# 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 $\partial D_{ij}$. If $\varepsilon_i$ and $\mu_i$ designate the constant material parameters in region

$D_i$, the wave equations in that domain read as

$$(\nabla^2 + k_i^2)\,\underline{\mathbf{E}}_i \;=\; -\mathrm{i}\omega\mu_0\mu_i\,\underline{\mathbf{j}}_{0_i} + \frac{\nabla\underline{\rho}_{0_i}}{\varepsilon_0\,\varepsilon_i}\;, \tag{4.1}$$

$$(\nabla^2 + k_i^2)\,\underline{\mathbf{H}}_i \;=\; -\nabla\times\,\underline{\mathbf{j}}_{0_i}\;, \tag{4.2}$$

where $k_i = (\omega/c)\sqrt{\mu_i\varepsilon_i} = k_0 n_i$ is the wavenumber and $\underline{\mathbf{j}}_{0_i}$, $\underline{\rho}_{0_i}$ are the source currents and charges 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 (1.31) 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

Equations (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 $\partial D_{ij}$. 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 (1.15)–(1.18). 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 (1.15) 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

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

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. (1.15) becomes

$$\int_V \underline{\rho}_0(\mathbf{r})\,dV \;=\; A\,dL\,\underline{\rho}_{0_i}(\mathbf{r}) \;+\; A\,dL\,\underline{\rho}_{0_j}(\mathbf{r})\;, \tag{4.4}$$

which finally yields in the limit of $A, dL \to 0$

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

Here, $\underline{\sigma} = dL\,\underline{\rho}_{0_i} + dL\,\underline{\rho}_{0_j} = dL\underline{\rho}_0$ is the surface charge density. Equation (4.5) is the boundary condition for the displacement $\underline{\mathbf{D}}$. 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 $\underline{\sigma}$.

Let us now have a look at Faraday's law, *i.e.* Maxwell's second equation (1.16). We evaluate the path integral of the $\mathbf{E}$-field along the small rectangular path shown in Fig. 4.1(b)

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

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 $\partial D_{ij}$. For the right hand side of Eq. (1.16) we find

$$\lim_{A, dL \to 0} \left[ -\frac{\partial}{\partial t} \int_A \underline{\mathbf{B}}(\mathbf{r}) \cdot \mathbf{n}_\mathrm{s} \, da \right] \; = \; 0 \; , \tag{4.7}$$

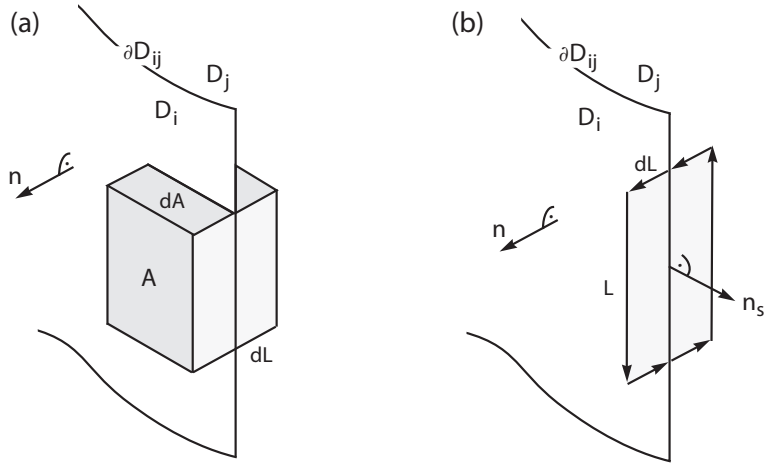


Figure 4.1: Integration paths for the derivation of the boundary conditions on the interface $\partial D_{ij}$ between two neighboring regions $D_i$ and $D_j$.

