Chapter 13: Plane waves (2): Physics at planar interfaces (13 Mar 2021).



## A. Overview.

In Ch. 13 we continue the discussion of solutions of the wave equation, now from the perspective of the physics that occurs at planar boundaries between dissimilar materials. The one-dimensional discontinuous change from one value to another across a planar interface is the simplest example of a spatially dependent dielectric function. While the dielectric response itself is not defined at the discontinuity, Stokes' and Gauss' Theorems can be applied in the usual manner to cast the evaluation of the relevant Maxwell equations into adjacent regions where everything is well-defined. Boundary conditions are the result. Although truly abrupt interfaces do not exist, if $\lambda$ is much larger than the width of the transition region, the mathematically sharp interface is a valid approximation. In practice these transition regions are typically of atomic dimensions, so the approximation is excellent indeed.

The steady-state solutions of the wave equation for plane waves incident on configurations consisting of planar interfaces between materials of different dielectric functions are the Fresnel Equations. These describe reflection, transmission, and interference. They also include elementary excitations, i.e., solutions of the homogeneous wave equation, which are called plasmons. For non-normal incidence we find transverse-electric (TE) and transverse-magnetic (TM) modes, Snell's Law, Brewster's angle, and under certain conditions, total internal reflection. Practical applications are legion, ranging from direct reflection to applications of interference that include but are not limited to laser mirrors and engineered coatings (decorative and otherwise), antireflection coatings, laser mirrors, etc.. Without current antireflection-coating technology, zoom lenses, including the entry-level 4-element variety standard on low-end single-lens-reflex cameras, could not exist.

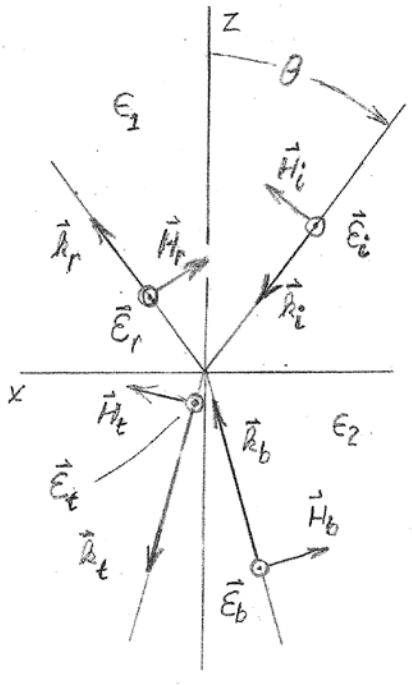
Surface plasmon polaritons, or more generally interface plasmons, are of considerable current interest owing to their small spatial extents relative to the wavelength of light. This has raised hope that circuits can be developed that operate at the speed of light but with dimensions much smaller than its wavelength. This technology continues to evolve.

Chapter 12 is built on Fourier analysis. Chapter 13 is built on matrix algebra. To simplify matters, we assume throughout that all dielectric functions are scalars, so the solutions of the wave equation are linearly polarized transverse plane waves with wave vectors satisfying the standard dispersion relation. Then the coefficients describing the incident and reflected waves on one side of an interface can be written as a $(2\times1)$ field vector that is connected to a similar $(2\times1)$ field vector describing the transmitted and back-reflected waves on the other side of the interface, the connection being made through a $2\times2$ reflection matrix. Another $2\times2$ matrix describes propagation of the waves transmitted to, and back-reflected from, the next interface, and the process repeats. The mathematical structure is similar to that of Jones matrices, although the meanings and interpretations are different.

As usual, we start from basic equations, get the math right, explain why we do what we do, then extract the physics. The development proceeds from a general description of boundary conditions, then to applications starting with the two-phase (ambient/substrate) model and adding complications.

B. Basic physics: the dispersion equation, continuity and kinematic conditions.

In this section we establish the constraints imposed by Maxwell's Equations and Stokes' Theorem on plane waves meeting at a planar interface. Most of the material in this section is a review of previous chapters, adapted to the planar configuration. The configuration consists of an ambient *a*, which we denote as material 1, that fills the half-space $z>0$ adjacent to a substrate, which we denote as material 2, that fills the half-space $z\le0$. The permeabilities and dielectric functions of the ambient and substrate are $\mu_1$, $\varepsilon_1$, $\mu_2$, and $\varepsilon_2$, respectively. "Ambient" generally refers to the transparent medium in which the initial plane wave arrives, but for boundary-condition calculations we take the more general perspective of the wave arriving at the interface of current interest. In later sections we usually set permeabilities equal to 1, but retain $\mu$ here to highlight the symmetry between electric and magnetic fields. It is also necessary for demonstrating that four continuity conditions on two fields do not overconstrain the system. We have already had significant experience with $\mu\neq1$, so retaining $\mu$ should not cause undue hardship. However, we do assume that the materials are isotropic, so $\varepsilon$ and $\mu$ are scalars, and therefore that $\vec{E}$, $\vec{H}$ and $\vec{k}$ form an orthogonal triple of vectors.




To begin, let a plane wave with an electric field $\vec{E}_i = E_i\hat{y}$ linearly polarized in the *y* direction be incident on the $z=0$ interface at an angle of incidence $\theta_i$ (see diagram). The

dielectric mismatch at $z = 0$ splits the incident wave *i* into a transmitted wave *t* and a reflected wave *r*. If additional discontinuities exist deeper in the material, a back-reflected wave *b* may also be present. We are thus dealing with four plane waves all polarized along $\hat{y}$:

$$\vec{E}_i = \hat{y}E_i e^{-ik_{ix}x - ik_{iz}z - i\omega t}; \tag{13.1a}$$

$$\vec{E}_r = \hat{y}E_r e^{-ik_{rx}x + ik_{rz}z - i\omega t} \tag{13.1b}$$

$$\vec{E}_t = \hat{y}E_t e^{-ik_{tx}x - ik_{tz}z - i\omega t} \tag{13.1c}$$

$$\vec{E}_b = \hat{y}E_b e^{-ik_{bx}x + ik_{bz}z - i\omega t} \tag{13.1d}$$

The incident wave drives the configuration, so the time dependence $e^{-i\omega t}$ is common to all and hence will generally be ignored in the following.

Our first task is to determine the constraints that these waves must satisfy. These constraints are established by the Faraday-Maxwell and Ampère Equations together with Stokes' Theorem. As shown below, constraints imposed by the remaining two Maxwell Equations are consistent with these, and so are not needed in the derivations.

