0$), r^{s} and r^{p} differ by a factor of -1. Notice that the transmitted waves can be either plane waves or evanescent waves. This aspect will be discussed in Chapter 4.4.

The Fresnel reflection and transmission coefficients have many interesting predictions. For example, if a plane wave is incident on a glass-air interface (incident from the optically denser medium, *i. e.* glass) then the incident field can be totally reflected if it is incident at an angle θ_1 that is larger than a critical angle θ_c. In this case, one speaks of *total internal reflection*. Also, there are situations for which the entire field is transmitted. According to Eqs. (4.39), for p-polarization this occurs when $\varepsilon_2 k_{z_1} = \varepsilon_1 k_{z_2}$. Using $k_{z_1} = k_1 \cos\theta_1$ and Snell's law (4.29) we obtain $k_{z_2} = k_2 \left[1 - (n_1/n_2)^2 \sin^2\theta_1\right]^{1/2}$. At optical frequencies, most materials are not magnetic ($\mu_1 = \mu_2 = 1$), which yields $\tan\theta_1 = n_2/n_1$. For a glass-air interface this occurs for $\theta_1 \approx 57°$. For s-polarization one cannot find such a condition. Therefore, reflections from surfaces at oblique angles are mostly s-polarized. Polarizing sunglasses exploit this fact in order to reduce glare. The Fresnel coefficients give

[1] For symmetry reasons, some textbooks omit the square root term in the coefficient t^{p}. In this case, t^{p} refers to the ratio of transmitted and incident *magnetic field*.

rise to even more phenomena if we allow for two or more interfaces. Examples are Fabry-Pérot resonances and waveguiding along the interfaces.

4.4 Evanescent Fields

As already discussed in Section 2.1.2, evanescent fields can be described by plane waves of the form $\mathbf{E}\mathrm{e}^{i(\mathbf{kr}-\omega t)}$. They are characterized by the fact that at least one component of the wavevector **k** describing the direction of propagation is imaginary. In the spatial direction defined by the imaginary component of **k** the wave does not propagate but rather decays exponentially.

Evanescent waves never occur in a homogeneous medium but are inevitably connected to the interaction of light with inhomogeneities, such as a plane interface. Let us consider a plane wave impinging on such a flat interface between two media characterized by optical constants ε_1, μ_1 and ε_2, μ_2. As discussed in Section 4.3, the presence of the interface will lead to a reflected wave and a refracted wave whose amplitudes and directions are described by Fresnel coefficients and by Snell's law, respectively.

To derive the evanescent wave generated by total internal reflection at the surface of a dielectric medium, we refer to the configuration shown in Fig. 4.2. We choose the x-axis to be in the plane of incidence. Using the symbols defined in Section 4.3, the complex transmitted field vector can be expressed as

$$\mathbf{E}_2 = \begin{bmatrix} -\mathrm{E}_1^{(\mathrm{p})}\, t^{\mathrm{p}}(k_x)\, k_{z_2}/k_2 \\ \mathrm{E}_1^{(\mathrm{s})}\, t^{\mathrm{s}}(k_x) \\ \mathrm{E}_1^{(\mathrm{p})}\, t^{\mathrm{p}}(k_x)\, k_x/k_2 \end{bmatrix} \mathrm{e}^{ik_x x + ik_{z_2} z}, \tag{4.41}$$

which can be expressed entirely by the angle of incidence θ_1 using $k_x = k_1 \sin\theta_1$. Note that we suppressed the harmonic time factor $\exp(-i\omega t)$. With this substitution the longitudinal wavenumbers can be written as (cf. Eq. (4.27))

$$k_{z_1} = k_1 \sqrt{1 - \sin^2\theta_1}, \qquad k_{z_2} = k_2 \sqrt{1 - \tilde{n}^2 \sin^2\theta_1}, \tag{4.42}$$

where we introduced the relative index of refraction

$$\tilde{n} = \frac{\sqrt{\varepsilon_1\mu_1}}{\sqrt{\varepsilon_2\mu_2}}\,. \tag{4.43}$$

For $\tilde{n} > 1$, with increasing θ_1 the argument of the square root in the expression of k_{z_2} gets smaller and smaller and eventually becomes negative. The critical angle θ_c can be defined by the condition

$$\left[1 - \tilde{n}^2 \sin^2\theta_1\right] = 0\,, \tag{4.44}$$

which describes a refracted plane wave with zero wavevector component in the z-direction ($k_{z_2} = 0$). Consequently, the refracted plane wave travels parallel to the interface. Solving for θ_1 yields

$$\theta_\mathrm{c} = \arcsin[1/\tilde{n}]\,. \tag{4.45}$$

For a glass/air interface at optical frequencies, we have $\varepsilon_2 = 1$, $\varepsilon_1 = 2.25$, and $\mu_1 = \mu_2 = 1$ yielding a critical angle $\theta_\mathrm{c} = 41.8°$.

For $\theta_1 > \theta_\mathrm{c}$, k_{z_2} becomes imaginary. Expressing the transmitted field as a function of the angle of incidence θ_1 results in

$$\mathbf{E}_2 \;=\; \begin{bmatrix} -i\mathrm{E}_1^{(\mathrm{p})}\, t^{\mathrm{p}}(\theta_1)\, \sqrt{\tilde{n}^2\, \sin^2\theta_1 \,-\, 1} \\ \mathrm{E}_1^{(\mathrm{s})}\, t^{\mathrm{s}}(\theta_1) \\ \mathrm{E}_1^{(\mathrm{p})}\, t^{\mathrm{p}}(\theta_1)\, \tilde{n} \sin\theta_1 \end{bmatrix} \mathrm{e}^{i \sin\theta_1 k_1 x}\, \mathrm{e}^{-\gamma z}, \tag{4.46}$$

where the decay constant γ is defined by

$$\gamma \;=\; k_2\, \sqrt{\tilde{n}^2\, \sin^2\theta_1 \,-\, 1}\,. \tag{4.47}$$

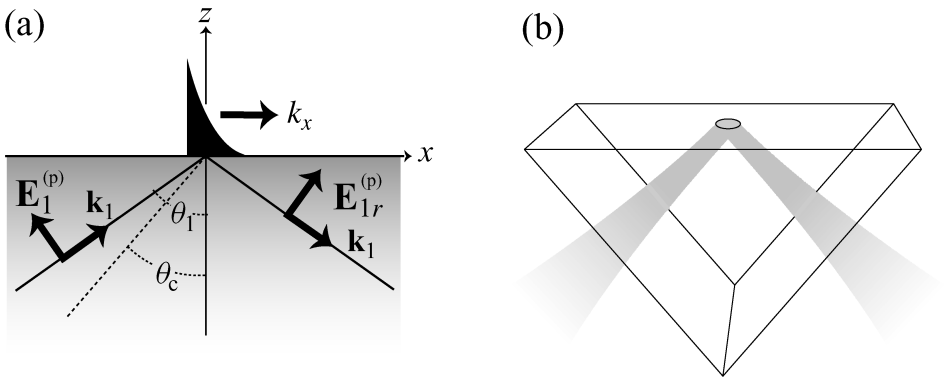

Figure 4.3: Excitation of an evanescent wave by total internal reflection. (a) An evanescent wave is created in a medium if the plane wave is incident at an angle $\theta_1 > \theta_\mathrm{c}$. (b) Actual experimental realization using a prism and a weakly focused Gaussian beam.

This equation describes a field that propagates along the surface but decays exponentially into the medium of transmittance. Thus, a plane wave incident at an angle $\theta_1 > \theta_c$ creates an evanescent wave. Excitation of an evanescent wave with a plane wave at supercritical incidence ($\theta_1 > \theta_c$) is referred to as *total internal reflection* (TIR). For the glass/air interface considered above and an angle of incidence of $\theta_i = 45°$, the decay constant is $\gamma = 2.22/\lambda$. This means that already at a distance of $\approx\lambda/2$ from the interface, the time-averaged field is a factor of e smaller than at the interface. At a distance of $\approx 2\lambda$ the field becomes negligible. The larger the angle of incidence θ_i the faster the decay will be. Note that the Fresnel coefficients depend on θ_1. For $\theta_1 > \theta_c$ they become complex numbers and, consequently, the phase of the reflected and transmitted wave is shifted relative to the incident wave. This phase shift is the origin of the so-called Goos–Hänchen shift. Furthermore, for p-polarized excitation, it results in elliptic polarization of the evanescent wave with the field vector rotating in the plane of incidence.

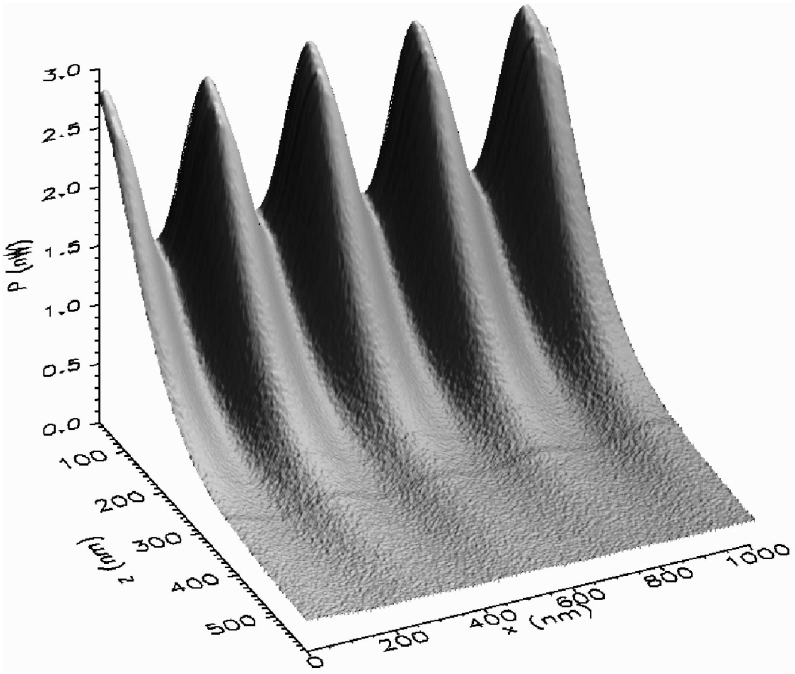

Figure 4.4: Spatial modulation of the standing evanescent wave along the propagation direction of two interfering waves (x-axis) and the decay of the intensity in the z-direction. The ordinate represents the measured optical power. From *Appl. Opt.* **33**, 7995 (1994).

Evanescent fields as described by Eq. (4.46) can be produced by directing a beam of light into a glass prism as sketched in Fig. 4.3(b). Experimental verification for the existence of this rapidly decaying field in the optical regime relies on approaching a transparent body to within less than $\lambda/2$ of the interface that supports the evanescent field (c.f. Fig. 4.4).

For p- and s-polarized evanescent waves, the intensity of the evanescent wave can be larger than that of the input beam. To see this we set $z = 0$ in Eq. (4.46) and we write for an s- and p-polarized plane wave separately the intensity ratio $|\mathbf{E}_2(z = 0)|^2/|\mathbf{E}_1(z = 0)|^2$. This ratio is equal to the absolute square of the Fresnel transmission coefficient $t^{p,s}$. These transmission coefficients are plotted in Fig. 4.5 for the example of a glass/air interface. For p-(s-)polarized light the transmitted evanescent intensity is up to a factor of 9 (4) larger than the incoming intensity. The maximum enhancement is found at the critical angle of TIR. The physical reason for this enhancement is a surface polarization that is induced by the incoming plane wave which is also represented by the boundary condition (4.17). A similar enhancement effect, but a much stronger one, can be obtained when the glass/air interface is covered by a thin layer of a noble metal. Here, so called surface plasmon polaritons can be excited.

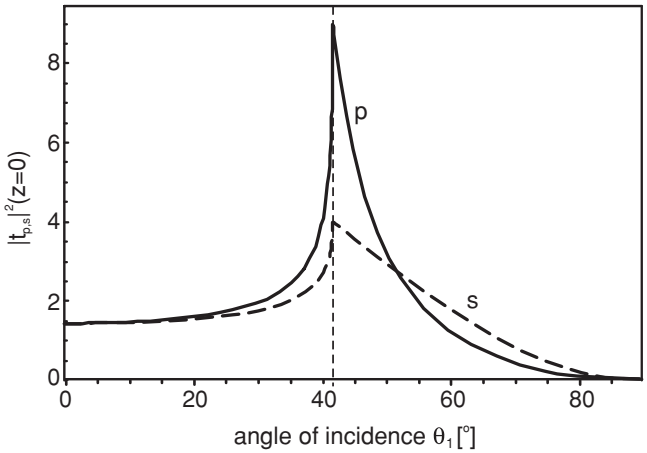

Figure 4.5: Intensity enhancement on top of a glass surface irradiated by a plane wave with variable angle of incidence θ_1. For p- and s-polarized waves, the enhancement peaks at the critical angle $\theta_c = 41.8°$ marked by the dotted line.

4.4.1 Frustrated total internal reflection

Evanescent fields can be converted into propagating radiation if they interact with matter. A plane interface will be used in order to create an evanescent wave by TIR as before. A second parallel plane interface is then advanced toward the first interface until the gap d is within the range of the typical decay length of the evanescent wave. A possible way to realize this experimentally is to close together two prisms with very flat or slightly curved surfaces as indicated in Fig. 4.6(b). The evanescent wave then interacts with the second interface and can be partly converted into propagating radiation. This situation is analogous to quantum mechanical tunneling through a potential barrier. The geometry of the problem is sketched in Fig. 4.6(a).

The fields are most conveniently expressed in terms of partial fields that are restricted to a single medium. The partial fields in media 1 and 2 are written as a superposition of incident and reflected waves, whereas for medium 3 there is only

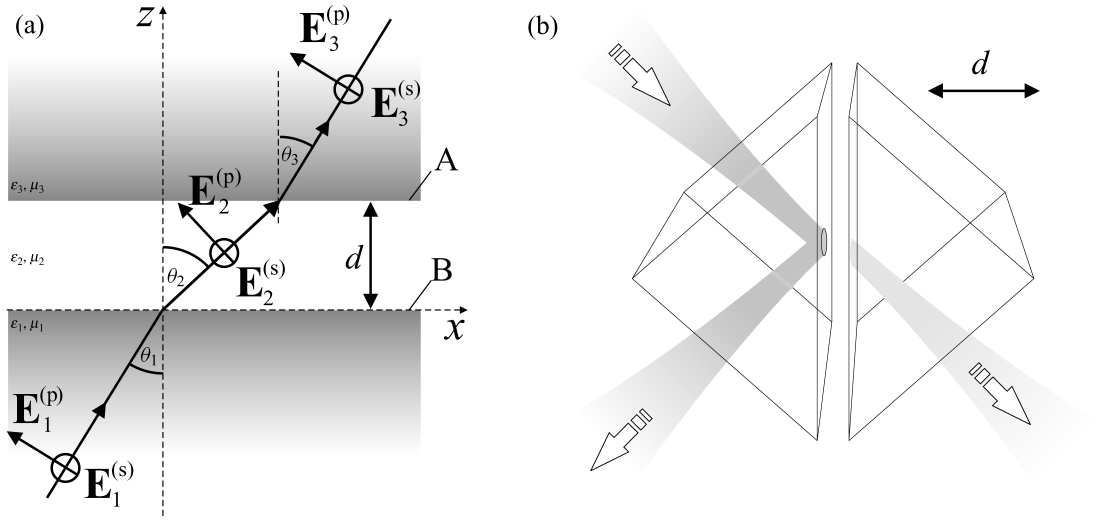

Figure 4.6: Transmission of a plane wave through a system of two parallel interfaces. In frustrated total internal reflection (FTIR), the evanescent wave created at interface B is partly converted into a propagating wave by the interface A of a second medium. (a) Configuration and definition of parameters. A, B: interfaces between media 2, 3 and 1, 2, respectively. The reflected waves are omitted for clarity. (b) Experimental set-up to observe frustrated total internal reflection.

a transmitted wave. The propagation character of these waves, i.e. whether they are evanescent or propagating in either of the three media, can be determined from the magnitude of the longitudinal wavenumber in each medium in analogy to Eq. (4.42). The longitudinal wavenumber in medium j reads

$$k_{j_z} \;=\; \sqrt{k_j^2 - k_\parallel^2} \;=\; k_j\sqrt{1 \,-\, (k_1/k_j)^2 \sin^2\theta_1}, \qquad j \in \{1,2,3\}\,, \tag{4.48}$$

where $k_j = n_j k_0 = n_j(\omega/c)$ and $n_j = \sqrt{\varepsilon_j \mu_j}$. In the following a layered system with $n_2 < n_3 < n_1$ will be discussed, which includes the system sketched in Fig. 4.6. This leads to three regimes for the angle of incidence in which the transmitted intensity as a function of the gap width d shows different behavior:

1. For $\theta_1 < \arcsin(n_2/n_1)$ or $k_\parallel < n_2 k_0$, the field is entirely described by propagating plane waves. The intensity transmitted to a detector far away from the second interface (far-field) will not vary substantially with gapwidth, but will only show rather weak interference undulations.

2. For $\arcsin(n_2/n_1) < \theta_1 < \arcsin(n_3/n_1)$ or $n_2 k_0 < k_\parallel < n_3 k_0$ the partial field in medium 2 is evanescent, but in medium (3) it is propagating. At the second interface evanescent waves are converted into propagating waves. The intensity transmitted to a remote detector will decrease strongly with increasing gapwidth. This situation is referred to as *frustrated total internal reflection* (FTIR).

3. For $\theta_1 > \arcsin\,(n_3/n_1)$ or $k_\parallel > n_3 k_0$ the waves in layer (2) *and* in layer (3) are evanescent and no intensity will be transmitted to a remote detector in medium (3).

If we chose θ_1 such that case 2 is realized (FTIR), the transmitted intensity $I(d)$ will reflect the steep distance dependence of the evanescent wave(s) in medium 2. However, as shown in Fig. 4.7, $I(d)$ deviates from a purely exponential behavior because the field in medium 2 is a superposition of two evanescent waves of the form

$$c_1\,\mathrm{e}^{-\gamma z} \;+\; c_2\,\mathrm{e}^{+\gamma z}\;. \tag{4.49}$$

The second term originates from the reflection of the primary evanescent wave (first term) at the second interface and its magnitude (c_2) depends on the material properties.

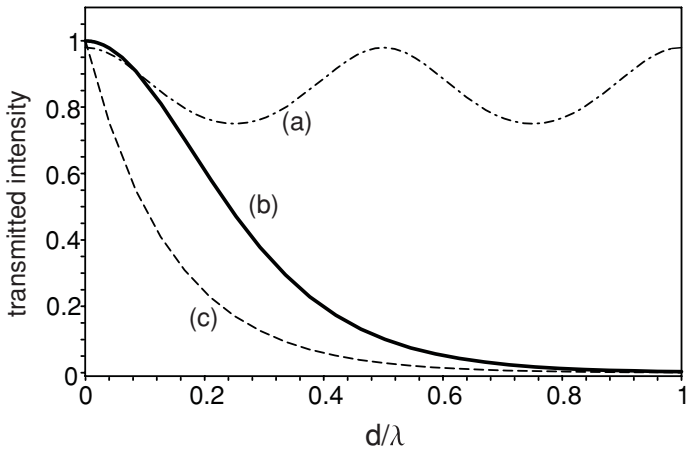

Figure 4.7: Transmission of a system of three media with parallel interfaces as a function of the gap d between the two interfaces. A p-polarized plane wave excites the system. The material constants are $n_1 = 2$, $n_2 = 1$, $n_3 = 1.51$. This leads to critical angles θ_c of 30° and 49.25°. For angles of incidence θ_i between (a) 0° and 30° the gap dependence shows interference-like behavior (here $\theta_1 = 0^\circ$, dash-dotted line), for angles between (b) 30° and 49.25° the transmission (monotonically) decreases with increasing gap width (here $\theta_1 = 35^\circ$, full line). (c) Intensity of the evanescent wave in the absence of the third medium.

The discussion of FTIR in this section contains most of the ingredients ti understand the physics of optical waveguides. We will return to this topic towards the end of this course.

Chapter 5

Energy and Momentum

The equations established so far describe the behavior of electric and magnetic fields. They are a direct consequence of Maxwell's equations and the properties of matter. Although the electric and magnetic fields were initially postulated to explain the forces in Coulomb's and Ampere's laws, Maxwell's equations do not provide any information about the energy content of an electromagnetic field. As we shall see, Poynting's theorem provides a plausible relationship between the electromagnetic field and its energy content.

5.1 Poynting's Theorem

If the scalar product of the field $\mathbf{E}$ and Eq. (1.34) is subtracted from the scalar product of the field $\mathbf{H}$ and Eq. (1.33) the following equation is obtained

$$\mathbf{H}\cdot(\nabla\times\mathbf{E})-\mathbf{E}\cdot(\nabla\times\mathbf{H}) \;=\; -\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t}-\mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t}-\mathbf{j}\cdot\mathbf{E}\,. \tag{5.1}$$

Noting that the expression on the left is identical to $\nabla\cdot(\mathbf{E}\times\mathbf{H})$, integrating both sides over space and applying Gauss' theorem the equation above becomes

$$\int_{\partial V}(\mathbf{E}\times\mathbf{H})\cdot\mathbf{n}\;da \;=\; -\int_{V}\left[\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t}+\mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t}+\mathbf{j}\cdot\mathbf{E}\right]dV \tag{5.2}$$

Although this equation already forms the basis of Poynting's theorem, more insight is provided when $\mathbf{B}$ and $\mathbf{D}$ are substituted by the generally valid equations (1.20).

Eq. (5.2) then reads as

$$\int_{\partial V} (\mathbf{E}\times\mathbf{H})\cdot\mathbf{n}\,da \;+\; \frac{\partial}{\partial t}\int_V \frac{1}{2}\Big[\mathbf{D}\cdot\mathbf{E}+\mathbf{B}\cdot\mathbf{H}\Big]\,dV \;=\; -\int_V \mathbf{j}\cdot\mathbf{E}\,dV \qquad (5.3)$$

$$-\frac{1}{2}\int_V\Big[\mathbf{E}\cdot\frac{\partial\mathbf{P}}{\partial t}-\mathbf{P}\cdot\frac{\partial\mathbf{E}}{\partial t}\Big]\,dV \;-\; \frac{\mu_0}{2}\int_V\Big[\mathbf{H}\cdot\frac{\partial\mathbf{M}}{\partial t}-\mathbf{M}\cdot\frac{\partial\mathbf{H}}{\partial t}\Big]\,dV.$$

This equation is a direct conclusion of Maxwell's equations and has therefore the same validity. Poynting's theorem is more or less an interpretation of the equation above. The left hand side has the general appearance of a *conservation law*, similar to the conservation of charge encountered previously in Eq. (1.1).

If we set $\mathbf{D} = \varepsilon_0\varepsilon\mathbf{E}$ and $\mathbf{B} = \mu_0\mu\mathbf{H}$ then the second integrand becomes $(1/2)\,[\varepsilon_0\varepsilon|\mathbf{E}|^2+\mu_0\mu|\mathbf{H}|^2]$, which is recognized as the sum of electric and magnetic energy density. Thus, the second term in Eq. (5.3) corresponds to the time rate of change of electromagnetic energy in the volume V and, accordingly, the first term is the flux of energy in or out of V. The remaining terms on the right side are equal to the rate of energy dissipation inside V. According to this interpretation

$$W \;=\; \frac{1}{2}\Big[\mathbf{D}\cdot\mathbf{E}+\mathbf{B}\cdot\mathbf{H}\Big] \qquad (5.4)$$

represents the density of electromagnetic energy, and

$$\mathbf{S} \;=\; (\mathbf{E}\times\mathbf{H}) \qquad (5.5)$$

is the energy flux density. $\mathbf{S}$ is referred to as the *Poynting vector*, discovered in 1883 by John Poynting and independently by Oliver Heaviside. In principle, the curl of any vector field can be added to $\mathbf{S}$ without changing the conservation law (5.3), but it is convenient to make the choice as stated in (5.5).

If the medium within V is linear and non-dispersive, the two last terms in Eq. (5.3) equal zero and the only term accounting for energy dissipation is $\mathbf{j}\cdot\mathbf{E}$. To understand this term, we consider the work done per unit time on a single charge q. In terms of the velocity $\mathbf{v}$ of the charge and the force $\mathbf{F}$ acting on it, the work per unit time is $dW/dt = \mathbf{F}\cdot\mathbf{v}$. Using the Lorentz force in Eq. (1) gives $dW/dt = q\,\mathbf{E}\cdot\mathbf{v} + q\,[\mathbf{v}\times\mathbf{B}]\cdot\mathbf{v}$. Because $[\mathbf{v}\times\mathbf{B}]\cdot\mathbf{v} = [\mathbf{v}\times\mathbf{v}]\cdot\mathbf{B} = 0$ we obtain

$\mathbf{F} = q\,\mathbf{v}\cdot\mathbf{E}$, which corresponds to the $\mathbf{j}\cdot\mathbf{E}$ term in Eq. (5.3).Thus, we find that the magnetic field does no work and that it is only the electric field that gives rise to dissipation of electromagnetic energy. The energy removed from the electromagnetic field is transferred to the charges in matter and ultimately to other forms of energy, such as heat.

In most practical applications we are interested in the mean value of **S**, that is, the value of $\mathbf{S}$ averaged over several oscillation periods. This quantity describes the net power flux density and is needed, for example, for the evaluation of radiation patterns. Assuming that the fields are harmonic in time, linear and non-dispersive, then the two last terms in Eq. (5.3) disappear. Furthermore, we assume that the energy density (5.4) only accounts for polarization and magnetization currents that a re loss-free, that is, all losses are associated with the $\mathbf{j}\cdot\mathbf{E}$ term. The time average of Eq. (5.3) then becomes

$$\int_{\partial V} \langle\mathbf{S}(\mathbf{r})\rangle\cdot\mathbf{n}\;da \;=\; -\frac{1}{2}\int_{V} \mathrm{Re}\left\{\mathbf{j}^{*}(\mathbf{r})\cdot\mathbf{E}(\mathbf{r})\right\}\;dV \tag{5.6}$$

where we have used complex notation. The term on the right defines the mean energy dissipation within the volume V. $\langle\mathbf{S}\rangle$ represents the time average of the Poynting vector

$$\langle\mathbf{S}(\mathbf{r})\rangle \;=\; \frac{1}{2}\,\mathrm{Re}\left\{\mathbf{E}(\mathbf{r})\times\mathbf{H}^{*}(\mathbf{r})\right\} \tag{5.7}$$

The magnitude of $\langle\mathbf{S}\rangle$ is called the intensity $I(\mathbf{r}) = |\langle\mathbf{S}(\mathbf{r})\rangle|$.

In the far-field, that is, far from sources and material boundaries, the electromagnetic field can be locally approximated by a plane wave (see Section 2.1.1). The electric and magnetic fields are in phase, perpendicular to each other, and the ratio of their amplitudes is constant. $\langle\mathbf{S}\rangle$ can then be expressed by the electric field alone as

$$\langle\mathbf{S}(\mathbf{r})\rangle \;=\; \frac{1}{2}\,\frac{1}{Z_i}\,|\mathbf{E}(\mathbf{r})|^{2}\,\mathbf{n}_r \;=\; \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\,n_i\,|\mathbf{E}(\mathbf{r})|^{2} \tag{5.8}$$

where $\mathbf{n}_r$ represents the unit vector in radial direction, $n_i = \sqrt{\varepsilon_i\,\mu_i}$ is the index of refraction, and Z_i is the wave impedance (4.33).

