

Chapter 2

Theoretical Foundations

Light embraces the most fascinating spectrum of electromagnetic radiation. This is mainly due to the fact that the energy of light quanta (photons) lie in the energy range of electronic transitions in matter. This gives us the beauty of color and is the reason why our eyes adapted to sense the optical spectrum.

Light is also fascinating because it manifests itself in forms of waves and particles. In no other range of the electromagnetic spectrum we are more confronted with the wave-particle duality than in the optical regime. While long wavelength radiation (radiofrequencies, microwaves) is well described by wave theory, short wavelength radiation (X-rays) exhibits mostly particle properties. The two worlds meet in the optical regime.

To describe optical radiation in nano-optics it is mostly sufficient to adapt the wave picture. This allows us to use classical field theory based on Maxwell's equations. Of course, in nano-optics, the systems with which the light fields interact are small (single molecules, quantum dots) which necessitates a quantum description of the material properties. Thus, in most cases we can use the framework of semiclassical theory which combines the classical picture of fields and the quantum picture of matter. However, occasionally, we have to go beyond the semiclassical description. For example the photons emitted by a quantum system can obey non-classical photon statistics in form of photon-antibunching.

This section summarizes the fundamentals of electromagnetic theory forming the necessary basis for this course. Only the basic properties are discussed and for more detailed treatments the reader is referred to standard textbooks on electromagnetism such as the books by Jackson [1], Stratton [2], and others. The starting point are Maxwell's equations established by James C. Maxwell in 1873.

2.1 Macroscopic Electrodynamics

In macroscopic electrodynamics the singular character of charges and their associated currents is avoided by considering charge densities ρ and current densities \mathbf{j} . In differential form and in *SI* units the *macroscopic* Maxwell's equations have the form

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}(\mathbf{r}, t)}{\partial t}, \quad (2.1)$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\partial \mathbf{D}(\mathbf{r}, t)}{\partial t} + \mathbf{j}(\mathbf{r}, t), \quad (2.2)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) = \rho(\mathbf{r}, t), \quad (2.3)$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0. \quad (2.4)$$

where \mathbf{E} denotes the electric field, \mathbf{D} the electric displacement, \mathbf{H} the magnetic field, \mathbf{B} the magnetic induction, \mathbf{j} the current density, and ρ the charge density. The components of these vector and scalar 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, Ampere, 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. It has to be emphasized 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'. Notice, that the macroscopic Maxwell equations deal with fields which are local spatial averages over microscopic fields associated with discrete charges. Hence, the microscopic nature of matter is not included in the macroscopic fields. Charge and current densities are considered as continuous functions of space. In order to describe the fields on an atomic scale it is necessary to use the microscopic Maxwell equations which consider all matter to be made of charged and uncharged particles.

The conservation of charge is implicitly contained in Maxwell's equations. Taking the divergence of Eq. (2.2), noting that $\nabla \cdot \nabla \times \mathbf{H}$ is identical zero, and substituting Eq. (2.3) for $\nabla \cdot \mathbf{D}$ one obtains the continuity equation

$$\nabla \cdot \mathbf{j}(\mathbf{r}, t) + \frac{\partial \rho(\mathbf{r}, t)}{\partial t} = 0. \quad (2.5)$$

The electromagnetic properties of the medium are most commonly discussed in terms of the macroscopic polarization \mathbf{P} and magnetization \mathbf{M} according to

$$\mathbf{D}(\mathbf{r}, t) = \varepsilon_o \mathbf{E}(\mathbf{r}, t) + \mathbf{P}(\mathbf{r}, t), \quad (2.6)$$

$$\mathbf{H}(\mathbf{r}, t) = \mu_o^{-1} \mathbf{B}(\mathbf{r}, t) - \mathbf{M}(\mathbf{r}, t), \quad (2.7)$$

where ε_o and μ_o are the permittivity and the permeability, respectively. These equations do not impose any conditions on the medium and are therefore always valid.

2.2 Wave equations

After substituting the fields \mathbf{D} and \mathbf{B} in Maxwell's *curl* equations by the expressions (2.6) and (2.7) and combining the two resulting equations we obtain the inhomogeneous wave equations

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_o \frac{\partial}{\partial t} \left(\mathbf{j} + \frac{\partial \mathbf{P}}{\partial t} + \nabla \times \mathbf{M} \right), \quad (2.8)$$

$$\nabla \times \nabla \times \mathbf{H} + \frac{1}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = \nabla \times \mathbf{j} + \nabla \times \frac{\partial \mathbf{P}}{\partial t} + \mu_o \frac{\partial \mathbf{M}}{\partial t}. \quad (2.9)$$

The constant c was introduced for $(\varepsilon_o \mu_o)^{-1/2}$ and is known as the vacuum speed of light. The expression in the brackets of Eq. (2.8) can be associated with the *total current density*

$$\mathbf{j}_t = \mathbf{j}_s + \mathbf{j}_c + \frac{\partial \mathbf{P}}{\partial t} + \nabla \times \mathbf{M}, \quad (2.10)$$

where \mathbf{j} has been split into a *source current density* \mathbf{j}_s and an induced *conduction current density* \mathbf{j}_c . The terms $\partial \mathbf{P} / \partial t$ and $\nabla \times \mathbf{M}$ are recognized as the *polarization current density* and the *magnetization current density*, respectively. The wave equations as stated in Eqs. (2.8), (2.9) do not impose any conditions on the media considered and hence are generally valid.

2.3 Constitutive relations

Maxwell's equations define the fields that are generated by currents and charges in matter. 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 which describe the behavior of matter under the influence of the fields. These material equations are known as . In a non-dispersive linear and isotropic medium they have the form

$$\mathbf{D} = \varepsilon_o \varepsilon \mathbf{E} \quad (\mathbf{P} = \varepsilon_o \chi_e \mathbf{E}), \quad (2.11)$$

$$\mathbf{B} = \mu_o \mu \mathbf{H} \quad (\mathbf{M} = \chi_m \mathbf{H}), \quad (2.12)$$

$$\mathbf{j}_c = \sigma \mathbf{E}. \quad (2.13)$$

For *non-linear* media, the right-hand sides can be supplemented by terms of higher power, but most materials behave linear if the fields are not too strong. *Anisotropic* media can be considered using tensorial forms for ε and μ . In order to account for general *bianisotropic* media, additional terms relating \mathbf{D} and \mathbf{E} to both \mathbf{B} and \mathbf{H} have to be introduced. For such complex media, solutions to the wave equations can be found for very special situations only. The constituent relations given above account for *inhomogeneous* media if the material parameters ε , μ and σ are functions of space. The medium is called *temporally dispersive* if the material parameters are functions of frequency, and *spatially dispersive* if the constitutive relations are convolutions over space. An electromagnetic field in a linear medium can be written as a superposition of monochromatic fields of the form*

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{k}, \omega) \cos(\mathbf{k} \cdot \mathbf{r} - \omega t), \quad (2.14)$$

where \mathbf{k} and ω are the wavevector and the angular frequency, respectively. In its most general form, the amplitude of the induced displacement $\mathbf{D}(\mathbf{r}, t)$ can be written as

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon_o \varepsilon(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega). \quad (2.15)$$

Since $\mathbf{E}(\mathbf{k}, \omega)$ is equivalent to the Fourier transform $\hat{\mathbf{E}}$ of an arbitrary time-dependent field $\mathbf{E}(\mathbf{r}, t)$, we can apply the inverse Fourier transform to Eq. (2.15) and obtain

$$\mathbf{D}(\mathbf{r}, t) = \varepsilon_o \iint \tilde{\varepsilon}(\mathbf{r} - \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t') d\mathbf{r}' dt'. \quad (2.16)$$

Here, $\tilde{\varepsilon}$ denotes the inverse Fourier transform (spatial and temporal) of ε . The above equation is a convolution in space and time. The displacement \mathbf{D} at time t depends on the electric field at all times t' previous to t (temporal dispersion). Additionally, the displacement at a point \mathbf{r} also depends on the values of the electric field at neighboring points \mathbf{r}' (spatial dispersion). A spatially dispersive medium is therefore also called a *nonlocal* medium. Nonlocal effects can be observed at interfaces between different media or in metallic objects with sizes comparable with the mean-free path of electrons. In general, it is very difficult to account for spatial dispersion in field calculations. 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 accurately take it into account.

2.4 Spectral representation of time-dependent fields

The spectrum $\hat{\mathbf{E}}(\mathbf{r}, \omega)$ of an arbitrary time-dependent field $\mathbf{E}(\mathbf{r}, t)$ is defined by the Fourier transform

$$\hat{\mathbf{E}}(\mathbf{r}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, t) e^{i\omega t} dt. \quad (2.17)$$

*In an anisotropic medium the dielectric constant $\varepsilon = \vec{\varepsilon}$ is a second-rank tensor.