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

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

We obtain a similar result for the Ampère-Maxwell law (1.17), 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

$$\int_A \underline{\mathbf{j}}_0(\mathbf{r}) \cdot \mathbf{n}_\mathrm{s}\, da = L\, dL \left[\underline{\mathbf{j}}_{0_i}(\mathbf{r}) \cdot \mathbf{n}_s\right] + L\, dL \left[\underline{\mathbf{j}}_{0_j}(\mathbf{r}) \cdot \mathbf{n}_s\right] \,. \tag{4.9}$$

Using an equation similar to (4.6) for $\mathbf{H}$ and inserting into Eq. (1.17) yields

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

where $\underline{\mathbf{K}} = dL\, \underline{\mathbf{j}}_{0_i} + dL\, \underline{\mathbf{j}}_{0_j} = dL\, \underline{\mathbf{j}}_0$ being the surface current density. The fourth Maxwell equation (1.18) yields a boundary condition similar to Eq. (4.5), but with no surface charge, that is $\mathbf{n} \cdot \left[\underline{\mathbf{B}}_i - \underline{\mathbf{B}}_j\right] = 0$.

Taken all equations together we obtain

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

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

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

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

The surface charge density $\underline{\sigma}$ and the surface current density $\underline{\mathbf{K}}$ are confined to the very interface between the regions $D_i$ and $D_j$. Such confinement is not found in practice and hence, $\underline{\sigma}$ and $\underline{\mathbf{K}}$ are mostly of theoretical significance. The existence of $\sigma$ 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 $\underline{\sigma} = 0$ and $\underline{\mathbf{K}} = 0$. Any currents and charges confined to the boundary $D_{ij}$ are accounted for by the imaginary part of the dielectric functions $\varepsilon_i$ and $\varepsilon_j$.

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 $\partial D_{ij}$ 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 ($\underline{\sigma} = 0$ and $\underline{\mathbf{K}} = 0$), we can represent Eqs. (4.11)–(4.14) as

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

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

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

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

Note that these boundary conditions assume that the fields exist on both sides of the interface. In the limiting case of a perfect conductor, the fields are zero inside the conductor, which gives rise to surface currents and surface charges. In this case we have to re-introduce $\underline{\sigma}$ and $\underline{\mathbf{K}}$ according to Eqs. (4.12) and (4.14).

## 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 $\theta_i$ 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

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

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

$$\mathbf{E}_{1r}(\mathbf{r},t) \ = \ \mathrm{Re}\{\underline{\mathbf{E}}_{1r}\,\mathrm{e}^{\mathrm{i}\mathbf{k}_{1r}\cdot\mathbf{r}-\mathrm{i}\omega t}\} \quad \mathbf{E}_2(\mathbf{r},t) \ = \ \mathrm{Re}\{\underline{\mathbf{E}}_2\,\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot\mathbf{r}-\mathrm{i}\omega t}\} \ , \tag{4.22}$$

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 $\mathbf{k}_2 = [k_{x_2},\ k_{y_2},\ k_{z_2}]$, respectively. All fields have the same frequency $\omega$ 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

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

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

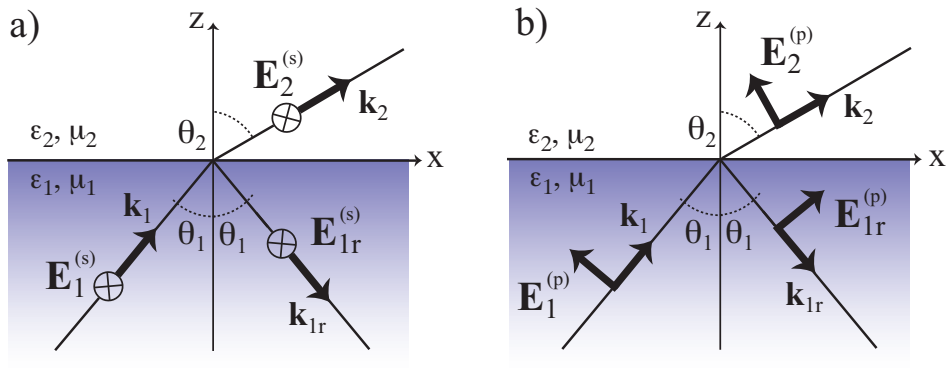


Figure 4.2: Reflection and refraction of a plane wave at a plane interface. (a) s-polarization, and (b) p-polarization.

is, $k_{x_1} = k_{x_{1r}} = k_{x_2}$ and $k_{y_1} = k_{y_{1r}} = k_{y_2}$. 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 since the reflected field is propagating away from the interface. To summarize, the three waves (incident, reflected, transmitted) can be represented as

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

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

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

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 $\theta_1$ 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 $\theta_1$.