The first constraint follows from the wave equation

$$\nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} + \frac{\mu\varepsilon}{c^2}\frac{\partial^2}{\partial t^2}\vec{E} = 0, \tag{13.2}$$

which follows by combining the Faraday-Maxwell and Ampère Equations. With $\vec{E}$, $\vec{H}$, and $\vec{k}$ mutually orthogonal, Eq. (13.2) reduces to the dispersion equation

$$\left(k^2 - \frac{\omega^2\mu\varepsilon}{c^2}\right)\vec{E} = 0. \tag{13.3}$$

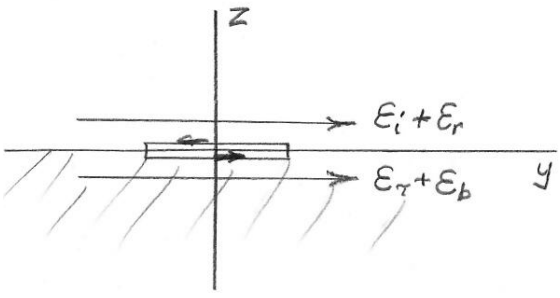
For medium *j* this can be written as

$$k_{xj}^2 + k_{zj}^2 = \frac{\omega^2\mu_j\varepsilon_j}{c^2} \tag{13.4a}$$

or defining the complex refractive index $n_j$, as

$$\frac{ck_j}{\omega} = \sqrt{\mu_j\varepsilon_j} = n_j. \tag{13.4b}$$

As discussed in Ch. 12, the $\pm$ sign that would ordinarily appear before the radical in Eq. (13.4b) is reserved for the direction of propagation, so the phase of $n_j$ is unambiguously connected to the phase of $\mu_j\varepsilon_j$. This connection among $k_x$, $k_z$, $\mu$,



and $\varepsilon$ is fundamental: a wave cannot exist unless it satisfies Eqs. (13.4).

Next, consider restrictions related to the interface. We start by applying Stokes' Theorem to the Maxwell-Faraday Equation using the contour indicated in the diagram at the bottom of the previous page. This is a *yz* cross-section that straddles the interface, where *y* is in the direction of the field. The rectangular path has a length $\Delta y$ and a height $\Delta z$. Applying Stokes' Theorem to the Faraday-Maxwell Equation we obtain in the limit $\Delta z \to 0$

$$\int_S d^2 r\, \hat{x} \cdot \nabla \times \vec{E} = \oint_C \vec{E} \cdot d\vec{l} = \left( E_\tau + E_b - E_r - E_i \right) \Delta l \tag{13.5a}$$

$$= -\frac{1}{c} \int_S d^2 r \frac{\partial \vec{B}}{\partial t} = 0 . \tag{13.5b}$$

Thus the tangential component of the total electric fields on the two sides of the interface are equal. This can be written more generally as

$$\hat{z} \times (\vec{E}_i + \vec{E}_r) = \hat{z} \times (\vec{E}_\tau + \vec{E}_b) . \tag{13.6}$$

As usual, by capitalizing on Stokes' Theorem we avoid having to deal with the interface itself, where the derivatives $\partial / \partial z$ are undefined, casting the calculation into regions where all functions are regular. While in the present context the eigenfunctions are plane waves, Eq. (13.6) is general, also applicable to the more complex modes encountered in crystal optics.

In the current plane-wave context, Eq. (13.6) imposes several restrictions, some of which are not immediately obvious. Evaluating Eqs. (13.1) at $z = 0$ explicitly and ignoring the common factor $e^{-i\omega t}$, Eq. (13.6) becomes

$$E_i e^{ik_{ix}x} + E_r e^{ik_{rx}x} = E_\tau e^{ik_{tx}x} + E_b e^{ik_{bx}x} . \tag{13.7}$$

Since Eq. (13.7) must hold at the interface everywhere and at all times, all phases $e^{ik_j x}$ must be the same, not only across a boundary, but everywhere throughout a configuration. Hence the $\hat{x}$ components of all wavevectors in the configuration must be equal, or

$$\hat{x} \cdot \vec{k}_r = k_{rx} = \hat{x} \cdot \vec{k}_t = k_{tx} = \hat{x} \cdot \vec{k}_b = k_{bx} = \hat{x} \cdot \vec{k}_i = k_{ix} = k_x . \tag{13.8a-g}$$

Except when dealing with elementary excitations, the specific value of $k_x$ is determined by the wave incident on the configuration, and is given in terms of its angle of incidence as $k_x = k_i \sin\theta_i$. Clearly, if these *x* components were not equal, there would be no hope of finding coefficients that would match fields across the boundary at all locations on the boundary.

This restriction on $k_x$ is called the *kinematic condition*, and has additional consequences of significance. These are:

(a) The vectors $\vec{k}_i$, $\vec{k}_r$, $\vec{k}_\tau$ and $\vec{k}_b$ must be coplanar. Equations (13.8) therefore also define the plane of incidence.

(b) In the incident medium, the angle of incidence equals the angle of reflection. This follows by combining Eqs. (13.4) and (13.8a). We have

$$k_x^2 + k_{iz}^2 = k_x^2 + k_{rz}^2 . \tag{13.9}$$

Thus $k_x / k_{iz} = k_x / k_{rz}$ .

(c) Again combining Eqs. (13.4) and (13.8), we find that the *z* components of all vectors $\vec{k}$ are given by the relation

$$k_{jz}^2 = k_j^2 - k_x^2 \tag{13.10}$$

for any material *j* in the configuration. Reserving in this case $\varepsilon_a$ to be the dielectric function of the transparent ambient where the incident plane wave arrives, we have

$$n_{jz} = \sqrt{\varepsilon_j - \varepsilon_a \sin^2 \theta_i} \tag{13.11}$$

for any medium *j*. In plasmonic applications we shall find situations where $k_{ix}^2 > k_i^2$, leading to wavevectors $k_{az}$ that are pure imaginary, necessary for plasmonic behavior.

(d) Because the dispersion equation requires

$$c^2 k^2 / \omega^2 = \varepsilon , \tag{13.12}$$

The *z* component of any wave vector is given by

$$k_z^2 = k^2 - k_a^2 \sin^2 \theta , \tag{13.13}$$

which we usually write as

$$n_{jz} = \frac{c k_{jz}}{\omega} = \sqrt{\varepsilon_j - \varepsilon_a \sin^2 \theta} \tag{13.14}$$

where $n_j = \sqrt{\varepsilon_j}$ is the complex index of refraction.