In order that $\mathbf{E}(\mathbf{r}, t)$ is a real valued field we have to require that

$$\hat{\mathbf{E}}(\mathbf{r}, -\omega) = \hat{\mathbf{E}}^*(\mathbf{r}, \omega) . \quad (2.18)$$

Applying the Fourier transform to the time-dependent Maxwell equations (2.1) - (2.4) gives

$$\nabla \times \hat{\mathbf{E}}(\mathbf{r}, \omega) = i\omega \hat{\mathbf{B}}(\mathbf{r}, \omega) , \quad (2.19)$$

$$\nabla \times \hat{\mathbf{H}}(\mathbf{r}, \omega) = -i\omega \hat{\mathbf{D}}(\mathbf{r}, \omega) + \hat{\mathbf{j}}(\mathbf{r}, \omega) , \quad (2.20)$$

$$\nabla \cdot \hat{\mathbf{D}}(\mathbf{r}, \omega) = \hat{\rho}(\mathbf{r}, \omega) , \quad (2.21)$$

$$\nabla \cdot \hat{\mathbf{B}}(\mathbf{r}, \omega) = 0 , \quad (2.22)$$

Once the solution for $\hat{\mathbf{E}}(\mathbf{r}, \omega)$ has been determined, the time-dependent field is calculated by the inverse transform as

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

Thus, the time-dependence of a non-harmonic electromagnetic field can be Fourier transformed and every spectral component can be treated separately as a monochromatic field. The general time dependence is obtained from the inverse transform.

2.5 Time harmonic fields

The time dependence in the wave equations can be easily separated to obtain a harmonic differential equation. A monochromatic field can then be written as[†]

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}\{\mathbf{E}(\mathbf{r}) e^{-i\omega t}\} = \frac{1}{2} [\mathbf{E}(\mathbf{r}) e^{-i\omega t} + \mathbf{E}^*(\mathbf{r}) e^{i\omega t}] , \quad (2.24)$$

with similar expressions for the other fields. Notice that $\mathbf{E}(\mathbf{r}, t)$ is real, whereas the spatial part $\mathbf{E}(\mathbf{r})$ is complex. 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. It is convenient to represent the fields of a time-harmonic field by their complex amplitudes. Maxwell's equations can then be written as

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega \mathbf{B}(\mathbf{r}) , \quad (2.25)$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega \mathbf{D}(\mathbf{r}) + \mathbf{j}(\mathbf{r}) , \quad (2.26)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r}) , \quad (2.27)$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0 , \quad (2.28)$$

[†]This can also be written as $\mathbf{E}(\mathbf{r}, t) = \text{Re}\{\mathbf{E}(\mathbf{r})\} \cos \omega t + \text{Im}\{\mathbf{E}(\mathbf{r})\} \sin \omega t = |\mathbf{E}(\mathbf{r})| \cos [\omega t + \varphi(\mathbf{r})]$, where the phase is determined by $\varphi(\mathbf{r}) = \arctan [\text{Im}\{\mathbf{E}(\mathbf{r})\}/\text{Re}\{\mathbf{E}(\mathbf{r})\}]$

which is equivalent to Maxwell's equations (2.19 - 2.22) for the spectra of arbitrary time-dependent fields. Thus, the solution for $\mathbf{E}(\mathbf{r})$ is equivalent to the spectrum $\hat{\mathbf{E}}(\mathbf{r}, \omega)$ of an arbitrary time-dependent field. It is obvious that the complex field amplitudes depend on the angular frequency ω , i.e. $\mathbf{E}(\mathbf{r}) = \mathbf{E}(\mathbf{r}, \omega)$. However, ω is usually not included in the argument. Also the material parameters ε , μ , and σ are functions of space and frequency, i.e. $\varepsilon = \varepsilon(\mathbf{r}, \omega)$, $\sigma = \sigma(\mathbf{r}, \omega)$, $\mu = \mu(\mathbf{r}, \omega)$. For simpler notation, we will often drop the argument in the fields and material parameters. It is the context of the problem which determines which of the fields $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{E}(\mathbf{r})$, or $\hat{\mathbf{E}}(\mathbf{r}, \omega)$ is being considered.

2.6 Complex dielectric constant

With the help of the linear constitutive relations we can express Maxwell's curl equations (2.25) and (2.26) in terms of $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$. We then multiply both sides of the first equation by μ^{-1} and then apply the curl operator to both sides. After the expression $\nabla \times \mathbf{H}$ is substituted by the second equation we obtain

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E} - \frac{\omega^2}{c^2} [\varepsilon + i\sigma/(\omega\varepsilon_0)] \mathbf{E} = i\omega\mu_0 \mathbf{j}_s. \quad (2.29)$$

It is common practice to replace the expression in the brackets on the left hand side by a complex dielectric constant, i.e.

$$[\varepsilon + i\sigma/(\omega\varepsilon_0)] \rightarrow \varepsilon. \quad (2.30)$$

In this notation one does not distinguish between conduction currents and polarization currents. Energy dissipation is associated with the imaginary part of the dielectric constant. With the new definition of ε , the wave equations for the complex fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ in linear, isotropic, but *inhomogeneous* media are

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E} - k_o^2 \varepsilon \mathbf{E} = i\omega\mu_0 \mathbf{j}_s, \quad (2.31)$$

$$\nabla \times \varepsilon^{-1} \nabla \times \mathbf{H} - k_o^2 \mu \mathbf{H} = \nabla \times \varepsilon^{-1} \mathbf{j}_s, \quad (2.32)$$

where $k_o = \omega/c$ denotes the vacuum wave number. These equations are also valid for anisotropic media if the substitutions $\varepsilon \rightarrow \vec{\varepsilon}$ and $\mu \rightarrow \vec{\mu}$ are performed. The complex dielectric constant will be used throughout this book.

2.7 Piecewise homogeneous media

In many physical situations the medium is piecewise homogeneous. In this case the entire space is divided into subdomains 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. (2.31), (2.32). However, the inhomogeneities are entirely confined to the boundaries and it is convenient to formulate the solution for each subdomain 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 domains D_i and D_j be denoted as ∂D_{ij} . If ε_i and μ_i designate the constant material parameters in subdomain D_i , the wave equations in that domain read as

$$(\nabla^2 + k_i^2) \mathbf{E}_i = -i\omega\mu_o\mu_i \mathbf{j}_i + \frac{\nabla\rho_i}{\varepsilon_o\varepsilon_i}, \quad (2.33)$$

$$(\nabla^2 + k_i^2) \mathbf{H}_i = -\nabla \times \mathbf{j}_i, \quad (2.34)$$

where $k_i = (\omega/c)\sqrt{\mu_i\varepsilon_i}$ is the wavenumber and \mathbf{j}_i , ρ_i the sources in domain D_i . To obtain these equations, the identity $\nabla \times \nabla \times = -\nabla^2 + \nabla\nabla\cdot$ was used and Maxwell's equation (2.3) was applied. Eqs. (2.33) and (2.34) 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.

2.8 Boundary conditions

Since the material properties are discontinuous on the boundaries, Eqs. (2.33) and (2.34) are only valid in the interior of the subdomains. However, Maxwell's equations must also hold for the boundaries. Due to the discontinuity it turns out to be difficult to apply the differential forms of Maxwell's equations but there is no problem with the corresponding integral forms. The latter can be derived by applying the theorems of Gauss and Stokes to the differential forms (2.1 - 2.4) and read as

$$\int_{\partial S} \mathbf{E}(\mathbf{r}, t) \cdot d\mathbf{s} = - \int_S \frac{\partial}{\partial t} \mathbf{B}(\mathbf{r}, t) \cdot \mathbf{n}_s da, \quad (2.35)$$

$$\int_{\partial S} \mathbf{H}(\mathbf{r}, t) \cdot d\mathbf{s} = \int_S \left[\mathbf{j}(\mathbf{r}, t) + \frac{\partial}{\partial t} \mathbf{D}(\mathbf{r}, t) \right] \cdot \mathbf{n}_s da, \quad (2.36)$$

$$\int_{\partial V} \mathbf{D}(\mathbf{r}, t) \cdot \mathbf{n}_s da = \int_V \rho(\mathbf{r}, t) dV, \quad (2.37)$$

$$\int_{\partial V} \mathbf{B}(\mathbf{r}, t) \cdot \mathbf{n}_s da = 0. \quad (2.38)$$

In these equations, da denotes a surface element, \mathbf{n}_s the normal unit vector to the surface, $d\mathbf{s}$ a line element, ∂V the surface of the volume V , and ∂S the border of the surface S . The integral forms of Maxwell equations lead to the desired boundary conditions if they are applied to a sufficiently small part of the considered boundary.

In this case the boundary looks flat and the fields are homogeneous on both sides [Fig. 2.1]. Consider a small rectangular path ∂S along the the boundary as shown in Fig. 2.1a. As the area S (enclosed by the path ∂S) is arbitrarily reduced, the electric and magnetic flux through S become zero. This does not necessarily apply for the source current, since a surface current density \mathbf{K} might be present. The first two Maxwell equations then lead to the boundary conditions for the tangential field components[‡]

$$\mathbf{n} \times (\mathbf{E}_i - \mathbf{E}_j) = \mathbf{0} \quad \text{on } \partial D_{ij}, \quad (2.39)$$

$$\mathbf{n} \times (\mathbf{H}_i - \mathbf{H}_j) = \mathbf{K} \quad \text{on } \partial D_{ij}, \quad (2.40)$$

where \mathbf{n} is the unit normal vector on the boundary. A relation for the normal components can be obtained by considering an infinitesimal rectangular box with volume V and surface ∂V according to Fig. 2.1b. If the fields are considered to be homogeneous on both sides and if a surface charge density σ is assumed, Maxwell's third and fourth equation lead to the boundary conditions for the normal field components

$$\mathbf{n} \cdot (\mathbf{D}_i - \mathbf{D}_j) = \sigma \quad \text{on } \partial D_{ij} \quad (2.41)$$

$$\mathbf{n} \cdot (\mathbf{B}_i - \mathbf{B}_j) = 0 \quad \text{on } \partial D_{ij}. \quad (2.42)$$

In most practical situations there are no sources in the individual domains, and \mathbf{K} and σ consequently vanish. The four boundary conditions (2.39 - 2.41) are not independent from each other since the fields on both sides of ∂D_{ij} are linked by Maxwell's

[‡]Notice, that \mathbf{n} and \mathbf{n}_s are different unit vectors: \mathbf{n}_s is perpendicular to the surfaces S and ∂V , whereas \mathbf{n} is perpendicular to the boundary ∂D_{ij} .