If we denote the angle of propagation of the refracted wave as $\theta_2$ (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 $(\mathrm{s})$ and $(\mathrm{p})$ stand for the German "senkrecht" (perpendicular) and "parallel" (parallel), respectively, and refer to the plane of incidence. The plane of incidence is the plane defined by the $\mathbf{k}$-vector and the surface normal $\mathbf{n}_z$. 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

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

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

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

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

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.* $\nabla\cdot\underline{\mathbf{E}} = 0$ and $\nabla\cdot\underline{\mathbf{H}} = 0$.

These conditions restrict the k-vector to directions perpendicular to the field vectors ($\mathbf{k}\cdot\underline{\mathbf{E}} = \mathbf{k}\cdot\underline{\mathbf{H}} = 0$).

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 $\underline{\mathrm{E}}_{1\mathrm{r}}^{(\mathrm{s})}$ and $\underline{\mathrm{E}}_{2}^{(\mathrm{s})}$. $\underline{\mathrm{E}}_{1}^{(\mathrm{s})}$ 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

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

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

The two other tangential boundary conditions (continuity of $\underline{\mathrm{E}}_{\mathrm{x}}$ and $\underline{\mathrm{H}}_{\mathrm{y}}$) are trivially fulfilled since $\underline{\mathrm{E}}_{\mathrm{x}} = \underline{\mathrm{H}}_{\mathrm{y}} = 0$. 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 $\underline{\mathrm{E}}_{1\mathrm{r}}^{(\mathrm{s})}$ yields

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

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

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

where $t^{\mathrm{s}}$ 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

$$\underline{\mathrm{E}}_{1r}^{(\mathrm{p})} = \underline{\mathrm{E}}_{1}^{(\mathrm{p})}\, r^{\mathrm{p}}(k_x, k_y) \qquad \underline{\mathrm{E}}_{2}^{(\mathrm{p})} = \underline{\mathrm{E}}_{1}^{(\mathrm{p})}\, t^{\mathrm{p}}(k_x, k_y)\,, \tag{4.38}$$

where $r^{\mathrm{p}}$ and $t^{\mathrm{p}}$ 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 ($\theta_1 = 0$), $r^{\mathrm{s}}$ and $r^{\mathrm{p}}$ 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 $\theta_1$ that is larger than a critical angle $\theta_c$. 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 air-glass interface this

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

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 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 $\underline{\mathbf{E}}\mathrm{e}^{\mathrm{i}(\mathbf{kr}-\omega t)}$. 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 $\varepsilon_1$, $\mu_1$ and $\varepsilon_2$, $\mu_2$. 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

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

which can be expressed entirely by the angle of incidence $\theta_1$ using $k_x = k_1 \sin\theta_1$. With this substitution the longitudinal wavenumbers can be written as (cf. Equation (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 $\theta_1$ the argument of the square root in the expression of $k_{z_2}$ gets smaller and smaller and eventually becomes negative. The critical angle $\theta_\mathrm{c}$ 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 $\theta_1$ 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

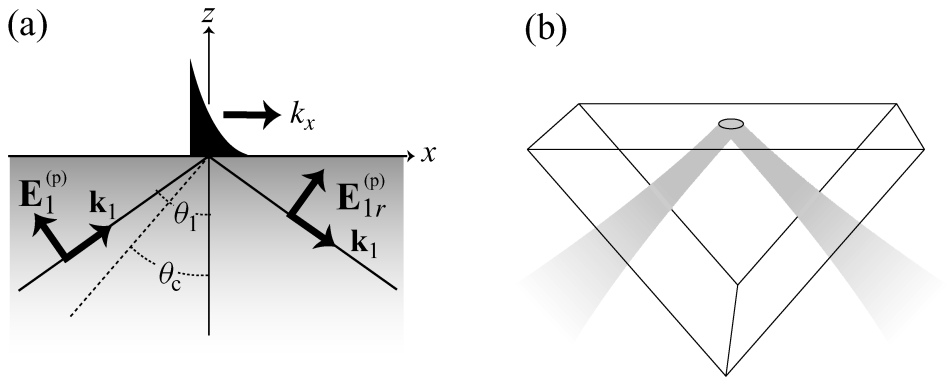


Figure 4.3: Excitation of an evanescent wave by total internal reflection. (a) An evanescent wave is created in a medium if the plane wave is incident at an angle $\theta_1 > \theta_\mathrm{c}$. (b) Actual experimental realization using a prism and a weakly focused Gaussian beam.