The final condition follows from Stokes' Theorem applied to Ampère's Equation. Determining first the magnetic fields $\vec{H}_j$ associated with the $\vec{E}_j$ at any region *j* in the configuration, the Faraday-Maxwell Equation yields

$$\vec{k}_j \times \vec{E}_j = -\frac{\omega}{c} \mu_j \vec{H}_j . \tag{13.15}$$

Therefore

$$\vec{H}_j = -\sqrt{\frac{\varepsilon_j}{\mu_j}} \hat{k}_j \times \vec{E}_j . \tag{13.16}$$

Given $\vec{H}_j$, the derivation leading to Eq. (13.6) now yields

$$\hat{z}\times(\vec{H}_i+\vec{H}_r)=\hat{z}\times(\vec{H}_\tau+\vec{H}_b)\,. \tag{13.17}$$

As with Eq. (13.6), Eq. (13.17) is completely general, applicable to any mode that can propagate in a material.

If the substrate is a perfect conductor, then $(\vec{H}_\tau+\vec{H}_b)=0$ and the surface-current term on the right side of in Ampère's Equation must be used. In this situation Eq. (13.13) becomes

$$\hat{z}\times(\vec{H}_i+\vec{H}_r)=\frac{4\pi}{c}\vec{K}\,, \tag{13.18}$$

where $\vec{K}$ is the surface current. This special case is easily recognized.

That the tangential components of $\vec{E}$ and $\vec{H}$ are continuous across the interface, and not some other combination of fields, is no accident. We recall that

$$\vec{S}=\frac{c}{4\pi}(\mathrm{Re}\,\vec{E})\times(\mathrm{Re}\,\vec{H})\,. \tag{13.19}$$

Hence with the tangenti.al components of $\vec{E}$ and $\vec{H}$ continuous at $z=0$, then the normal component of $\vec{S}$ is also continuous at $z=0$. Hence energy cannot accumulate at the interface. This of course places no restrictions on the tangential components of $\vec{S}$.

We now repeat the entire development for perpendicular fields $\vec{H}=\hat{y}H$. Nothing changes: the phase constraints remain, and the tangential components of the total $\vec{E}$ and $\vec{H}$ fields are again continuous across the interface. However, for any region *j* we now have

$$\vec{E}_j=\sqrt{\frac{\mu_j}{\varepsilon_j}}\hat{k}_j\times\vec{H}_j\,. \tag{13.20}$$

This expression is different from Eq. (13.16). Thus at non-normal incidence the reflection/transmission problem divides into two separate and distinct normal modes: a transverse-electric (TE) with $\vec{E}$ perpendicular to the plane of incidence, and a transverse-magnetic (TM) with $\vec{H}$ perpendicular. In the older literature these are denoted *s* and *p*-polarized modes, respectively, for the German *senkrecht* and *parallel*, which translate to perpendicular and parallel, respectively, in English. By now the *s*- and *p*-polarization nomenclature has nearly died out.

While the TE and TM solutions could also be described by the components of the fields in the plane of incidence, the advantages of working with single components should be obvious. Finally, the symmetry of the above results with respect to $\vec{E}$ and $\vec{H}$, and $\mu$ and $\varepsilon$, is evident.

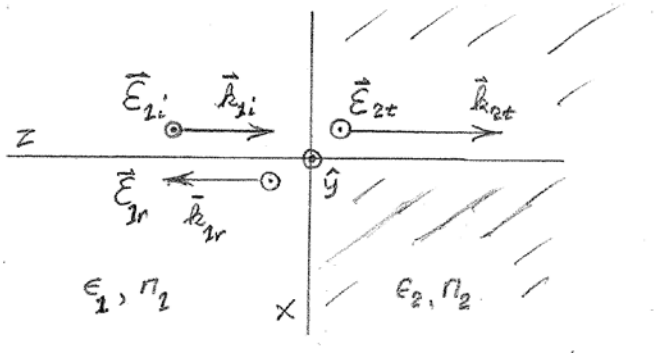
We next consider the implications of $\nabla \cdot \vec{D} = 0$ and $\nabla \cdot \vec{B} = 0$. These equations together with Gauss' Theorem show that the normal components of $\vec{D}$ and $\vec{B}$ must also be continuous across an interface. The question is therefore whether the boundary conditions are overdetermined. By Ampère's Equation

$$\hat{z} \cdot \vec{D}_j = \hat{z} \cdot \vec{k}_j \times \vec{H}_j = \vec{k}_j \times \vec{H}_j \cdot \hat{z} = \vec{k}_j \cdot \vec{H}_j \times \hat{z}, \tag{13.21}$$

where in the third version we have used the commutative property of the scalar product, and in the last expression we have interchanged the scalar and cross products. Now $\vec{H}_j \times \hat{z}$ is tangential to the interface, and hence is continuous across the interface. Being tangential, its scalar product with $\hat{k}_j$ extracts its tangential component of $\vec{k}_j$. Now by Eq. (13.8), the tangential component of $\vec{k}$ is the same everywhere throughout the configuration. Therefore, Eq. (13.21) shows that the normal component of $\vec{D}$ must also be continuous across an interface. The proof that the normal component of $\vec{B}$ is continuous across an interface follows by the same procedure starting from Eq. (13.12) and is left as a homework assignment. Thus overdetermination does not occur.

C. Reflectance and transmittance, the two-phase model at normal incidence.

The two-phase (ambient/substrate) model with propagation occurring at normal incidence to the interface is the simplest configuration involving reflection and transmission. It is important not only because it illustrates basic procedures but also defines the two-phase transmittance and reflectance expressions that simplify calculations for multilayer stacks. As shown in the diagram, let the dielectric functions on the left and right be $\varepsilon_1 = n_1^2$ and $\varepsilon_2 = n_2^2$ where the subscripts denote material 1 and material 2, respectively. For simplicity we assume $\mu_1 = \mu_2 = 1$, since this describes most systems of interest for optical applications. Let the incident beam be described as $\vec{E}_{1i}(z,t) = \hat{y} E_{1i} e^{-ik_1 z - i\omega t}$. The objective is to determine the reflected and transmitted beams $\vec{E}_{1r}(z,t) = \hat{y} E_{1r} e^{ik_1 z - i\omega t}$ and $\vec{E}_{2t}(z,t) = \hat{y} E_{2t} e^{-ik_s z - i\omega t}$, respectively. At $z = 0$ Eqs. (13.7), (13.11), and (13.12) give




$$E_{1i} + E_{1r} = E_{2t}; \tag{13.22a}$$

$$n_1 E_{1i} - n_1 E_{1r} = n_2 E_{2t}. \tag{13.22b}$$

The solution is

$$E_{1r} = \frac{n_1 - n_2}{n_1 + n_2} E_{1i}; \quad E_{2t} = \frac{2n_1}{n_1 + n_2} E_{1i}. \tag{13.23a,b}$$