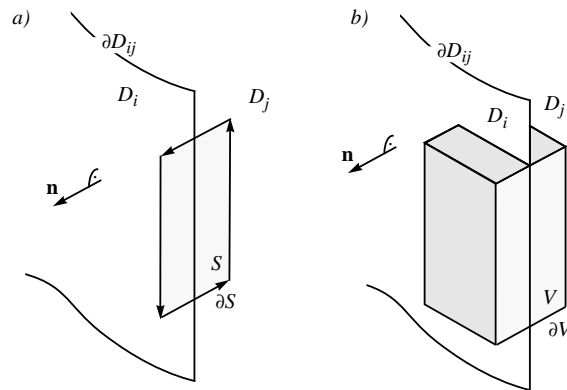


Figure 2.1: Integration paths for the derivation of the boundary conditions on the interface ∂D_{ij} between two adjacent domains D_i and D_j .

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 domains.

2.8.1 Fresnel reflection and transmission coefficients

Applying the boundary conditions to a simple plane wave incident on a single planar interface leads to the familiar Fresnel reflection and transmission coefficients. A detailed derivation can be found in many textbooks, e.g. [3], page 36 ff. We only mention here briefly the results.

An arbitrarily polarized plane wave $\mathbf{E}_1 \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t)$ can always be written as the superposition of two orthogonally polarized plane waves. It is convenient to choose these polarizations parallel or perpendicular to the *plane of incidence* defined by the \mathbf{k} -vector of the plane wave and the surface normal \mathbf{n} of the plane interface

$$\mathbf{E}_1 = \mathbf{E}_1^{(s)} + \mathbf{E}_1^{(p)}. \quad (2.43)$$

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

As shown in Fig. 2.2, we denote the dielectric constants of the medium of incidence and the medium of transmittance as ε_1 and ε_2 , respectively. The same designation applies to the magnetic permeability μ . Similarly, we distinguish between incident and

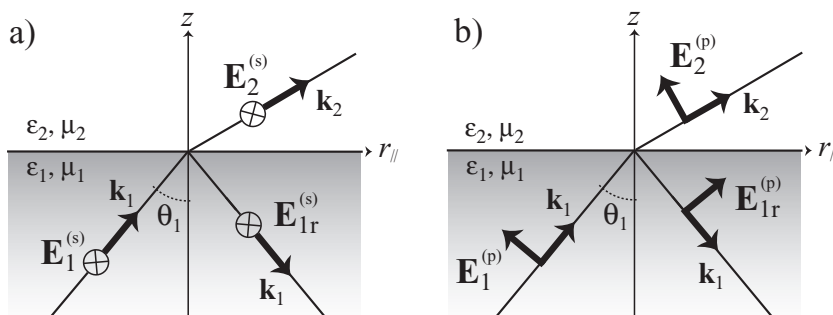


Figure 2.2: Reflection and refraction of a plane wave at a plane interface. a) s-polarization, and b) p-polarization.

transmitted wavevectors \mathbf{k}_1 and \mathbf{k}_2 . Using the coordinate system shown in Fig. 2.2, it follows from the boundary conditions that

$$\mathbf{k}_1 = (k_x, k_y, k_{z_1}), \quad |\mathbf{k}_1| = k_1 = \frac{\omega}{c} \sqrt{\varepsilon_1 \mu_1} \quad (2.44)$$

$$\mathbf{k}_2 = (k_x, k_y, k_{z_2}), \quad |\mathbf{k}_2| = k_2 = \frac{\omega}{c} \sqrt{\varepsilon_2 \mu_2}. \quad (2.45)$$

Thus, the transverse components of the wavevector (k_x, k_y) are conserved and the magnitude of the longitudinal wavenumbers are given by

$$k_{z_1} = \sqrt{k_1^2 - (k_x^2 + k_y^2)}, \quad k_{z_2} = \sqrt{k_2^2 - (k_x^2 + k_y^2)}. \quad (2.46)$$

The transverse wavenumber $k_{\parallel} = \sqrt{k_x^2 + k_y^2}$ can be conveniently expressed in terms of the angle of incidence θ_1 as

$$k_{\parallel} = \sqrt{k_x^2 + k_y^2} = k_1 \sin \theta_1, \quad (2.47)$$

which, according to Eqs. (2.46), also allows us to express k_{z_1} and k_{z_2} in terms of θ_1 .

It follows from the boundary conditions that the amplitudes of the reflected and transmitted waves can be represented as

$$\begin{aligned} \mathbf{E}_{1r}^{(s)} &= \mathbf{E}_1^{(s)} r^s(k_x, k_y), & \mathbf{E}_{1r}^{(p)} &= \mathbf{E}_1^{(p)} r^p(k_x, k_y) \\ \mathbf{E}_2^{(s)} &= \mathbf{E}_1^{(s)} t^s(k_x, k_y), & \mathbf{E}_2^{(p)} &= \mathbf{E}_1^{(p)} t^p(k_x, k_y), \end{aligned} \quad (2.48)$$

where the Fresnel reflection and transmission coefficients are defined as[§]

$$r^s(k_x, k_y) = \frac{\mu_2 k_{z_1} - \mu_1 k_{z_2}}{\mu_2 k_{z_1} + \mu_1 k_{z_2}}, \quad r^p(k_x, k_y) = \frac{\varepsilon_2 k_{z_1} - \varepsilon_1 k_{z_2}}{\varepsilon_2 k_{z_1} + \varepsilon_1 k_{z_2}} \quad (2.49)$$

$$t^s(k_x, k_y) = \frac{2\mu_2 k_{z_1}}{\mu_2 k_{z_1} + \mu_1 k_{z_2}}, \quad t^p(k_x, k_y) = \frac{2\varepsilon_2 k_{z_1}}{\varepsilon_2 k_{z_1} + \varepsilon_1 k_{z_2}} \sqrt{\frac{\mu_2 \varepsilon_1}{\mu_1 \varepsilon_2}}. \quad (2.50)$$

As indicated by the superscripts, these coefficients depend on the polarization of the incident plane wave. The coefficients are functions of k_{z_1} and k_{z_2} which can be expressed in terms of k_x, k_y and thus in terms of the angle of incidence θ_1 . The sign of the Fresnel coefficients depends on the definition of the electric field vectors shown in Fig. 2.2. 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 section 2.11.

[§]For symmetry reasons, some authors 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*. We adopt the definition from Born&Wolf [3].

2.9 Conservation of energy

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 or forces in a system. The basic Lorentz' law describes the forces acting on moving charges only. As the Abraham-Minkowski controversy shows, the forces acting on an arbitrary object cannot be extracted from a given electrodynamic field in a consistent way. It is also interesting, that Coulomb's and Ampere's laws were sufficient to establish Lorentz' force law. While later the field equations have been completed by adding the Maxwell displacement current, the Lorentz law remained unchanged. There is less controversy regarding the energy. Although also not a direct consequence of Maxwell's equations, Poynting's theorem provides a plausible relationship between the electromagnetic field and its energy content. For later reference, Poynting's theorem shall be outlined below.

If the scalar product of the field \mathbf{E} and Eq. (2.2) is subtracted from the scalar product of the field \mathbf{H} and Eq. (2.1) the following equation is obtained

$$\mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) = -\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \mathbf{j} \cdot \mathbf{E}. \quad (2.51)$$

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

$$\int_{\partial V} (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{n} \, da = - \int_V \left[\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{j} \cdot \mathbf{E} \right] dV. \quad (2.52)$$

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 (2.6) and (2.7). Eq. (2.52) then reads

$$\begin{aligned} \int_{\partial V} (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{n} \, da + \frac{1}{2} \frac{\partial}{\partial t} \int_V [\mathbf{D} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{H}] dV = \\ - \int_V \mathbf{j} \cdot \mathbf{E} dV - \frac{1}{2} \int_V \left[\mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} - \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial t} \right] dV - \frac{\mu_0}{2} \int_V \left[\mathbf{H} \cdot \frac{\partial \mathbf{M}}{\partial t} - \mathbf{M} \cdot \frac{\partial \mathbf{H}}{\partial t} \right] dV. \end{aligned} \quad (2.53)$$

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. It states that the first term is equal to the net energy flow in or out of the volume V , the second term is equal to the time rate of change of electromagnetic energy inside V and the remaining terms on the right side are equal to the rate of energy dissipation inside V . According to this interpretation

$$\mathbf{S} = (\mathbf{E} \times \mathbf{H}) \quad (2.54)$$

represents the energy flux density and

$$W = \frac{1}{2} [\mathbf{D} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{H}] \quad (2.55)$$

is the density of electromagnetic energy. If the medium within V is linear, the two last terms equal zero and the only term accounting for energy dissipation is $\mathbf{j} \cdot \mathbf{E}$. Hence, the two last terms can be associated with non-linear losses. The vector \mathbf{S} is denoted as the Poynting vector. In principle, the curl of any vector field can be added to \mathbf{S} without changing the conservation law (2.53), but it is convenient to make the choice as stated in (2.54). Notice that the current \mathbf{j} in Eq. (2.53) is the current associated with energy dissipation and therefore does not include polarization and magnetization currents.