function of the angle of incidence $\theta_1$ results in

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

where the decay constant $\gamma$ is defined by

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

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 $\theta_1 > \theta_{\mathrm{c}}$ creates an evanescent wave. Excitation of an evanescent wave with a plane wave at supercritical incidence ($\theta_1 > \theta_{\mathrm{c}}$) 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 $\theta_i$ the faster the decay will be. Note that the Fresnel coefficients depend on $\theta_1$. For $\theta_1 > \theta_{\mathrm{c}}$ 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.

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 $|\underline{\mathbf{E}}_2(z=0)|^2/|\underline{\mathbf{E}}_1(z=0)|^2$. 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

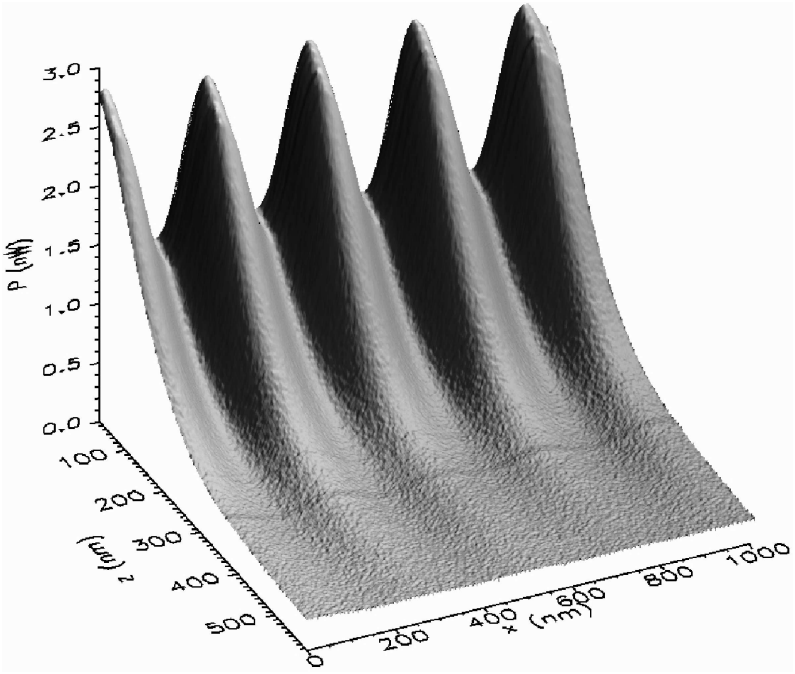


Figure 4.4: Spatial modulation of the standing evanescent wave along the propagation direction of two interfering waves ($x$-axis) and the decay of the intensity in the $z$-direction. The ordinate represents the measured optical power. From *Appl. Opt.* **33**, 7995 (1994).

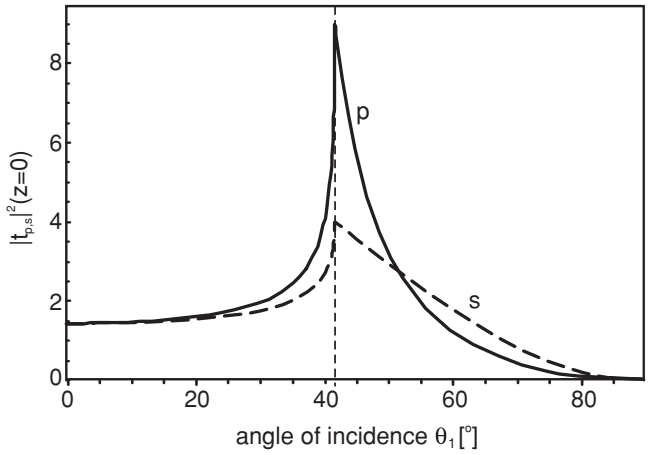


Figure 4.5: Intensity enhancement on top of a glass surface irradiated by a plane wave with variable angle of incidence $\theta_1$. For $p$- and $s$-polarized waves, the enhancement peaks at the critical angle $\theta_c = 41.8°$ marked by the dotted line.