The surface integral of $\langle \mathbf{S} \rangle$ correspond to the total power generated or dissipated inside the enclosed surface, that is,

$$\bar{P} = \int_{\partial V} \langle \mathbf{S}(\mathbf{r}) \rangle \cdot \mathbf{n} \, da = \int_{\partial V} I(\mathbf{r}) \, da \tag{5.9}$$

5.1.1 Example: Energy Transport by Evanescent Waves

Let us consider a single dielectric interface that is irradiated by a plane wave under conditions of *total internal reflection* (TIR) (c.f. Section 4.4). For non-absorbing media and for supercritical incidence, all the power of the incident wave is reflected. One can anticipate that because no losses occur upon reflection at the interface there is no net energy transport into the medium of transmittance. In order to prove this fact we have to investigate the time-averaged energy flux across a plane parallel to the interface. This can be done by considering the z-component of the Poynting vector (cf. Eq. (5.7))

$$\langle S \rangle_z = \frac{1}{2} \mathrm{Re}\big(E_x H_y^* - E_y H_x^* \big) \ , \tag{5.10}$$

where all fields are evaluated in the upper medium, i.e. the medium of transmittance. Applying Maxwell's equation (2.32) to the special case of a plane or evanescent wave, allows us to express the magnetic field in terms of the electric field as

$$\mathbf{H} = \sqrt{\frac{\varepsilon_0 \varepsilon}{\mu_0 \mu}} \left[\left(\frac{\mathbf{k}}{k} \right) \times \mathbf{E} \right] \ . \tag{5.11}$$

Introducing the expressions for the transmitted field components of **E** and **H** into Eq. (5.10), it is straightforward to prove that $\langle \mathbf{S} \rangle_z$ vanishes and that there is no net energy transport in the direction *normal* to the interface.

On the other hand, when considering the energy transport along the interface ($\langle \mathbf{S} \rangle_x$), a non-zero result is found:

$$\langle \mathbf{S} \rangle_{\mathrm{x}} = \frac{1}{2} \sqrt{\frac{\varepsilon_2 \mu_2}{\varepsilon_1 \mu_1}} \sin \theta_1 \left(|t^{\mathrm{s}}|^2 \left| \mathbf{E}_1^{(\mathrm{s})} \right|^2 + |t^{\mathrm{p}}|^2 \left| \mathbf{E}_1^{(\mathrm{p})} \right|^2 \right) \mathrm{e}^{-2\gamma z} \, . \tag{5.12}$$

Thus, an evanescent wave transports energy along the surface, in the direction of the transverse wavevector.

5.1.2 Energy density in dispersive and lossy media

The two last terms in Eq. (5.3) strictly vanish only in a linear medium with no dispersion and no losses. The only medium fulfilling these conditions is vacuum. For all other media, the last two terms only vanish approximately. In this section we consider a linear medium with a frequency-dependent and complex ε and μ.

Let us return to the Poynting theorem stated in Eq. (5.2). While the left hand side denotes the power flowing in or out of the volume V, the right hand side denotes the power dissipated or generated in the volume V. The three terms on the right hand side are of similar form and so we start by considering first the electric energy term $\mathbf{E} \cdot (\partial \mathbf{D}/\partial t)$. The electric energy density w_E at the time t is

$$w_E(\mathbf{r}, t) \;=\; \int_{-\infty}^{t} \mathbf{E}(\mathbf{r}, t') \cdot \frac{\partial \mathbf{D}(\mathbf{r}, t')}{\partial t'}\, dt' \;. \tag{5.13}$$

We now express the fields $\mathbf{E}$ and $\mathbf{D}$ in terms of their Fourier transforms as $\mathbf{E}(t') = \int \hat{\mathbf{E}}(\omega) \exp[-\mathrm{i}\omega t']\, \mathrm{d}\omega$ and $\mathbf{D}(t') = \int \hat{\mathbf{D}}(\omega) \exp[-\mathrm{i}\omega t']\, \mathrm{d}\omega$, respectively. In the last expression we substitute $\omega = -\omega'$ and obtain $\mathbf{D}(t') = \int \hat{\mathbf{D}}^*(\omega') \exp[\mathrm{i}\omega' t']\, \mathrm{d}\omega'$, where we used $\hat{\mathbf{D}}(-\omega') = \hat{\mathbf{D}}^*(\omega')$ since $\mathbf{D}(t)$ is real (c.f. Eq. (2.22)). Using the linear relation $\hat{\mathbf{D}} = \varepsilon_0\, \varepsilon \hat{\mathbf{E}}$ and inserting the Fourier transforms in Eq. (5.13) yields

$$w_E(\mathbf{r}, t) \;=\; \varepsilon_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\omega'\, \varepsilon^*(\omega')}{\omega' - \omega}\, \hat{\mathbf{E}}(\omega) \cdot \hat{\mathbf{E}}^*(\omega')\, \mathrm{e}^{i(\omega' - \omega)t}\, d\omega'\, d\omega \;, \tag{5.14}$$

where we have carried out the differentiation and integration over time and assumed that the fields were zero at $t \to -\infty$. For later purposes it is advantageous to represent the above result in different form. Using the substitutions $u' = -\omega$ and $u = -\omega'$ and making use of $\hat{\mathbf{E}}(-u) = \hat{\mathbf{E}}^*(u)$ and $\varepsilon(-u) = \varepsilon^*(u)$ gives an expression similar to Eq. (5.14) but in terms of u and u'. Finally, we add this expression to Eq. (5.14) and take one half of the resulting sum, which yields

$$w_E(\mathbf{r}, t) \;=\; \frac{\varepsilon_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\omega' \varepsilon^*(\omega') - \omega \varepsilon(\omega)}{\omega' - \omega} \right] \hat{\mathbf{E}}(\omega) \cdot \hat{\mathbf{E}}^*(\omega')\, \mathrm{e}^{i(\omega' - \omega)t}\, d\omega'\, d\omega \;. \tag{5.15}$$

Similar expressions are obtained for the magnetic term $\mathbf{H} \cdot (\partial \mathbf{B}/\partial t)$ and the dissipative term $\mathbf{j} \cdot \mathbf{E}$ in Eq. (5.2).

If $\varepsilon(\omega)$ is a complex function then w_E not only accounts for the energy density built up in the medium but also for the energy transferred to the medium, such as

heat dissipation. This contribution becomes indistinguishable from the term $\mathbf{j}\cdot\mathbf{E}$ in Eq. (5.2). Thus, the imaginary part of ε can be included in the conductivity σ (c.f. Eq. (3.12)) and accounted for in the term $\mathbf{j}\cdot\mathbf{E}$ through the linear relationship $\hat{\mathbf{j}}=\sigma\hat{\mathbf{E}}$. Therefore, to discuss the energy density it suffices to consider only the real part of ε, which we're going to denote as ε'.

Let us now consider a monochromatic field represented by $\hat{\mathbf{E}}(\mathbf{r},\omega)=\mathbf{E}_0(\mathbf{r})[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]/2$. Inserting in Eq. (5.15) yields four terms: two that are constant in time and two that oscillate in time. Upon averaging over an oscillation period $2\pi/\omega_0$ the oscillatory terms vanish and only the constant terms survive. For these terms we must view the expression in brackets in Eq. (5.15) as a limit, that is,

$$\lim_{\omega'\to\omega}\left[\frac{\omega'\varepsilon'(\omega')-\omega\varepsilon'(\omega)}{\omega'-\omega}\right] \;=\; \left.\frac{d\left[\omega\,\varepsilon'(\omega)\right]}{d\omega}\right|_{\omega=\omega_0} . \tag{5.16}$$

Thus, the cycle average of Eq. (5.15) yields

$$\bar{w_E}(\mathbf{r}) \;=\; \frac{\varepsilon_0}{4}\left.\frac{d\left[\omega\,\varepsilon'(\omega)\right]}{d\omega}\right|_{\omega=\omega_0}|\mathbf{E}_0(\mathbf{r})|^2 . \tag{5.17}$$

A similar result can be derived for the magnetic term $\mathbf{H}\cdot(\partial\mathbf{B}/\partial t)$.

It can be shown that Eq. (5.17) also holds for quasi-monochromatic fields which have frequency components ω only in a narrow range about a center frequency ω_0. Such fields can be represented as

$$\mathbf{E}(\mathbf{r},t)=\mathrm{Re}\{\tilde{\mathbf{E}}(\mathbf{r},t)\}=\mathrm{Re}\{\mathbf{E}_0(\mathbf{r},t)\,\mathrm{e}^{-i\omega_0 t}\}\;, \tag{5.18}$$

which is known as the *slowly varying amplitude approximation*. Here, $\mathbf{E}_0(\mathbf{r},t)$ is the slowly varying (complex) amplitude and ω_0 is the 'carrier' frequency. The envelope $\mathbf{E}_0$ spans over many oscillations of frequency ω_0.

Expressing the field amplitudes in terms of time-averages, that is $|\mathbf{E}_0|^2=2\,\langle\mathbf{E}(t)\cdot\mathbf{E}(t)\rangle$, we can express the total cycle-averaged energy density $\bar{W}$ as

$$\bar{W} \;=\; \left[\varepsilon_0\frac{d\left[\omega\,\varepsilon'(\omega)\right]}{d\omega}\left\langle\mathbf{E}\cdot\mathbf{E}\right\rangle \;+\; \mu_0\frac{d\left[\omega\,\mu'(\omega)\right]}{d\omega}\left\langle\mathbf{H}\cdot\mathbf{H}\right\rangle\right] \tag{5.19}$$

where $\mathbf{E}=\mathbf{E}(\mathbf{r},t)$ and $\mathbf{H}=\mathbf{H}(\mathbf{r},t)$ are the time-dependent fields. Notice, that ω is the center frequency of the spectra of $\mathbf{E}$ and $\mathbf{H}$. For a medium with negligible dispersion this expression reduces to the familiar $\bar{W}\;=\;(1/2)\,[\varepsilon_o\varepsilon'|\mathbf{E}_0|^2+\;\mu_o\mu'\,|\mathbf{H}_0|^2]$,

which follows from Eq. (5.4) using the dispersion-free constitutive relations. Because of $d(\omega\varepsilon')/d\omega > 0$ and $d(\omega\mu')/d\omega > 0$ the energy density is always positive, even for metals with $\varepsilon' < 0$.

5.2 The Maxwell Stress Tensor

In this section we use Maxwell's equations to derive the conservation law for linear momentum in an electromagnetic field. The net force exerted on an arbitrary object is entirely determined by Maxwell's stress tensor. In the limiting case of an infinitely extended object, the formalism renders the known formulas for radiation pressure.

The general law for forces in electromagnetic fields is based on the conservation law for linear momentum. To derive this conservation law we will consider Maxwell's equations in vacuum. In this case we have $\mathbf{D} = \varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu_0\mathbf{H}$. Later we will relax this constraint. The conservation law for linear momentum is entirely a consequence of Maxwell's equations (1.16) - (1.19) and of the Lorentz force law (5), which connects the electromagnetic world with the mechanical one.

If we operate on Maxwell's first equation by $\times\varepsilon_0\mathbf{E}$, on the second equation by $\times\mu_0\mathbf{H}$, and then add the two resulting equations we obtain

$$\varepsilon_0(\nabla\times\mathbf{E})\times\mathbf{E}+\mu_0(\nabla\times\mathbf{H})\times\mathbf{H}=\mathbf{j}\times\mathbf{B}-\frac{1}{c^2}\left[\frac{\partial\mathbf{H}}{\partial t}\times\mathbf{E}\right]+\frac{1}{c^2}\left[\frac{\partial\mathbf{E}}{\partial t}\times\mathbf{H}\right] \tag{5.20}$$

We have omitted the arguments $(\mathbf{r},t)$ for the different fields and we used $\varepsilon_0\mu_0 = 1/c^2$. The last two expressions in Eq. (5.20) can be combined to $(1/c^2)\,\mathrm{d}/\mathrm{d}t\,[\mathbf{E}\times\mathbf{H}]$. For the first expression in Eq. (5.20) we can write

$$\begin{aligned}&\varepsilon_0(\nabla\times\mathbf{E})\times\mathbf{E} \;= \qquad\qquad (5.21)\\ &\varepsilon_0\begin{bmatrix}\partial/\partial x(E_x^2-E^2/2) & +\partial/\partial y(E_xE_y) & +\partial/\partial z(E_xE_z)\\ \partial/\partial x(E_xE_y) & +\partial/\partial y(E_y^2-E^2/2) & +\partial/\partial z(E_yE_z)\\ \partial/\partial x(E_xE_z) & +\partial/\partial y(E_yE_z) & +\partial/\partial z(E_z^2-E^2/2)\end{bmatrix}-\varepsilon_0\mathbf{E}\,\nabla\cdot\mathbf{E}\\ &=\;\nabla\cdot[\varepsilon_0\mathbf{E}\mathbf{E}\;-\;(\varepsilon_0/2)E^2\,\overleftrightarrow{\mathbf{I}}]-\rho\mathbf{E}\;.\end{aligned}$$

where Eq. (1.32) has been used in the last step. The notation $\mathbf{EE}$ denotes the outer product, $E^2 = E_x^2 + E_y^2 + E_z^2$ is the electric field strength, and $\overleftrightarrow{\mathbf{I}}$ denotes the unit tensor. A similar expression can be derived for $\mu_0(\nabla \times \mathbf{H}) \times \mathbf{H}$. Using these two expressions in Eq. (5.20) we obtain

$$\nabla \cdot \left[\varepsilon_0 \mathbf{EE} - \mu_0 \mathbf{HH} - \frac{1}{2}\left(\varepsilon_0 E^2 + \mu_0 H^2\right) \overleftrightarrow{\mathbf{I}}\right] = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{c^2}[\mathbf{E} \times \mathbf{H}] + \rho\,\mathbf{E} + \mathbf{j} \times \mathbf{B}\,. \quad (5.22)$$

The expression in brackets on the left hand side is called Maxwell's stress tensor in vacuum, usually denoted as $\overleftrightarrow{\mathbf{T}}$. In Cartesian components it reads as

$$\overleftrightarrow{\mathbf{T}} = \left[\varepsilon_0 \mathbf{EE} - \mu_0 \mathbf{HH} - \frac{1}{2}\left(\varepsilon_0 E^2 + \mu_0 H^2\right) \overleftrightarrow{\mathbf{I}}\right] = \quad (5.23)$$

$$\begin{bmatrix} \varepsilon_0(E_x^2 - E^2/2) + \mu_0(H_x^2 - H^2/2) & \varepsilon_0 E_x E_y + \mu_0 H_x H_y & \varepsilon_0 E_x E_z + \mu_0 H_x H_z \\ \varepsilon_0 E_x E_y + \mu_0 H_x H_y & \varepsilon_0(E_y^2 - E^2/2) + \mu_0(H_y^2 - H^2/2) & \varepsilon_0 E_y E_z + \mu_0 H_y H_z \\ \varepsilon_0 E_x E_z + \mu_0 H_x H_z & \varepsilon_0 E_y E_z + \mu_0 H_y H_z & \varepsilon_0(E_z^2 - E^2/2) + \mu_0(H_z^2 - H^2/2) \end{bmatrix}$$

After integration of Eq. (5.22) over an arbitrary volume V which contains all sources ρ and $\mathbf{j}$ we obtain

$$\int_V \nabla \cdot \overleftrightarrow{\mathbf{T}}\; \mathrm{d}V = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{c^2}\int_V [\mathbf{E} \times \mathbf{H}]\,\mathrm{d}V + \int_V [\rho\,\mathbf{E} + \mathbf{j} \times \mathbf{B}]\,\mathrm{d}V\,. \quad (5.24)$$

The last term is recognized as the mechanical force (cf. Eq. (5)). The volume integral on the left can be transformed to a surface integral using Gauss's integration law

$$\int_V \nabla \cdot \overleftrightarrow{\mathbf{T}}\; \mathrm{d}V = \int_{\partial V} \overleftrightarrow{\mathbf{T}} \cdot \mathbf{n}\,\mathrm{d}a\,. \quad (5.25)$$

∂V denotes the surface of V, $\mathbf{n}$ the unit vector perpendicular to the surface, and $\mathrm{d}a$ an infinitesimal surface element. We then finally arrive at the *conservation law for linear momentum*

$$\int_{\partial V} \overleftrightarrow{\mathbf{T}}\,(\mathbf{r}, t) \cdot \mathbf{n}(\mathbf{r})\,\mathrm{d}a = \frac{\mathrm{d}}{\mathrm{d}t}\left[\mathbf{G}_{\mathrm{field}} + \mathbf{G}_{\mathrm{mech}}\right] \quad (5.26)$$

Here, $\mathbf{G}_{\mathrm{mech}}$ and $\mathbf{G}_{\mathrm{field}}$ denote the mechanical momentum and the field momentum, respectively. In Eq. (5.26) we have used Newton's expression of the mechanical

force $\mathbf{F} = \mathrm{d}/\mathrm{d}t\,\mathbf{G}_{\mathrm{mech}}$ and the definition of the *field momentum* (Abraham density)

$$\mathbf{G}_{\mathrm{field}} = \frac{1}{c^2}\int_V [\mathbf{E}\times\mathbf{H}]\,\mathrm{d}V \tag{5.27}$$

This is the momentum carried by the electromagnetic field within the volume V. It is created by the dynamic terms in Maxwell's curl equations. The time-derivative of the field momentum is zero when it is averaged over one oscillation period and hence the average mechanical force becomes

$$\langle\mathbf{F}\rangle = \int_{\partial V}\langle\overleftrightarrow{\mathbf{T}}\,(\mathbf{r},t)\rangle\cdot\mathbf{n}(\mathbf{r})\;\mathrm{d}a \tag{5.28}$$

with $\langle ...\rangle$ denoting the time average. Equation (5.28) is of general validity. It allows the mechanical force acting on an arbitrary body within the closed surface ∂V to be calculated (see Figure 5.1). The force is entirely determined by the electric and magnetic fields on the surface ∂V. It is interesting to note that no material properties enter the expression for the force; the entire information is contained in the electromagnetic field. The only material constraint is that the body is rigid. If the body deforms when it is subject to an electromagnetic field we have to include electrostrictive and magnetostrictive forces. Since the enclosing surface is

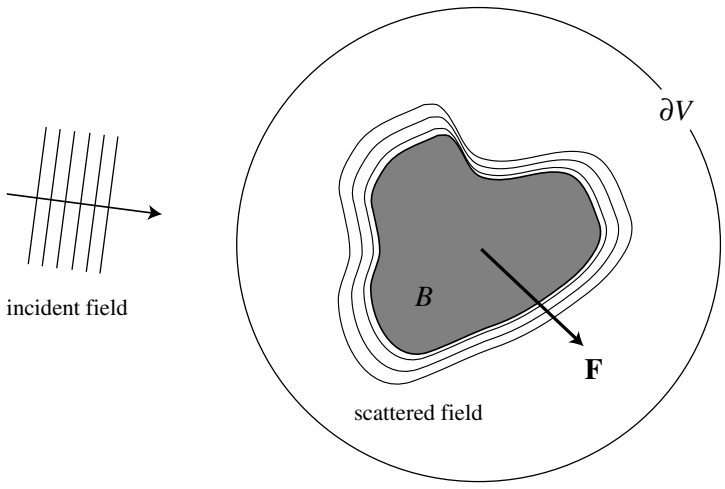

Figure 5.1: The mechanical force $\mathbf{F}$ acting on the object B is entirely determined by the electric and magnetic fields at an arbitrary surface ∂V enclosing B.

arbitrary the same results are obtained whether the fields are evaluated at the surface of the body or in the far-field. It is important to note that the fields used to calculate the force are the self-consistent fields of the problem, which means that they are a superposition of the incident and the scattered fields. Therefore, prior to calculating the force, one has to solve for the electromagnetic fields. If the object B is surrounded by a medium that can be represented accurately enough by the dielectric constant ε and magnetic susceptibility μ, the mechanical force can be calculated in the same way if we replace Maxwell's stress tensor Eq. (5.23) by

$$\overleftrightarrow{\mathbf{T}} = [\varepsilon_0\varepsilon\mathbf{EE} - \mu_0\mu\mathbf{HH} - \frac{1}{2}(\varepsilon_0\varepsilon E^2 + \mu_0\mu H^2)\,\overleftrightarrow{\mathbf{I}}] \tag{5.29}$$

5.3 Radiation pressure

Here, we consider the radiation pressure on a medium with an infinitely extended planar interface as shown in Fig. 5.2. The medium is irradiated by a monochromatic plane wave at normal incidence to the interface. Depending on the material properties of the medium, part of the incident field is reflected at the interface. Introducing the complex reflection coefficient r, the electric field outside the medium can be written as the superposition of two counter-propagating plane waves

$$\mathbf{E}(\mathbf{r},t) = E_0\,\mathrm{Re}\Big\{[\mathrm{e}^{\mathrm{i}kz} + r\,\mathrm{e}^{-\mathrm{i}kz}]\,\mathrm{e}^{-\mathrm{i}\omega t}\Big\}\,\mathbf{n}_x\,. \tag{5.30}$$

Using Maxwell's curl equation (1.33) we find for the magnetic field

$$\mathbf{H}(\mathbf{r},t) = \sqrt{\varepsilon_0/\mu_0}\,E_0\,\mathrm{Re}\Big\{[\mathrm{e}^{\mathrm{i}kz} - r\,\mathrm{e}^{-\mathrm{i}kz}]\,\mathrm{e}^{-\mathrm{i}\omega t}\Big\}\,\mathbf{n}_y\,. \tag{5.31}$$

To calculate the radiation pressure P we integrate Maxwell's stress tensor on an infinite planar surface A parallel to the interface as shown in Fig. 5.2. The radiation pressure can be calculated by using Eq. (5.28) as

$$P\,\mathbf{n}_z = \frac{1}{A}\int_A \langle\overleftrightarrow{\mathbf{T}}\,(\mathbf{r},t)\rangle\cdot\mathbf{n}_z\,\mathrm{d}a\,. \tag{5.32}$$

We do not need to consider a closed surface ∂V since we are interested in the pressure exerted on the interface of the medium and not in the mechanical force

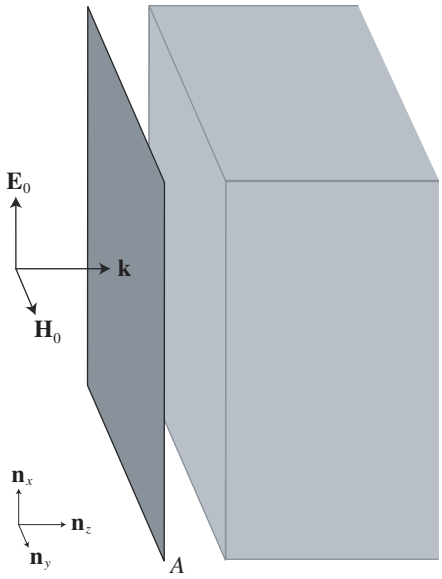

Figure 5.2: Configuration used to derive the radiation pressure.

acting on the medium. Using the fields of Eqs. (5.30) and (5.31) we find that the first two terms in Maxwell's stress tensor Eq. (5.23) give no contribution to the radiation pressure. The third term yields

$$\langle \overleftrightarrow{\mathbf{T}}\,(\mathbf{r},t)\rangle \cdot \mathbf{n}_z \;=\; -\frac{1}{2}\,\langle \varepsilon_0 E^2 + \mu_0 H^2\rangle\,\mathbf{n}_z \;=\; \frac{\varepsilon_0}{2}\,E_0^2\,[1+|r|^2]\,\mathbf{n}_z\;. \tag{5.33}$$

Using the definition of the intensity of a plane wave $I_0 = (\varepsilon_0/2)\,c\,E_0^2$, c being the vacuum speed of light, we can express the radiation pressure as

$$P \;=\; \frac{I_0}{c}\,[1\,+\,R]\;, \tag{5.34}$$

with $R = |r|^2$ being the reflectivity. For a perfectly absorbing medium we have $R = 0$, whereas for a perfectly reflecting medium $R = 1$. Therefore, the radiation pressure on a perfectly reflecting medium is twice as high as for a perfectly absorbing medium.

To conclude this chapter we should emphasize the importance of electromagnetic energy and momentum. Energy and momentum are fundamental concepts of physics that make transitions between different fields feasible. For example, electromagnetic energy can be transferred to heat, which is a concept of thermodynamics, and electromagnetic momentum can be transferred to mechanical

forces, which is a concept of classical mechanics. Electromagnetic energy and momentum are also used in Lagrangian and Hamiltonian formalisms, which form the stepping stones to quantum mechanics and quantum electrodynamics. Thus, energy and momentum make it possible to transition between different fields of physics. Such transitions cannot be accomplished by more standard electrical concepts such as voltage and current.

Chapter 6

Radiation

In this chapter we will discuss the emission of electromagnetic radiation from elementary sources. A stationary charge yields a static electric field, but it does not radiate. Similarly, a uniformly moving charge (a current) yields a static magnetic field, but it does not radiate. It is the *acceleration* of charge that gives rise to radiation. The smallest radiating unit is a dipole, an electromagnetic point source. According to linear response theory, a point source excitation yields the *system response function*, from which we can calculate the fields of more complicated sources by using the superposition principle. The system response function is also referred to as the *Green function*.

6.1 Green functions

Before calculating the fields radiated by elementary sources let us discuss an important mathematical concept, namely the concept of the Green function. Consider the following general, inhomogeneous equation:

$$\mathcal{L}\,\mathbf{A}(\mathbf{r}) \;=\; \mathbf{B}(\mathbf{r})\;. \tag{6.1}$$

$\mathcal{L}$ is a linear operator acting on the vectorfield $\mathbf{A}$ representing the unknown response of the system. The vectorfield $\mathbf{B}$ is a known source function and makes the differential equation inhomogeneous. A well-known theorem for linear differential equations states that the general solution is equal to the sum of the complete

homogeneous solution $(\mathbf{B}=0)$ and a particular inhomogeneous solution. Here, we assume that the homogeneous solution $(\mathbf{A}_0)$ is known. We thus need to solve for an arbitrary particular solution.

Usually it is difficult to find a solution of Eq. (6.1) and it is easier to consider the special inhomogeneity $\delta(\mathbf{r}-\mathbf{r}')$, which is zero everywhere, except in the point $\mathbf{r}=\mathbf{r}'$. Then, the linear equation reads as

$$\mathcal{L}\,\mathbf{G}_i(\mathbf{r},\mathbf{r}') \;=\; \mathbf{n}_i\,\delta(\mathbf{r}-\mathbf{r}') \qquad (i=x,y,z)\;, \tag{6.2}$$

where $\mathbf{n}_i$ denotes an arbitrary constant unit vector. In general, the vectorfield $\mathbf{G}_i$ is dependent on the location $\mathbf{r}'$ of the inhomogeneity $\delta(\mathbf{r}-\mathbf{r}')$. Therefore, the vector $\mathbf{r}'$ has been included in the argument of $\mathbf{G}_i$. The three equations given by Eq. (6.2) can be written in closed form as

$$\mathcal{L}\,\overleftrightarrow{\mathbf{G}}(\mathbf{r},\mathbf{r}') \;=\overleftrightarrow{\mathbf{I}}\,\delta(\mathbf{r}-\mathbf{r}')\;, \tag{6.3}$$

where the operator $\mathcal{L}$ acts on each column of $\overleftrightarrow{\mathbf{G}}$ separately and $\overleftrightarrow{\mathbf{I}}$ is the unit tensor. The dyadic function $\overleftrightarrow{\mathbf{G}}$ fulfilling Eq. (6.3) is known as the dyadic Green function.

In a next step, assume that Eq. (6.3) has been solved and that $\overleftrightarrow{\mathbf{G}}$ is known. Postmultiplying Eq. (6.3) with $\mathbf{B}(\mathbf{r}')$ on both sides and integrating over the volume V in which $\mathbf{B}\neq 0$ gives

$$\int_V \mathcal{L}\,\overleftrightarrow{\mathbf{G}}(\mathbf{r},\mathbf{r}')\,\mathbf{B}(\mathbf{r}')\,\mathrm{d}V' \;=\; \int_V \mathbf{B}(\mathbf{r}')\,\delta(\mathbf{r}-\mathbf{r}')\,\mathrm{d}V'\;. \tag{6.4}$$

The right hand side simply reduces to $\mathbf{B}(\mathbf{r})$ and with Eq. (6.1) it follows that

$$\mathcal{L}\,\mathbf{A}(\mathbf{r}) \;=\; \int_V \mathcal{L}\,\overleftrightarrow{\mathbf{G}}(\mathbf{r},\mathbf{r}')\,\mathbf{B}(\mathbf{r}')\,\mathrm{d}V'\;. \tag{6.5}$$

If on the right hand side the operator $\mathcal{L}$ is taken out of the integral, the solution of Eq. (6.1) can be expressed as

$$\mathbf{A}(\mathbf{r}) \;=\; \int_V \overleftrightarrow{\mathbf{G}}(\mathbf{r},\mathbf{r}')\,\mathbf{B}(\mathbf{r}')\,\mathrm{d}V'\;. \tag{6.6}$$

Thus, the solution of the original equation can be found by integrating the product of the dyadic Green function and the inhomogeneity $\mathbf{B}$ over the source volume V.