Of special interest is the mean time value of \mathbf{S} . This quantity describes the net power flux density and is needed for the evaluation of radiation patterns. Assuming that the fields are harmonic in time and that the media are linear, the time average of Eq. (2.53) becomes

$$\int_{\partial V} \langle \mathbf{S} \rangle \cdot \mathbf{n} \, da = -\frac{1}{2} \int_V \text{Re} \{ \mathbf{j}^* \cdot \mathbf{E} \} \, dV, \quad (2.56)$$

where 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

$$\langle \mathbf{S} \rangle = \frac{1}{2} \text{Re} \{ \mathbf{E} \times \mathbf{H}^* \}. \quad (2.57)$$

In the far-field, the electromagnetic field is purely transverse. Furthermore, the electric and magnetic fields are in phase and the ratio of their amplitudes is constant. In this case $\langle \mathbf{S} \rangle$ can be expressed by the electric field alone as

$$\langle \mathbf{S} \rangle = \frac{1}{2} \sqrt{\frac{\epsilon_o \epsilon}{\mu_o \mu}} |\mathbf{E}|^2 \mathbf{n}_r, \quad (2.58)$$

where \mathbf{n}_r represents the unit vector in radial direction and the inverse of the square root denotes the wave impedance.

2.10 Dyadic Green's functions

An important concept in field theory are Green's functions: the fields due to a point source. In electromagnetic theory, the dyadic Green's function $\vec{\mathbf{G}}$ is essentially defined by the electric field \mathbf{E} at the fieldpoint \mathbf{r} generated by a radiating electric dipole $\boldsymbol{\mu}$ located at the source point \mathbf{r}' . In mathematical terms this reads as

$$\mathbf{E}(\mathbf{r}) = \omega^2 \mu_o \mu \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \boldsymbol{\mu} . \quad (2.59)$$

To understand the basic idea of Green's functions we will first consider a general mathematical point of view.

2.10.1 Mathematical basis of Green's functions

Consider the following general, inhomogeneous equation:

$$\mathcal{L} \mathbf{A}(\mathbf{r}) = \mathbf{B}(\mathbf{r}) . \quad (2.60)$$

\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}_o) is known. We thus need to solve for an arbitrary particular solution.

Usually it is difficult to find a solution of Eq. (2.60) 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

$$\mathcal{L} \mathbf{G}_i(\mathbf{r}, \mathbf{r}') = \mathbf{n}_i \delta(\mathbf{r} - \mathbf{r}') \quad (i=x, y, z) , \quad (2.61)$$

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. (2.61) can be written in closed form as

$$\mathcal{L} \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \vec{\mathbf{I}} \delta(\mathbf{r} - \mathbf{r}') , \quad (2.62)$$

where the operator \mathcal{L} acts on each column of $\vec{\mathbf{G}}$ separately and $\vec{\mathbf{I}}$ is the unit dyad. The dyadic function $\vec{\mathbf{G}}$ fulfilling Eq. (2.62) is known as the dyadic Green's function.

In a next step, assume that Eq. (2.62) has been solved and that $\vec{\mathbf{G}}$ is known. Postmultiplying Eq. (2.62) with $\mathbf{B}(\mathbf{r}')$ on both sides and integrating over the volume

V in which $\mathbf{B} \neq 0$ gives

$$\int_V \mathcal{L} \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{B}(\mathbf{r}') dV' = \int_V \mathbf{B}(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') dV' . \quad (2.63)$$

The right hand side simply reduces to $\mathbf{B}(\mathbf{r})$ and with Eq. (2.60) it follows that

$$\mathcal{L} \mathbf{A}(\mathbf{r}) = \int_V \mathcal{L} \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{B}(\mathbf{r}') dV' . \quad (2.64)$$

If on the right hand side the operator \mathcal{L} is taken out of the integral, the solution of Eq. (2.60) can be expressed as

$$\mathbf{A}(\mathbf{r}) = \int_V \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{B}(\mathbf{r}') dV' . \quad (2.65)$$

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

The assumption that the operators \mathcal{L} and $\int dV'$ can be interchanged is not strictly valid and special care must be applied if the integrand is not well behaved. Most often $\vec{\mathbf{G}}$ is singular at $\mathbf{r} = \mathbf{r}'$ and an infinitesimal exclusion volume surrounding $\mathbf{r} = \mathbf{r}'$ has to be introduced [4, 5]. Depolarization of the principal volume must be treated separately resulting in a term ($\vec{\mathbf{L}}$) depending on the geometrical shape of the volume. Furthermore, in numerical schemes the principal volume has a finite size giving rise to a second correction term designated by $\vec{\mathbf{M}}$. 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. However, the topic of the principal volume will be taken up in later chapters.

2.10.2 Derivation of the Green's function for the electric field

The derivation of the Green's function for the electric field is most conveniently accomplished by considering the time-harmonic vector potential \mathbf{A} and the scalar potential ϕ in an infinite and homogeneous space which is characterized by the constants ε and μ . In this case, \mathbf{A} and ϕ are defined by the relationships

$$\mathbf{E}(\mathbf{r}) = i\omega \mathbf{A}(\mathbf{r}) - \nabla \phi(\mathbf{r}) \quad (2.66)$$

$$\mathbf{H}(\mathbf{r}) = \frac{1}{\mu_o \mu} \nabla \times \mathbf{A}(\mathbf{r}) . \quad (2.67)$$

We can insert these equations into Maxwell's second equation Eq. (2.2) and obtain

$$\nabla \times \nabla \times \mathbf{A}(\mathbf{r}) = \mu_o \mu \mathbf{j}(\mathbf{r}) - i\omega \mu_o \mu \varepsilon_o \varepsilon [i\omega \mathbf{A}(\mathbf{r}) - \nabla \phi(\mathbf{r})] , \quad (2.68)$$

where we used $\mathbf{D} = \varepsilon_o \varepsilon \mathbf{E}$. The potentials \mathbf{A} , ϕ are not uniquely defined by Eqs. (2.66, 2.67). We are still free to define the value of $\nabla \cdot \mathbf{A}$ which we choose as

$$\nabla \cdot \mathbf{A}(\mathbf{r}) = i\omega\mu_o\mu\varepsilon_o\varepsilon\phi(\mathbf{r}). \quad (2.69)$$

A condition which fixes the redundancy of Eqs. (2.66, 2.67) is called a gauge condition. The gauge chosen through Eq. (2.69) is the so-called Lorentz gauge. Using the mathematical identity ' $\nabla \times \nabla \times = -\nabla^2 + \nabla \nabla \cdot$ ' together with the Lorentz gauge we can rewrite Eq. (2.68) as

$$[\nabla^2 + k^2] \mathbf{A}(\mathbf{r}) = -\mu_o\mu\mathbf{j}(\mathbf{r}), \quad (2.70)$$

which is the inhomogeneous Helmholtz equation. It holds independently for each component A_i of \mathbf{A} . A similar equation can be derived for the scalar potential ϕ

$$[\nabla^2 + k^2] \phi(\mathbf{r}) = -\rho(\mathbf{r})/\varepsilon_o\varepsilon. \quad (2.71)$$

Thus, we obtain four scalar Helmholtz equations of the form

$$[\nabla^2 + k^2] f(\mathbf{r}) = -g(\mathbf{r}). \quad (2.72)$$

To derive the *scalar* Green's function $G_o(\mathbf{r}, \mathbf{r}')$ for the Helmholtz operator we replace the source term $g(\mathbf{r})$ by a single point source $\delta(\mathbf{r} - \mathbf{r}')$ and obtain

$$[\nabla^2 + k^2] G_o(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \quad (2.73)$$

The coordinate \mathbf{r} denotes the location of the field point, i.e. the point in which the fields are to be evaluated, whereas the coordinate \mathbf{r}' designates the location of the point source. Once we have determined G_o we can state the particular solution for the vector potential in Eq. (2.70) as

$$\mathbf{A}(\mathbf{r}) = \mu_o\mu \int_V \mathbf{j}(\mathbf{r}') G_o(\mathbf{r}, \mathbf{r}') dV'. \quad (2.74)$$

A similar equation holds for the scalar potential. Both solutions require the knowledge of the Green's function defined through Eq. (2.73). In free space, the only physical solution of this equation is [1]

$$\boxed{G_o(\mathbf{r}, \mathbf{r}') = \frac{e^{\pm ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}}. \quad (2.75)$$

The solution with the plus sign denotes a spherical wave that propagates out of the origin whereas the solution with the minus sign is a wave that converges towards the origin. In the following we only retain the outwards propagating wave. The scalar Green's function can be introduced into Eq. (2.74) and the vector potential can be

calculated by integrating over the source volume V . Thus, we are in a position to calculate the vector potential and scalar potential for any given current distribution \mathbf{j} and charge distribution ρ . Notice, that the Green's function in Eq. (2.75) applies only to a homogeneous three-dimensional space. The Green's function of a two-dimensional space or a half-space will have a different form.