The reflectance $r_{12}$ and transmittance $t_{12}$ are therefore

$$r_{12} = \frac{E_{1r}}{E_{1i}} = \frac{n_1 - n_2}{n_1 + n_2}; \quad t_{12} = \frac{E_{2t}}{E_{1i}} = \frac{2n_1}{n_1 + n_2}. \tag{13.24a,b}$$

The physical implications of Eq. (13.24) are straightforward. These expressions reduce correctly for $n_2 = n_1$, where $r_{12} = 0$. By analogy to the situation for electrostatics, this interface is termed virtual. However, had we retained $\mu$, we would find that the $r_{12} = 0$ condition in air would occur at any wavelength where $\mu_2 = \varepsilon_2$. If $n_1$ and $n_2$ are real and $n_1 < n_2$, the usual situation for an object in air, the phase of the incoming wave is inverted on reflection. If $n_1$ is real and $n_2$ is purely imaginary, as for a Drude metal for $\omega < \omega_p$, then $|r_{12}|$ is equal to 1 regardless of the values of $n_1$ and $|n_2|$.

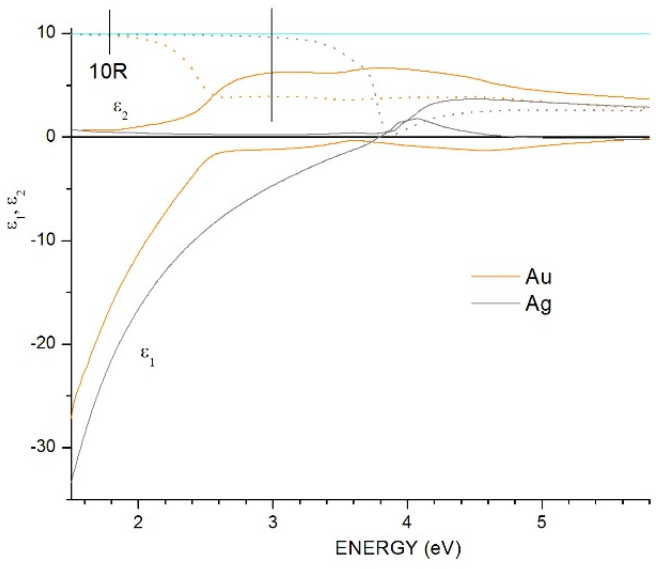
From the discussion of the previous chapter, the important quantity for slow detectors is power. The power reflectance is

$$R_{12} = \frac{I_{1r}}{I_{1i}} = |r_{12}|^2 = \left| \frac{n_1 - n_2}{n_1 + n_2} \right|^2. \tag{13.25}$$

For real $n_1$ and complex $\tilde{n}_2 = n_2 + i\kappa_2$ this can be written

$$R_{12} = \frac{(n_1 - n_2)^2 + \kappa_2^2}{(n_1 + n_2)^2 + \kappa_2^2}. \tag{13.26}$$

In the context of Eq. (13.26), $n_2$ is called the *ordinary refractive index* and $\kappa_2$ the *extinction coefficient*. For most transparent materials $n_2 \approx 1.5$, in which case $R_{12}$ in air is of the order of 0.05.

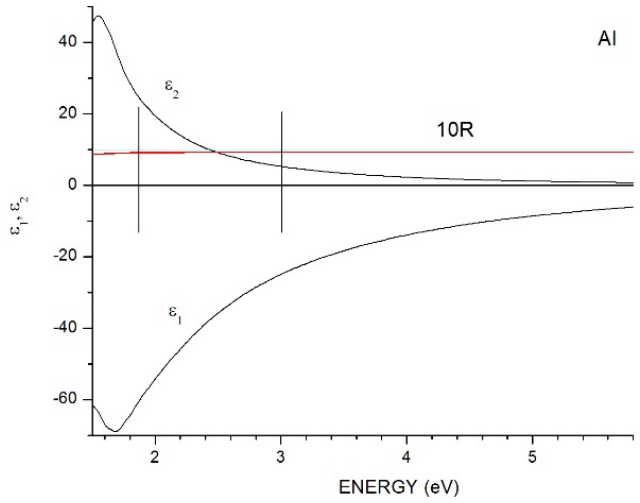



In metals and doped semiconductors, at sufficiently low frequencies free-carrier effects dominate the dielectric response. For the ideal Drude metal $R_{12} = 1$ for

$\omega < \omega_p$, defining the "plasma edge". In practice some losses are present, so the transition to $R_{12} = 1$ at $\omega = \omega_p$ is not abrupt. The existence of a plasma edge is the explanation of the "gold" color of gold, with $R_{12}$ being significantly larger in the yellow and red part of the spectrum than in the green and blue part. Owing to its electronic structure the plasma edge in copper occurs at longer wavelengths, leading to a more reddish color. A similar transition occurs in silver, but this occurs in the ultraviolet and so cannot be observed directly.

These points are illustrated in the figure on the previous page, which shows the dielectric functions of Au and Ag from 1.5 to 5.8 eV. The visible range from 700 to 400 nm is indicated by the short vertical lines. As can be seen, $R_{12}$ approaches unity where the real parts of the respective dielectric functions are substantially negative. The reflectance of silver is nearly equal to 1 over the entire visible range, and with a low absorption loss, silver represents the best approximation to the optical properties of the ideal Drude metal. Single-crystal Ag is also our best approximation of a perfect reflector in the visible range for natural materials, although we can do better at specific wavelengths with multilayer stacks of transparent materials, as discussed in Sec. D.

A second example is provided by Al, which is shown in the figure at the right. The plasma edge for Al is about 15 eV, well outside the energy range shown. Al exhibits substantial absorption in the visible range, but the reflectance remains uniformly high. In contrast, transition metals such as Ni and Cr exhibit strong absorption over the visible spectrum, which results in a power reflectance only of the order of 60%. Thus while Ni and Cr unquestionably have a shiny metallic appearance, they have a distinctly gray appearance relative to Al.