The assumption that the operators $\mathcal{L}$ and $\int \mathrm{d}V'$ can be interchanged is not strictly valid and special care must be applied if the integrand is not well behaved. Most often $\overleftrightarrow{\mathbf{G}}$ is singular at $\mathbf{r} = \mathbf{r}'$ and an infinitesimal exclusion volume surrounding $\mathbf{r} = \mathbf{r}'$ has to be introduced. As long as we consider field points outside of the source volume V, i.e. $\mathbf{r} \notin V$, we do not need to consider these tricky issues.

6.2 Scalar and Vector Potentials

The $\mathbf{E}$ and $\mathbf{B}$ fields define a total of six functions in space and time. It turns out, that these fields are not independent and that one needs fewer functions to uniquely determine the electromagnetic field. The vector potential $\mathbf{A}$ and the scalar potential ϕ constitute a set of only four functions which, depending on the type of problem, can be reduced to even fewer functions. These potentials are also of key importance in quantum mechanics.

Let's consider Maxwell's equation $\nabla \cdot \mathbf{B} = 0$ and replace $\mathbf{B}$ by another function. Because, $\nabla \cdot \nabla \times = 0$ we choose $\mathbf{B} = \nabla \times \mathbf{A}$. Next, we consider Faraday's law $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$ and replace $\mathbf{B}$ by $\nabla \times \mathbf{A}$. We obtain $\nabla \times [\mathbf{E} + \partial \mathbf{A}/\partial t] = 0$. Considering that $\nabla \times \nabla = 0$, we set $[\mathbf{E} + \partial \mathbf{A}/\partial t] = -\nabla \phi$, which yields $\mathbf{E} = -\partial \mathbf{A}/\partial t - \nabla \phi$. To summarize,

$$\mathbf{E}(\mathbf{r}, t) = -\frac{\partial}{\partial t}\mathbf{A}(\mathbf{r}, t) - \nabla \phi(\mathbf{r}, t) \tag{6.7}$$

$$\mathbf{B}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t) \tag{6.8}$$

It turns out that these definitions of vector potential $\mathbf{A}$ and scalar potential ϕ are not unique. If the potentials are replaced by new potentials $\tilde{\mathbf{A}}$, $\tilde{\phi}$ according to

$$\mathbf{A} \to \tilde{\mathbf{A}} + \nabla \chi \quad \text{and} \quad \phi \to \tilde{\phi} - \partial \chi / \partial t \ , \tag{6.9}$$

with $\chi(\mathbf{r}, t)$ being an arbitrary gauge function, Maxwell's equations remain unaffected. This is easily seen by introducing the above substitutions in the definitions of $\mathbf{A}$ and ϕ.

6.2.1 The Gauges

Any vectorfield $\mathbf{F}$ is specified by the definition of $\nabla \cdot \mathbf{F}$ and $\nabla \times \mathbf{F}$. A vectorfield with $\nabla \cdot \mathbf{F} = 0$ is called *transverse*, whereas $\nabla \times \mathbf{F} = 0$ defines a *longitudinal* field.

So far, we have defined the curl of $\mathbf{A}$, i.e. $\nabla \times \mathbf{A} = \mathbf{B}$. However, we did not specify $\nabla \cdot \mathbf{A}$. The choice of $\nabla \cdot \mathbf{A}$ does not affect the fields $\mathbf{E}$ and $\mathbf{B}$. Typically one chooses $\nabla \cdot \mathbf{A}$ such that the wave equation for $\mathbf{A}$ assumes a simple form or that favorable symmetries can be exploited. To demonstrate this, we consider Maxwell's equation $\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t + \mathbf{j}$. Using the relations (1.20) we obtain $\nabla \times \mathbf{B} - (1/c^2)\partial \mathbf{E}/\partial t = \mu_0[\nabla \times \mathbf{M} + \partial \mathbf{P}/\partial t + \mathbf{j}]$, where the expression in brackets is the total current density $\mathbf{j}_{\rm tot}$. Inserting Eqs. (6.7) and (6.8) yields

$$\nabla \times \nabla \times \mathbf{A} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} + \frac{1}{c^2}\nabla\frac{\partial \phi}{\partial t} = \mu_0 \, \mathbf{j}_{\rm tot} \; , \tag{6.10}$$

which can be rewritten as

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = -\mu_0 \, \mathbf{j}_{\rm tot} + \nabla\left[\nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t}\right] \; . \tag{6.11}$$

The expression in brackets contains a $\nabla \cdot \mathbf{A}$ term, which we can choose as we wish. Finally, we also express Gauss' law $\nabla \cdot \mathbf{D} = \rho$ in terms of $\mathbf{A}$ and ϕ and obtain

$$\nabla \cdot (\partial \mathbf{A}/\partial t + \nabla \phi) = -\rho_{\rm tot}/\varepsilon_0 \; . \tag{6.12}$$

There is again a $\nabla \cdot \mathbf{A}$ term that can be arbitrarily chosen.

Lorenz Gauge

In the Lorenz gauge one chooses $\nabla \cdot \mathbf{A} = -(1/c^2)\,\partial \phi/\partial t$[1], which yields

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\mathbf{A} = -\mu_0 \, \mathbf{j}_{\rm tot} \tag{6.13}$$

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\phi = -\frac{1}{\varepsilon_0}\,\rho_{\rm tot} \tag{6.14}$$

[1] This has the form of a continuity equation ($\mathbf{A}$ is the current density and ϕ/c^2 is the charge density).

Thus, we obtain two decoupled partial differential equations of the *same* form for $\mathbf{A}$ and ϕ. Note, that one ends up with the same differential equations by a proper choice of the gauge function χ (6.9).

The advantage of the Lorenz gauge is that the vectorial differential equation for $\mathbf{A}$ is decoupled into a set of three independent scalar differential equations, that is, each vector component A_i depends only on the source component j_{tot_i}. There is no mixing of components $i \in [x, y, z]$.

Coulomb Gauge

In the Coulomb gauge one chooses $\nabla \cdot \mathbf{A} = 0$. This gauge is also referred to as the transverse gauge or the minimal coupling gauge. With this choice of gauge Eqs. (6.11) and (6.12) reduce to

$$\begin{aligned} \left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right] \mathbf{A} &= -\mu_0 \mathbf{j}_{\mathrm{tot}} + \frac{1}{c^2}\nabla\frac{\partial \phi}{\partial t} \\ \nabla^2 \phi &= -\frac{1}{\varepsilon_0}\rho_{\mathrm{tot}} \,. \end{aligned} \tag{6.15}$$

Here, the scalar potential ϕ is determined by a *Poisson equation*, that is, there is no retardation and ϕ is an instantaneous function. The Coulomb gauge is mostly used for problems in quantum optics and is less important for this course. There are many more gauges, which we won't discuss here. Among them are the Poincaré gauge, the Landau gauge, and the Weyl gauge. We will be mostly dealing with the Lorenz gauge.

Note that by going from $\mathbf{E}$, $\mathbf{H}$ to $\mathbf{A}$, ϕ we reduced the field parameters from six to four (three per vector and one per scalar). It turns out that the four parameters are still redundant and that they can be reduced even more. One way is to introduce the so-called Hertz potential $\mathbf{\Pi}(\mathbf{r})$, which has only three components. The vector and scalar potentials are related to $\mathbf{\Pi}$ as $\mathbf{A} = (1/c^2)\,\partial\mathbf{\Pi}/\partial t$ and $\phi = -\nabla \cdot \mathbf{\Pi}$, respectively. Using so-called Debye potentials is yet another representation of fields, but these won't be discussed here.

6.3 Dipole Radiation

In this section we will derive the electromagnetic field of a dipole source, the smallest radiating system. Mathematically, the dipole source corresponds to a delta excitation, and the response to a delta excitation is the *Green function* discussed previously. Any source can be thought of as being composed of individual point sources with different origins. In other words, any macroscopic source volume can be chopped up into little infinitesimal cubes, each of which carries a current that is represented by a delta function.

As shown in Fig. 6.1, a dipole is a separation of a pair of charges by an infinitesimal distance $d\mathbf{s} = \mathbf{n}_s\, ds$. The *dipole moment* **p** is defined as

$$\mathbf{p}(t) = q(t)\, d\mathbf{s}\,. \tag{6.16}$$

The time derivative of the dipole moment is

$$\frac{\partial}{\partial t}\,\mathbf{p}(t) \;=\; \left[\frac{\partial\, q(t)}{\partial t}\,\mathbf{n}_s\right]\, ds \;=\; [\mathbf{j}\, da]\; ds \;=\; \mathbf{j}\, dV\,, \tag{6.17}$$

where $[\mathbf{j}\cdot\mathbf{n}_s]da$ is the current flowing through the cross-sectional area da. The product of da and ds defines the infinitesimal source volume dV.

Let us now consider an arbitrary macroscopic current density $\mathbf{j}(\mathbf{r})$ that is entirely contained within the volume V. We can express this current density in terms of a sum of microscopic point current densities. In terms of the Dirac delta function δ

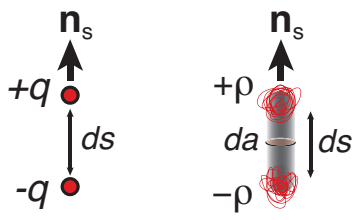

Figure 6.1: Illustration of a dipole with moment $\mathbf{p} = q\, d\mathbf{s} \;=\; q\, ds\, \mathbf{n}_s$. Left: in terms of discrete point charges q; Right: in terms of charge distributions ρ.

this sum becomes

$$\mathbf{j}(\mathbf{r}) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{j}(x',y',z')\;\delta(x-x')\,\delta(y-y')\,\delta(z-z')\;dx'\,dy'\,dz' \qquad (6.18)$$

$$\;=\; \int_V \mathbf{j}(\mathbf{r}')\;\delta(\mathbf{r}-\mathbf{r}')\;dV'\;, \qquad (6.19)$$

Here, $\mathbf{j}(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')dV'$ are elementary point currents. Using Eq. (6.17) we obtain the current density of an elementary dipole source

$$\mathbf{j}(\mathbf{r},t) \;=\; \frac{\partial}{\partial t}\,\mathbf{p}(t)\;\delta(\mathbf{r}-\mathbf{r}_0) \qquad (6.20)$$

where $\mathbf{r}_0$ is the dipole location.

In principle, we can now determine the fields $\mathbf{E}$ and $\mathbf{H}$ radiated by a dipole $\mathbf{p}$ in free space by inserting (6.20) into the wave equation (2.1). This task is, however, easier accomplished by replacing $\mathbf{E}$ and $\mathbf{H}$ by the potentials $\mathbf{A}$ and ϕ, as discussed previously.

6.3.1 Vector Potential of a Time-Harmonic Dipole

To calculate the fields of a dipole we will use the Lorenz gauge. The appeal of the Lorenz gauge is its symmetry, that is, there is a *scalar* wave equation of the form $[\nabla^2-(1/c^2)\partial^2/\partial t^2]\,\Psi=\Theta$ for any of the field components A_x, A_y, A_z, ϕ.

Let us start with deriving the vectorfield of a dipole with a harmonic time dependence. In this case, $\mathbf{p}(t) \;=\; \mathrm{Re}\{\mathbf{p}\,\exp[-i\omega t]\}$, with $\mathbf{p}$ being a complex amplitude. Thus, we can use complex notation and the equations for the components of $\mathbf{A}$ become

$$\left[\nabla^2 \;+\; k^2\right] A_i(\mathbf{r}) \;\;=\;\; i\omega\mu_0\,p_i\,\delta(\mathbf{r}-\mathbf{r}')\;. \qquad (6.21)$$

where we used Eq. (6.20) for the current density of a dipole field. Let us now define the *scalar Green function* as $G_0 = i\,Ai/\omega\mu_0\,p_i$. Then, Eq. (6.21) turns into

$$\left[\nabla^2 \;+\; k^2\right] G_0(\mathbf{r},\mathbf{r}') \;\;=\;\; -\delta(\mathbf{r}-\mathbf{r}') \qquad (6.22)$$

We have included the origin of the the point source ($\mathbf{r}'$) in the argument of G_0 to remind us where the origin of the point source is. In other words $G_0(\mathbf{r},\mathbf{r}')$ is the response at $\mathbf{r}$ to to a dipole source at $\mathbf{r}'$.

In free space, $G_0(\mathbf{r},\mathbf{r}')$ must be point symmetric, because the source depends only on the radial distance $R = |\mathbf{r}-\mathbf{r}'|$ and things don't change if the coordinate system is rotated around the origin $\mathbf{r} = \mathbf{r}'$. To solve Eq. (6.22) we will try the following ansatz

$$R\,G_0 \;=\; a_1\,\mathrm{e}^{ikR} \;+\; a_2\,\mathrm{e}^{-ikR}\;, \tag{6.23}$$

which is a superposition of an outgoing and an incoming wave. After inserting into Eq. (6.22) and integrating on both sides over a small spherical volume ΔV centered at $R = 0$ and with radius r_o. We obtain

$$\int_{\Delta V} \nabla^2 \frac{1}{R}\,dV \;+\; k^2 \int_{\Delta V} \frac{1}{R}\,dV \;=\; \frac{1}{a_1 + a_2} \tag{6.24}$$

The second term integrates to $2\pi k^2 r_o^2$ and the first term is calculated as

$$\int\limits_{\Delta V} \nabla\cdot\left[\nabla\frac{1}{R}\right] dV \;=\; \int\limits_{\partial\Delta V} \left[\nabla\frac{1}{R}\right]\cdot\mathbf{n}_R\,da \;=\; -\int\limits_{\partial\Delta V} \frac{\mathbf{n}_R\cdot\mathbf{n}_R}{R^2}\,da \;=\; -4\pi\;. \tag{6.25}$$

where we used Gauss' theorem (1.28). Thus, for $r_o \to 0$ we obtain $(a_1 + a_2) = 1\,/\,4\pi$. Finally, in free-space the radiation released by the point source is not coming back, which implies that we can drop the incoming wave in (6.23) or, equivalently, set $a_2 = 0$. The solution for the scalar Green function becomes

$$G_0(\mathbf{r},\mathbf{r}') \;=\; \frac{\mathrm{e}^{i\,k\,|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{6.26}$$

G_0 defines the vector potential at $\mathbf{r}$ due to a dipole $\mathbf{p}$ at $\mathbf{r}'$ according to

$$\mathbf{A}(\mathbf{r}) \;=\; -i\,\omega\mu_0\,\frac{\mathrm{e}^{i\,k\,|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}\,\mathbf{p}\;. \tag{6.27}$$

What if the source is not a dipole but an arbitrary current distribution? In this case we go back to Eq. (6.22) and multiply both sides with $\mu_0\,j_{\mathrm{tot}_i}(\mathbf{r}')$, where j_{tot_i} is the i-th vector component of the total current density $\mathbf{j}_{\mathrm{tot}}$. Integrating both sides over the source volume V yields

$$\begin{aligned}\mu_0\int_V \left[\nabla^2 + k^2\right] G_0(\mathbf{r},\mathbf{r}')\,j_{\mathrm{tot}_i}(\mathbf{r}')\,dV' \;&=\; -\mu_0\int_V \delta(\mathbf{r}-\mathbf{r}')\,j_{\mathrm{tot}_i}(\mathbf{r}')\,dV' \\ &=\; = -\mu_0\,j_{\mathrm{tot}_i}(\mathbf{r})\;,\end{aligned} \tag{6.28}$$

where we used the definition of the delta function. We now assume that the observation point $\mathbf{r}$ is outside the source volume described by the coordinate $\mathbf{r}'$. In this case, we can swap the sequence of integration and differentiation in Eq. (6.28) and obtain

$$\left[\nabla^2 + k^2\right] \mu_0 \int_V G_0(\mathbf{r}, \mathbf{r}')\, j_{\mathrm{tot}_i}(\mathbf{r}')\, dV' = -\mu_0\, j_{\mathrm{tot}_i}(\mathbf{r})\ . \tag{6.29}$$

Comparing this equation with Eq. (6.13) we conclude that

$$\mathbf{A}(\mathbf{r}) = \mu_0 \int_V G_0(\mathbf{r}, \mathbf{r}')\, \mathbf{j}_{\mathrm{tot}}(\mathbf{r}')\, dV' \tag{6.30}$$

Thus, the solution for $\mathbf{A}$ turns out to be the linear superposition of dipole fields with different origins $\mathbf{r}'$ and different weights $\mathbf{j}_{\mathrm{tot}}$.

6.3.2 Electric and Magnetic Dipole Fields

Now that we have derived the vector potential $\mathbf{A}$ of an oscillation dipole, we find the magnetic field using $\mathbf{B} = \nabla \times \mathbf{A}$ and the electric field using Maxwell's equation $\mathbf{E} = i(\omega/k^2)\nabla \times \mathbf{B}$. Skipping the details of the calculation, we find

$$\mathbf{E}(\mathbf{r}) = \omega^2 \mu_0\, \overleftrightarrow{\mathbf{G}}_0(\mathbf{r}, \mathbf{r}')\, \mathbf{p} \tag{6.31}$$

$$\mathbf{H}(\mathbf{r}) = -\mathrm{i}\omega \left[\nabla \times \overleftrightarrow{\mathbf{G}}_0(\mathbf{r}, \mathbf{r}')\right] \mathbf{p} \tag{6.32}$$

where we introduced the so-called *dyadic Green function* $\overleftrightarrow{\mathbf{G}}_0$ defined as

$$\overleftrightarrow{\mathbf{G}}_0(\mathbf{r}, \mathbf{r}') = \left[\overleftrightarrow{\mathbf{I}} + \frac{1}{k^2}\nabla\nabla\right] G_0(\mathbf{r}, \mathbf{r}') \tag{6.33}$$

with G_0 being the scalar Green function (6.26) and $\overleftrightarrow{\mathbf{I}}$ being the unit tensor. Notice that $\overleftrightarrow{\mathbf{G}}_0$ is a tensor. It is straightforward to calculate $\overleftrightarrow{\mathbf{G}}_0$ in the major three coordinate systems. In a Cartesian system $\overleftrightarrow{\mathbf{G}}_0$ can be written as

$$\overleftrightarrow{\mathbf{G}}_0(\mathbf{r}, \mathbf{r}') = \frac{\exp(\mathrm{i}kR)}{4\pi R}\left[\left(1 + \frac{\mathrm{i}kR - 1}{k^2R^2}\right)\overleftrightarrow{\mathbf{I}} + \frac{3 - 3\mathrm{i}kR - k^2R^2}{k^2R^2}\frac{\mathbf{RR}}{R^2}\right] \tag{6.34}$$

where R is the absolute value of the vector $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ and $\mathbf{RR}$ denotes the outer product of $\mathbf{R}$ with itself. Equation (6.34) defines a symmetric 3×3 matrix

$$\overleftrightarrow{\mathbf{G}}_0 = \begin{bmatrix} G_{xx} & G_{xy} & G_{xz} \\ G_{xy} & G_{yy} & G_{yz} \\ G_{xz} & G_{yz} & G_{zz} \end{bmatrix} , \tag{6.35}$$

which, together with Eqs. (6.31) and (6.32), determines the electromagnetic field of an arbitrary electric dipole $\mathbf{p}$ with Cartesian components p_x, p_y, p_z. The tensor $[\nabla\times\overleftrightarrow{\mathbf{G}}_0]$ can be expressed as

$$\nabla\times\overleftrightarrow{\mathbf{G}}_0\,(\mathbf{r},\mathbf{r}') \;=\; \frac{\exp(\mathrm{i}kR)}{4\pi R}\,\frac{k\left(\mathbf{R}\times\overleftrightarrow{\mathbf{I}}\right)}{R}\left(\mathrm{i}-\frac{1}{kR}\right) \;, \tag{6.36}$$

where $\mathbf{R}\times\overleftrightarrow{\mathbf{I}}$ denotes the matrix generated by the cross-product of $\mathbf{R}$ with each column vector of $\overleftrightarrow{\mathbf{I}}$.

Near-fields and Far-fields

The Green function $\overleftrightarrow{\mathbf{G}}_0$ has terms in $(kR)^{-1}$, $(kR)^{-2}$ and $(kR)^{-3}$. In the *far-field*, for which $R\gg\lambda$, only the terms with $(kR)^{-1}$ survive. On the other hand, the dominant terms in the *near-field*, for which $R\ll\lambda$, are the terms with $(kR)^{-3}$. The terms with $(kR)^{-2}$ dominate the *intermediate-field* at $R\approx\lambda$. To distinguish these three ranges it is convenient to write

$$\overleftrightarrow{\mathbf{G}}_0 = \overleftrightarrow{\mathbf{G}}_{\mathrm{NF}} + \overleftrightarrow{\mathbf{G}}_{\mathrm{IF}} + \overleftrightarrow{\mathbf{G}}_{\mathrm{FF}} \;, \tag{6.37}$$

where the near-field (G_{NF}), intermediate-field (G_{IF}) and far-field (G_{FF}) Green functions are given by

$$\overleftrightarrow{\mathbf{G}}_{\mathrm{NF}} \;=\; \frac{\exp(\mathrm{i}kR)}{4\pi R}\,\frac{1}{k^2R^2}\left[-\overleftrightarrow{\mathbf{I}} + 3\mathbf{RR}/R^2\right] \;, \tag{6.38}$$

$$\overleftrightarrow{\mathbf{G}}_{\mathrm{IF}} \;=\; \frac{\exp(\mathrm{i}kR)}{4\pi R}\,\frac{\mathrm{i}}{kR}\left[\overleftrightarrow{\mathbf{I}} - 3\mathbf{RR}/R^2\right] \;, \tag{6.39}$$

$$\overleftrightarrow{\mathbf{G}}_{\mathrm{FF}} \;=\; \frac{\exp(\mathrm{i}kR)}{4\pi R}\left[\overleftrightarrow{\mathbf{I}} - \mathbf{RR}/R^2\right] \;. \tag{6.40}$$

Notice that the intermediate-field is 90° out of phase with respect to the near- and far-field.

Because the dipole is located in a homogeneous environment, all three dipole orientations lead to fields that are identical upon suitable frame rotations. We therefore choose a coordinate system with origin at $\mathbf{r} = \mathbf{r}_0$ and a dipole orientation along the dipole axis, i.e. $\mathbf{p} = |\mathbf{p}|\mathbf{n}_z$ (see Fig. 6.2). It is most convenient to represent the dipole fields in spherical coordinates $\mathbf{r} = (r, \vartheta, \varphi)$ and in spherical vector components $\mathbf{E} = (E_r, E_\vartheta, E_\varphi)$. In this system the field components E_φ and H_r, H_ϑ are identical to zero and the only non-vanishing field components are

$$E_r = \frac{|\mathbf{p}|\cos\vartheta}{4\pi\varepsilon_0\varepsilon}\,\frac{\exp(\mathrm{i}kr)}{r}\,k^2\left[\frac{2}{k^2r^2} - \frac{2\mathrm{i}}{kr}\right], \tag{6.41}$$

$$E_\vartheta = \frac{|\mathbf{p}|\sin\vartheta}{4\pi\varepsilon_0\varepsilon}\,\frac{\exp(\mathrm{i}kr)}{r}\,k^2\left[\frac{1}{k^2r^2} - \frac{\mathrm{i}}{kr} - 1\right], \tag{6.42}$$

$$H_\varphi = \frac{|\mathbf{p}|\sin\vartheta}{4\pi\varepsilon_0\varepsilon}\,\frac{\exp(\mathrm{i}kr)}{r}\,k^2\left[-\frac{\mathrm{i}}{kr} - 1\right]\sqrt{\frac{\varepsilon_0\varepsilon}{\mu_0\mu}}\,. \tag{6.43}$$

The fact that E_r has no far-field term ensures that the far-field is purely transverse. Furthermore, since the magnetic field has no terms in $(kr)^{-3}$ the near-field is dominated by the electric field (see Fig. 6.3). This justifies a quasi-electrostatic consideration.

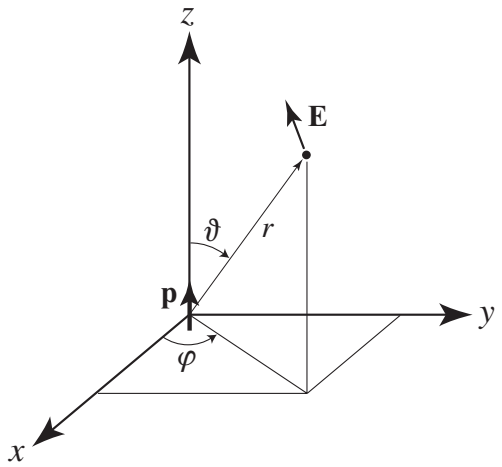

Figure 6.2: The fields of a dipole are most conveniently represented in a spherical coordinate system (r, ϑ, φ) in which the dipole points along the z-axis ($\vartheta = 0$).

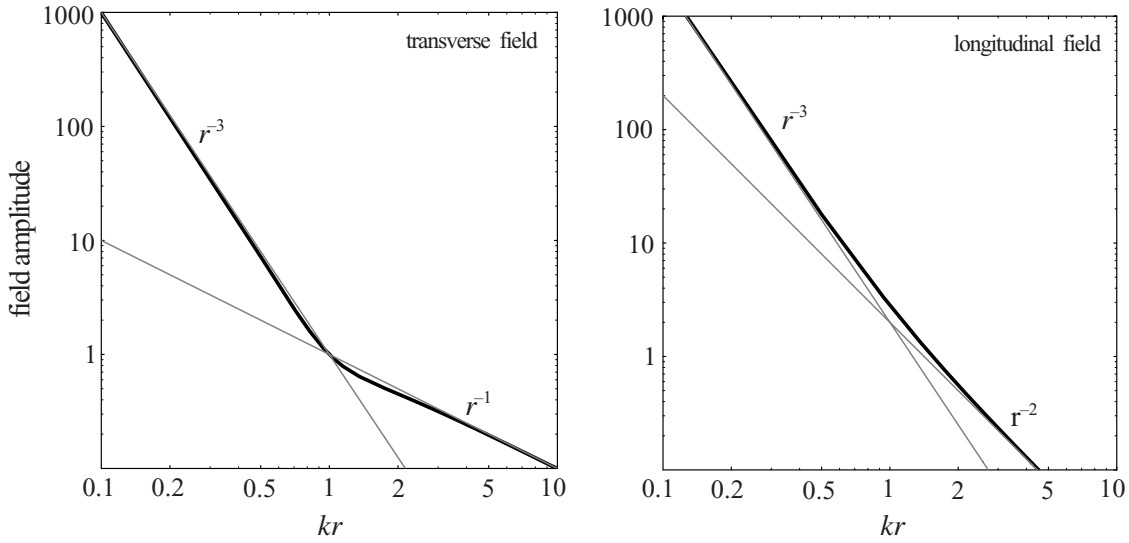

Figure 6.3: Radial decay of the dipole's transverse and longitudinal fields. The curves correspond to the absolute value of the expressions in brackets of Eqs. (6.41) and (6.42), respectively. While both the transverse and the longitudinal field contribute to the near-field, only the transverse field survives in the far-field. Notice that the intermediate-field with $(kr)^{-2}$ does not really show up for the transverse field. Instead the near-field dominates for $(kr) < 1$ and the far-field for $(kr) > 1$.