So far we reduced the treatment of Green's functions to the potentials \mathbf{A} and ϕ because it allows us to work with scalar equations. The formalism becomes more involved when we consider the electric and magnetic fields. The reason for this is that a source current in x -direction leads to an electric and magnetic field with x , y , and z -components. This is different for the vector potential: a source current in x gives only rise to a vector potential with a x component. Thus, in the case of the electric and magnetic fields we need a Green's function which relates all components of the source with all components of the fields, or, in other words, the Green's function must be a tensor. This type of Green's function is denoted as *dyadic Green's function* and has been introduced in the previous section. To determine the dyadic Green's function we start with the waveequation for the electric field Eq. (2.31). In a homogeneous space it reads as

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k^2 \mathbf{E}(\mathbf{r}) = i\omega\mu_0\mu \mathbf{j}(\mathbf{r}) . \quad (2.76)$$

We can define for each component of \mathbf{j} a corresponding Green's function. For example, for j_x we have

$$\nabla \times \nabla \times \mathbf{G}_x(\mathbf{r}, \mathbf{r}') - k^2 \mathbf{G}_x(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \mathbf{n}_x , \quad (2.77)$$

where \mathbf{n}_x is the unit vector in x -direction. A similar equation can be formulated for a point source in y and z -direction. In order to account for all orientations we write

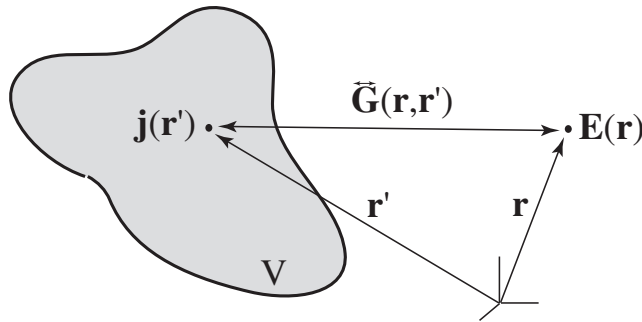


Figure 2.3: Illustration of the dyadic Green's function $\vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$. The Green's function renders the electric field at the field point \mathbf{r} due to a single point source \mathbf{j} at the source point \mathbf{r}' . Since the field at r depends on the orientation of \mathbf{j} the Green's function must account for all possible orientations in the form of a tensor.

as the general definition of the dyadic Green's function for the electric field [6]

$$\nabla \times \nabla \times \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - k^2 \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \vec{\mathbf{I}} \delta(\mathbf{r} - \mathbf{r}'), \quad (2.78)$$

$\vec{\mathbf{I}}$ being the unit dyad (unit tensor). The first column of the tensor $\vec{\mathbf{G}}$ corresponds to the field due to a point source in x -direction, the second column to the field due to a point source in y -direction, and the third column is the field due to a point source in z -direction. Thus a dyadic Green's function is just a compact notation for three vectorial Green's functions.

As before, we can view the source current in Eq. (2.76) as a superposition of point currents. Thus, if we know the Greens' function $\vec{\mathbf{G}}$ we can state a particular solution of Eq. (2.76) as

$$\mathbf{E}(\mathbf{r}) = i\omega\mu\mu_o \int_V \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{j}(\mathbf{r}') dV'. \quad (2.79)$$

However, this is a particular solution and we need to add any homogeneous solutions \mathbf{E}_o . Thus, the general solution turns out to be

$$\boxed{\mathbf{E}(\mathbf{r}) = \mathbf{E}_o(\mathbf{r}) + i\omega\mu_o\mu \int_V \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{j}(\mathbf{r}') dV' \quad \mathbf{r} \notin V}. \quad (2.80)$$

The corresponding magnetic field reads as

$$\boxed{\mathbf{H}(\mathbf{r}) = \mathbf{H}_o(\mathbf{r}) + \int_V \left[\nabla \times \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \right] \mathbf{j}(\mathbf{r}') dV' \quad \mathbf{r} \notin V}. \quad (2.81)$$

These equations are denoted as *volume integral equations*. They are very important since they form the basis for various formalisms such as the 'method of moments', the 'Lippmann-Schwinger equation', or the 'coupled dipole method'. We have limited the validity of the volume integral equations to the space outside the source volume V in order to avoid the apparent singularity of $\vec{\mathbf{G}}$ at $\mathbf{r} = \mathbf{r}'$. This limitation will be relaxed in later chapters.

In order to solve Eqs. (2.80) and (2.81) for a given distribution of currents, we still need to determine the value of $\vec{\mathbf{G}}$. Introducing the Lorentz gauge Eq. (2.69) into Eq. (2.66) leads to

$$\mathbf{E}(\mathbf{r}) = i\omega \left[1 + \frac{1}{k^2} \nabla \nabla \cdot \right] \mathbf{A}(\mathbf{r}). \quad (2.82)$$

The first column vector of $\vec{\mathbf{G}}$, i.e. \mathbf{G}_x , defined in Eq. (2.77) is simply the electric field due to a point source current $\mathbf{j} = (i\omega\mu_o)^{-1} \delta(\mathbf{r} - \mathbf{r}') \mathbf{n}_x$. The vector potential originating from this source current is according to Eq. (2.74)

$$\mathbf{A}(\mathbf{r}) = (i\omega)^{-1} G_o(\mathbf{r}, \mathbf{r}') \mathbf{n}_x. \quad (2.83)$$

Inserting this vector potential into Eq. (2.82) we find

$$\mathbf{G}_x(\mathbf{r}, \mathbf{r}') = \left[1 + \frac{1}{k^2} \nabla \nabla \cdot \right] G_o(\mathbf{r}, \mathbf{r}') \mathbf{n}_x, \quad (2.84)$$

with similar expressions for \mathbf{G}_y and \mathbf{G}_z . The only remaining thing to do is to tie the three solutions together to form a dyad. With the definition $\nabla \cdot [G_o \vec{\mathbf{I}}] = \nabla G_o$ the dyadic Green's function $\vec{\mathbf{G}}$ can be calculated from the scalar Green's function G_o in Eq. (2.75) as

$$\vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \left[\vec{\mathbf{I}} + \frac{1}{k^2} \nabla \nabla \right] G_o(\mathbf{r}, \mathbf{r}'). \quad (2.85)$$

2.10.3 Time dependent Green's functions

The time dependence in the wave equations can be separated and the resulting harmonic differential equation for the time behavior is easily solved. A monochromatic field can be represented in the form of Eq. (2.24) and any other time-dependent field can be generated by a Fourier transform (sum of monochromatic fields). However, for the study of ultrafast phenomena it is of advantage to retain the explicit time behavior. In this case we have to generalize the definition of \mathbf{A} and ϕ as[¶]

$$\mathbf{E}(\mathbf{r}, t) = -\frac{\partial}{\partial t} \mathbf{A}(\mathbf{r}, t) - \nabla \phi(\mathbf{r}, t) \quad (2.86)$$

$$\mathbf{H}(\mathbf{r}, t) = \frac{1}{\mu_o \mu} \nabla \times \mathbf{A}(\mathbf{r}, t), \quad (2.87)$$

from which we find the time-dependent Helmholtz equation in the Lorentz gauge [c.f. Eq. (2.70)]

$$\left[\nabla^2 - \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} \right] \mathbf{A}(\mathbf{r}, t) = -\mu_o \mu \mathbf{j}(\mathbf{r}, t). \quad (2.88)$$

A similar equation holds for the scalar potential ϕ . The definition of the *scalar* Green's function is now generalized to

$$\left[\nabla^2 - \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} \right] G_o(\mathbf{r}, \mathbf{r}'; t, t') = -\delta(\mathbf{r} - \mathbf{r}') \delta(t - t'). \quad (2.89)$$

The point source is now defined with respect to space and time. The solution for G_o is [1]

$$G_o(\mathbf{r}, \mathbf{r}'; t, t') = \frac{\delta\left(t' - \left[t \mp \frac{n}{c} |\mathbf{r} - \mathbf{r}'|\right]\right)}{4\pi |\mathbf{r} - \mathbf{r}'|}, \quad (2.90)$$

[¶]We assume a nondispersive medium, i.e. $\varepsilon(\omega) = \varepsilon$ and $\mu(\omega) = \mu$.

where the minus sign is associated with the response at a time t later than t' . Using G_o we can construct the time-dependent dyadic Green's function $\vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}'; t, t')$ similar to the previous case. Since we will mostly work with time-independent Green's functions we avoid further details and refer the interested reader to specialized books on electrodynamics. Working with time-dependent Green's functions accounts for arbitrary time behavior but it is very difficult to incorporate dispersion. Time dependent processes in dispersive media are more conveniently solved using Fourier transforms of monochromatic fields.

2.11 Evanescent fields

Evanescent fields play a central role in nano-optics. The word *evanescent* derives from the latin word *evanescere* and has meanings like *vanishing from notice* or *imperceptible*. Evanescent fields can be described e.g. by plane waves of the form $\mathbf{E}e^{i(\mathbf{k}\mathbf{r}-\omega t)}$. They are characterized by the fact that at least one component of the wave vector describing the direction of propagation \mathbf{k} is imaginary. In the spatial direction defined by the imaginary component of \mathbf{k} the wave does not propagate but rather decays exponentially. Evanescent fields are of major importance for the understanding of optical fields that are confined to subwavelength dimensions. This section discusses the basic properties of evanescent waves and introduces simple experimental arrangements for their creation and measurement.