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.

### 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 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 choose $\theta_1$ such that case 2 is realized (FTIR), the transmitted intensity $I(d)$

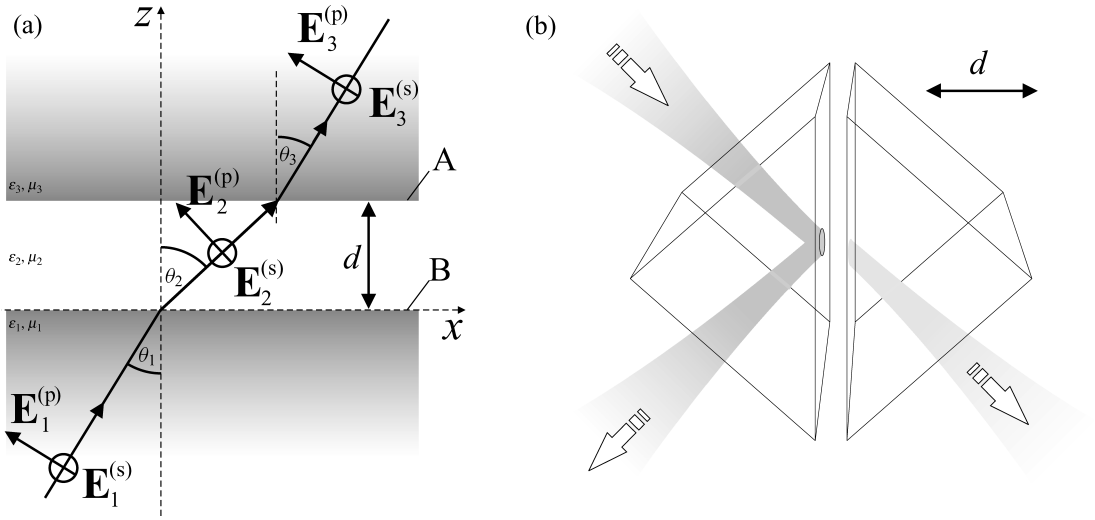


Figure 4.6: Transmission of a plane wave through a system of two parallel interfaces. In frustrated total internal reflection (FTIR), the evanescent wave created at interface B is partly converted into a propagating wave by the interface A of a second medium. (a) Configuration and definition of parameters. 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.

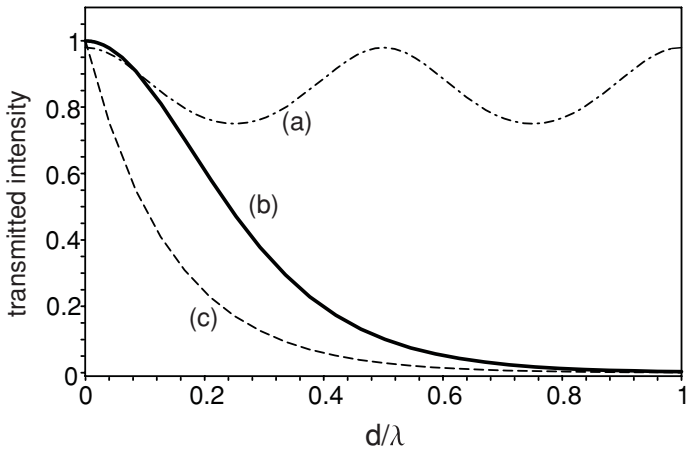


Figure 4.7: Transmission of a system of three media with parallel interfaces as a function of the gap $d$ between the two interfaces. A p-polarized plane wave excites the system. The material constants are $n_1 = 2$, $n_2 = 1$, $n_3 = 1.51$. This leads to critical angles $\theta_c$ of $30^\circ$ and $49.25^\circ$. For angles of incidence $\theta_i$ between (a) $0^\circ$ and $30^\circ$ the gap dependence shows interference-like behavior (here $\theta_1 = 0^\circ$, dash-dotted line), for angles between (b) $30^\circ$ and $49.25^\circ$ 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.

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.

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