The Phase of the Dipole Field

It is instructive to also have a look at the phase of the dipole field since close to the origin it deviates considerably from the familiar phase of a spherical wave $\exp[ikr]$. The phase of the field is defined relative to the oscillation of the dipole p_z. In Fig. 6.4 we plot the phase of the field E_z along the x-axis and along the z-axis (c.f. Fig. 6.2). Interestingly, at the origin the phase of the transverse field is 180° out of phase with the dipole oscillation (Fig. 6.4(a)). The phase of the transverse field then drops to a minimum value at a distance of $x \sim \lambda/5$ after which it increases and then asymptotically approaches the phase of a spherical wave with origin at the dipole (dashed line). On the other hand, the phase of the longitudinal field, shown in Fig. 6.4(b), starts out to be the same as for the oscillating dipole, but it runs 90° out of phase for distances $z \gg \lambda$. The reason for this behavior is the missing far-field term in the longitudinal field (c.f. Eq. (6.41). The 90° phase shift is

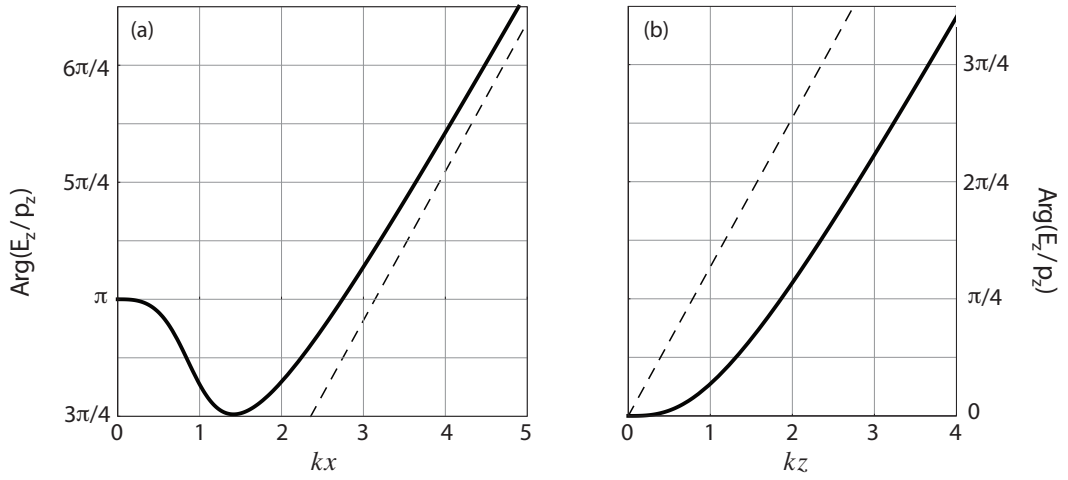

Figure 6.4: Phase of the electric field near the origin. (a) Phase of the transverse field E_z evaluated along the x-axis. At the origin, the electric field is 180° out of phase with the dipole. The phase drops to a minimum at a distance of $x \sim \lambda/5$. For larger distances, the phase approaches that of a spherical wave $\exp[ikr]$ (dashed line). (b) Phase of the longitudinal field E_z evaluated along the z-axis. At the origin, the electric field is in phase with the dipole. At larger distances, the phase is 90° out of phase with a spherical wave $\exp[ikr]$ (dashed line).

due to the intermediate field represented by the Green function in Eq. (6.39). The same intermediate field is also responsible for the dip near $x \sim \lambda/5$ in Fig. 6.4(a). This phase dip is of relevance for the design of multi-element antennas, such as the Yagi-Uda antennas. It is important to remember that close to the source the phase of the field *does not* evolve linearly with distance and that the phase can be advanced or delayed by small distance variations.

6.3.3 Radiation Patterns and Power Dissipation

To calculate the power radiated by the dipole $\mathbf{p}$ we consider a fictitious spherical surface ∂V of radius R_o centered at the origin of the dipole. According to Poynting's theorem discussed in Section 5.1, the net power $\bar{P}$ radiated corresponds to the flux of the time-averaged Poynting vector through the enclosing spherical surface (see

Eq. 5.9)

$$\bar{P} \;=\; \frac{1}{2}\int_{\partial V} \mathrm{Re}\left\{\mathbf{E}(\mathbf{r}) \times \mathbf{H}^{*}(\mathbf{r})\right\}\cdot\mathbf{n}\; da\,. \tag{6.44}$$

Because we chose a spherical surface, the normal vector n is a radial vector and hence we only need to calculate the radial component of $\langle \mathbf{S} \rangle$. Using Eqs. (6.42) and (6.43) we find

$$\bar{P} \;=\; \frac{1}{2}\int_{\partial V} \mathrm{Re}\left\{E_\vartheta\, H_\varphi^{*}\right\}\; \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi\;, \tag{6.45}$$

which yields

$$\bar{P} \;=\; \frac{|\mathbf{p}|^2}{4\pi\varepsilon_0\varepsilon}\,\frac{n^3\omega^4}{3\,c^3} \;=\; \frac{|\mathbf{p}|^2\omega\,k^3}{12\pi\varepsilon_0\varepsilon} \tag{6.46}$$

We find that the radiated power scales with the fourth power of the frequency and that only the far-field of the dipole contributes to the net energy transport.

To determine the radiation pattern we calculate the power $\bar{P}(\vartheta,\varphi)$ radiated into an infinitesimal unit solid angle $\mathrm{d}\Omega = \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$ and divide by the total radiated power $\bar{P}$

$$\frac{\bar{P}(\vartheta,\varphi)}{\bar{P}} \;=\; \frac{3}{8\pi}\sin^2\vartheta\;. \tag{6.47}$$

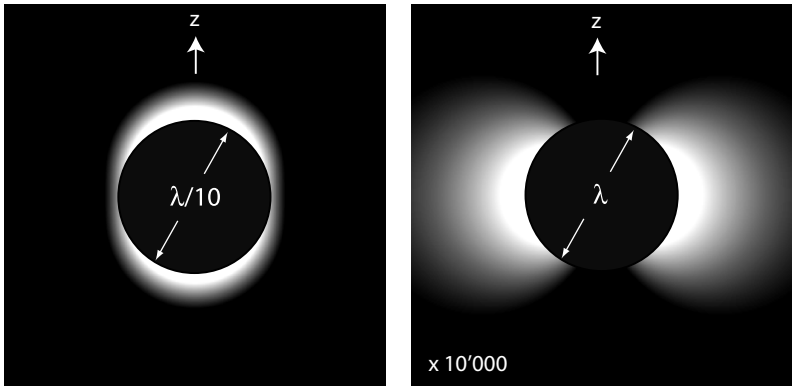

Figure 6.5: Electric energy density outside a fictitious sphere enclosing a dipole $\mathbf{p} = p_z$. (Left) Close to the dipole's origin the field distribution is elongated along the dipole axis (near-field). (Right) At larger distances the field spreads transverse to the dipole axis (far-field).

Thus, in the far-field most of the energy is radiated perpendicular to the dipole moment (see Fig. 6.47) and there is no radiation at all in the direction of the dipole.

6.4 Dipole Radiation in Arbitrary Environments

So far we have considered a dipole in a homogeneous space characterized by μ and ε. What happens if we place the dipole near a material boundary or enclose it in a box? Will the dipole still dissipate the same amount of power? The answer is no. The environment acts back on the dipole and influences its ability to radiate.

According to Poynting's theorem (cf. Eq. 5.6) the radiated power $\bar{P}$ of any current distribution with a harmonic time dependence has to be identical to the rate of energy dissipation $\mathrm{d}W/\mathrm{d}t$ given by

$$\frac{\mathrm{d}W}{\mathrm{d}t} \;=\; -\frac{1}{2}\int_V \mathrm{Re}\{\mathbf{j}^*\cdot\mathbf{E}\}\,\mathrm{d}V, \tag{6.48}$$

V being the source volume. The current density $\mathbf{j}$ is either a source current that generates the fields, or a loss current that is associated with thermal losses. Either way, $\mathbf{j}$ represents both energy sources and energy sinks. If we introduce the dipole's current density from Eq. (6.20) we obtain the important result

$$\bar{P} \;=\; \frac{\omega}{2}\,\mathrm{Im}\left\{\mathbf{p}^*\cdot\mathbf{E}(\mathbf{r}_0)\right\} \tag{6.49}$$

where the field $\mathbf{E}$ is evaluated at the dipole's origin $\mathbf{r}_0$. This equation can be rewritten in terms of the Green function by using Eq. (6.31) as

$$\bar{P} \;=\; \frac{\omega^3\,|\mathbf{p}|^2}{2\,c^2\varepsilon_0\varepsilon}\left[\mathbf{n}_\mathrm{p}\cdot\mathrm{Im}\left\{\overleftrightarrow{\mathbf{G}}\,(\mathbf{r}_0,\mathbf{r}_0)\right\}\cdot\mathbf{n}_\mathrm{p}\right]\quad, \tag{6.50}$$

with $\mathbf{n}_\mathrm{p}$ being the unit vector in the direction of the dipole moment.

At first sight it seems not possible to evaluate Eq. (6.49) since $\exp(\mathrm{i}kR)/R$ appears to be infinite at $\mathbf{r}=\mathbf{r}_0$. As we shall see this is not the case. We first note that due to the dot product between $\mathbf{p}$ and $\mathbf{E}$ we need only to evaluate the component of $\mathbf{E}$ in the direction of $\mathbf{p}$. Choosing $\mathbf{p}=|\mathbf{p}|\,\mathbf{n}_z$ we calculate E_z as

$$E_z \;=\; \frac{|\mathbf{p}|}{4\pi\,\varepsilon_0\,\varepsilon}\,\frac{\mathrm{e}^{\mathrm{i}kR}}{R}\left[k^2\,\sin^2\!\vartheta \;+\; \frac{1}{R^2}\,(3\cos^2\!\vartheta-1) \;-\; \frac{\mathrm{i}k}{R}\,(3\cos^2\!\vartheta-1)\right]. \tag{6.51}$$

Since the interesting part is the field at the origin of the dipole, the exponential term is expanded into a series [$\exp(\mathrm{i}kR) = 1 + \mathrm{i}kR + (1/2)(\mathrm{i}kR)^2 + (1/6)(\mathrm{i}kR)^3 + \cdots$] and the limiting case $R \to 0$ is considered. Thus,

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \lim_{R\to 0} \frac{\omega}{2} \, |\mathbf{p}| \, \mathrm{Im}\{E_z\} = \frac{\omega \, |\mathbf{p}|^2}{8\pi\,\varepsilon_0\,\varepsilon} \lim_{R\to 0} \left\{ \frac{2}{3}\,k^3 + R^2\,(..) + .. \right\} = \frac{|\mathbf{p}|^2}{12\,\pi} \, \frac{\omega}{\varepsilon_0\,\varepsilon}\, k^3, \tag{6.52}$$

which is identical with Eq. (6.46). Thus, Eq. (6.49) leads to the correct result despite the apparent singularity at $R = 0$.

The importance of Eq. (6.49) becomes obvious if we consider a dipole in an inhomogeneous environment, such as an antenna next to the earth surface. The rate at which energy is released can still be calculated by integrating the Poynting vector over a surface enclosing the dipole. However, to do this, we need to know the electromagnetic field everywhere on the enclosing surface. Because of the inhomogeneous environment, this field is not equal to the dipole field alone! Instead, it is the self-consistent field, that is, the field $\mathbf{E}$ generated by the superposition of the dipole field and the scattered field from the environment (see Fig. 6.6). Thus, to determine the energy dissipated by the dipole we first need to determine the electromagnetic field everywhere on the enclosing surface. However, by using Eq. (6.49) we can do the same job by only evaluating the total field at the dipole's origin $\mathbf{r}_0$.

As illustrated in Fig. 6.6, we decompose the electric field at the dipole's position

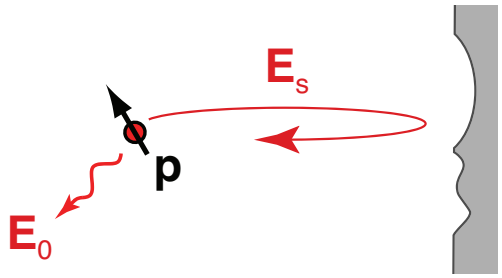

Figure 6.6: Illustration of dipole radiation in inhomogeneous environments. The total field is composed of a primary field $\mathbf{E}_0$ directly radiated by the dipole and a secondary field $\mathbf{E}_\mathrm{s}$ that is emitted by the dipole and then scattered at inhomogeneities in the environment.

as

$$\mathbf{E}(\mathbf{r}_0) \;=\; \mathbf{E}_0(\mathbf{r}_0) \;+\; \mathbf{E}_\mathrm{s}(\mathbf{r}_0)\;, \tag{6.53}$$

where $\mathbf{E}_0$ and $\mathbf{E}_\mathrm{s}$ are the primary dipole field and the scattered field, respectively. Introducing Eq. (6.53) into Eq. (6.49) allows us to split the rate of energy dissipation $P = \mathrm{d}W/\mathrm{d}t$ into two parts. The contribution of $\mathbf{E}_0$ has been determined in Eqs. (6.46) and (6.52) as

$$\bar{P}_0 \;=\; \frac{|\mathbf{p}|^2}{12\pi}\,\frac{\omega}{\varepsilon_0\varepsilon}\,k^3\;, \tag{6.54}$$

which allows us to write for the normalized rate of energy radiation

$$\frac{\bar{P}}{\bar{P}_0} \;=\; 1 \;+\; \frac{6\pi\varepsilon_0\varepsilon}{|\mathbf{p}|^2}\,\frac{1}{k^3}\,\mathrm{Im}\{\mathbf{p}^*\cdot\mathbf{E}_\mathrm{s}(\mathbf{r}_0)\} \tag{6.55}$$

Thus, the change of energy dissipation depends on the *secondary field* of the dipole. This field corresponds to the dipole's own field emitted at a former time. It arrives at the position of the dipole after it has been scattered in the environment.

6.5 Fields Emitted by Arbitrary Sources

In Section 6.3.1 we have derived the vector potential $\mathbf{A}$ of a time-harmonic dipole $\mathbf{p}$. Using the scalar free-space Green function G_0 we have then found a solution for the vector potential of an arbitrary current distribution (see Eq. 6.30). The same procedure can be applied to the electric field vector $\mathbf{E}$.

According to Eq. (6.31) the $\mathbf{E}$-field can be expressed in terms of a dyadic (tensorial) Green function as $\mathbf{E}(\mathbf{r}) = \omega^2\mu_0\,\overleftrightarrow{\mathbf{G}}_0\,(\mathbf{r},\mathbf{r}')\,\mathbf{p}$, where $\mathbf{r}'$ is the origin of the dipole. We can rewrite this equation as

$$\mathbf{E}(\mathbf{r}) \;=\; \omega^2\mu_0\int_V \overleftrightarrow{\mathbf{G}}_0\,(\mathbf{r},\mathbf{r}'')\,\mathbf{p}\,\delta(\mathbf{r}'-\mathbf{r}'')\,dV'' \tag{6.56}$$

Using Eq. (6.20) for the current density of a dipole [$\mathbf{j}(\mathbf{r}'') = -i\omega\,\mathbf{p}\,\delta(\mathbf{r}'-\mathbf{r}'')$] and substituting into Eq. (6.56) above, yields

$$\mathbf{E}(\mathbf{r}) \;=\; i\,\omega\mu_0\int_V \overleftrightarrow{\mathbf{G}}_0\,(\mathbf{r},\mathbf{r}')\,\mathbf{j}(\mathbf{r}')\,dV' \tag{6.57}$$

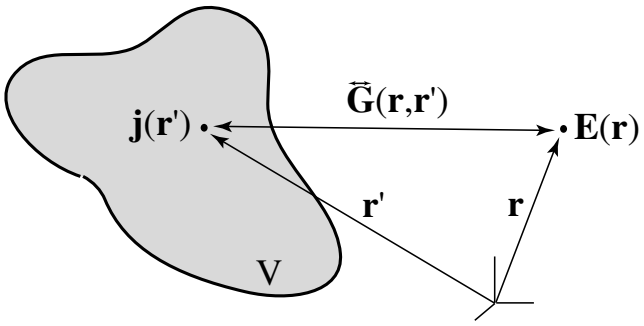

Figure 6.7: Illustration of the dyadic Green function $\overleftrightarrow{\mathbf{G}}(\mathbf{r},\mathbf{r}')$. The Green function renders the electric field at the field point $\mathbf{r}$ due to a single point source $\mathbf{j}$ at the source point $\mathbf{r}'$. Since the field at $\mathbf{r}$ depends on the orientation of $\mathbf{j}$ the Green function must account for all possible orientations in the form of a tensor.

where $\mathbf{j}$ now is an arbitrary current density distribution within the source volume V. We could have derived Eq. (6.57) also by following a more formal way using the definition of the Green function as described in Section 6.1. Fig. 6.7 illustrates the meaning of Eq. (6.57): The volume V is subdivided into infinitesimal units, each of which occupied by a point source with weight $\mathbf{j}(\mathbf{r}')$. In a similar way we find the solution for the magnetic field as

$$\mathbf{H}(\mathbf{r}) \;=\; \int_V \left[\nabla\times\, \overleftrightarrow{\mathbf{G}}_0\,(\mathbf{r},\mathbf{r}')\right]\,\mathbf{j}(\mathbf{r}')\,dV' \tag{6.58}$$

Note that mathematically, the fields $\mathbf{E}$ and $\mathbf{H}$ above are *particular* solutions of the differential equations (2.1) and (2.2). For a complete solution we need to superimpose the *homogeneous* solutions, which are solutions of (2.1) and (2.2) with the right sides being zero. These homogeneous solutions are fields that are present even in absence of the sources $\mathbf{j}$.

6.6 Sources with Arbitrary Time-Dependence

So far we have considered the fields generated by a source that is oscillating harmonically in time with angular frequency ω. But what if the time dependence is

arbitrary, for example, a short pulse? In these cases we can employ Fourier transforms, which describe an arbitrary time dependence by a superposition of time harmonic dependences (see Section 2.2).

Let us go back to the time-harmonic solution (6.30) of the vector potential $\mathbf{A}$. We have pointed out in Section 2.2 that Maxwell's equations for the Fourier transforms of he fields ($\hat{\mathbf{E}}$, $\hat{\mathbf{H}}$, ..) are formally the same as Maxwell's equations for the complex amplitudes ($\mathbf{E}(\mathbf{r})$, $\mathbf{H}(\mathbf{r})$, ..). Therefore, Eq, (6.30) implies that

$$\hat{\mathbf{A}}(\mathbf{r},\omega) = \mu_0 \int_V \hat{G}_0(\mathbf{r},\mathbf{r}',\omega)\, \hat{\mathbf{j}}_{\rm tot}(\mathbf{r}',\omega)\, dV' \tag{6.59}$$

where, according to Eq. (6.26), $\hat{G}_0 = \exp(i\,k(\omega)\,|\mathbf{r}-\mathbf{r}'|)\,/\,(4\pi|\mathbf{r}-\mathbf{r}'|)$, with $k(\omega) = n(\omega)\,\omega/c$. $\hat{\mathbf{j}}_{\rm tot}$ denotes the Fourier transform of an arbitrary time-dependent current density $\mathbf{j}(\mathbf{r},t)$, that is,

$$\hat{\mathbf{j}}_{\rm tot}(\mathbf{r},\omega) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mathbf{j}}_{\rm tot}(\mathbf{r},t)\, {\rm e}^{i\omega t}\, {\rm d}t \;. \tag{6.60}$$

The time-dependent vector potential $\mathbf{A}(\mathbf{r},t)$ of this current density is found by Fourier transforming Eq. (6.59), which yields

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{2\pi} \int_V G_0(\mathbf{r},\mathbf{r}',t) * \mathbf{j}_{\rm tot}(\mathbf{r}',t)\, dV' \;, \tag{6.61}$$

where $*$ denotes convolution in time and $G_0(\mathbf{r},\mathbf{r}',t)$ is given by

$$G_0(\mathbf{r},\mathbf{r}',t) \;=\; \int_{-\infty}^{\infty} \hat{G}_0(\mathbf{r},\mathbf{r}'\omega)\, {\rm e}^{-{\rm i}\omega t}\, {\rm d}\omega \;. \tag{6.62}$$

Inserting the expression for $\hat{G}_0$ yields

$$G_0(\mathbf{r},\mathbf{r}',t) \;=\; \int_{-\infty}^{\infty} \frac{{\rm e}^{i\,k(\omega)\,|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}\, {\rm e}^{-{\rm i}\omega t}\, {\rm d}\omega \;\;=\;\; \frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|} \int_{-\infty}^{\infty} {\rm e}^{-{\rm i}\omega[t-n(\omega)|\mathbf{r}-\mathbf{r}'|/c]}\, {\rm d}\omega \;. \tag{6.63}$$

In order to solve this integral we need to know the dependence of the index of refraction n on frequency ω, which is referred to as *dispersion*. We assume that $n(\omega) = n$ and obtain[2]

$$G_0(\mathbf{r},\mathbf{r}',t) \;=\; \frac{1}{2}\, \frac{\delta[t-|\mathbf{r}-\mathbf{r}'|n/c]}{|\mathbf{r}-\mathbf{r}'|} \;. \tag{6.64}$$

[2] $\int \exp[ixy]\,dy = 2\pi\delta[x]$.

Thus, the Green function in time domain is a simple delta function evaluated at the earlier time $t' = t - nR/c$, where t is the current time and R is the distance between source point $\mathbf{r}'$ and observation point $\mathbf{r}$.

We now insert $G_0(\mathbf{r}, \mathbf{r}', t)$ into Eq. (6.61) and obtain

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) &= \frac{\mu_0}{4\pi} \int_V \int_{t'} \frac{\delta[t' - |\mathbf{r} - \mathbf{r}'|n/c]}{|\mathbf{r} - \mathbf{r}'|} \mathbf{j}_{\text{tot}}(\mathbf{r}', t - t')\, dt'\, dV' \\ &= \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{j}_{\text{tot}}(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|n/c)}{|\mathbf{r} - \mathbf{r}'|}\, dV' \ . \end{aligned} \tag{6.65}$$

A similar equation can be derived for the scalar potential $\phi(\mathbf{r})$. Taken both together we have

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{j}_{\text{tot}}(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|n/c)}{|\mathbf{r} - \mathbf{r}'|}\, dV' \tag{6.66}$$

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon_0} \int_V \frac{\rho_{\text{tot}}(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|n/c)}{|\mathbf{r} - \mathbf{r}'|}\, dV' \tag{6.67}$$

These equations state that the fields $\mathbf{A}$ and ϕ at the location $\mathbf{r}$ and time t are determined by the sources $\mathbf{j}_{\text{tot}}$ and ρ_{tot} at location $\mathbf{r}'$ at the earlier time $t - |\mathbf{r} - \mathbf{r}'|n/c$. The earlier time is a consequence of the speed of light c: it takes a time of $|\mathbf{r} - \mathbf{r}'|n/c$ for the fields to travel a distance of $|\mathbf{r} - \mathbf{r}'|$ in a medium with index of refraction n. Thus, Maxwell's equations explain the mysterious "action-at-distance" phenomenon discussed in the introduction of this course (see Fig. 1). It has to be emphasized that the index of refraction n is assumed to be dispersion-free, which is an approximation. The only material that is truly dispersion-free is vacuum ($n = 1$).

To find the fields $\mathbf{E}$ and $\mathbf{H}$ we insert the solutions of $\mathbf{A}$ and ϕ into Eqs. (6.7) and (6.8). The calculation is not straightforward because $\mathbf{A}$ and ϕ depend on the retarded time $t - |\mathbf{r} - \mathbf{r}'|n/c$. Therefore, an operation on the spatial coordinates (e.g. $\nabla\times$) is implicitly also an operation on time. We will not go through this exercise and only mention that the solution is identical with Eq. (2) if we express ρ and $\mathbf{j}$ with the charge and current densities of a discrete charge (c.f. Eq. 3 and 4). The result has three terms: the first term depends on the charge, the second term on the velocity of the charge, and the third term on the acceleration of charge. It is the latter that is associated with electromagnetic radiation.

The expression of fields in terms of retarded time is of limited practical value. They help us to understand the physical origin of radiation but carrying out the integrals in Eqs. (6.66) and (6.67) is nearly impossible for realistic sources. Furthermore, the time-domain approach taken here is not able to accommodate dispersive materials. Therefore, it is generally more favorable to process the fields in Fourier space, that is, first calculate the spectrum of the source via Fourier transformation, then calculate the spectra of the fields, and finally taking the inverse Fourier transform to express the fields in time domain. This procedure is shown in Fig. 6.8.

6.6.1 Dipole Fields in Time Domain

We have calculated the fields of a dipole with harmonic time dependence in Section 6.3.2. These fields were expressed in spherical vector components (see Eq. 6.41 - 6.43). Note that these fields are the complex amplitudes and that the time-dependent fields are arrived at by multiplying with $\exp[-i\omega t]$ and taking the real part.

Remember, that Maxwell's equations for the complex amplitudes of time harmonic fields (Eq. 2.31 - 2.34) are identical with Maxwell's equations for the Fourier transforms of fields with arbitrary time dependence (Eq. 2.25 - 2.28) . Therefore, the solutions are identical as well. For example, the spectrum of the dipole's E_ϑ field is

$$\hat{E}_\vartheta(\mathbf{r},\omega) \;=\; \hat{p}(\omega)\,\frac{\sin\vartheta}{4\pi\varepsilon_0\varepsilon(\omega)}\,\frac{\exp[\mathrm{i}k(\omega)\,r]}{r}\,k^2(\omega)\left[\frac{1}{k^2(\omega)\,r^2}-\frac{\mathrm{i}}{k(\omega)\,r}-1\right]\,, \qquad (6.68)$$

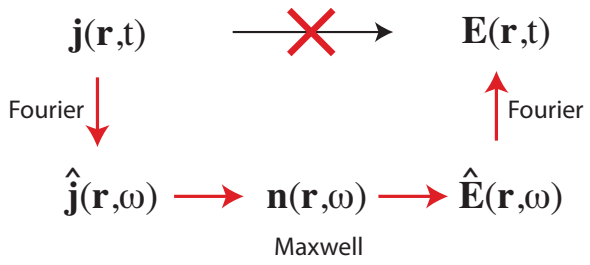

Figure 6.8: Calculating the field $\mathbf{E}$ of a time-dependent source $\mathbf{j}$. Because of dispersion and retardation it is favorable to solve for the fields in frequency space.

in analogy to the corresponding complex amplitude in Eq. (6.42). We used $\hat{p}(\omega) = |\hat{\mathbf{p}}(\omega)|$. Note that dispersion is fully accounted for through $\varepsilon(\omega)$ and $k(\omega)$. Using the time-dependent field $E_\vartheta(\mathbf{r}, t)$ can simply be calculated using the Fourier transform (2.23).