Evanescent waves never occur in a homogeneous medium but are inevitably connected to the interaction of light with inhomogeneities [7]. The simplest case of an inhomogeneity is 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 2.8.1, the presence of the interface will lead to a reflected 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. 2.2. We choose the x -axis to be in the plane of incidence. Using the symbols defined in Section 2.8.1, the complex transmitted field vector can be expressed as

$$\mathbf{E}_2 = \begin{bmatrix} -\mathbf{E}_1^{(p)} t^p(k_x) k_{z2}/k_2 \\ \mathbf{E}_1^{(s)} t^s(k_x) \\ \mathbf{E}_1^{(p)} t^p(k_x) k_x/k_2 \end{bmatrix} e^{ik_x x + ik_{z2} z} \quad (2.91)$$

which can be expressed entirely by the angle of incidence θ_1 using $k_x = k_1 \sin \theta_1$. With this substitution the longitudinal wavenumbers can be written as [c.f. Eq. (2.46)]

$$k_{z1} = k_1 \sqrt{1 - \sin^2 \theta_1}, \quad k_{z2} = k_2 \sqrt{1 - \tilde{n}^2 \sin^2 \theta_1}. \quad (2.92)$$

where we introduced the relative index of refraction

$$\tilde{n} = \frac{\sqrt{\varepsilon_1 \mu_1}}{\sqrt{\varepsilon_2 \mu_2}}. \quad (2.93)$$

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

$$[1 - \tilde{n}^2 \sin^2 \theta_1] = 0, \quad (2.94)$$

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

$$\theta_c = \arcsin [1/\tilde{n}]. \quad (2.95)$$

For a glass/air interface in the optical frequency regime, we have $\varepsilon_2 = 1$, $\varepsilon_1 = 2.25$, and $\mu_1 = \mu_2 = 1$ yielding a critical angle $\theta_c = 41.8^\circ$.

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

$$\mathbf{E}_2 = \begin{bmatrix} -i\mathbf{E}_1^{(p)} t^p(\theta_1) \sqrt{\tilde{n}^2 \sin^2 \theta_1 - 1} \\ \mathbf{E}_1^{(s)} t^p(\theta_1) \\ \mathbf{E}_1^{(p)} t^p(\theta_1) \tilde{n} \sin \theta_1 \end{bmatrix} e^{i \sin \theta_1 k_1 x} e^{-\gamma z} \quad (2.96)$$

where the decay constant γ is defined by

$$\gamma = k_1 \sqrt{\tilde{n}^2 \sin^2 \theta_1 - 1}. \quad (2.97)$$

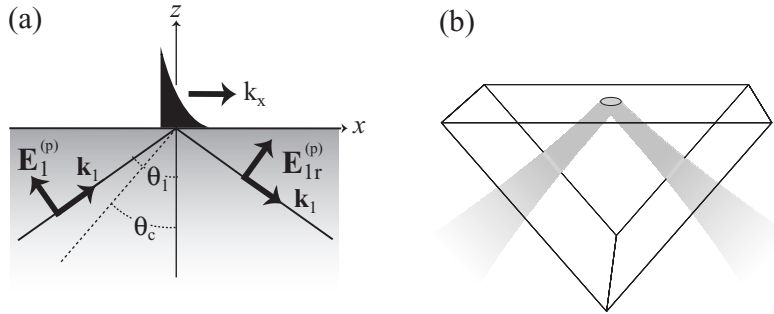


Figure 2.4: Excitation of an evanescent wave by total internal reflection. (a) An evanescent wave is created in medium if the plane wave is incident at an angle $\theta_1 > \theta_c$. (b) Actual experimental realization using a prism and a weakly focused Gaussian beam.

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 $\theta_i=45^\circ$, 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 by 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 (see e.g. [8], and problem 2.5).

Evanescent fields as described by Eq. (2.96) can be produced by directing a beam of light into a glass prism as sketched in Fig. 2.4 (b). Experimental verification for the existence of this rapidly decaying field in the optical regime relies on approaching a transparent body within less than $\lambda/2$ to the interface that supports the evanescent field. As shown in Fig. 2.5, this can be accomplished, for example, by a sharp transparent fiber which converts the evanescent field at its tip into a guided mode propagating along the fiber [9]. This measurement technique is called *photon scanning tunneling microscopy* and will be discussed later in Chapter ??.

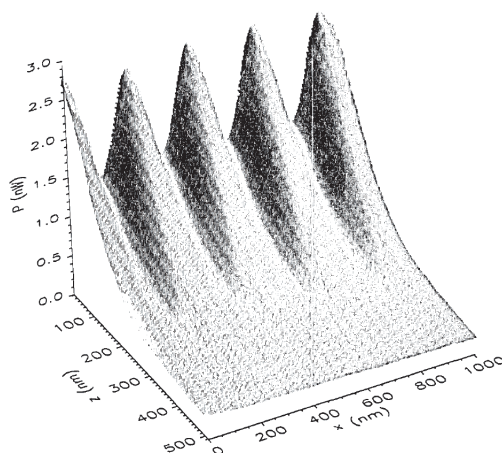


Figure 2.5: Spatial modulation of the standing evanescent wave along the propagation direction of two interfering waves (x -axis) and the decay of the intensity in z -direction. The ordinate represents the measured optical power. From [9].

For p -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. (2.96) and we write for an s - and p -polarized plane wave separately the intensity ratio $|\mathbf{E}_2(z = 0)|/|\mathbf{E}_1(z = 0)|$. This ratio is equal to the absolute square of the Fresnel transmission coefficient $t^{p,s}$. These transmission coefficients are plotted in Fig. 2.6 for the example of a glass/air interface. While for s -polarized light no field enhancement is observed, for p -polarized light the transmitted evanescent intensity is up to a factor of 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 (2.41). 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. We will discuss this and similar effects in more detail in Chapter ??.

2.11.1 Energy transport by evanescent waves

For non-absorbing media and for supercritical incidence, all the power of the incident wave is reflected. This effect coins the term *total internal reflection* (TIR). 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 proof 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 (c.f. Eq. 2.57)

$$\langle S \rangle_z = \frac{1}{2} \text{Re} (E_x H_y^* - E_y H_x^*) , \quad (2.98)$$

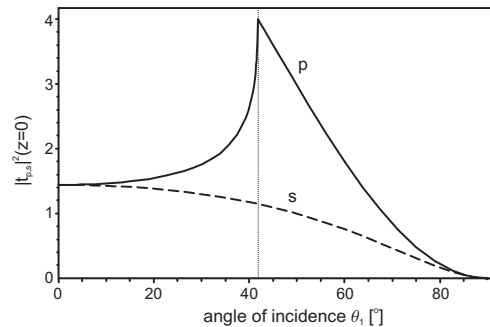


Figure 2.6: Intensity enhancement on top of a glass surface irradiated by a plane wave with variable angle of incidence θ_1 . For a p -polarized wave, the enhancement peaks at the critical angle $\theta_c = 41.8^\circ$ marked by the dotted line.

where all fields are evaluated in the upper medium, i.e. the medium of transmittance. Applying Maxwell's equation (2.25) to the special case of a plane or evanescent wave, allows us to express the magnetic field in terms of the electric field as

$$\mathbf{H} = \sqrt{\frac{\varepsilon_o \varepsilon}{\mu_o \mu}} \left[\left(\frac{\mathbf{k}}{k} \right) \times \mathbf{E} \right]. \quad (2.99)$$

Introducing the expressions for the transmitted field components of \mathbf{E} and \mathbf{H} into Eq. (2.98) it is straightforward to prove that $\langle \mathbf{S} \rangle_z$ vanishes (problem 2.4) 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 nonzero result is found

$$\langle \mathbf{S} \rangle_x = \frac{1}{2} \sqrt{\frac{\varepsilon_2 \mu_2}{\varepsilon_1 \mu_1}} \sin \theta_1 \left(|t^s|^2 |\mathbf{E}_1^{(s)}|^2 + |t^p|^2 |\mathbf{E}_1^{(p)}|^2 \right) e^{-2\gamma z}. \quad (2.100)$$

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

The absence of a net energy flow normal to the surface does not mean that there is no energy contained in an evanescent wave. For example, the local field distribution can be mapped out by using the fluorescence of a single molecule as a local probe^{||} The rate R at which the fluorophore emits photons when excited by the optical electric field is given by

$$R \sim |\boldsymbol{\mu} \cdot \mathbf{E}|^2, \quad (2.101)$$

where $\boldsymbol{\mu}$ is the absorption dipole moment of the molecule. As an example, for s -polarized fields the fluorescence rate of a molecule with a nonzero dipole component along the y -axis at a distance z above the interface will be

$$R(z) \sim |t^s \mathbf{E}_1^{(s)}|^2 e^{-2\gamma z}, \quad (2.102)$$

decaying twice as fast as the electric field itself. Notice, that a molecule can be excited even though the average Poynting vector vanishes. This applies also to other interactions and will be discussed in a later chapter.