If additional interfaces are present deeper in the material, a back-reflected wave is generally present present. In this situation Eqs. (12.22) are inadequate. The equivalent interface expressions in this case are

$$E_{1i} + E_{1r} = E_{2t} + E_{2b}; \tag{13.27a}$$

$$n_1 E_{1i} - n_1 E_{1r} = n_2 E_{2t} - n_2 E_{2b}; \tag{13.27b}$$

where $E_{2b}(z,t) = E_{2b} e^{ik_2 z - i\omega t}$ is the field of the back-reflected wave. The solution of Eqs. (13.27) is best written in matrix form:

$$\begin{pmatrix} E_{1i} \\ E_{1r} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(1+\frac{n_2}{n_1}\right) & \frac{1}{2}\left(1-\frac{n_2}{n_1}\right) \\ \frac{1}{2}\left(1-\frac{n_2}{n_1}\right) & \frac{1}{2}\left(1+\frac{n_2}{n_1}\right) \end{pmatrix} \begin{pmatrix} E_{2t} \\ E_{2b} \end{pmatrix} \tag{13.28a}$$

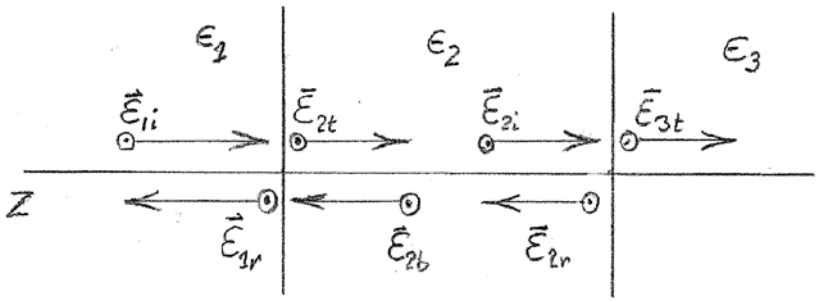
$$= \frac{1}{t_{12}} \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} E_{2t} \\ E_{2b} \end{pmatrix}, \tag{13.28b}$$

recalling the definitions of $t_{12}$ and $r_{12}$. The mathematics is similar in form to the Jones-matrix calculus discussed earlier, but the meanings of the two-vectors are different. By taking advantage of the definitions of $r_{12}$ and $t_{12}$, the connections of the fields on either side of the interface take a particularly simple form. Writing the incident/reflected fields in terms of the transmitted/back-reflected fields is a form of planning ahead, since the typical boundary condition is $E_{\nu b} = 0$, where $\nu$ represents the $\nu^{th}$ material in the stack. This final layer is termed the substrate.

More generally, Eqs. (13.28) represent the coupling of a linearly polarized incident wave to a linearly polarized reflected wave and two linearly polarized waves in the substrate. For materials with tensorial dielectric functions, the normal modes are generally not linearly polarized plane waves but have a more complicated structure. The elliptically polarized normal modes of optically active materials such as α-quartz is an example. In this case the boundary conditions associated with an incoming linearly polarized plane wave can only be satisfied if both modes are activated, so the transmitted beam consists of two waves, which in turn generates reflected beams of both polarizations (or a linear combination.) In this case a 2-vector wave representation is not possible, and all 4 waves must be considered. We do not pursue this point further, but mention it for completeness.

D. The three-phase model: the Airy Equation.

The previous section nearly sets us up for multilayer stacks. We are missing only the propagation matrix, which describes the phase delays that occur in propagating waves from one interface to the next. This is conveniently introduced by considering the so-called 3-phase model, which consists of an ambient, and overlayer, and a substrate, as indicated in the figure. A beam is incident from the left, generating reflected, transmitted, and back-reflected waves that in turn interact with the deeper interface. The boundary condition is that there is no back-reflected wave in the substrate.



The connections across interfaces are described by Eq. (13.28b). For the isotropic materials considered here, the connection between the 2-vectors on the transmission/back-reflected side and the incidence/reflected side is established directly from the waves themselves. If the thickness of layer 2 in the figure is $d_2$, the connection is

$$\begin{pmatrix} E_{2t} \\ E_{2b} \end{pmatrix} = \begin{pmatrix} e^{-ik_2 d_2} & 0 \\ 0 & e^{ik_2 d_2} \end{pmatrix} \begin{pmatrix} E_{2i} \\ E_{2r} \end{pmatrix}, \tag{13.29}$$

where the $(2\times 2)$ matrix in Eq. (13.29) is the propagation matrix. This matrix follows directly either by substituting the location $z = -d_2$ of the deeper interface in the plane waves themselves, or simply by noting that the phase must accumulate in the direction of propagation.

The complete description of the three-phase model is therefore

$$\begin{pmatrix} E_{1i} \\ E_{1r} \end{pmatrix} = \frac{1}{t_{12}} \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} e^{-ik_2 d_2} & 0 \\ 0 & e^{ik_2 d_2} \end{pmatrix} \begin{pmatrix} 1 & r_{23} \\ r_{23} & 1 \end{pmatrix} \begin{pmatrix} E_{3t} \\ 0 \end{pmatrix}, \tag{13.30}$$

where $r_{23}$ and $t_{23}$ are the equivalents of Eqs. (13.24a) and (13.24b) for materials 2 and 3. The direction of information flow in the figure is consistent with the ordering of the matrices in the equation describing the system.

Equation (13.30) is easily solved for the reflectance $r_{123}$ of the three-layer stack. The closed-form expression is

$$r_{123} = \frac{E_{1r}}{E_{1i}} = \frac{r_{12} + r_{23} e^{i2k_2 d_2}}{1 + r_{12} r_{23} e^{i2k_2 d_2}}. \tag{13.31}$$

This is the famous Airy Equation, named after the George Airy of Airy-function fame. Looking ahead to Sec. E, the non-normal-incidence version is identical except $r_{12}$ and $r_{23}$ are replaced with their non-normal-incidence equivalents, and $k_2$ with its component normal to the surface. To lowest order it has the physical characteristic that we expect: if material 2 is absorbing, then for sufficient thickness the exponential terms vanish, and we are left with $r_{12}$. This is precisely what should happen if back reflection vanishes.