To illustrate the transformation from frequency to time domain, we assume that the dipole is in vacuum ($k = \omega/c$ and $\varepsilon = 1$). Also, we will only consider the far-field term in Eq. (6.68). The time-dependent far-field E^f_ϑ is calculated as

$$E^f_\vartheta(\mathbf{r}, t) \;=\; \int_\infty^\infty \hat{E}^f_\vartheta(\mathbf{r}, \omega)\, \mathrm{e}^{-i\omega t}\, d\omega \;=\; -\frac{\sin\vartheta}{4\pi\varepsilon_0}\frac{1}{c^2\, r}\int_\infty^\infty \omega^2\, \hat{p}(\omega)\, \mathrm{e}^{-i\omega(t-r/c)}\, d\omega\ . \qquad (6.69)$$

To solve this integral we set $\omega^2\, \hat{p}(\omega) = \hat{\psi}(\omega)$, where the Fourier transform of $\hat{\psi}(\omega)$ is $\psi(t)$. Using the translation property of Fourier-transforms we obtain

$$\int_\infty^\infty \hat{\psi}(\omega)\, \mathrm{e}^{-i\omega(t-r/c)}\, d\omega \;=\; \psi(t - r/c)\ . \qquad (6.70)$$

Thus, it remains to solve for $\psi(t)$:

$$\psi(t) \;=\; \int_\infty^\infty \omega^2\, \hat{p}(\omega)\, \mathrm{e}^{-i\omega t}\, d\omega \;=\; -\frac{d^2}{dt^2}\, p(t)\ . \qquad (6.71)$$

Putting the pieces together we finally find

$$E^f_\vartheta(\mathbf{r}, t) \;=\; \frac{\sin\vartheta}{4\pi\varepsilon_0}\frac{1}{c^2\, r}\ \left.\frac{d^2\, p(\tau)}{d\tau^2}\right|_{\tau=t-r/c} \qquad (6.72)$$

Thus, the field at $\mathbf{r} = [r, \vartheta, \varphi]$ and time t is determined by the dipole at $\mathbf{r}' = 0$ at the earlier time $t - r/c$. As before, we find that it takes a time r/c for the "action" to travel from the dipole to the observation point. Other terms of the dipole fields (Eqs. 6.41 - 6.43) can be calculated following the same procedure. The result is

$$E_r(t) \;=\; \frac{\cos\vartheta}{4\pi\varepsilon_0}\left[\frac{2}{r^3} + \frac{2}{c\, r^2}\frac{\mathrm{d}}{\mathrm{d}\tau}\right] p(\tau)\,\Big|_{\tau=t-r/c} \qquad (6.73)$$

$$E_\vartheta(t) \;=\; -\frac{\sin\vartheta}{4\pi\varepsilon_0}\left[\frac{1}{r^3} + \frac{1}{c\, r^2}\frac{\mathrm{d}}{\mathrm{d}\tau} + \frac{1}{c^2 r}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\right] p(\tau)\,\Big|_{\tau=t-r/c} \qquad (6.74)$$

$$H_\varphi(t) \;=\; -\frac{\sin\vartheta}{4\pi\varepsilon_0}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left[\frac{1}{c\, r^2}\frac{\mathrm{d}}{\mathrm{d}\tau} + \frac{1}{c^2 r}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\right] p(\tau)\,\Big|_{\tau=t-r/c} \qquad (6.75)$$

We see that the far-field is generated by the acceleration of the charges that constitute the dipole moment. Similarly, the intermediate-field and the near-field are generated by the speed and the position of the charges, respectively.

6.7 The Lorentzian Power Spectrum

The spectrum of various physical processes is characterized by narrow lines described by Lorentzian line shape functions. Examples are the spontaneous emission by atoms or molecules, laser radiation, or microwave resonators. To understand the origin of Lorentzian line shapes we consider a dipole located at $\mathbf{r}_0 = 0$ that starts to oscillate at time $t = 0$. The observer is assumed to be at large distance from the dipole, which allows us to restrict the discussion to the dipole's far-field $E^f_\vartheta(\mathbf{r}, t)$.

The equation of motion for an undriven harmonically oscillating dipole is

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbf{p}(t) \,+\, \gamma_0 \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}(t) \,+\, \omega_0^2 \mathbf{p}(t) \;=\; 0\,. \tag{6.76}$$

The natural frequency of the oscillator is ω_0 and its damping constant is γ_0. The solution for $\mathbf{p}$ is

$$\mathbf{p}(t) \;=\; \mathrm{Re}\left\{\mathbf{p}_0\, \mathrm{e}^{-\mathrm{i}\omega_0\sqrt{1-(\gamma_0^2/4\omega_0^2)}\,t}\, \mathrm{e}^{\gamma_0\, t/2}\right\}\,. \tag{6.77}$$

Typically, the damping constant is much smaller than the oscillation frequency ($\gamma_0 \ll \omega_0$), which implies $\sqrt{1-(\gamma_0^2/4\omega_0^2)} \approx 1$.

The spectrum $\hat{E}_\vartheta(\omega)$ detected by the observer is (cf. Eq. (2.24))

$$\hat{E}_\vartheta(\omega) \;=\; \frac{1}{2\pi}\int_{r/c}^{\infty} E_\vartheta(t)\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}t. \tag{6.78}$$

Here we set the lower integration limit to $t = r/c$ because the dipole starts emitting at $t = 0$ and it takes the time $t = r/c$ for the radiation to propagate to the observation point. Therefore $E_\vartheta(t < r/c) = 0$. Inserting the solution for the dipole moment from Eq. (6.77) and making use of $\gamma_0 \ll \omega_0$ we obtain after integration

$$\hat{E}_\vartheta(\omega) \;=\; \frac{1}{2\pi}\frac{|\mathbf{p}|\,\sin\vartheta\,\omega_0^2}{8\pi\varepsilon_0 c^2 r}\left[\frac{\exp(\mathrm{i}\omega r/c)}{\mathrm{i}(\omega+\omega_0)-\gamma_0/2} + \frac{\exp(\mathrm{i}\omega r/c)}{\mathrm{i}(\omega-\omega_0)-\gamma_0/2}\right]\,. \tag{6.79}$$

The energy radiated into the unit solid angle $\mathrm{d}\Omega = \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$ is calculated as

$$\frac{\mathrm{d}W}{\mathrm{d}\Omega} = \int_{-\infty}^{\infty} I(\mathbf{r}, t)\, r^2\, \mathrm{d}t \;=\; r^2\sqrt{\frac{\varepsilon_0}{\mu_0}}\int_{-\infty}^{\infty}|E_\vartheta(t)|^2\,\mathrm{d}t \;=\; 4\pi r^2\sqrt{\frac{\varepsilon_0}{\mu_0}}\int_0^{\infty}|\hat{E}_\vartheta(\omega)|^2\,\mathrm{d}\omega\,, \tag{6.80}$$

where we applied Parseval's theorem and used the definition of the intensity $I = \sqrt{\varepsilon_0/\mu_0}\,|E_\vartheta|^2$ of the emitted radiation. The total energy per unit solid angle $\mathrm{d}\Omega$ and per unit frequency interval $\mathrm{d}\omega$ can now be expressed as

$$\frac{\mathrm{d}W}{\mathrm{d}\Omega\,\mathrm{d}\omega} = \frac{1}{4\pi\varepsilon_0}\,\frac{|\mathbf{p}|^2\,\sin^2\vartheta\,\omega_0^2}{4\pi^2 c^3\gamma_0^2}\left[\frac{\gamma_0^2/4}{(\omega-\omega_0)^2+\gamma_0^2/4}\right] \tag{6.81}$$

The spectral shape of this function is determined by the expression in the brackets known as the *Lorentzian lineshape function*. The function is shown in Fig. 6.9. The width of the curve measured at half its maximum height is $\Delta\omega = \gamma_0$, and is called "radiative linewidth."

Integrating the lineshape function over the entire spectral range yields a value of $\pi\gamma_0/2$. Integrating Eq. (6.81) over all frequencies and all directions leads to the totally radiated energy

$$W = \frac{|\mathbf{p}|^2}{4\pi\varepsilon_0}\,\frac{\omega_0^4}{3\,c^3\,\gamma_0}\,. \tag{6.82}$$

This value is equal to the average power $\bar{P}$ radiated by a driven harmonic oscillator divided by the linewidth γ_0 (cf. Eq. 6.46).

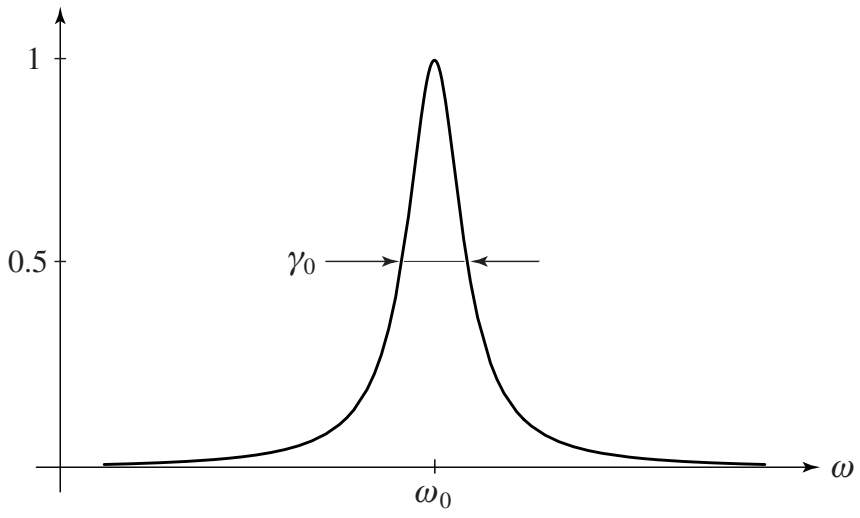

Figure 6.9: Lorentzian lineshape function as defined by the expression in brackets in Eq. (6.81).

Chapter 7

Angular Spectrum Representation

The angular spectrum representation is a mathematical technique to describe optical fields in homogeneous media. Optical fields are described as a superposition of plane waves and evanescent waves which are physically intuitive solutions of Maxwell's equations. The angular spectrum representation is found to be a very powerful method for the description of laser beam propagation and light focusing. Furthermore, in the paraxial limit, the angular spectrum representation becomes identical with the framework of Fourier optics which extends its importance even further.

In this chapter we will consider purely monochromatic fields of angular frequency ω, which can be represented as complex fields $\mathbf{E}(\mathbf{r})$ according to (c.f. Chapter 2)

$$\mathbf{E}(\mathbf{r},t) \;=\; \mathrm{Re}\{\mathbf{E}(\mathbf{r})\,\mathrm{e}^{-i\omega t}\}\;. \tag{7.1}$$

In situations where the field is not time-harmonic, we we simply replace the complex field $\mathbf{E}(\mathbf{r})$ by the spectrum $\hat{\mathbf{E}}(\mathbf{r},\omega)$ and obtain the time-dependent field by Fourier transformation (see Section 2.2).

By the angular spectrum representation we understand the series expansion of an arbitrary field in terms of plane (and evanescent) waves with variable amplitudes and propagation directions (see Section 2.1.3). Assume we know the electric field $\mathbf{E}(\mathbf{r})$ at any point $\mathbf{r} = (x, y, z)$ in space. For example, $\mathbf{E}(\mathbf{r})$ can be the solution of an optical scattering problem, as shown in Fig. 7.1, for which $\mathbf{E} = \mathbf{E}_{\mathrm{inc}} + \mathbf{E}_{\mathrm{scatt}}$. In the angular spectrum picture, we draw an arbitrary axis z and consider the field $\mathbf{E}$

in a plane $z = \text{const.}$ transverse to the chosen axis. In this plane we can evaluate the two-dimensional Fourier transform of the complex field $\mathbf{E}(\mathbf{r}) = \mathbf{E}(x, y, z)$ as

$$\hat{\mathbf{E}}(k_x, k_y;\, z) \;=\; \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}(x, y, z)\, \mathrm{e}^{-i\,[k_x x + k_y y]}\, \mathrm{d}x\, \mathrm{d}y\ , \tag{7.2}$$

where x, y are the Cartesian transverse coordinates and k_x, k_y the corresponding spatial frequencies or reciprocal coordinates. Similarly, the inverse Fourier transform reads as

$$\mathbf{E}(x, y, z) \;=\; \iint_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x, k_y;\, z)\, \mathrm{e}^{i\,[k_x x + k_y y]}\, \mathrm{d}k_x\, \mathrm{d}k_y\ . \tag{7.3}$$

Notice that in the notation of Eqs. (7.2) and (7.3) the field $\mathbf{E} = (E_x, E_y, E_z)$ and its Fourier transform $\hat{\mathbf{E}} = (\hat{E}_x, \hat{E}_y, \hat{E}_z)$ represent vectors. Thus, the Fourier integrals hold separately for each vector component.

So far we have made no requirements about the field $\mathbf{E}$, but we will assume that in the transverse plane the medium is homogeneous, isotropic, linear and source-free. Then, a time-harmonic, optical field with angular frequency ω has to satisfy

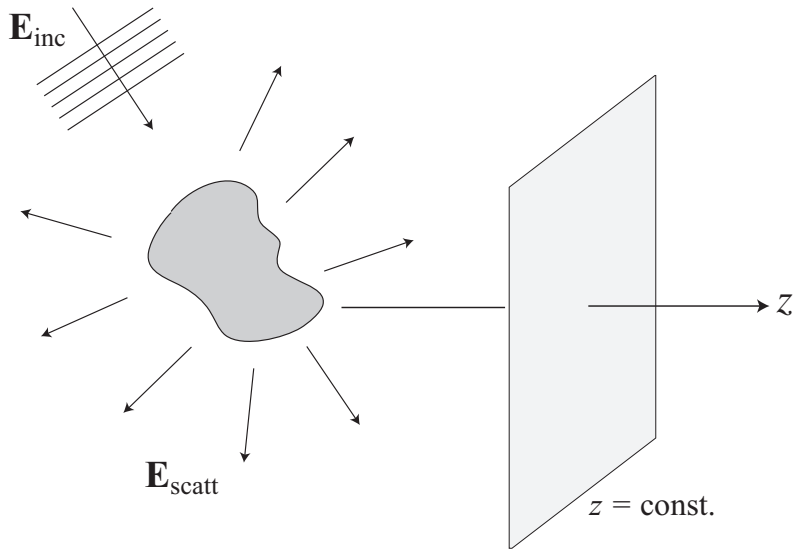

Figure 7.1: In the angular spectrum representation the fields are evaluated in planes ($z = \text{const.}$) perpendicular to an arbitrarily chosen axis z.

the vector Helmholtz equation (c.f. Eq. 2.12)

$$(\nabla^2 + k^2)\mathbf{E}(\mathbf{r}) \;=\; 0\;, \tag{7.4}$$

where k is determined by $k = (\omega/c)\,n$ and $n = \sqrt{\mu\varepsilon}$ is the index of refraction. Inserting the Fourier representation of $\mathbf{E}(\mathbf{r})$ (Eq. 7.3) into the Helmholtz equation and defining

$$k_z \;\equiv\; \sqrt{(k^2 - k_x^2 - k_y^2)} \qquad \text{with} \quad \mathrm{Im}\{k_z\} \geq 0\;, \tag{7.5}$$

we find that the Fourier spectrum $\hat{\mathbf{E}}$ evolves along the z-axis as

$$\hat{\mathbf{E}}(k_x, k_y;\, z) \;=\; \hat{\mathbf{E}}(k_x, k_y;\, 0)\; \mathrm{e}^{\pm i\, k_z\, z}\;. \tag{7.6}$$

The $\pm$ sign specifies that we have two solutions that need to be superimposed: the $+$ sign refers to a wave propagating into the half-space $z > 0$ whereas the $-$ sign denotes a wave propagating into $z < 0$. Equation (7.6) tells us that the Fourier spectrum of $\mathbf{E}$ in an arbitrary *image plane* located at $z = \mathrm{const.}$ can be calculated by multiplying the spectrum in the *object plane* at $z = 0$ by the factor $\exp(\pm i\, k_z z)$. This factor is called the *propagator* in reciprocal space. In Eq. (7.5) we defined that the square root leading to k_z renders a result with positive imaginary part. This ensures that the solutions remain finite for $z \to \pm\infty$. Inserting the result of Eq. (7.6) into Eq. (7.3) we finally find for arbitrary z

$$\mathbf{E}(x, y, z) \;=\; \iint_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x, k_y;\, 0)\; \mathrm{e}^{\,i\,[k_x x + k_y y \pm k_z z]}\; \mathrm{d}k_x\, \mathrm{d}k_y \tag{7.7}$$

which is known as the *angular spectrum representation*. In a similar way, we can also represent the magnetic field $\mathbf{H}$ by an angular spectrum as

$$\mathbf{H}(x, y, z) \;=\; \iint_{-\infty}^{\infty} \hat{\mathbf{H}}(k_x, k_y;\, 0)\; \mathrm{e}^{\,i\,[k_x x + k_y y \pm k_z z]}\; \mathrm{d}k_x\, \mathrm{d}k_y\;. \tag{7.8}$$

By using Maxwell's equation $\mathbf{H} = (i\omega\mu\mu_0)^{-1}\,(\nabla \times \mathbf{E})$ we find the following relationship between the Fourier spectra $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$

$$\begin{aligned}
\hat{H}_x &= Z_{\mu\varepsilon}^{-1}\,[(k_y/k)\,\hat{E}_z - (k_z/k)\,\hat{E}_y]\;, \\
\hat{H}_y &= Z_{\mu\varepsilon}^{-1}\,[(k_z/k)\,\hat{E}_x - (k_x/k)\,\hat{E}_z]\;, \\
\hat{H}_z &= Z_{\mu\varepsilon}^{-1}\,[(k_x/k)\,\hat{E}_y - (k_y/k)\,\hat{E}_x]\;,
\end{aligned} \tag{7.9}$$

where $Z_{\mu\varepsilon} = \sqrt{(\mu_0\mu)/(\varepsilon_0\varepsilon)}$ is the wave impedance of the medium. Although the angular spectra of $\mathbf{E}$ and $\mathbf{H}$ fulfill Helmholtz equation they are not yet rigorous solutions of Maxwell's equations. We still have to require that the fields are divergence free, i.e. $\nabla\cdot\mathbf{E} = 0$ and $\nabla\cdot\mathbf{H} = 0$. These conditions restrict the k-vector to directions perpendicular to the spectral amplitudes ($\mathbf{k}\cdot\hat{\mathbf{E}} = \mathbf{k}\cdot\hat{\mathbf{H}} = 0$).

For the case of a purely dielectric medium with no losses the index of refraction n is a real and positive quantity. The wavenumber k_z is then either real or imaginary and turns the factor $\exp(\pm i\, k_z\, z)$ into an oscillatory or exponentially decaying function. For a certain $(k_x,\, k_y)$ pair we then find two different characteristic solutions: plane waves with $k_x^2 + k_y^2 \leq k^2$ and evanescent waves with $k_x^2 + k_y^2 > k^2$ (see Section 2.1.2).

7.1 Propagation and Focusing of Fields

We have established that, in a homogeneous space, the spatial spectrum $\hat{\mathbf{E}}$ of an optical field $\mathbf{E}$ in a plane $z =$ const. (image plane) is uniquely defined by the spatial spectrum in a different plane $z = 0$ (object plane) according to the linear relationship

$$\hat{\mathbf{E}}(k_x, k_y;\, z) \;=\; \hat{H}(k_x, k_y;\, z)\, \hat{\mathbf{E}}(k_x, k_y;\, 0)\;, \tag{7.10}$$

where $\hat{H}$ is the so-called *propagator in reciprocal space*

$$\hat{H}(k_x, k_y;\, z) \;=\; \mathrm{e}^{\pm \mathrm{i}\, k_z\, z} \tag{7.11}$$

also referred to as the *optical transfer function* (OTF) in free space. Remember that the longitudinal wavenumber is a function of the transverse wavenumber, i.e. $k_z = [k^2 - (k_x^2 + k_y^2)]^{1/2}$, where $k = n\, k_0 = n\, \omega/c = n\, 2\pi/\lambda$. The $\pm$ sign indicates that the field can propagate in positive and/or negative z direction. Equation (7.10) can be interpreted in terms of linear response theory: $\hat{\mathbf{E}}(k_x, k_y;\, 0)$ is the input, $\hat{H}$ is a filter function, and $\hat{\mathbf{E}}(k_x, k_y;\, z)$ is the output. The filter function describes the propagation of an arbitrary spectrum through space. $\hat{H}$ can also be regarded as the response function because it describes the field at z due to a point source at $z = 0$. In this sense, it is directly related to the Green's function $\overleftrightarrow{\mathbf{G}}_0$.

The filter $\hat{H}$ is an oscillating function for $(k_x^2 + k_y^2) < k^2$ and an exponentially decreasing function for $(k_x^2 + k_y^2) > k^2$. Thus, if the image plane is sufficiently separated from the object plane, the contribution of the decaying parts (evanescent waves) is zero and the integration can be reduced to the circular area $(k_x^2 + k_y^2) \leq k^2$. In other words, the image at z is a *low pass filtered* representation of the original field at $z = 0$. The spatial frequencies $(k_x^2 + k_y^2) > k^2$ of the original field are filtered out during propagation and the information on high spatial variations gets lost. Hence, there is always a loss of information on propagating from near- to far-field and only structures with lateral dimensions larger than

$$\Delta x \;\approx\; \frac{1}{k} = \frac{\lambda}{2\pi n} \tag{7.12}$$

can be imaged with sufficient accuracy. Here, n is the index of refraction. This equation is qualitative and we will provide a more detailed discussion later. In general, higher resolution can be obtained by a higher index of refraction of the embodying system (substrate, lenses, etc.) or by shorter wavelengths. Theoretically, resolutions down to a few nanometers can be achieved by using far-ultraviolet radiation or X-rays.

Let us now determine how the fields themselves evolve. For this purpose we denote the transverse coordinates in the object plane at $z = 0$ as (x', y') and in the image plane at $z =$ const. as (x, y). The fields in the image plane are described by the angular spectrum (7.7). We just have to express the Fourier spectrum $\hat{\mathbf{E}}(k_x, k_y; 0)$ in terms of the fields in the object plane. Similarly to Eq. (7.2) this Fourier spectrum can be represented as

$$\hat{\mathbf{E}}(k_x, k_y; 0) \;=\; \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}(x', y', 0)\, \mathrm{e}^{-\mathrm{i}\,[k_x x' + k_y y']}\, \mathrm{d}x'\, \mathrm{d}y' \;. \tag{7.13}$$

After inserting into Eq. (7.7) we find the following expression for the field $\mathbf{E}$ in the image plane $z =$ const.

$$\begin{aligned} \mathbf{E}(x, y, z) &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}(x', y';\, 0) \iint_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,[k_x (x - x') + k_y (y - y') \pm k_z z]}\, \mathrm{d}x'\, \mathrm{d}y' \mathrm{d}k_x\, \mathrm{d}k_y \\ &= \mathbf{E}(x, y; 0) \;*\; H(x, y; z) \;. \end{aligned} \tag{7.14}$$

This equation describes an invariant filter with the following impulse response

(*propagator in direct space*)

$$H(x,\,y;\,z) \;=\; \iint_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,[k_x x + k_y y \pm k_z z]}\;\mathrm{d}k_x\,\mathrm{d}k_y\;. \tag{7.15}$$

H is simply the inverse Fourier transform of the propagator in reciprocal space $\hat{H}$ (7.11). The field at $z =$ const. is represented by the convolution of H with the field at $z = 0$.

7.1.1 Paraxial Approximation

In many optical problems the light fields propagate along a certain direction z and spread out only slowly in the transverse direction. Examples are laser beam propagation or optical waveguide applications. In these examples the wavevectors $\mathbf{k} = (k_x, k_y, k_z)$ in the angular spectrum representation are almost parallel to the z-axis and the transverse wavenumbers (k_x, k_y) are small compared to k. We can then expand the square root of Eq. (7.5) in a series as

$$k_z \;=\; k\,\sqrt{1-(k_x^2+k_y^2)/k^2} \;\approx\; k \;-\; \frac{(k_x^2+k_y^2)}{2\,k}\;. \tag{7.16}$$

This approximation is called the paraxial approximation and it considerably simplifies the analytical integration of the Fourier integrals. In the following we shall apply the paraxial approximation to find a description for weakly focused laser beams.

7.1.2 Gaussian Beams

We consider a fundamental laser beam with a linearly polarized, Gaussian field distribution in the beam waist

$$\mathbf{E}(x',\,y',\,0) \;=\; \mathbf{E}_o\,\mathrm{e}^{-\frac{x'^2+y'^2}{w_0^2}}\;, \tag{7.17}$$

where $\mathbf{E}_0$ is a constant field vector in the transverse $(x,\,y)$ plane. We have chosen $z = 0$ at the beam waist. The parameter w_0 denotes the beam waist radius. We

can calculate the spatial Fourier spectrum at $z = 0$ as[1]

$$\begin{aligned}\hat{\mathbf{E}}(k_x, k_y; 0) &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}_0 \, \mathrm{e}^{-\frac{x'^2+y'^2}{w_0^2}} \, \mathrm{e}^{-\mathrm{i}\,[k_x x' + k_y y']} \, \mathrm{d}x' \, \mathrm{d}y' \\ &= \mathbf{E}_0 \, \frac{w_0^2}{4\pi} \, \mathrm{e}^{-(k_x^2+k_y^2)\frac{w_0^2}{4}}, \end{aligned} \tag{7.18}$$

which is again a Gaussian function. We now insert this spectrum into the angular spectrum representation Eq. (7.7) and replace k_z by its paraxial expression in Eq. (7.16)

$$\mathbf{E}(x, y, z) \;=\; \mathbf{E}_0 \, \frac{w_0^2}{4\,\pi} \, \mathrm{e}^{\mathrm{i}\,kz} \iint_{-\infty}^{\infty} \mathrm{e}^{-(k_x^2+k_y^2)(\frac{w_0^2}{4}+\frac{\mathrm{i}\,z}{2\,k})} \, \mathrm{e}^{\mathrm{i}\,[k_x x + k_y y]} \, \mathrm{d}k_x \, \mathrm{d}k_y \;, \tag{7.19}$$

This equation can be integrated and gives as a result the familiar paraxial representation of a Gaussian beam

$$\mathbf{E}(x, y, z) \;=\; \frac{\mathbf{E}_0 \, \mathrm{e}^{\mathrm{i}\,kz}}{(1 \,+\, 2\,\mathrm{i}\,z/kw_0^2)} \, \mathrm{e}^{-\frac{(x^2+y^2)}{w_0^2} \frac{1}{(1\,+\,2\,\mathrm{i}\,z/kw_0^2)}} \;. \tag{7.20}$$

To get a better feeling for a paraxial Gaussian beam we set $\rho^2 = x^2 + y^2$, define a new parameter z_0 as

$$z_0 \;=\; \frac{k\,w_0^2}{2} \;, \tag{7.21}$$

and rewrite Eq. (7.20) as

$$\mathbf{E}(\rho, z) \;=\; \mathbf{E}_0 \, \frac{w_0}{w(z)} \, \mathrm{e}^{-\frac{\rho^2}{w^2(z)}} \, \mathrm{e}^{\mathrm{i}\,[kz - \eta(z) + k\rho^2/2R(z)]} \tag{7.22}$$

with the following abbreviations

$$\begin{aligned} w(z) &= w_0(1 + z^2/z_0^2)^{1/2} && \text{beam radius} \\ R(z) &= z(1 + z_0^2/z^2) && \text{wavefront radius} \\ \eta(z) &= \arctan z/z_0 && \text{phase correction} \end{aligned} \tag{7.23}$$

The transverse size of the beam is usually defined by the value of $\rho = \sqrt{x^2 + y^2}$ for which the electric field amplitude has decreased to a value of $1/\mathrm{e}$ of its center value

$$|\mathbf{E}(x, y, z)| \,/\, |\mathbf{E}(0, 0, z)| \;=\; 1/\mathrm{e} \;. \tag{7.24}$$

It can be shown that the surface defined by this equation is a hyperboloid whose asymptotes enclose an angle

$$\theta = \frac{2}{k\,w_0} \tag{7.25}$$

with the z-axis. From this equation we can directly find the correspondence between the numerical aperture (NA $= n \sin\theta$) and the beam angle as NA $\approx 2n/kw_0$. Here we used the fact that in the paraxial approximation, θ is restricted to small beam angles. Another property of the paraxial Gaussian beam is that close to the focus, the beam stays roughly collimated over a distance $2z_0$. z_0 is called the *Rayleigh range* and denotes the distance from the beam waist to where the beam radius has increased by a factor of $\sqrt{2}$. It is important to notice that along the z-axis ($\rho = 0$) the phases of the beam deviate from those of a plane wave. If at $z \to -\infty$ the beam was in phase with a reference plane wave, then at $z \to +\infty$ the beam will be exactly out of phase with the reference wave. This phase shift is called *Gouy phase shift*. The $180°$ phase change happens gradually as the beam propagates through its focus. The phase variation is described by the factor $\eta(z)$ in Eq. (7.23). The tighter the focus the faster the phase variation will be.