2.11.2 Frustrated total internal reflection

Evanescent fields can be converted into propagating radiation if they interact with matter [7]. This phenomenon is among the most important effects in near-field optical microscopy since it explains how information about subwavelength structures is

^{||}Excitation of fluorescence using evanescent waves is quite popular in biological imaging. Since only a thin slice of the sample is illuminated, background is drastically reduced. The technique is known as total internal reflection fluorescence microscopy (TIRFM) [10].

transported into the farfield. We will discuss the physics behind this conversion by considering again a very simple model. A plane interface will be used in order to create an evanescent wave by TIR as before. A second parallel plane interface is then approached to 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 approach two prisms with very flat or slightly curved surfaces as indicated in Fig. 2.7 (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. 2.7 (a).

single medium. The partial fields in media 1,2 are written as a superposition of incident and reflected waves whereas for medium 3 there is only 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. (2.92). The longitudinal wavenumber in medium j reads

$$k_{jz} = \sqrt{k_j^2 - k_{\parallel}^2} = k_j \sqrt{1 - (k_1/k_j)^2 \sin^2 \theta_1}, \quad j \in \{1, 2, 3\}, \quad (2.103)$$

where $k_j = n_j k_o = n_j(\omega/c)$ and $n_j = \sqrt{\varepsilon_j \mu_j}$. In the following a layered system with

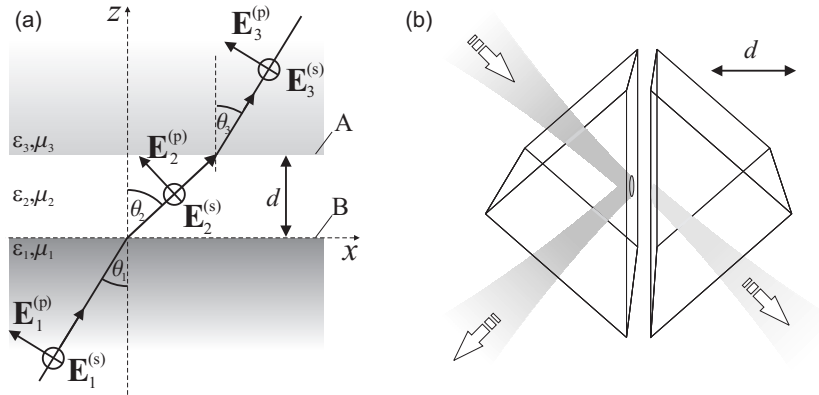


Figure 2.7: Transmission of a plane wave through a system of two parallel interfaces. In frustrated total internal reflection (FTIR), the evanescent wave created at interface A is partly converted into a propagating wave by the interface B of a second medium. (a) Configuration and definition of parameters. A,B: interfaces between media 1,2 and 2,3, respectively. The reflected waves are omitted for clarity. (b) Experimental setup to observe frustrated total internal reflection.

$n_2 < n_3 < n_1$ shall be discussed, which includes the system sketched in Fig. 2.7. 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 $k_{\parallel} < n_2 k_o$, the field is entirely described by propagating plane waves. The intensity transmitted to a detector far away from the second interface (farfield) will not vary substantially with gap width, but will only show rather weak interference undulations.
2. For $\arcsin(n_2/n_1) < \theta_1 < \arcsin(n_3/n_1)$ or $n_2 k_o < k_{\parallel} < n_3 k_o$ 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 gap width. This situation is referred to as *frustrated total internal reflection* (FTIR).
3. For $\theta_1 > \arcsin(n_3/n_1)$ or $k_{\parallel} > n_3 k_o$ 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).

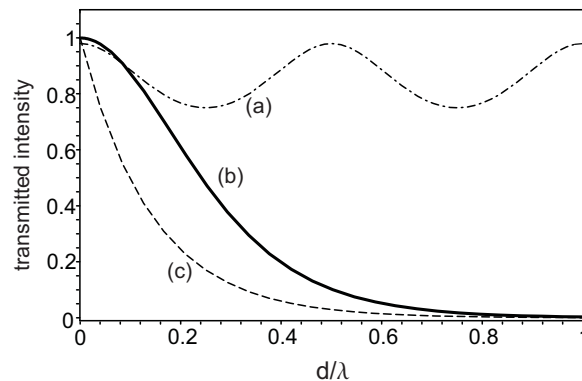


Figure 2.8: Transmission of a system of three media with parallel interfaces as a function of the gap d between the two interfaces. A p -polarized plane wave excites the system. The material constants are $n_1 = 2$, $n_2 = 1$, $n_3 = 1.51$. This leads to critical angles θ_c of 30° and 49.25° . For angles 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 absence of the third medium.

However, as shown in Fig. 2.8, $I(d)$ deviates from a purely exponential behavior because the field in medium (2) is a superposition of two evanescent waves of the form

$$c_1 e^{-\gamma z} + c_2 e^{+\gamma z} . \quad (2.104)$$

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. This simple experiment illustrates the fact that in near-field optical experiments the effect of the probe on the field distribution must never be neglected. Fig. 2.8 shows typical transmission curves for two different angles of incidence. The figure also shows that the decay measured in FTIR deviates from a simple exponential decay. In the next section it will be discussed that evanescent waves are important for the rigorous theoretical description of arbitrary optical fields near sources or material boundaries. Mathematically, they are more difficult to deal with than plane waves because they are not orthogonal functions.

2.12 Angular Spectrum Representation of Optical Fields

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.

Under angular spectrum representation we understand the series expansion of an arbitrary field in terms of plane waves with variable amplitudes and propagation directions. 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. 2.9 for which $\mathbf{E} = \mathbf{E}_{inc} + \mathbf{E}_{scatt}$. In the angular spectrum picture, we draw an arbitrary axis z and consider the field \mathbf{E} in a plane $z = const$ transverse to the chosen axis. In this plane we can evaluate the two dimensional Fourier transform of the field \mathbf{E} as

$$\hat{\mathbf{E}}(k_x, k_y; z) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}(x, y, z) e^{-i[k_x x + k_y y]} dx dy , \quad (2.105)$$

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

$$\mathbf{E}(x, y, z) = \iint_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x, k_y; z) e^{i[k_x x + k_y y]} dk_x dk_y . \quad (2.106)$$

Notice, that in the notation of Eqs. (2.105) and (2.106) 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 the vector Helmholtz equation

$$(\nabla^2 + k^2) \mathbf{E}(\mathbf{r}) = 0 , \quad (2.107)$$

where k is determined by $k = (\omega/c)n$ and $n = \sqrt{\mu\epsilon}$ is the index of refraction. In order to get the time-dependent field $\mathbf{E}(\mathbf{r}, t)$ we use the convention

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}\{\mathbf{E}(\mathbf{r}) e^{-i\omega t}\} . \quad (2.108)$$

Inserting the Fourier representation of $\mathbf{E}(\mathbf{r})$ [Eq. (2.106)] into the Helmholtz equation and defining

$$k_z \equiv \sqrt{(k^2 - k_x^2 - k_y^2)} \quad \text{with} \quad \text{Im}\{k_z\} \geq 0 , \quad (2.109)$$

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

$$\hat{\mathbf{E}}(k_x, k_y; z) = \hat{\mathbf{E}}(k_x, k_y; 0) e^{\pm i k_z z} . \quad (2.110)$$

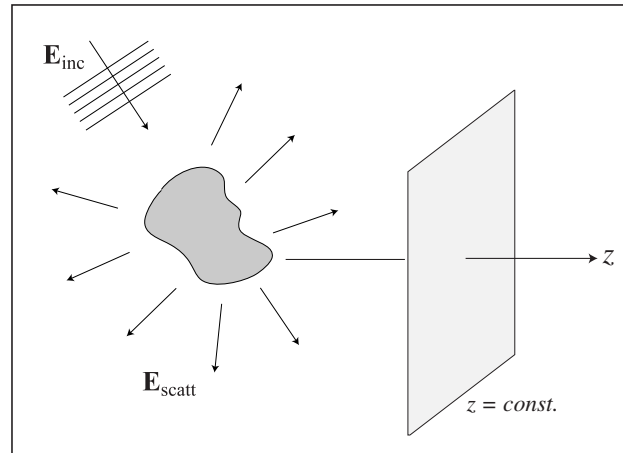


Figure 2.9: In the angular spectrum representation the fields are evaluated in planes ($z = \text{const}$) perpendicular to an arbitrarily chosen axis z .

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 $z < 0$. Eq. (2.110) tells us that the Fourier spectrum of \mathbf{E} in an arbitrary *image plane* located at $z = \text{const}$ 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 *propagator* in reciprocal space. In Eq. (2.109) 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 $z \rightarrow \pm\infty$. Inserting the result of Eq. (2.110) into Eq. (2.106) we finally find for arbitrary z

$$\mathbf{E}(x, y, z) = \iint_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x, k_y; 0) e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y \quad (2.111)$$

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

$$\mathbf{H}(x, y, z) = \iint_{-\infty}^{\infty} \hat{\mathbf{H}}(k_x, k_y; 0) e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y, \quad (2.112)$$

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

$$\begin{aligned} \hat{H}_x &= Z_{\mu\varepsilon}^{-1} [(k_y/k) \hat{E}_z - (k_z/k) \hat{E}_y], \\ \hat{H}_y &= Z_{\mu\varepsilon}^{-1} [(k_z/k) \hat{E}_x - (k_x/k) \hat{E}_z], \\ \hat{H}_z &= Z_{\mu\varepsilon}^{-1} [(k_x/k) \hat{E}_y - (k_y/k) \hat{E}_x], \end{aligned} \quad (2.113)$$

where $Z_{\mu\varepsilon} = \sqrt{(\mu_o\mu)/(\varepsilon_o\varepsilon)}$ 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 field vectors ($\mathbf{k} \cdot \mathbf{E} = \mathbf{k} \cdot \mathbf{H} = 0$).