Equation (13.31) describes a wide range of phenomena, from antireflection and decorative coatings through oil slicks to soap bubbles. The antireflection calculation is straightforward. Given $n_3$ and $n_1$, what are the values of $n_2$ and $d_2$ that minimize $r_{123}$? Considering $n_1$, $n_2$, and $n_3$ to be real, and $n_1 < n_2 < n_3$, then both terms in the numerator of Eq. (13.29) are negative. Thus if the numerator is to vanish, then $e^{2ik_2 d_2} = -1$. The second requirement is that $r_{12} = r_{23}$. This leads to $n_2 = \sqrt{n_1 n_3}$. Finally,

$$d_o = d_{o\nu} = \frac{1}{2k_o}(\pi + 2\pi\nu) = \frac{\lambda_a}{4n_o}(1+2\nu)\; d_{2\nu} = \frac{1}{2k_2}(\pi + 2\pi\nu) = \frac{\lambda_1}{4n_2}(1+2\nu), \tag{13.32a}$$

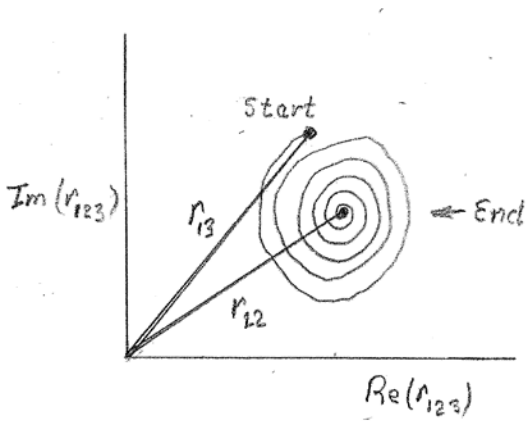
$$= \frac{\lambda_2}{4}, \qquad (13.32b)$$

where $\nu$ is an integer and $\lambda_2 = \lambda_1 / n_2$ is the wavelength of light in the layer. With $\nu = 0$ the change of phase with $d_2$ is as small as possible, making the antireflection coating effective over the largest wavelength range possible. The maximum reflectance condition occurs for $e^{2ik_2 d_2} = 1$, with a reflectance the same as if the overlayer were not present. This can be verified by direct calculation.

In informal situations $\nu = 0$ is not always realized. For example, the distribution of somewhat washed-out colors seen for an oil slick on water occurs because several zero- and maximum-reflectance conditions generally overlap and the difference between the refractive indices of oil $(n_2 \sim 1.5)$ and water $(n_3 = 1.33)$ is small. Similar effects are seen in soap bubbles, where the difference between $(n_1 = n_3 = 1)$ and $n_2 = 1.33$ is again small. Because the water in soap bubbles evaporates, their color changes continually with time. The reflectance vanishes completely just before breakage, because at $d_2 = 0$, because at $d = 0$ the reflectance at all wavelengths is minimized simultaneously. The physics is analogous to that of a Fourier-transform infrared spectrometer with a moving mirror, but the investment in laboratory equipment is much less.

A technologically important application is the Fabry-Perot interferometer, which is a slab of transparent material with parallel sides. With no losses, the antireflection condition translates to a transmission of 100%. Hence the device can be used for wavelength filtering and/or wavelength calibration. Fabry-Perot interferometers are usually fabricated by coating both surfaces with a thin metallic film to increase $r_{12}$, thereby artificially increasing contrast at some cost of transmitted intensity.

If the overlayer is optically absorbing, then for sufficiently large *d* the terms containing the exponentials vanish, and the reflectance reduces to that of the two-phase model with the overlayer playing the role of the substrate. Left as a homework assignment is the demonstration that for $d_2 = 0$, the three-phase model reduces to the reflectance $r_{12}$ of the two-phase model, and $n_2$ disappears. Finally, if the overlayer is weakly absorbing and $r_{23}$ is small, then to a good approximation the denominator can be set equal to 1, and $r_{123}$ traces out an exponential spiral starting at $r_{13}$ and ending at $r_{12}$, as shown in the diagram. The power reflectance is the square of the length of a line drawn between the origin and the appropriate point on the spiral, as shown in the figure. As $d_2$ increases this length increases and decreases, leading to the well-known interference oscillations.



The exponential-spiral limit has also been used in a "virtual-interface" approach to assess compositions of materials being deposited in real time. In this application the substrate is treated as an unknown, to be determined if the rate of deposition is known. The result has been used to control compositions of alloys during growth.

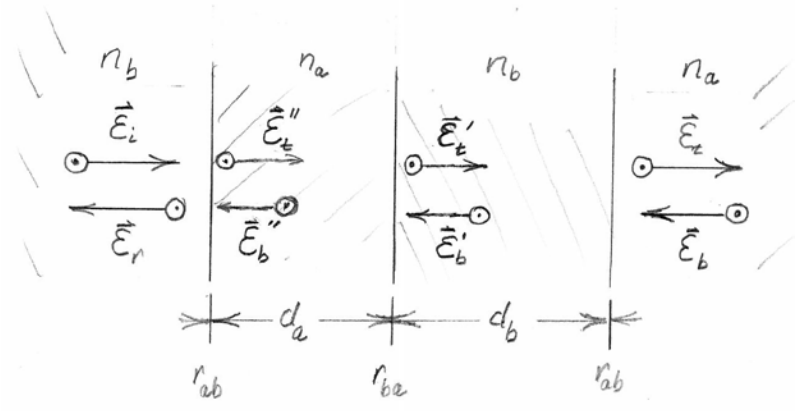
As a matter of interest, in older texts the three-phase model is treated as the limit of an infinite sum, where a portion of the initial transmitted wave is back-reflected, returns to be reflected off the first interface, then back-reflected again, etc., etc. While this approach gives the same result, it is clearly impractical if second film were added. The recognition that these multilayer-stack problems could be solved using normal modes, as done above, was only achieved surprisingly late, by Abèles in 1950.

E. Multilayer stacks: one-dimensional photonic crystals.

Taken together, the above sections reduce the multilayer-stack problem to bookkeeping. However, with few exceptions, evaluating the matrix product of a multilayer stack is a computer problem. Good practice dictates that after solving the configuration starting with the boundary condition $E_{b\nu} = 0$, that the computation is reversed to ensure that $E_{b\nu}$ is indeed zero.

An exception to the computer calculation for multilayer stacks is the configuration currently termed a one-dimensional photonic crystal. This is a repeating two-layer pair. The configurations can be solved analytically. The simplest practical example is the end-cavity mirror in a gas laser, which consists of alternating layers of transparent materials of different refractive indices, each layer being a quarter-wave thick. In the asymptotic limit the reflectance is 100%.

The reason that these relatively "exotic" mirrors are used can be understood by considering loss. For lasers to function, laser-cavity mirrors must be as close to ideal as possible. While some metals exhibit high reflectance, as seen in the figures in Sec. C, metal mirrors are far too lossy for use in laser cavities even though they may be more than adequate for casual applications. Because the materials from which the mirrors are fabricated are transparent, absorption losses are zero, so within a range of frequencies the mirrors can be considered ideal. We analyze them here, not only for their technical importance, but also because the analysis is a good example of why accurate bookkeeping is important, how self-consistent concepts are used in repeating systems in general (electrical filters for example), the math is interesting in its own right (the solution has some surprises), and they represent a good opportunity to pull the physics out of the math.