A qualitative picture of a paraxial Gaussian beam and some of its characteristics are shown in Fig. 7.2. It has to be emphasized that once the paraxial approx-

[1] $\int_{-\infty}^{\infty} \exp(-ax^2 + \mathrm{i}bx)\,\mathrm{d}x = \sqrt{\pi/a}\,\exp(-b^2/4a)$ and $\int_{-\infty}^{\infty} x\exp(-ax^2 + \mathrm{i}bx)\,\mathrm{d}x = \mathrm{i}b\sqrt{\pi}\exp(-b^2/4a)/\,(2a^{3/2})$

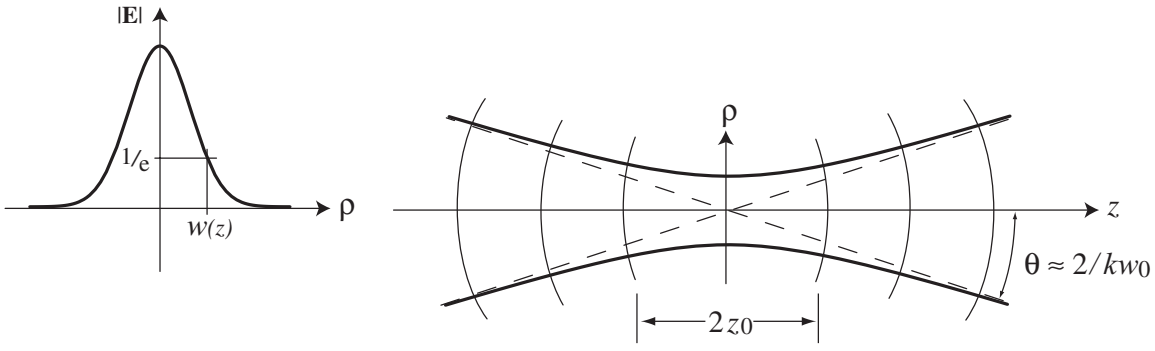

Figure 7.2: Illustration and main characteristics of a paraxial Gaussian beam. The beam has a Gaussian field distribution in the transverse plane. The surfaces of constant field strength form a hyperboloid along the z-axis.

imation is introduced, the field $\mathbf{E}$ fulfills no longer Maxwell's equations. The error becomes larger the smaller the beam waist radius w_0 is. Another important aspect of Gaussian beams is that they do not exist, no matter how rigorous the theory that describes them! The reason is that a Gaussian beam profile demands a Gaussian spectrum. However, the Gaussian spectrum is infinite and contains evanescent components that are not available in a realistic situation. Thus, a Gaussian beam must always be regarded as an approximation. The tighter the focus, the broader the Gaussian spectrum and the more contradictory the Gaussian beam profile will be. The angular spectrum representation can be used to derive a rigorous description of focussed fields (e.g. Novotny, Principles of Nano-Optics).

7.2 Far-field Approximation

In this section we will derive the important result that Fourier Optics and Geometrical Optics naturally emerge from the angular spectrum representation.

Consider a particular (localized) field distribution in the plane $z = 0$. The angular spectrum representation tells us how this field propagates and how it is mapped onto other planes $z = z_0$. Here, we ask what the field will be in a very remote plane. Vice versa, we can ask what field will result when we focus a particular far-field onto an image plane. Let us start with the familiar angular spectrum representation of an optical field

$$\mathbf{E}(x, y, z) \;=\; \iint_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x, k_y;\, 0)\; \mathrm{e}^{\mathrm{i}\,[k_x x + k_y y \pm k_z z]}\; \mathrm{d}k_x\, \mathrm{d}k_y \;. \tag{7.26}$$

We are interested in the asymptotic far-zone approximation of this field, i.e. in the evaluation of the field in a point $\mathbf{r} = \mathbf{r}_\infty$ at infinite distance from the object plane. The dimensionless unit vector $\mathbf{s}$ in the direction of $\mathbf{r}_\infty$ is given by

$$\mathbf{s} = (s_x, s_y, s_z) = \left(\frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right) \;, \tag{7.27}$$

where $r = (x^2 + y^2 + z^2)^{1/2}$ is the distance of $\mathbf{r}_\infty$ from the origin. To calculate the far-field $\mathbf{E}_\infty$ we require that $r \to \infty$ and rewrite Eq. (7.26) as

$$\mathbf{E}_\infty(s_x, s_y) \;=\; \lim_{kr\to\infty} \iint\limits_{(k_x^2+k_y^2)\le k^2} \hat{\mathbf{E}}(k_x, k_y;\, 0)\; \mathrm{e}^{\mathrm{i}\,kr\,[\frac{k_x}{k}\,s_x + \frac{k_y}{k}\,s_y \pm \frac{k_z}{k}\,s_z]}\; \mathrm{d}k_x\, \mathrm{d}k_y \;, \tag{7.28}$$

where $s_z = \sqrt{1-(s_x^2+s_y^2)}$. Because of their exponential decay, evanescent waves do not contribute to the fields at infinity. We therefore rejected their contribution and reduced the integration range to $(k_x^2+k_y^2) \leq k^2$. The asymptotic behavior of the double integral as $kr \to \infty$ can be evaluated by the method of *stationary phase*. A clear outline of this method can be found in other textbooks (e.g. Born & Wolf, Principles of Optics). Without going into details, the result of Eq. (7.28) can be expressed as

$$\mathbf{E}_\infty(s_x, s_y) \;=\; -2\pi \mathrm{i} k\, s_z\, \hat{\mathbf{E}}(ks_x, ks_y;\, 0)\, \frac{\mathrm{e}^{\mathrm{i}kr}}{r}\;. \tag{7.29}$$

This equation tells us that the far-fields are entirely defined by the Fourier spectrum of the fields $\hat{\mathbf{E}}(k_x, k_y;\, 0)$ in the object plane if we replace $k_x \to ks_x$ and $k_y \to ks_y$. This simply means that the unit vector $\mathbf{s}$ fulfills

$$\mathbf{s} = (s_x, s_y, s_z) = \left(\frac{k_x}{k}, \frac{k_y}{k}, \frac{k_z}{k}\right)\;, \tag{7.30}$$

which implies that only one plane wave with the wavevector $\mathbf{k} = (k_x, k_y, k_z)$ of the angular spectrum at $z = 0$ contributes to the far-field at a point located in the direction of the unit vector $\mathbf{s}$ (see Fig. 7.3). The effect of all other plane waves is cancelled by destructive interference. This beautiful result allows us to treat the field in the far-zone as a collection of rays with each ray being characterized by a particular plane wave of the original angular spectrum representation (Geometrical optics). Combining Eqs. (7.29) and (7.30) we can express the Fourier spectrum $\hat{\mathbf{E}}$ in terms of the far-field as

$$\hat{\mathbf{E}}(k_x, k_y;\, 0) \;=\; \frac{\mathrm{i}r\,\mathrm{e}^{-\mathrm{i}kr}}{2\pi k_z}\, \mathbf{E}_\infty(\frac{k_x}{k}, \frac{k_y}{k})\,, \tag{7.31}$$

keeping in mind that the vector $\mathbf{s}$ is entirely defined by k_x, k_y. This expression can be substituted into the angular spectrum representation (Eq. 7.26) as

$$\mathbf{E}(x,y,z) \;=\; \frac{\mathrm{i}r\,\mathrm{e}^{-\mathrm{i}kr}}{2\pi} \iint\limits_{(k_x^2+k_y^2)\leq k^2} \mathbf{E}_\infty(\frac{k_x}{k}, \frac{k_y}{k})\; \mathrm{e}^{\mathrm{i}\,[k_x x + k_y y \pm k_z z]}\; \frac{1}{k_z}\, \mathrm{d}k_x\, \mathrm{d}k_y \tag{7.32}$$

Thus, as long as evanescent fields are not part of our system then the field $\mathbf{E}$ and its far-field $\mathbf{E}_\infty$ form essentially a Fourier transform pair at $z = 0$. The only deviation is given by the factor $1/k_z$. In the approximation $k_z \approx k$, the two fields form a perfect Fourier transform pair. This is the limit of *Fourier Optics*.

As an example consider the diffraction at a rectangular aperture with sides $2L_x$ and $2L_y$ in an infinitely thin conducting screen which we choose to be our object plane ($z=0$). A plane wave illuminates the aperture at normal incidence from the back. For simplicity we assume that the field in the object plane has a constant field amplitude $\mathbf{E}_0$ whereas the screen blocks all the field outside of the aperture. The Fourier spectrum at $z=0$ is then

$$\begin{aligned}\hat{\mathbf{E}}(k_x, k_y; 0) &= \frac{\mathbf{E}_0}{4\pi^2}\int_{-L_y}^{+L_y}\int_{-L_x}^{+L_x} \mathrm{e}^{-\mathrm{i}\,[k_x x' + k_y y']}\,\mathrm{d}x'\,\mathrm{d}y' \\ &= \mathbf{E}_0\,\frac{L_x L_y}{\pi^2}\,\frac{\sin(k_x\,L_x)}{k_x\,L_x}\,\frac{\sin(k_y\,L_y)}{k_y\,L_y}\,, \end{aligned} \tag{7.33}$$

With Eq. (7.29) we now determine the far-field as

$$\mathbf{E}_\infty(s_x, s_y) \;=\; -\mathrm{i}ks_z\mathbf{E}_0\,\frac{2L_xL_y}{\pi}\,\frac{\sin(ks_x\,L_x)}{ks_x\,L_x}\,\frac{\sin(ks_y\,L_y)}{ks_y\,L_y}\,\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\,, \tag{7.34}$$

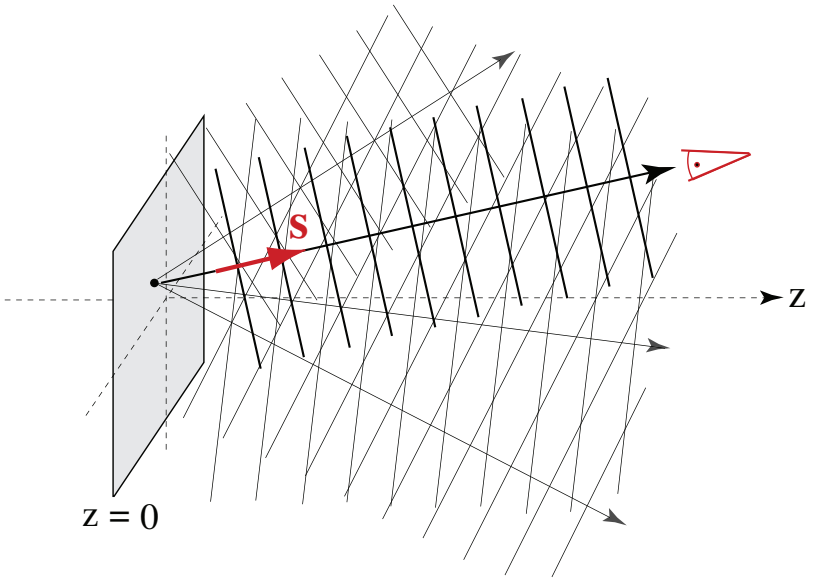

Figure 7.3: Illustration of the far-field approximation. According to the angular spectrum representation, a point in the source plane $z=0$ emits plane waves in all possible directions. However, a distant detector ($kr \gg 1$) measures only the plane wave that propagates towards it (in direction of unit vector s). The fields of all other plane waves are canceled by destructive interference.

which, in the paraxial limit $k_z \approx k$, agrees with Fraunhofer diffraction.

Equation (7.29) is an important result. It links the near-fields of an object with the corresponding far-fields. While in the near-field a rigorous description of fields is necessary, the far-fields are well approximated by the laws of Geometrical Optics.

7.3 Fresnel and Fraunhofer Diffraction

Diffraction refers to the observation that light rays break away from their geometrical paths, which is to say, that the wave nature of radiation becomes relevant. In this section we will discuss two important regimes of diffraction theory, Fresnel and Fraunhofer diffraction.

The far-field approximation derived in the previous chapter has its limitations. It has been assumed that the observation point is at infinite distance from the source plane. However, how far do we have to go to be approximately at infinity? The best

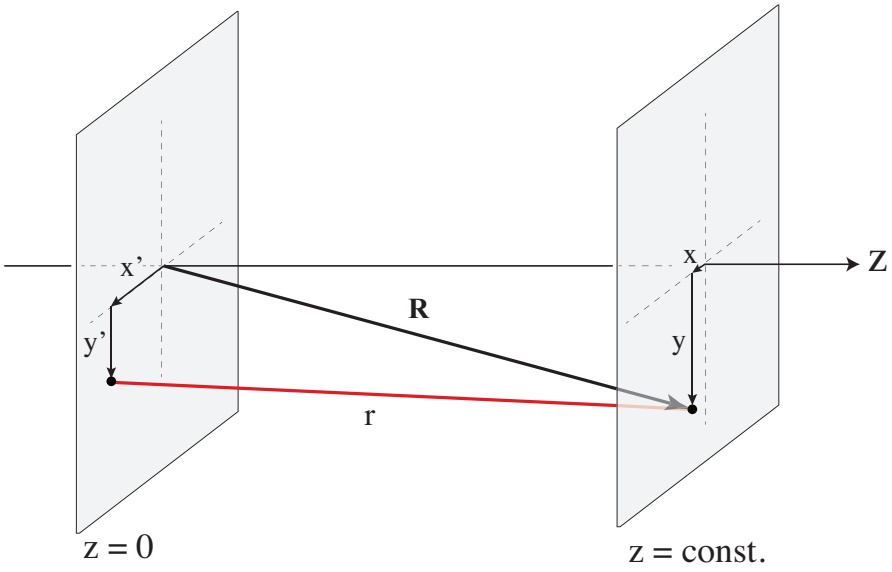

Figure 7.4: Coordinates used in discussion of Fresnel and Fraunhofer diffraction.

is to look at this question geometrically and to consider the situation illustrated in Fig. 7.4. R is the observation distance defined as the distance from the origin in the source plane (e.g. center of an aperture) to the observer. On the other hand, r is the true distance between source point $(x', y', 0)$ and the observation point (x, y, z). The square of r calculated as

$$r^2 = (x - x')^2 + (y - y')^2 + z^2 \;=\; R^2 \left[1 - \frac{2(xx' + yy')}{R^2} + \frac{x'^2 + y'^2}{R^2}\right] . \qquad (7.35)$$

After taking the square root on both sides and invoking the paraxial approximation we obtain

$$r(x', y') = = \; R \; - \; [x'(x/R) + y'(y/R)] \; + \; \frac{x'^2 + y'^2}{2\,R} \; + \; .. \; . \qquad (7.36)$$

To determine the field at the observation point, we have to sum up the waves emanating from different locations in the source plane (x', y'). This yields integrals of the form

$$\int_{z=0} A(x', y') \, \frac{\exp[-ikr(x', y')]}{r(x', y')} \, dx' \, dy' \, , \qquad (7.37)$$

where $A(x', y')$ is some amplitude function. The "summing up" of elementary spherical waves is referred to as *Huygens' principle*. Because of the large distance between source and observer we can safely replace $r(x', y')$ in the denominator by R or z. However, we cannot apply this approximation to the exponent since we would eliminate the effects of interference. Thus, we need to retain at least one of the additional terms in the expansion of r in (7.36).

Let us denote the maximum extent of the sources at $z = 0$ as D, that is $D/2 = \mathrm{Max}\{\sqrt{x'^2 + y'^2}\}$. If the observation distance is sufficiently large ($R \gg D$), we can neglect the last term in Eq. (7.36) and we end up with *Fraunhofer diffraction*. On the other hand, if the last term is not negligible, we speak of *Fresnel diffraction*. Fresnel diffraction becomes considerably more complicated because the exponent depends on the square of the source plane coordinates.

The transition from Fresnel to Fraunhofer diffraction happens at a distance z that roughly corresponds to the Rayleigh range z_0 of a Gaussian beam (c.f. Eq. 7.21), that is, the distance where the beam transitions into a spherical wave. Expressing the beam waist as $w_0 = D/2$ we obtain

$$z_0 \;=\; \frac{1}{8} \, k \, D^2 \, . \qquad (7.38)$$

As an example, let us consider a laser beam with beam diameter of $D = 3\,\text{mm}$ and wavelength $\lambda = 532\,\text{nm}$ (green). It turns out that $z_0 = 13\,\text{m}$, which is quite a distance to reach the far field!

7.4 The Point-Spread Function

The point-spread function is a measure of the resolving power of an imaging system. The narrower the point-spread function the better the resolution will be. As the name implies, the point-spread function defines the spread of a point source. If

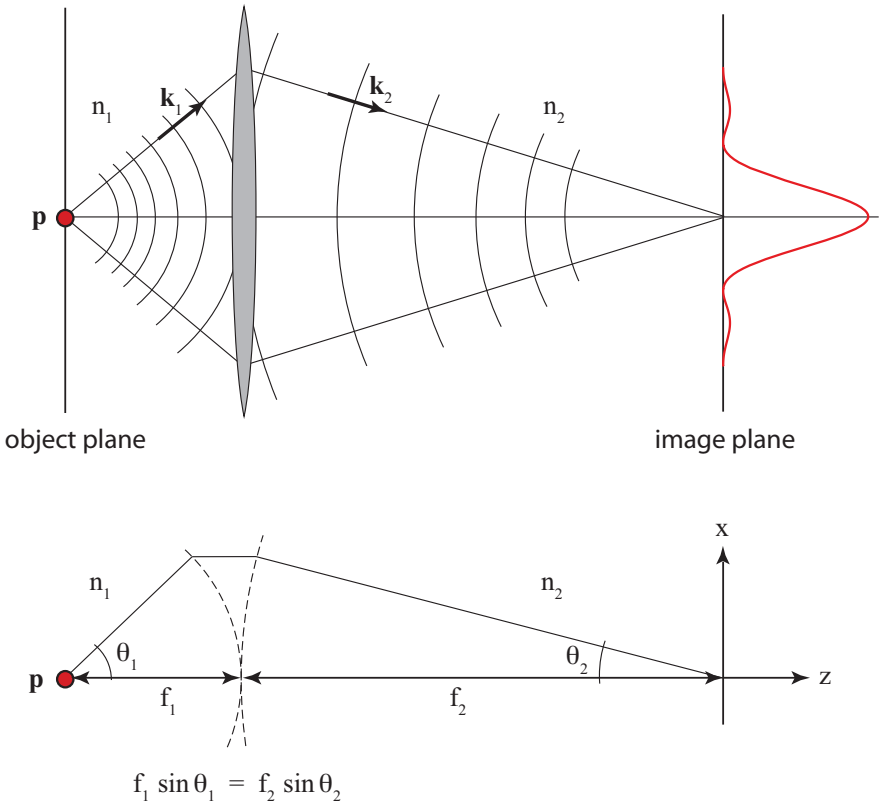

Figure 7.5: Calculation of the point-spread-function (PSF). The fields of a point source are projected onto an image plane. Because of the loss of evanescent waves and the finite angular collection angle of the imaging system, the point appears as a function with finite width.

we have a radiating point source then the image of that source will appear to have a finite size. This broadening is a direct consequence of spatial filtering. A point in space is characterized by a delta function that has an infinite spectrum of spatial frequencies k_x, k_y. On propagation from the source to the image, high-frequency components are filtered out. Usually the entire spectrum $(k_x^2+k_y^2) > k^2$ associated with the evanescent waves is lost. Furthermore, not all plane wave components can be collected, which leads to a further reduction in bandwidth. The reduced spectrum is not able to accurately reconstruct the original point source and the image of the point will have a finite size.

The smallest radiating electromagnetic unit is a dipole. As shown in Fig. 7.5, to calculate the point-spread function (PSF) we have to trace the dipole's field through an imaging system that focuses it onto an image plane. We will choose the origin of coordinates $(x, y, z) = (0, 0, 0)$ at the focus and use the angular spectrum representation of Eq. (7.32) to calculate the fields in the image plane. It is convenient to represent Eq. (7.32) in spherical coordinates by using the substitutions

$$k_x \;=\; k_2\,\sin\theta_2\,\cos\phi, \qquad k_y \;=\; k_2\,\sin\theta_2\,\sin\phi, \qquad k_z \;=\; k_2\,\cos\theta_2\;. \tag{7.39}$$

Furthermore, due to the symmetry of our problem it is favorable to express the transverse coordinates (x, y) of the field point as

$$x \;=\; \rho\cos\varphi \qquad y \;=\; \rho\sin\varphi\;. \tag{7.40}$$

Finally, we note that the integration in Eq. (7.32) runs over a plane, which is not a constant-coordinate surface in spherical coordinates. We therefore transform the planar integration surface into a spherical one using

$$\frac{1}{k_z}\,\mathrm{d}k_x\,\mathrm{d}k_y \;=\; k_2\,\sin\theta_2\,\mathrm{d}\theta_2\,\mathrm{d}\phi\,, \tag{7.41}$$

which is illustrated in Fig. 7.6.

Taken all together, the focal field represented by Eq. (7.32) can be written as

$$\mathbf{E}(\rho,\varphi,z) = \frac{\mathrm{i}k_2 f_2\,\mathrm{e}^{-\mathrm{i}k_2 f_2}}{2\pi}\int_0^{\mathrm{Max}[\theta_2]}\int_0^{2\pi}\mathbf{E}_\infty(\theta_2,\phi)\;\mathrm{e}^{\mathrm{i}k_2 z\cos\theta_2}\;\mathrm{e}^{\mathrm{i}k_2\rho\sin\theta_2\cos(\phi-\varphi)}\sin\theta_2\,\mathrm{d}\phi\,\mathrm{d}\theta_2 \tag{7.42}$$

Here, we have replaced the distance r_∞ between the focal point and the surface of the reference sphere of the lens by the focal length f_2. We have also limited the integration over θ_2 to the finite range $[0\,..\,\mathrm{Max}[\theta_2]]$ because any lens will have a finite size. Furthermore, since all fields propagate in the positive z-direction we retained only the $+$ sign in the exponent of Eq. (7.32).

To evaluate Eq. (7.42) we need to insert the field $\mathbf{E}_\infty$ of the dipole after it has been refracted by the lens. To simplify the analysis we will ignore the vectorial nature of dipole and its fields. Furthermore, we will assume that the imaging system can be treated in the paraxial approximation, or small angle limit, for which $\sin\theta_1 \approx \theta_1$ and $\sin\theta_2 \approx \theta_2$. Using the far-field term of Eq. (6.42) and ignoring the angular dependence (scalar point source), the dipole field before refraction at the lens is

$$E_1 \;=\; -\frac{p\,k_1^2}{4\pi\varepsilon_0\varepsilon_1}\,\frac{\exp(\mathrm{i}k_1 f_1)}{f_1}\;. \tag{7.43}$$

We trace this field through the lens and then insert it as E_∞ into Eq. (7.42) above. In essence, the dipole field is a field that uniformly illuminates the focusing lens, that is $E_\infty(\theta_2,\phi) \approx \mathrm{const.}$. Using the mathematical relation

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}x\cos(\phi-\varphi)}\,\mathrm{d}\phi \;=\; 2\pi\,J_0(x)\;, \tag{7.44}$$

we can carry out the integration over ϕ analytically. Here, J_0 is the 0th-order Bessel

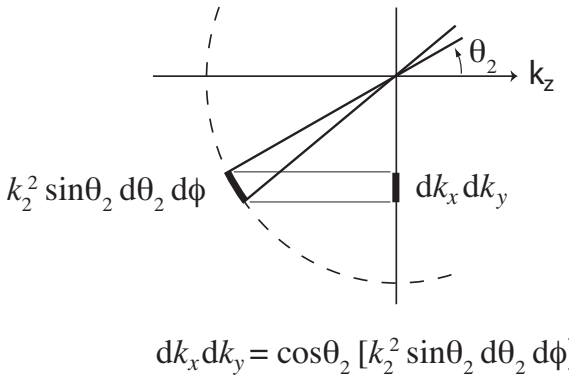

Figure 7.6: Illustration of the substitution $(1/k_z)\,\mathrm{d}k_x\,\mathrm{d}k_y = k\,\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$. The factor $1/k_z = 1/(k\cos\theta)$ ensures that the differential areas on the plane and the sphere stay equal.

function. The final expression for the focal field now contains a single integration over the variable θ_2. Skipping the constant prefactors we obtain

$$E(\rho,\varphi,z) \propto \int_0^{\mathrm{Max}[\theta_2]} \mathrm{e}^{\mathrm{i}k_2 z\cos\theta_2}\, J_0(k_2\rho\sin\theta_2)\,\sin\theta_2\,\mathrm{d}\phi\,\mathrm{d}\theta_2\ . \tag{7.45}$$

Using $\sin\theta_2 \approx \theta_2$, setting $z=0$ (image plane), and using

$$\int x\,J_0(x)\mathrm{d}x \;=\; x\,J_1(x)\ , \tag{7.46}$$

we find the following result for the intensity in the image plane

$$\lim_{\theta_{\max}\ll\pi/2} \left|\mathbf{E}(\rho,z=0)\right|^2 \;=\; \frac{\pi^4}{\varepsilon_0^2 n_1 n_2}\,\frac{p^2}{\lambda^6}\,\frac{\mathsf{NA}^4}{M^2}\left[2\,\frac{J_1(2\pi\tilde\rho)}{(2\pi\tilde\rho)}\right]^2, \quad \tilde\rho = \frac{\mathsf{NA}\,\rho}{M\lambda} \tag{7.47}$$

Here, we have used a normalized coordinate $\tilde\rho$, which is expressed in terms of the magnification $M=(n_1/n_2)(f_2/f_1)$ and the numerical aperture

$$\mathsf{NA} \;=\; n_1\,\sin(\mathrm{Max}[\theta_1])\ . \tag{7.48}$$

Note that $\sin(\mathrm{Max}[\theta_1]) = (f_2/f_1)\sin(\mathrm{Max}[\theta_2])$. The result (7.47) is the *point-spread-function*, first derived by Abbe in 1873. Its functional form is given by the term in

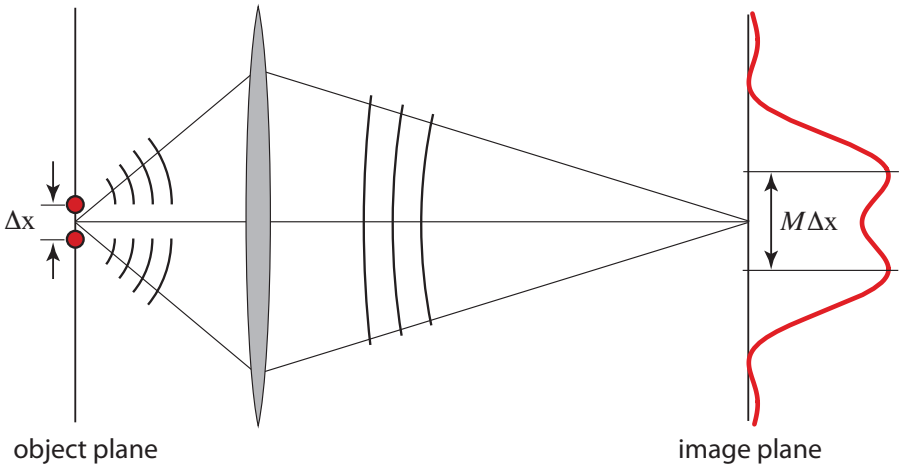

Figure 7.7: Illustration of the resolution limit. Two simultaneously radiating point sources separated by $\Delta r_{||}$ in the object plane generate a combined point-spread function in the image plane. The two point sources are resolved if they can be distinguished based on their image pattern.

brackets which is known as the *Airy function*. It tells us that the image of a point is no longer a point, but a function with a finite width. This width determines the resolution of an imaging system. In essence, two points in the object plane have to be separated by more than the width of the PSF in order to be distinguishable. This is illustrated in Fig. 7.7.

The point-spread function can be measured by using a single quantum emitter, such as a single molecule of quantum dot, as a point emitter. Fig. 7.8 shows such a measurement together with a fit according to Eq. (7.47). The point-spread-function has been recorded by using a NA $= 1.3$ lens to collect the fluorescence photons from a single DiI molecule with center wavelength of $\lambda \approx 580$ nm.