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

$$\begin{aligned} \text{Plane waves :} & \quad e^{i[k_x x + k_y y]} e^{\pm i|k_z|z}, & k_x^2 + k_y^2 \leq k^2 \\ \text{Evanescent waves :} & \quad e^{i[k_x x + k_y y]} e^{-|k_z||z|}, & k_x^2 + k_y^2 > k^2 \end{aligned} \quad (2.114)$$

Hence, we find that the angular spectrum is indeed a superposition of *plane waves* and *evanescent waves*. Plane waves are oscillating functions in z and are restricted

by the condition $k_x^2 + k_y^2 \leq k^2$. 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. Fig. 2.10 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 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 achieved by evanescent waves. In principle, an infinite bandwidth of spatial frequencies can be achieved. However, the higher the spatial frequencies of an evanescent wave are, the stronger the field decay along the z axis will be. Therefore, practical limitations make the bandwidth finite.

2.12.1 Angular spectrum representation of a dipole

Strongly localized sources such as dipoles are most conveniently described in a spherical coordinate system. The corresponding solutions of the wave equation are called multipoles. In order to couple these solutions with the angular spectrum picture we need to express the localized sources in terms of plane waves and evanescent waves. Let us start with the vector potential \mathbf{A} of an oscillating dipole with its axis aligned along an arbitrary z -axis. The vector potential can be expressed as a one-component vector field as [c.f. Eq. (2.83)]

$$\mathbf{A}(x, y, z) = A(x, y, z) \mathbf{n}_z = \frac{-ikZ_{\mu\varepsilon}}{4\pi} \frac{e^{ik\sqrt{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}} \mathbf{n}_z, \quad (2.115)$$

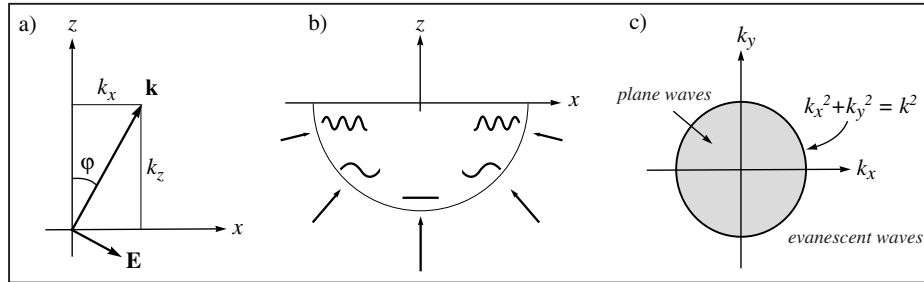


Figure 2.10: a) Representation of a plane wave propagating at an angle φ to the z axis. b) Illustration of the transverse spatial frequencies of plane waves incident from different angles. The transverse wavenumber $(k_x^2 + k_y^2)^{1/2}$ depends on the angle of incidence and is limited to the interval $[0..k]$. c) The transverse wavenumbers k_x, k_y of plane waves are restricted to a circular area with radius k . Evanescent waves fill the space outside the circle.

where $Z_{\mu\varepsilon} = \sqrt{\mu_o\mu/\varepsilon_o\varepsilon}$ is the wave impedance of the medium. Besides a constant factor, the expression on the right hand side corresponds to the scalar Green's function (2.75). According to Eq. (2.67) and (2.82) the electric and magnetic fields are obtained from \mathbf{A} as

$$\mathbf{E}(x, y, z) = i\omega\left(1 + \frac{1}{k^2}\nabla\nabla\cdot\right)\mathbf{A}(x, y, z) \quad (2.116)$$

$$\mathbf{H}(x, y, z) = \frac{1}{\mu_o\mu}\nabla\times\mathbf{A}(x, y, z). \quad (2.117)$$

Thus, the electromagnetic field of the dipole can be constructed from the function $\exp(ikr)/r$, where $r = (x^2 + y^2 + z^2)^{1/2}$ is the radial distance from the dipole's origin. To find an angular spectrum representation of the dipole's electric and magnetic field we need first to find the angular spectrum of the function $\exp(ikr)/r$. This is not a trivial task and cannot be derived using the angular spectrum representation. The reason is that the function $\exp(ikr)/r$ is singular at $r=0$ and therefore not divergence free at its origin. The homogeneous Helmholtz equation is therefore not valid in the present case. Nevertheless, using complex contour integration it is possible to derive an angular spectrum representation of the function $\exp(ikr)/r$. Since the derivation can be found in other textbooks [11] we state here only the result which is

$$\frac{e^{ik\sqrt{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}} = \frac{i}{2\pi} \iint_{-\infty}^{\infty} \frac{e^{ik_x x + ik_y y + ik_z |z|}}{k_z} dk_x dk_y. \quad (2.118)$$

We have to require that the real and imaginary parts of k_z stay positive for all values of k_x, k_y in the integration. The result (2.118) is known as the Weyl identity [12]. In Chapter ?? we will use the Weyl identity to calculate dipole emission near planar interfaces.

Problems

Problem 2.1 Derive the dyadic Green's function $\vec{\mathbf{G}}$ by substituting the scalar Green's function G_o into Eq. (2.85). Discuss the distance dependence.

Problem 2.2 Consider an interface between two media (1) and (2) with dielectric constants $\varepsilon_1 = 2.25$ and $\varepsilon = 1$, respectively. The magnetic permeabilities are equal to one. A p -polarized plane wave with wavelength $\lambda = 532\text{nm}$ is incident from medium (1) at an angle of incidence of θ_1 . Express the Fresnel reflection coefficient in terms of amplitude A and phase Φ . Plot A and Φ as a function of θ_1 . What are the consequences for the reflected wave?

Problem 2.3 Consider the refraction of a plane wave at a plane interface and derive Snell's law by using the invariance of the transverse wavevector \mathbf{k}_{\parallel} .

Problem 2.4 Show that the z -component of the time-averaged Poynting vector $\langle \mathbf{S} \rangle_z$ vanishes for an evanescent field propagating in the x -direction.

Problem 2.5 Analyze the polarization state of an evanescent field propagating in x -direction created by total internal reflection of a p -polarized plane wave. Calculate the time-dependent electric field $\mathbf{E}_2(x, t) = (E_{2,x}(x, t), 0, E_{2,z}(x, t))$ just on top of the interface ($z = 0$). For a fixed position x , the electric field vector \mathbf{E}_2 defines a curve in the (x, z) plane as the time runs from 0 to λ/c . Determine and plot the shape of these curves as a function of the position x . For numerical values choose $\theta_1 = 60^\circ$, $\tilde{n} = 1.5$.

Problem 2.6 Calculate the transmitted intensity for a system of two glass half-spaces ($n = 1.5$) separated by an air gap (d) and as a function of the angle of incidence θ_1 . Determine the transmission function for s -polarized excitation. Normalize the transmission function with the value obtained for $\theta_1 = 0^\circ$. Repeat for p -polarized excitation.

Problem 2.7 Derive Eq. (2.110) by inserting the inverse Fourier transform in Eq. (2.106) into the Helmholtz equation (2.107). Assume that the Fourier spectrum is known in the plane $z = 0$.

Problem 2.8 Using the Weyl identity (2.118), derive the spatial spectrum $\hat{E}(k_x, k_y; z)$ of an electric dipole at $\mathbf{r}_o = (0, 0, z_o)$ with dipole moment $\boldsymbol{\mu} = (\mu, 0, 0)$. Consider the asymptotic limit $z \rightarrow \infty$ and solve for the electric field \mathbf{E} .

References

- [1] J. D. Jackson, *Classical Electrodynamics*. New York: Wiley, 2nd ed. (1975).
- [2] J. A. Stratton, *Electromagnetic Theory*. New York: McGraw-Hill, 1st ed. (1941).
- [3] M. Born and E. Wolf *Principles of Optics*. New York: Cambridge University Press, 7th ed. (1999).
- [4] A. D. Yaghjian, "Electric dyadic Green's functions in the source region," *Proc. IEEE* **68**, 248–263 (1980).
- [5] J. V. Bladel, "Some remarks on green's dyadic for infinite space," *IRE Trans. Antennas Propagat.* **9**, 563–566 (1961).
- [6] C. T. Tai, *Dyadic Green's Functions in Electromagnetic Theory*. New York: IEEE Press, 2nd ed. (1993).
- [7] E. Wolf and M. Nieto-Vesperinas, "Analyticity of the angular spectrum amplitude of scattered fields and some of its consequences," *J. Opt. Soc. Am. A* **2**, 886–889 (1985).
- [8] S. Sund, J. Swanson, and D. Axelrod, "Cell Membrane Orientation Visualized by Polarized Total Internal Reflection Fluorescence," *Biophys. J.* **77**, 2266–2283 (1999).
- [9] A. Meixner, M. Bopp, and G. Tarrach, "Direct Measurement of Standing Evanescent Waves with a Photon Scanning Tunneling Microscope," *Appl. Opt.* **33**, 7995 (1994).
- [10] D. Axelrod, N. Thompson, and T. Burghardt, "Total internal reflection fluorescent microscopy," *J. Microsc.* **129**, 19–28 (1983).
- [11] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. New York: Cambridge University Press (1995).
- [12] H. Weyl, "Ausbreitung elektromagnetischer Wellen über einem ebenen Leiter," *Ann. der Physik* **60**, 481–500 (1919).