The theory is a straightforward extension of Eq. (13.30). Consider

the fundamental repeating unit from an interior segment of the stack, which is shown schematically in the figure at the bottom of the preceding page. The layers are described by refractive indices $n_a$ and $n_b$ and thicknesses $d_a$ and $d_b$. Consistent with the previous section, we assume that the source is on the left. Accordingly, the calculation starts with the fields $\vec{E}_t$ and $\vec{E}_b$ adjacent to the *b*/*a* boundary in medium *a* on the right, working as usual from right to left, and completing the unit cell with the fields $\vec{E}_t$" and $\vec{E}_b$"adjacent to the *b*/*a* boundary on the left. The fields $\vec{E}_t$' and $\vec{E}_b$' at the interior *a*/*b* boundary are used below in interpreting the solution. By extending Eq. (13.30) we find

$$\begin{pmatrix} E_t'' \\ E_b'' \end{pmatrix} = \begin{pmatrix} e^{-ik_a d_a} & 0 \\ 0 & e^{ik_a d_a} \end{pmatrix} \frac{1}{t_{ba}} \begin{pmatrix} 1 & r_{ba} \\ r_{ba} & 1 \end{pmatrix} \begin{pmatrix} e^{-ik_b d_b} & 0 \\ 0 & e^{ik_b d_b} \end{pmatrix} \frac{1}{t_{ab}} \begin{pmatrix} 1 & r_{ab} \\ r_{ab} & 1 \end{pmatrix} \begin{pmatrix} E_t \\ E_b \end{pmatrix}. \tag{13.33}$$

The rightmost $(2\times 2)$ matrix and its prefactor $t_{ab}$ describe the result of crossing the *b*/*a* boundary on the right, where

$$r_{ab} = \frac{n_b - n_a}{n_b + n_a}; \;\; t_{ab} = \frac{2n_b}{n_b + n_a}. \tag{13.34}$$

The subscripts are reversed for the *a*/*b* interface in the interior. Verification that these definitions are correct is left as a homework exercise.

If $d_b = 0$ the rightmost propagation matrix and the product of the two reflectance matrices both reduce to the unit matrix. This leaves only the leftmost propagation matrix, as expected from the physics involved. Therefore, we can anticipate that the greatest contrast will occur when the difference between the propagation terms for the rightward- and leftward-propagating waves is a maximum, or $e^{ik_a d_a} = e^{ik_b d_b} = e^{i\pi/2} = i$. Substituting these values and carrying out the multiplications explicitly yields

$$\begin{pmatrix} E_t'' \\ E_b'' \end{pmatrix} = -\frac{1}{2n_b n_a} \begin{pmatrix} n_b^2 + n_a^2 & n_b^2 - n_a^2 \\ n_b^2 - n_a^2 & n_b^2 + n_a^2 \end{pmatrix} \begin{pmatrix} E_t \\ E_b \end{pmatrix}. \tag{13.35}$$

We now assume that the stack is infinitely thick, so that end effects are irrelevant. In that case all unit cells are equivalent, so the vector on the left must be a scaled version of the vector on the right. Then write

$$\begin{pmatrix} E_t'' \\ E_b'' \end{pmatrix} = C_0 \begin{pmatrix} E_t \\ E_b \end{pmatrix}, \tag{13.36}$$

where $C_0$ is a constant to be determined. Substituting Eq. (13.34) into Eq. (13.33) yields

$$0 = -\frac{1}{2n_b n_a} \begin{pmatrix} (n_b^2 + n_a^2) + 2n_b n_a C_0 & (n_b^2 - n_a^2) \\ (n_b^2 - n_a^2) & (n_b^2 + n_a^2) + 2n_b n_a C_0 \end{pmatrix} \begin{pmatrix} E_t \\ E_b \end{pmatrix}. \tag{13.37}$$

This is a standard eigenvalue/eigenvector calculation with a nontrivial solution if the determinant of the $2\times 2$ matrix equals zero. After some algebra we find

$$2n_b n_a C_0 = -(n_b^2 + n_a^2) \pm (n_b^2 - n_a^2), \tag{13.38}$$

and $E_b = \mp E_t$. For the upper sign

$$C_0^+ = -\frac{n_a}{n_b};\ \begin{pmatrix} E_t' \\ E_b' \end{pmatrix} = -\frac{n_a}{n_b}\begin{pmatrix} 1 \\ -1 \end{pmatrix} E_t, \tag{13.39a,b}$$

and for the lower sign

$$C_0^- = -\frac{n_b}{n_a};\ \begin{pmatrix} E_t' \\ E_b' \end{pmatrix} = -\frac{n_b}{n_a}\begin{pmatrix} 1 \\ 1 \end{pmatrix} E_t. \tag{13.40a,b}$$

If we stop at this point, the solution cannot be interpreted (I tried.) It is necessary to evaluate the fields. Considering the upper sign first and placing $z = 0$ at the rightmost *b*/*a* boundary, more math yields

$z \le 0$:

$$\vec{E}_-(z,t) = -2i\hat{y}E_t \sin(k_a z)e^{-i\omega t};\ \ \vec{H}_-(z,t) = 2n_a\hat{x}E_t \cos(k_a z)e^{-i\omega t}; \tag{13.41a,b}$$

$z \ge 0$:

$$\vec{E}_+(z,t) = -2i\hat{y}\frac{n_a}{n_b}E_t \sin(k_b z)e^{-i\omega t};\ \ \vec{H}_+(z,t) = 2n_a\hat{x}E_t \cos(k_b z)e^{-i\omega t}. \tag{13.42c,d}$$

Thus instead of running waves, as we might guess from the eigenvalue/eigenvector solution, the fields occur in standing waves. Hence the average Poynting vector is zero, which we could have anticipated given that a perfect mirror reflects all incoming radiation. The radiation in the layers themselves is trapped.

What is more interesting is how the solution avoids a potential violation of the boundary conditions on tangential $\vec{E}$ and $\vec{H}$. Considering Eqs. (13.41), the tangential $\vec{H}$ field is clearly continuous, but the prefactors of $\vec{E}_-(z,t)$ and $\vec{E}_+(z,t)$ do not agree. However, the equations deal with this by using the sine function to ensure that the electric fields vanish at $z = 0$. Advancing the solution to the interior boundary *a*/*b* at $k_a d_a = \pi/4$, we find

$$\vec{E}(0^-,t) = \vec{E}(0^+,t) = -2i\hat{y}\frac{n_a}{n_b}e^{-i\omega t}; \tag{13.43a}$$

$$\vec{H}(0^-,t) = \vec{H}(0^+,t) = 0. \tag{13.43b}$$

Here the continuity condition on $\vec{E}$ is nontrivial, and the potentially contradictory magnetic-field condition is avoided because $\vec{H} = 0$. Thus the roles of $\vec{E}$ and $\vec{H}$ have interchanged. Note also $\vec{E}$ has picked up a prefactor $\frac{n_a}{n_b}$, as predicted. The overall reversal of sign occurs because $r_{ab}$ and $r_{ba}$ have opposite signs.