The width of the point-spread function Δx is usually defined as the radial dis-

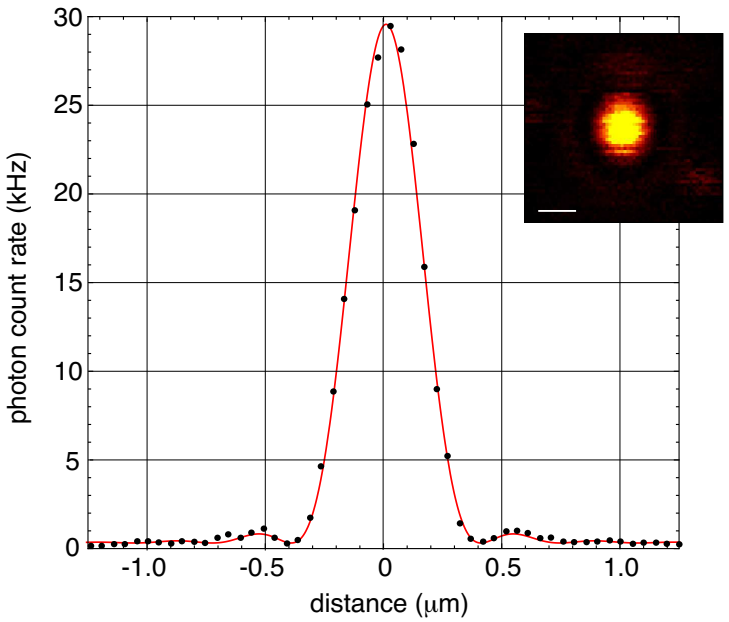

Figure 7.8: Point-spread function measured with a single molecule point source. Fluorescence photons emitted by a DiI molecule are collected with a NA $= 1.3$ objective lens. The center wavelength is $\lambda \approx 580$ nm. The data points correspond to a horizontal line cut through the center of the fluorescence rate image shown in the inset. The solid curve corresponds to the Airy function.

tance for which the value of the paraxial point-spread function becomes zero, or

$$\Delta x \;=\; 0.6098\,\frac{M\,\lambda}{\mathrm{NA}}\;. \tag{7.49}$$

This width is also denoted as the *Airy disk radius*. It depends in a simple manner on the numerical aperture, the wavelength and the magnification of the system.

Chapter 8

Waveguides and Resonators

The objective of resonators is to *confine* electromagnetic energy. On the other hand, the purpose of waveguides is to *guide* electromagnetic energy. In both cases, the desired functionality is achieved through material boundaries.

8.1 Resonators

Let us consider a rectangular box with sides L_x, L_y, and L_z, as shown in Fig. 8.1. The interior of the box is filled with a linear and isotropic material characterized by ε and μ. The walls of the box are perfectly reflecting, that is, the fields are not able to penetrate into the walls. Furthermore, there are no sources present, which implies that we're looking for homogeneous solutions of the wave equation, that is, solutions of the Helmholtz equation (3.15).

To solve the Helmholtz equation for the x component of the complex electric field vector $\mathbf{E}$ we write

$$E_x(x,y,z) = E_0^{(x)}\, X(x)\, Y(y)\, Z(z)\;, \tag{8.1}$$

which is referred to as *separation of variables*. X, Y, and Z are dimensionless functions and $E_0^{(x)}$ is a constant field amplitude. Inserting into $[\nabla^2 + k^2]E_x = 0$ yields

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} + k^2 = 0\;. \tag{8.2}$$

In order for this equation to hold for any x, y, and z we have to require that each of the terms is constant. We set these constants to be $-k_x^2$, $-k_y^2$ and $-k_z^2$, which implies

$$k_x^2 + k_y^2 + k_z^2 \;=\; k^2 \;=\frac{\omega^2}{c^2} n^2(\omega)\;. \tag{8.3}$$

We obtain three separate equations for X, Y, and Z

$$\begin{aligned} \partial^2 X \,/\, \partial x^2 \;+\; k_x^2 X \;&=\; 0 \\ \partial^2 Y \,/\, \partial y^2 \;+\; k_y^2 Y \;&=\; 0 \\ \partial^2 Z \,/\, \partial z^2 \;+\; k_z^2 Z \;&=\; 0\,, \end{aligned} \tag{8.4}$$

with the solutions $\exp[\pm i k_x x]$, $\exp[\pm i k_y y]$, and $\exp[\pm i k_z z]$. Thus, the solutions for E_x become

$$E_x(x,y,z) = E_0^{(x)} \left[c_{1,x}\,\mathrm{e}^{-ik_x x} + c_{2,x}\,\mathrm{e}^{ik_x x}\right]\left[c_{3,x}\,\mathrm{e}^{-ik_y y} + c_{4,x}\,\mathrm{e}^{ik_y y}\right]\left[c_{5,x}\,\mathrm{e}^{-ik_z z} + c_{6,x}\,\mathrm{e}^{ik_z z}\right]. \tag{8.5}$$

Electric fields cannot penetrate into a perfect conductor. Therefore, the boundary condition (4.13) implies $E_x(x=0) = E_x(x=L_x) = 0$, which turns Eq. (8.5) into

$$E_x(x,y,z) \;=\; E_0^{(x)}\;\sin[n\,\pi\,x/L_x]\left[c_{3,x}\,\mathrm{e}^{-ik_y y} + c_{4,x}\,\mathrm{e}^{ik_y y}\right]\left[c_{5,x}\,\mathrm{e}^{-ik_z z} + c_{6,x}\,\mathrm{e}^{ik_z z}\right]\,, \tag{8.6}$$

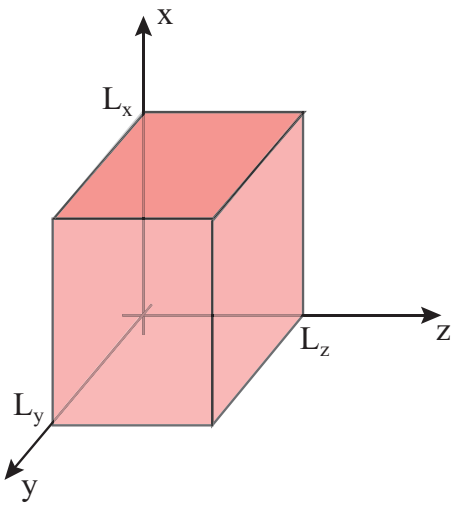

Figure 8.1: A resonator with perfectly reflecting walls and side lengths L_x, L_y, and L_z .

with n being an integer. Similar solutions are found for E_y and E_z, namely,

$$E_y(x,y,z) = E_0^{(y)} \left[c_{1,y}\, e^{-ik_x x} + c_{2,y}\, e^{ik_x x}\right] \sin[m\,\pi\, y/L_y] \left[c_{5,y}\, e^{-ik_z z} + c_{6,y}\, e^{ik_z z}\right] \quad (8.7)$$

$$E_z(x,y,z) = E_0^{(z)} \left[c_{1,z}\, e^{-ik_x x} + c_{2,z}\, e^{ik_x x}\right] \left[c_{3,z}\, e^{-ik_y y} + c_{4,z}\, e^{ik_y y}\right] \sin[l\,\pi\, z/L_z]. \quad (8.8)$$

Because $\nabla \cdot \mathbf{E} = 0$ for any x, y, and z we find

$$\begin{aligned} E_x(x,y,z) &= E_0^{(x)} \sin[n\,\pi\, x/L_x] \sin[m\,\pi\, y/L_y] \sin[l\,\pi\, z/L_z] \\ E_y(x,y,z) &= E_0^{(y)} \sin[n\,\pi\, x/L_x] \sin[m\,\pi\, y/L_y] \sin[l\,\pi\, z/L_z] \\ E_z(x,y,z) &= E_0^{(z)} \sin[n\,\pi\, x/L_x] \sin[m\,\pi\, y/L_y] \sin[l\,\pi\, z/L_z]\,, \end{aligned} \quad (8.9)$$

and

$$\frac{n}{L_x} E_0^{(x)} + \frac{m}{L_y} E_0^{(y)} + \frac{l}{L_z} E_0^{(z)} = 0 \quad (8.10)$$

Using Eq. (8.3) we find the dispersion relation or mode structure of the resonator

$$\pi^2 \left[\frac{n^2}{L_x^2} + \frac{m^2}{L_y^2} + \frac{l^2}{L_z^2}\right] = \frac{\omega_{nml}^2}{c^2}\, n^2(\omega_{nml}) \qquad n,m,l \in \{0,1,2,3,..\} \quad (8.11)$$

Thus, we find that the resonator supports only discrete frequencies ω_{nml}, each associated with a mode (n,m,l) of the resonator. Note that n is used both for the index of refraction and for the mode number in x.

8.1.1 The Density of States

Let us now consider a resonator with equal sides, that is, $L = L_x + L_y + L_z$. In this case,

$$n^2 + m^2 + l^2 = \left[\frac{\omega_{nml}\, L\, n(\omega_{nml})}{\pi\, c}\right]^2 . \quad (8.12)$$

If n, m, and l were real numbers, then this equation defines a sphere of radius $r_0 = [omega_{nml}\, L\, n(\omega_{nml})/(\pi\, c)]$. Indeed, for large numbers we can approximate these numbers by real numbers. Let us now consider a specific mode given by $[n,m,l]$ and with angular frequency $\omega = \omega_{nml}$ and count the number of modes N with frequencies smaller than ω. According to Eq. (8.12) this corresponds to the

interior of a sphere with radius r_0, and because n, m, and l are positive numbers we have to count only $1/8$-th of the sphere. Thus,

$$N(\omega) = \frac{1}{8}\left[\frac{4\pi}{3} r_0^3\right] \; 2 \; . \tag{8.13}$$

The '2' at the end has been added because for each $[n, m, l]$ there are two solutions with different polarizations. This follows from Eq. (8.10). Inserting the expression for r_0 we find

$$N(\omega) = V \, \frac{\omega^3 \, n^3(\omega)}{3\pi^2 \, c^3} \; , \tag{8.14}$$

where V is the volume of the resonator. The number of different resonator modes in the frequency interval $[\omega \; .. \; \omega + \Delta\omega]$ becomes

$$\frac{d \, N(\omega)}{d\omega} \, \Delta\omega \;\; = \;\; V \, \frac{\omega^2 \, n^3(\omega)}{\pi^2 \, c^3} \, \Delta\omega \; , \tag{8.15}$$

which states that there a many more modes for high frequencies ω. We now define the number of modes $\rho(\omega)$ per unit volume V and unit frequency $\Delta\omega$. We obtain

$$\rho(\omega) \;\; = \;\; \frac{\omega^2 \, n^3(\omega)}{\pi^2 \, c^3} \tag{8.16}$$

which is generally referred to as the *density of states* (DOS). The number of modes N in the volume V and in the frequency range $[\omega_1 \; .. \; \omega_2]$ is then calculated as

$$N(\omega) \;\; = \;\; \int_V \int_{\omega_1}^{\omega_2} \rho(\omega) \, d\omega \, dV \; . \tag{8.17}$$

The density of states is of importance in blackbody radiation and is an important concept to describe how efficient radiation interacts with matter. For example, the power $\bar{P}$ emitted by a dipole (c.f. Eq 6.46) can be expressed in terms of the density of states as

$$\bar{P} \;\; = \;\; \frac{\pi \, \omega^2}{12 \, \varepsilon_0 \, \varepsilon} \; |\mathbf{p}| \; \rho(\omega) \tag{8.18}$$

As we discussed in Section 6.4, the amount of radiation released by a dipole depends on the environment, and this dependence can be accounted for by the density of states rho. In other words, ρ depends on the specific environment and takes on a value of $\omega^2 \, / \, \pi^2 c^3$ in empty space (vacuum). Any objects placed into the empty space will influence ρ and the ability of a dipole source to radiate.

The resonator that we have analyzed possesses discrete frequencies ω_{mnl}. In reality, any resonator has losses, for example due to the absorption of electromagnetic energy at the boundaries or due to radiation. As a consequence, the discrete frequencies broaden and assume a finite line width $\Delta\omega = 2\gamma$. The quality factor, or Q-factor of a line is defined as $Q = \omega/(2\gamma)$. It is a measure for how long electromagnetic energy can be stored in a resonator. The line shape of a mode is generally a Lorentzian (see Section 6.7). The electric field in the cavity is therefore an exponentially damped oscillation of the form

$$\mathbf{E}(\mathbf{r},t) \;=\; \mathrm{Re}\left\{\mathbf{E}_0(\mathbf{r})\,\exp\left[(\mathrm{i}\omega_0 - \frac{\omega_0}{2\,Q})\,t\right]\right\}\;, \tag{8.19}$$

where ω_0 represents one of the resonance frequencies ω_{mnl}. The spectrum of the stored energy density becomes

$$W_\omega(\omega) \;=\; \frac{\omega_0^2}{4Q^2}\,\frac{W_\omega(\omega_0)}{(\omega-\omega_0)^2 \;+\; (\omega_0\,/\,2Q)^2}\;. \tag{8.20}$$

8.1.2 Cavity Perturbation

A sharp resonance is a key requirement for ultrasensitive detection in various applications. For example, watches and clocks use high-Q quartz crystals to measure time, some biosensing schemes make use of oscillating cantilevers to detect adsorption of molecules, and atomic clocks use atomic resonances as frequency standards. A perturbation of the resonator (cavity), for example due to particle adsorption or a change of the index of refraction, leads to a shift of the resonance frequency, which can be measured and used as a control signal for sensing.

To establish an understanding of cavity perturbation we consider the system depicted in Fig. 8.2. A leaky cavity and its environment are characterized by a spatially varying permittivity $\varepsilon(\mathbf{r})$ and permeability $\mu(\mathbf{r})$. In the absence of any perturbation the system assumes a resonance at frequency ω_0 and the the fields are described by

$$\nabla\times\mathbf{E}_0 \;=\; i\omega_0\mu_0\mu(\mathbf{r})\mathbf{H}_0\,,\quad \nabla\times\mathbf{H}_0 \;=\; -i\omega_0\varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}_0\,, \tag{8.21}$$

with $\mathbf{E}_0(\mathbf{r},\omega_0)$ and $\mathbf{H}_0(\mathbf{r},\omega_0)$ denoting the unperturbed complex field amplitudes. A particle with material parameters $\Delta\varepsilon(\mathbf{r})$, $\Delta\mu(\mathbf{r})$ constitutes a perturbation and gives

rise to a new resonance frequency ω. Maxwell's curl equations for the perturbed system read as

$$\nabla \times \mathbf{E} \;=\; i\omega\mu_0 \left[\mu(\mathbf{r})\mathbf{H} + \Delta\mu(\mathbf{r})\mathbf{H}\right] \tag{8.22}$$

$$\nabla \times \mathbf{H} \;=\; -i\omega\varepsilon_0 \left[\varepsilon(\mathbf{r})\mathbf{E} + \Delta\varepsilon(\mathbf{r})\mathbf{E}\right] . \tag{8.23}$$

Notice that both $\Delta\varepsilon$ and $\Delta\mu$ are zero outside of the volume occupied by the perturbation. Using $\nabla \cdot (\mathbf{A} \times \mathbf{B}) = (\nabla \times \mathbf{A}) \cdot \mathbf{B} - (\nabla \times \mathbf{B}) \cdot \mathbf{A}$ we find

$$\begin{aligned} \nabla \cdot \left[\mathbf{E}_0^* \times \mathbf{H} - \mathbf{H}_0^* \times \mathbf{E}\right] = {} & i(\omega - \omega_0) \left[\varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}_0^* \cdot \mathbf{E} \;+\; \mu_0\mu(\mathbf{r})\mathbf{H}_0^* \cdot \mathbf{H}\right] \\ & +\; i\omega \left[\mathbf{E}_0^*\varepsilon_0\Delta\varepsilon(\mathbf{r})\mathbf{E} \;+\; \mathbf{H}_0^*\mu_0\Delta\mu(\mathbf{r})\mathbf{H}\right] . \end{aligned} \tag{8.24}$$

We now consider a fictitious spherical surface ∂V at very large distance from the cavity and integrate Eq. (8.24) over the enclosed volume V (c.f. Fig. 8.2). Using Gauss' theorem, the left hand side of Eq. (8.24) becomes

$$\int_{\partial V} \left[\mathbf{H} \cdot (\mathbf{n} \times \mathbf{E}_0^*) \;+\; \mathbf{H}_0^* \cdot (\mathbf{n} \times \mathbf{E})\right] da \;=\; 0 \tag{8.25}$$

where $\mathbf{n}$ is a unit vector normal to the surface ∂V. The above expression vanishes because of the transversality of the field, i.e. $(\mathbf{n} \times \mathbf{E}_0^*) = (\mathbf{n} \times \mathbf{E}) = 0$ on the surface

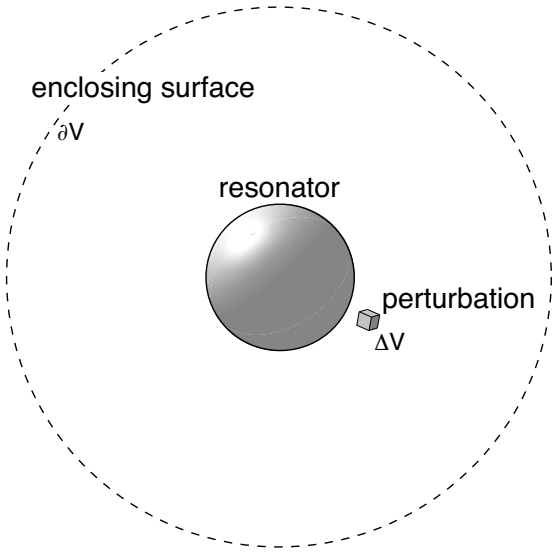

Figure 8.2: A resonator with resonance frequency ω_0 interacts with an external perturbation giving rise to a new resonance frequency ω. The calculation makes use of a fictitious spherical surface at infinity.

of the spherical surface. We thus arrive at the equation

$$\frac{\omega-\omega_0}{\omega} \;=\; -\frac{\int_V[\mathbf{E}_0^*\varepsilon_0\Delta\varepsilon(\mathbf{r})\mathbf{E} \;+\; \mathbf{H}_0^*\mu_0\Delta\mu(\mathbf{r})\mathbf{H}]\,dV}{\int_V\left[\varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}_0^*\cdot\mathbf{E} \;+\; \mu_0\,\mu(\mathbf{r})\mathbf{H}_0^*\cdot\mathbf{H}\right]\,dV}\,, \tag{8.26}$$

which is known as the Bethe-Schwinger cavity perturbation formula. Eq. (8.26) is an exact formula, but because $\mathbf{E}$ and $\mathbf{H}$ are not known the equation cannot be used in its form. Notice that because $\Delta\varepsilon$ and $\Delta\mu$ are zero outside of the volume occupied by the perturbation the integral in the nominator runs only over the volume of the perturbation ΔV. For situations where there are no radiation losses and all the energy is contained inside the boundaries of a resonator the surface ∂V can be chosen to coincide with the boundaries.

We assume that the perturbation has a small effect on the cavity. Therefore we write as a first-order approximation $\mathbf{E}=\mathbf{E}_0$ and $\mathbf{H}=\mathbf{H}_0$. After performing these substitutions in Eq. (8.26) we find

$$\frac{\omega-\omega_0}{\omega} \;\approx\; -\frac{\int_{\Delta V}[\mathbf{E}_0^*\varepsilon_0\Delta\varepsilon(\mathbf{r})\mathbf{E}_0 \;+\; \mathbf{H}_0^*\mu_0\Delta\mu(\mathbf{r})\mathbf{H}_0]\,dV}{\int_V\left[\varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}_0^*\cdot\mathbf{E}_0 \;+\; \mu_0\,\mu(\mathbf{r})\mathbf{H}_0^*\cdot\mathbf{H_0}\right]\,dV} \tag{8.27}$$

For a high-Q resonator the radiation losses are small and the integration volume V can be taken over the boundaries of the resonator. To evaluate Eq. (8.27) we first must solve for the fields $\mathbf{E}_0(\mathbf{r}),\mathbf{H}_0(\mathbf{r})$ of the unperturbed cavity. Interestingly, for a weakly-dispersive medium the denominator of Eq. (8.27) denotes the total energy of the unperturbed cavity (W_0) whereas the nominator accounts for the energy introduced by the perturbation (ΔW). Hence, $(\omega-\omega_0)/\omega = -\Delta W/W_0$. An increase of energy by ΔW causes the resonance frequency to *red-shift* to $\omega=\omega_0\,[W_0/(W_0+\Delta W)]$. A *blue-shift* is possible by perturbing the cavity volume, i.e. by removing ΔW from the cavity.

As an example let us consider a planar cavity with perfectly reflecting end faces of area A and separated by a distance L. The fundamental mode $\lambda=2L$ has a resonance frequency $\omega_0=\pi c/L$, and the electric and magnetic fields inside the cavity are calculated to be $E_0\sin[\pi z/L]$ and $-i\sqrt{\varepsilon_0/\mu_0}E_0\cos[\pi z/L]$, respectively. The coordinate z is perpendicular to the surfaces of the end faces. The denominator of Eq. (8.27) is easily determined to be $V\,\varepsilon_0\,E_0^2$, where $V=LA$. We place a spherical nanoparticle with dielectric constant $_\Delta\varepsilon$ and volume $_\Delta V$ in the center of

the cavity and assume that the field is homogeneous across the dimensions of the particle. The nominator of Eq. (8.27) is calculated to be $\Delta V\, \Delta\varepsilon\, \varepsilon_0\, E_0^2$ and the frequency shift is determined to be $(\omega - \omega_0)/\omega = -\Delta\varepsilon\, \Delta V/V$. A better approximation retains the perturbed fields $\mathbf{E}$ and $\mathbf{H}$ in the nominator of Eq. (8.26). Making use of the quasi-static solution for a small spherical particle we write $\mathbf{E} = 3\mathbf{E}_0/(2 + \Delta\varepsilon)$ and obtain a frequency shift of $(\omega - \omega_0)/\omega = -[3\Delta\varepsilon/(2 + \Delta\varepsilon)]\, \Delta V/V$. In both cases the resonance shift scales with the ratio of resonator and perturbation volumes.

8.2 Waveguides

Waveguides are used to carry electromagnetic energy from location A to location B. They exist in form of wires, coaxial cables, parallel plates, or optical fibers. In general, the transmission of radiation through free-space is subject to diffraction, which results in spreading out of energy. Waveguides avoid this problem at the expense of material structures that have to connect the locations A and B.

The simplest waveguide system is the two-wire transmission line. It can be formulated in terms of distributed circuit elements, such as capacitances C and inductances L. One can derive a one-dimensional wave equation for the current and the voltage carried along the transmission line. The speed of propagation is given by $c = 1/\sqrt{L\,C}$. Detailed discussions of transmission lines can be found in many textbooks and we won't derive or analyze it here. Instead, we will concentrate on parallel-plate waveguide and on hollow metal waveguides and then make a transition to all-dielectric waveguides used in fiber-optic communication.

8.2.1 Parallel-Plate Waveguides

The parallel-plate waveguide that we consider here is illustrated in Fig. 8.3. It consists of a medium characterized by the index of refraction $n(\omega)$ sandwiched between two ideally conducting plates. We will align our coordinate system such that the wave propagates in direction of the z axis. As we shall see, one distinguishes between two different solutions: 1) Waves that have no electric field in propagation

direction z (TE modes), and 2) waves that have no magnetic field in propagation direction z (TM modes).

TE Modes

For TE modes the electric field is parallel to the surface of the plates. As an ansatz we choose a plane wave propagating at an angle θ towards the bottom plate

$$\mathbf{E}_1(x,y,z) \;=\; E_0\,\mathbf{n}_y\,\exp[-ikx\cos\theta + ikz\sin\theta]\;, \tag{8.28}$$

with $k = n(\omega)\,\omega/c$. The wave reflects at the bottom plate and then propagates towards the top plate

$$\mathbf{E}_2(x,y,z) \;=\; -E_0\,\mathbf{n}_y\,\exp[ikx\cos\theta + ikz\sin\theta]\;, \tag{8.29}$$

where the sign change is due to the boundary condition, that is, the total parallel $\mathbf{E}$-field has to vanish. The sum of the fields becomes

$$\mathbf{E}(x,y,z) \;=\; \mathbf{E}_1(x,y,z) + \mathbf{E}_2(x,y,z) \;=\; -2i\,E_0\,\mathbf{n}_y\,\exp[ikz\sin\theta]\sin[kx\cos\theta]\;. \tag{8.30}$$

While this field satisfies the boundary condition at the bottom surface, it does not yet at the top plate $x = d$. Requiring that $\mathbf{E}(d,y,z) = 0$ yields $\sin[kx\cos\theta] = 0$, which is fulfilled if

$$kd\cos\theta \;=\; n\pi \qquad n \in \{1,2,..\}\;, \tag{8.31}$$

which corresponds to a 'quantization' of the normal wavenumber $k_x = k\cos\theta$

$$k_{x_n} \;=\; n\frac{\pi}{d} \qquad n \in \{1,2,..\}\;. \tag{8.32}$$

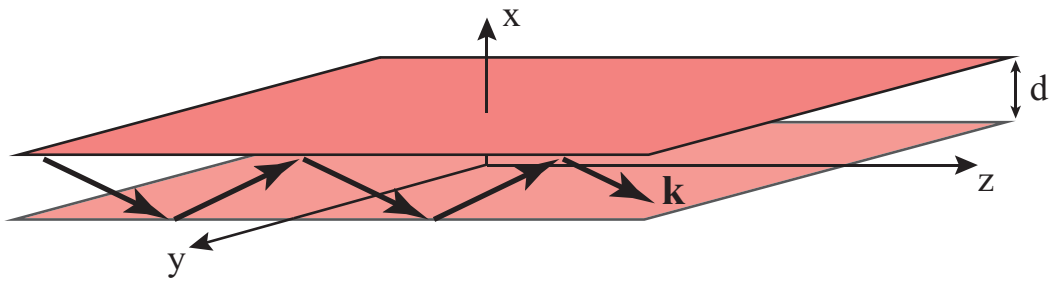

Figure 8.3: A parallel-plate waveguide with plate separation d.

Note that $n = 0$ is excluded because it yields a zero-field (trivial solution). Using $k^2 = k_z^2 + k_x^2$, we find the the propagation constant k_z of the field propagating in between of the plates in z direction is

$$k_{z_n} = \sqrt{k^2 - n^2\,[\pi/d]^2} \qquad n \in \{1, 2, ..\} \tag{8.33}$$

As long as k_{z_n} is real the field will propagate along the z direction. However, when $n\pi/d > k$ the propagation constant k_{z_n} is imaginary and the field will exponential decay in direction of z, similar to evanescent waves discussed in Section 4.4. Because $k \propto \omega$ it turns out that the waveguide acts as a high-pass filter. Frequencies below the *cut-off frequency*

$$\omega_c = \frac{n\,\pi\,c}{d\,n(\omega_c)} \qquad n \in \{1, 2, ..\} \tag{8.34}$$

cannot propagate. In summary, the solutions of the wave equation for a system of two parallel plates are characterized by a mode number n. We refer to the solutions that have no electric field in propagation direction as TE_n modes. Solutions are found only for discrete frequencies ω_n and for each mode there is a cut-off frequency below which no propagation is possible. Note, that $n(\omega_c)$ is the dispersive index of refraction of the medium between the two plates.

For propagating fields with $\omega > \omega_c$, the *phase velocity* is defined by the phase factor $\exp[ik_{z_n}z - i\omega t]$ as $v_{ph} = \omega/k_{z_n}$. On the other hand, the energy of the field and thus any information associated with it is transported at the *group velocity* $v_g = d\omega/dk_{z_n}$.

TM Modes

Let us now repeat the analysis for the case where the magnetic field is parallel to the surface of the plates. Similar to Eq. (8.28) we write

$$\mathbf{H}_1(x, y, z) = H_0\,\mathbf{n}_y\,\exp[-ikx\cos\theta + ikz\sin\theta]\;, \tag{8.35}$$

The wave reflects at the bottom plate and then propagates towards the top plate

$$\mathbf{H}_2(x, y, z) = H_0\,\mathbf{n}_y\,\exp[ikx\cos\theta + ikz\sin\theta]\;, \tag{8.36}$$

In contrast to Eq. (8.36) there is no sign change of H_0 upon reflection at the boundary. This follows from the boundary conditions (4.11) - (4.14).[1] The sum of the fields becomes

$$\mathbf{H}(x,y,z) \;=\; \mathbf{H}_1(x,y,z)+\mathbf{H}_2(x,y,z) \;=\; 2\,H_0\,\mathbf{n}_y\,\exp[ikz\sin\theta]\cos[kx\cos\theta]\;, \quad (8.37)$$

and the boundary conditions at the top interface $z = d$ lead to the condition

$$kd\cos\theta \;=\; n\pi \qquad n\in\{0,1,2\,..\}\;. \quad (8.38)$$

In contrast to the quantization condition (8.31) for TE modes, we now also find solutions for $n = 0$ ($\cos(0) \neq 0$). Thus,

$$k_{z_n} \;=\; \sqrt{k^2-n^2\,[\pi/d]^2} \qquad n\in\{0,1,..\} \quad (8.39)$$

The fundamental mode TM_0 has no cut-off frequency whereas all higher order TM do have a cut-off, similar to TE modes discussed above. The absence of a cut-off

[1] The electric field associated with $\mathbf{H}_1$ is $\mathbf{E}_1 \sim E_0[\sin\theta, 0, \cos\theta]^T$, which upon reflection becomes $\mathbf{E}_2 \sim E_0[\sin\theta, 0, -\cos\theta]^T$ because the parallel $\mathbf{E}$-field at the boundary has to vanish. Consequently, H_0 remains unchanged upon reflection.

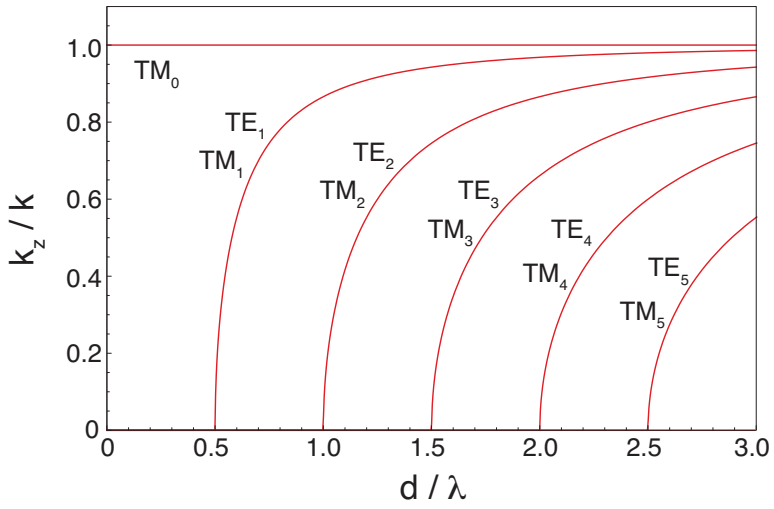

Figure 8.4: Mode structure of a parallel-plate waveguide with plate separation d. The vertical axis shows the real part of the normalized propagation constant. All modes, with the exception of TM_0 run into cut-off as the wavelength λ is increased or the plate separation d is decreased.

for the TM mode is a finding that is not restricted to two parallel plates, but holds for any waveguide made of two electrically isolated metal electrodes. For example, a coaxial cable has no cut-off, but a hollow metal waveguide (a pipe) does have a cut-off.

8.2.2 Hollow Metal Waveguides

We now confine the lateral extent of the waveguide modes. The parallel-plate waveguide then turns into a hollow metal waveguide as illustrated in Fig. 8.5. To solve for the fields in such a waveguide structure we write

$$\mathbf{E}(x,y,z) \;=\; \mathbf{E}^{xy}(x,y)\,\mathrm{e}^{-ik_z z}\;. \tag{8.40}$$

Inserting into the the Helmholtz equation (3.15) leads to

$$\left[\nabla_t^2 + k_t^2\right]\mathbf{E}^{xy}(x,y) \;=\; 0\;, \tag{8.41}$$

where $\nabla_t^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse Laplacian and and $k_t = [k_x^2 + k_y^2]^{1/2} = [k^2 - k_z^2]^{1/2}$ is the transverse wavenumber.

Next, we write the electric field in terms of a transverse vector in the (x,y) plane and a longitudinal vector that points along the z axis, that is,

$$\mathbf{E} \;=\; \mathbf{E}_t + \mathbf{E}_z\;. \tag{8.42}$$

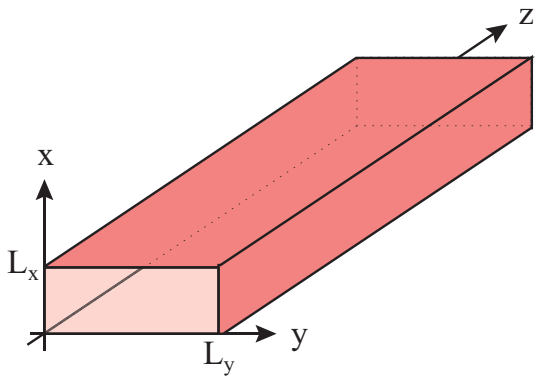

Figure 8.5: A rectangular hollow metal waveguide.

Here, $\mathbf{E}_t = \mathbf{E} \times \mathbf{n}_z$ and $\mathbf{E}_z = (\mathbf{E} \cdot \mathbf{n}_z)\mathbf{n}_z$. A similar expression can be written for the magnetic field $\mathbf{H}$. Inserting both expressions into Maxwell's equations (2.31) - (2.34), and assuming that the waveguide materials is linear and source-free, we find

$$E_x^{xy} = -Z \frac{ik}{k_t^2}\frac{\partial H_z^{xy}}{\partial y} + \frac{ik_z}{k_t^2}\frac{\partial E_z^{xy}}{\partial x} \tag{8.43}$$

$$E_y^{xy} = Z \frac{ik}{k_t^2}\frac{\partial H_z^{xy}}{\partial x} + \frac{ik_z}{k_t^2}\frac{\partial E_z^{xy}}{\partial y} \tag{8.44}$$

$$H_x^{xy} = Z^{-1}\frac{ik}{k_t^2}\frac{\partial E_z^{xy}}{\partial y} + \frac{ik_z}{k_t^2}\frac{\partial H_z^{xy}}{\partial x} \tag{8.45}$$

$$H_y^{xy} = -Z^{-1}\frac{ik}{k_t^2}\frac{\partial E_z^{xy}}{\partial x} + \frac{ik_z}{k_t^2}\frac{\partial H_z^{xy}}{\partial y}\ . \tag{8.46}$$

where Z is the wave impedance defined in Eq. (4.33). These equations show that the transverse field components E_x^{xy}, E_y^{xy} and H_x^{xy}, H_y^{xy} derive from the longitudinal field components E_z^{xy} and H_z^{xy}. Thus, it is sufficient to solve for E_z^{xy} and H_z^{xy}. So far, the discussion was general, not restricted to any particular waveguide geometry. We next discuss the particular case of a rectangular hollow metal waveguide, as illustrated in Fig. 8.5. Such waveguides are commonly used in the microwave regime, which spans the frequency range of 1-100 GHz.

TE Modes

For TE waves the field $\mathbf{E}_z^{xy}$ is zero. Thus, according to Eqs. (8.43) - (8.46) all field components can be derived from $\mathbf{H}_z^{xy}$. Following the discussion in Section 8.2.1 we find

$$\begin{aligned} H_z^{xy}(x,y) &= H_{0_z}\,\cos[k_x x]\,\cos[k_y y] \\ &= H_{0_z}\,\cos[\frac{n\pi}{L_x}x]\,\cos[\frac{m\pi}{L_y}y] \qquad n,m \in \{0,1,2..\}\ , \end{aligned} \tag{8.47}$$

with the transverse wavenumber given by

$$k_t^2 = [k_x^2 + k_y^2] = \left[\frac{n^2\pi^2}{L_x^2} + \frac{m^2\pi^2}{L_y^2}\right] \qquad n,m \in \{0,1,2..\}\ , \tag{8.48}$$

similar to the mode structure of a resonator (Eq. 8.11). Accordingly, the frequencies of the TE_{nm} modes are

$$\omega_{nm} \;=\; \frac{\pi\, c}{n(\omega_{nm})}\sqrt{\frac{n^2}{L_x^2}+\frac{m^2}{L_y^2}} \qquad n,m\in\{0,1,2,..\} \tag{8.49}$$

with $n(\omega_{nm})$ being the index of refraction. It turns out that n and m cannot both be zero because the TE_{00} does not exist. Thus, the lowest frequency mode is the TE_{01} or the TE_{10} mode. While the modes of a parallel-plate waveguide are formed by the superposition of two plane waves, the fields of a hollow rectangular waveguide follow from the superposition of four plane waves. Note that the condition of a magnetic field parallel to the waveguide surfaces leads here to TE modes, whereas in the case of two parallel plates it leads to TM modes.

The propagation constant (longitudinal wavenumber) is calculated as

$$k_z=\sqrt{k^2-k_t^2} \;=\; \sqrt{\frac{\omega_{nm}^2}{c^2}\,n^2(\omega_{nm})-\left[\frac{n^2\pi^2}{L_x^2}+\frac{m^2\pi^2}{L_y^2}\right]} \qquad n,m\in\{0,1,2\,..\}\,. \tag{8.50}$$

Since n, m cannot both be zero we find that all modes run into cut-off for low enough frequencies ω. The lowest-order mode is generally referred to as the fundamental mode or the dominant mode.

Let us choose $L_x > L_y$. In this case, the fundamental mode is the TE_{01} mode for which, according to Eqs. (8.43) - (8.46), the fields are determined as

$$\begin{aligned}
H_z^{xy} &= H_{0_z}\cos[\frac{\pi}{L_y}y] && (8.51)\\
H_y^{xy} &= i(k_z/k_t^2)\,H_{0_z}\,(\pi/L_y)\sin[\frac{\pi}{L_y}y] && (8.52)\\
E_x^{xy} &= -(ikZ/k_t^2)\,H_{0_z}\,(\pi/L_y)\sin[\frac{\pi}{L_y}y] && (8.53)\\
H_x^{xy} &= E_y^{xy} \;=\; E_z^{xy} \;=\; 0\,.
\end{aligned}$$

According to Eq. (8.40), the total fields $\mathbf{E}$ and $\mathbf{H}$ are obtained by multiplying with $\exp[-ik_z z]$. The fields of the TE_{01} mode are illustrated in Fig. 8.6.

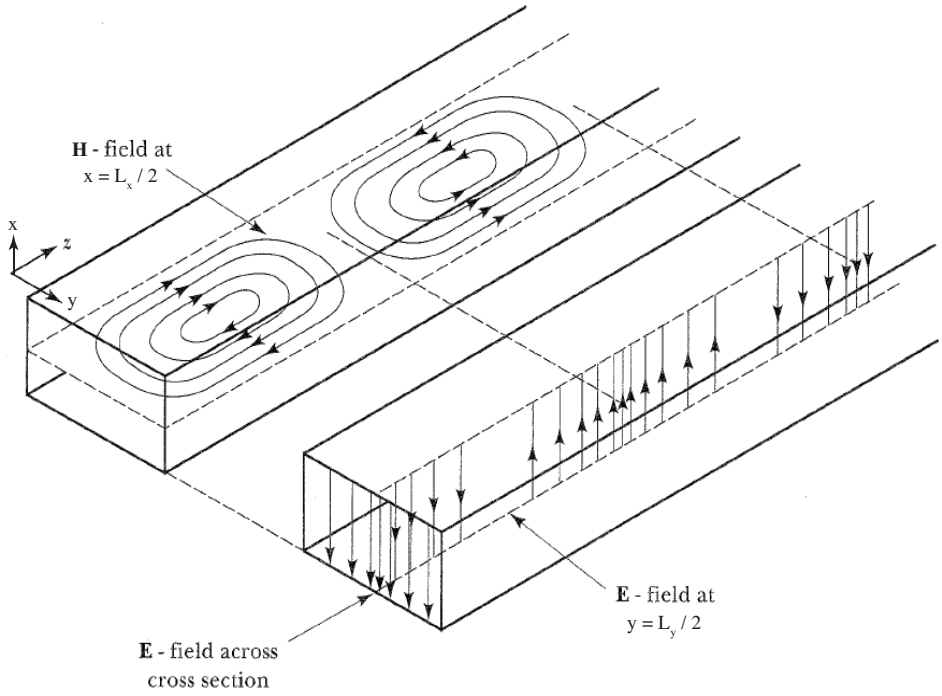

Figure 8.6: Fields of the fundamental TE_{01} mode. Adapted from *Heald & Marion, Classical Electromagnetic Radiation, Saunders College Publishing, 3rd ed. 1995.*

TM Modes

Following the discussion in Section 8.2.1, the electric field of a TM mode is

$$E_z^{xy}(x,y) \;=\; E_{0_z}\,\sin[\frac{n\pi}{L_x}x]\,\sin[\frac{m\pi}{L_y}y] \qquad n,m \in \{1,2\,..\}\,. \tag{8.54}$$

Mode indices $n = 0$ and $m = 0$ are forbidden because they lead to a zero-field solution. Thus, the lowest-order TM mode is the TM_{11} mode.

8.2.3 Optical Waveguides

Because of absorption, metal waveguides become lossy at very hight frequencies. Therefore, at optical frequencies ($200 - 800$ THz) it is more favorable to guide electromagnetic radiation in dielectric structures. The waveguiding principle is based on total internal reflection (TIR) between dielectric interfaces (see Section 4.4). In order for a wave to be totally reflected at a boundary between two dielectrics with

refractive indices n_1 and n_2 it must be incident from the optically denser medium ($n_1 > n_2$) and propagate at an angle larger than the critical angle $\theta_c = \arctan[n_2/n_1]$ measured from the surface normal. In contrast to metal waveguides, TIR in dielectric waveguides leads to evanescent fields that stretch out into the surrounding medium with the lower index of refraction n_2 (see Fig. 8.7). Thus, the boundary conditions at the interfaces become more complex.

Optical fibers are axially symmetric, that is, they consist of a dielectric rod of index n_1 (the core) surrounded by a medium of index n_2 (the cladding). For claddings of sufficient thickness the evanescent waves are strongly attenuated at the outer boundaries and it is reasonable to approximate the cladding radius as being infinite. The fields of an optical waveguide with circular cross section are described by cylindrical Bessel functions J_n of order n. To keep things simple, we will not analyze optical fibers with circular cross-sections. Instead, we focus on the dielectric slab waveguide that consists of a dielectric layer with index n_1 sandwiched between two infinite dielectrics of index $n_2 < n_1$. The mode structure is different but the physical principles are the same.

The waveguide fields have to satisfy the Helmholtz equation (3.15). Similar to Section 8.2.1 we will separate the fields into transverse electric (TE) and transverse magnetic (TM) solutions. In the former case the electric field has no vector component along the waveguide axis z, whereas in the latter case the same holds for the magnetic field. As illustrated in Fig. 8.7, a waveguide mode can be de-

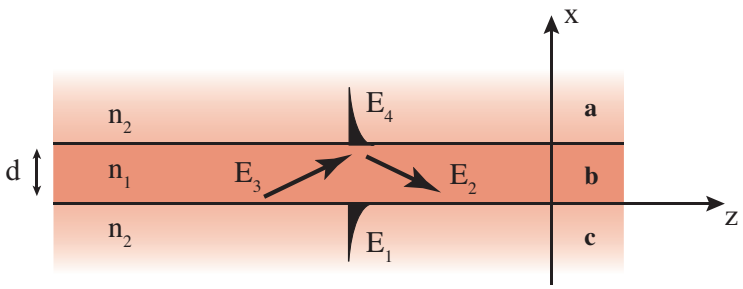

Figure 8.7: Optical step-index waveguide made of a dielectric with refractive index n_1 surrounded by a dielectric with lower index of refraction n_2. Waveguiding originates from a sequence of total internal reflections.

scribed by the superposition of four waves, namely two plane waves inside the waveguide ($0 < x < d$), an evanescent wave in the medium above the waveguide ($x > d$) and an evanescent wave below the waveguide ($x < 0$). The evanescent waves above and below the waveguide ensure that no energy is radiated away from the waveguide. Thus, the total field $\mathbf{E}$ is written in terms of partial fields $\mathbf{E}_1 .. \mathbf{E}_4$ as

$$\mathbf{E}(\mathbf{r}) = \begin{cases} \mathbf{E}_1(\mathbf{r}) & x < 0 \\ \mathbf{E}_2(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}) & 0 < x < d \\ \mathbf{E}_4(\mathbf{r}) & x > d \end{cases} \tag{8.55}$$

In the following we will discuss the TM and TE cases separately.

TM Modes

Denoting the $\mathbf{k}$ vector in the waveguide (medium 1) as $\mathbf{k}_1 = [k_{x_1}, 0, k_z]$ and outside the waveguide (medium 2) as $\mathbf{k}_2 = [k_{x_2}, 0, k_z]$, the partial fields of a TM mode are calculated as follows

$$\mathbf{E}_1 = E_1 \begin{pmatrix} k_z/k_2 \\ 0 \\ k_{x_2}/k_2 \end{pmatrix} \mathrm{e}^{-ik_{x_2}x + ik_z z}, \quad \mathbf{H}_1 = \frac{E_1}{Z_2} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \mathrm{e}^{-ik_{x_2}x + ik_z z} \tag{8.56}$$

$$\mathbf{E}_2 = E_2 \begin{pmatrix} k_z/k_1 \\ 0 \\ k_{x_1}/k_1 \end{pmatrix} \mathrm{e}^{-ik_{x_1}x + ik_z z}, \quad \mathbf{H}_2 = \frac{E_2}{Z_1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \mathrm{e}^{-ik_{x_1}x + ik_z z} \tag{8.57}$$

$$\mathbf{E}_3 = E_3 \begin{pmatrix} k_z/k_1 \\ 0 \\ -k_{x_1}/k_1 \end{pmatrix} \mathrm{e}^{ik_{x_1}x + ik_z z}, \quad \mathbf{H}_3 = \frac{E_3}{Z_1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \mathrm{e}^{ik_{x_1}x + ik_z z} \tag{8.58}$$

$$\mathbf{E}_4 = E_4 \begin{pmatrix} k_z/k_2 \\ 0 \\ -k_{x_2}/k_2 \end{pmatrix} \mathrm{e}^{ik_{x_2}x + ik_z z}, \quad \mathbf{H}_4 = \frac{E_4}{Z_2} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \mathrm{e}^{ik_{x_2}x + ik_z z} \tag{8.59}$$

Here we used the continuity of k_z and the transversality of the fields ($\nabla \cdot \mathbf{E} = i\mathbf{k} \cdot \mathbf{E} = 0$). To calculate the magnetic field we used Maxwell's curl equation (2.32) and assumed linear material equations (3.7). Note that in order for the fields to be evanescent outside the waveguide we require $k_z > k_2$. On the other hand, for the

fields to be propagating inside the waveguide we require $k_z < k_2$. Thus, waveguide modes will exist only in the interval $k_2 < k_z < k_1$.

Having defined the partial fields we know have to match them at the boundaries $x = 0$ and $x = d$. The continuity of the parallel components of electric and magnetic fields lead to

$$\begin{bmatrix} k_{x_2}/k_2 & -k_{x_1}/k_1 & k_{x_1}/k_1 & 0 \\ 1/Z_2 & -1/Z_1 & -1/Z_1 & 0 \\ 0 & k_{x_1}/k_1 \exp[-ik_{x_1}d] & -k_{x_1}/k_1 \exp[ik_{x_1}d] & k_{x_2}/k_2 \exp[ik_{x_2}d] \\ 0 & 1/Z_1 \exp[-ik_{x_1}d] & 1/Z_1 \exp[ik_{x_1}d] & -1/Z_2 \exp[ik_{x_2}d] \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix} = 0 \tag{8.60}$$

This is a homogeneous system of linear equations, that is, an eigenvalue problem. The reason why we end up with a *homogeneous* system of equations is the absence of an excitation, which means that there are no sources or incident fields. Thus, we are looking for solutions that the system supports in absence of an external driving force, similar to an undriven harmonic oscillator. A homogeneous system of equations has solutions only if the determinant of the matrix acting on the eigenvector $[E_1, E_2, E_3, E_4]^T$ vanishes, that is,

$$\begin{vmatrix} k_{x_2}/k_2 & -k_{x_1}/k_1 & k_{x_1}/k_1 & 0 \\ 1/Z_2 & -1/Z_1 & -1/Z_1 & 0 \\ 0 & k_{x_1}/k_1 \exp[-ik_{x_1}d] & -k_{x_1}/k_1 \exp[ik_{x_1}d] & k_{x_2}/k_2 \exp[ik_{x_2}d] \\ 0 & 1/Z_1 \exp[-ik_{x_1}d] & 1/Z_1 \exp[ik_{x_1}d] & -1/Z_2 \exp[ik_{x_2}d] \end{vmatrix} = 0 \tag{8.61}$$

Writing out the expression for the determinant and arranging terms we obtain

$$1 + \frac{(Z_2 k_{x_2}/k_2 - Z_1 k_{x_1}/k_1)}{(Z_2 k_{x_2}/k_2 + Z_1 k_{x_1}/k_1)} \frac{(Z_1\, k_{x_1}/k_1 - Z_2 k_{x_2}/k_2)}{(Z_1 k_{x_1}/k_1 + Z_2 k_{x_2}/k_2)} \exp[2ik_{x_1}d] = 0\,, \tag{8.62}$$

which can be written in the form

$$1 + r^p_{ab}(k_z)\, r^p_{bc}(k_z)\, \mathrm{e}^{2ik_{x_1}d} = 0 \tag{8.63}$$

Here, r^p_{ab} and r^p_{bc} are the Fresnel reflection coefficients for p polarization, as defined in Eq. (4.39). The subscript 'ab' indicates that the reflection is measured between the upper medium (medium a in Fig. 8.7) and the waveguide (medium b in Fig. 8.7). Similarly, the subscript 'bc' is the reflection coefficient measured

between the waveguide and the lower medium (medium 3 in Fig. 8.7). Note that

$$k_{x_1} = \sqrt{k_1^2 - k_z^2} = \sqrt{\frac{\omega^2}{c^2}\mu_1\varepsilon_1 - k_z^2}, \quad k_{x_2} = \sqrt{k_2^2 - k_z^2} = \sqrt{\frac{\omega^2}{c^2}\mu_2\varepsilon_2 - k_z^2}\,, \tag{8.64}$$

and hence Eq. (8.63) defines the characteristic equation for the eigenvalues k_z. Every solution for k_z defines a waveguide mode. It has to be emphasized that the sign of the square roots in the expressions for k_{x_1} and k_{x_2} has to be chosen such that the imaginary part is positive. This ensures that the fields decay exponentially with distance from the waveguide (evanescent fields). The other sign would imply an exponential increase, which would violate energy conservation.

TE Modes

A similar analysis can be applied for TE polarized fields, for which the electric field is parallel to the boundaries. The resulting characteristic equation for k_z is

$$1 \;+\; r^s_{ab}(k_z)\, r^s_{bc}(k_z)\, \mathrm{e}^{2ik_{x_1}d} \;=\; 0 \tag{8.65}$$

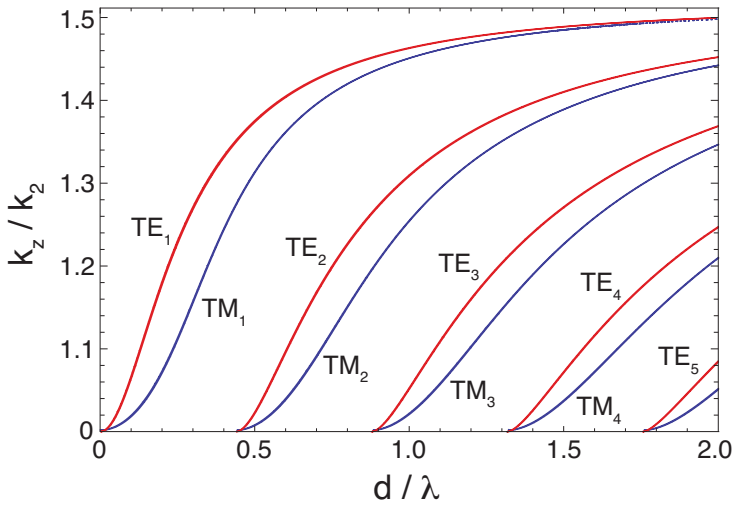

Figure 8.8: Mode structure of a dielectric slab waveguide with $n_1 = 1.5$ (glass) and $n_2 = 1$ (air). The vertical axis shows the real part of the normalized propagation constant k_z. There is no cut-off for the fundamental modes TM_1 and TE_1.

with r_{ab}^p and r_{bc}^p being the Fresnel reflection coefficients for s polarization.

As shown in Fig. 8.8, the mode structure for the dielectric waveguide is similar to the mode structure of a parallel-plate waveguide in Fig. 8.4. However, due to the different Fresnel reflection coefficients for TE and TM modes, the curves for TM and TE modes are no longer the same and split into separate curves. The fundamental modes TE_1 and TM_1 have no cut-off and in the limit $d/\lambda \to 0$ they become plane waves ($k_z = k_2$). Similarly, in the limit $d/\lambda \to \infty$ the modes become plane waves propagating in the higher dielectric ($k_z = k_1$).

8.2.4 Optical Fibers

In many regards the mode structure in optical fibers is similar to the mode structure of a dielectric slab waveguide. There are, however, also some relevant differences. First, an optical fiber is a waveguide that confines the electromagnetic field in two dimensions which, according to Section 8.2.2, leads to two mode indices nm. One of them specifies the angular dependence of the fields [$\sin m\phi$, $\cos m\phi$] and the other one the radial dependence [$J_n(\rho)$, $H_n^{(1)}(\rho)$]. Second, besides pure TE and

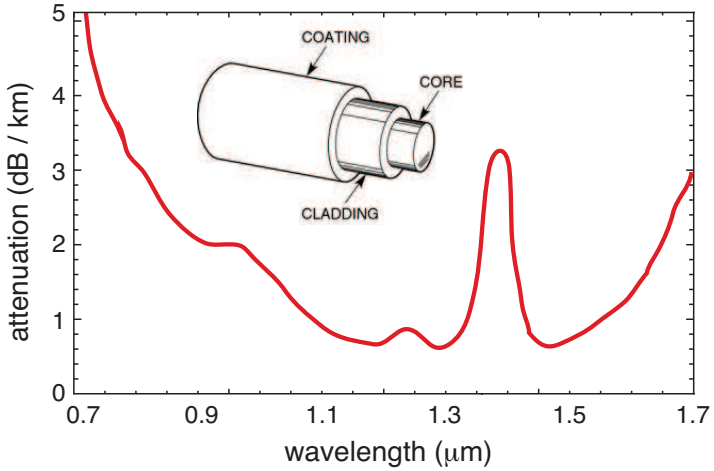

Figure 8.9: Attenuation in an optical fiber. Lowest propagation losses are obtained at wavelengths of $\sim 1.3\,\mu$m and $\sim 1.5\,\mu$m.

TM modes there are also hybrid modes, which are classified as HE and EH modes. HE modes have TE flavor whereas the EH modes have more TM character.

There are two basic versions of optical fibers: gradient index fibers and step-index fibers. The former has an index of refraction profile that varies gradually with radius [$n = n(\rho)$], whereas the latter exhibits an abrupt transitions between two refractive indices, similar to the slab waveguide. Step-index fibers are the most commonly used fibers and usually the index difference is very small, that is $(n_1 - n_2)/n_1 \ll 1$, which is referred to as the *weakly guiding* condition. Polarized weakly guided modes are denoted as LP_{nm} modes. A single-mode fiber supports only the fundamental LP_{01} mode and all higher-order modes are suppressed by the cut-off condition. Fig. 8.9 shows the propagation loss in a modern optical fiber. The losses on the short-wavelength side are due to Rayleigh scattering and the losses at the long wavelength end are due to infrared absorption. The bumps are absorption bands of molecular bonds, such as OH^-. The lowest attenuation is obtained near 1.3 and 1.5 μm and is the reason why fiber-optic communication is operated at these wavelengths.

Chapter 9

Classical Electron Theory

Chapter 10

Coherence Theory

Chapter 11

Relativistic Electromagnetics

Chapter 12

Antennas