We now consider the second solution given by the lower sign, Eqs. (13.40). Here, $C_0 = -n_b/n_a$ and $E_b = E_t$. Repeating the same calculations yields

$z \le 0$:

$$\vec{E}_-(z,t) = -2i\hat{y}E_t \sin(k_a z)e^{-i\omega t}; \quad \vec{H}_-(z,t) = 2n_a\hat{x}E_t \cos(k_a z)e^{-i\omega t}; \tag{13.44a,b}$$

$z \ge 0$:

$$\vec{E}_+(z,t) = -2i\hat{y}\frac{n_a}{n_b}E_t \sin(k_b z)e^{-i\omega t}; \quad \vec{H}_+(z,t) = 2n_a\hat{x}E_t \cos(k_b z)e^{-i\omega t}. \tag{13.44c,d}$$

The roles of $\vec{E}$ and $\vec{H}$ have reversed. Hence at the *b*/*a* boundary the two solutions can be considered TM and TE, respectively, at the *b*/*a* boundary, and TE and TM, respectively, at the *a*/*b* boundary. Evaluating these fields at the interior boundary we have

$$\vec{E}\,'(0^-,t) = \vec{E}\,'(0^+,t) = 0; \tag{13.45a}$$

$$\vec{H}\,'(0^-,t) = \vec{H}\,'(0^+,t) = -2i\hat{x}n_b E_t. \tag{13.45b}$$

This is the starting value multiplied by $(n_b/n_a)$, consistent with the eigenvalue. Again, the overall reversal of sign occurs because $r_{ab}$ and $r_{ba}$ have opposite signs.

Another interesting aspect of the configuration is that it is symmetric overall, but locally asymmetric. Because the mirror is a passive device, the standing waves mustincrease in amplitude as the source side is approached. Therefore, for the *b*/*a* boundaries, the source is on the left of the figure if $n_b > n_a$, otherwise on the right. The situation is reversed if $n_a > n_b$. The fact that the source can be on either side is a consequence of overall symmetry, but that it is on a specific side at a given boundary is a consequence of local asymmetry.

Finally, it is of interest to consider how the standing waves evolve starting from the true substrate, where $\vec{E}_b = 0$. Taking $n_b > n_a$ so the substrate is on the right, consistent with the order of matrix multiplication, the overall gain is $(n_b/n_a)^\nu$ for $\nu$ unit cells. To determine how this gain is achieved, write

$$\begin{pmatrix} E_t \\ E_b \end{pmatrix} = \begin{pmatrix} 1 \\ 1-\delta \end{pmatrix} E_t, \tag{13.46}$$

where $\delta$ is the discrepancy from the limiting value of 1. After some mathematics, we find

$$\begin{pmatrix} E_t\,' \\ E_b\,' \end{pmatrix} = -\frac{2n_b^2 - (n_b^2 - n_a^2)\delta}{2n_b n_a}\begin{pmatrix} 1 \\ 1 - \dfrac{2n_a^2}{2n_b^2 - (n_b^2 - n_a^2)}\delta \end{pmatrix} E_t$$

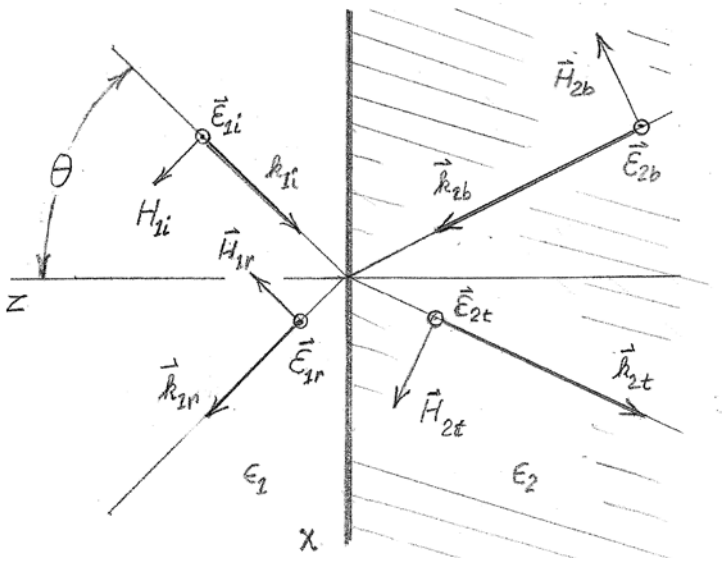
$$= -\frac{2n_b^2 - (n_b^2 - n_a^2)\delta}{2n_b n_a}\begin{pmatrix} 1 \\ 1 - \dfrac{2n_a^2}{n_b^2 + n_a^2}\delta \end{pmatrix} E_t. \tag{13.47}$$

Thus each new cell reduces the discrepancy between its values and those of the asymptotic solution. Reduction also occurs for $\delta = 1$, that is, for the first unit cell at the boundary-condition side.

The general formalism presented above remains valid even if the boundaries between layers are not abrupt. Although contrast is reduced, the important parameter is the lowest coefficient of the spatial Fourier transform of the configuration at the wave vector of the incident light. The mathematics is analogous to the propagation of photons in any periodic structure (photonics) or electrons in a crystal lattice (band structure). For certain wavelengths there exist stop bands analogous to the forbidden gaps in the energy band structure of solids. Thus the above derivation or its equivalent is found in a wide range of disciplines, although the usual treatments do not go into detail about how attenuation occurs.

F. Non-normal incidence.

At non-normal incidence, the new physics that arises includes the plane of incidence, Snell's Law, total internal reflection, Brewster's Angle, the Ewald-Oseen Extinction Theorem, and surface plasmon polaritons (interface plasmons). When vectors $\vec{k}$ are not perpendicular to the interface, the directional degeneracy of the electric and magnetic fields that occurs at normal incidence is removed, leading to normal modes that are either transverse-electric (TE) or transverse-magnetic (TM) according to whether $\vec{E}$ or $\vec{H}$ is perpendicular to the plane of incidence. In older literature the TM and TE modes are called *p*- and *s*-polarizations, respectively, for the German *parallel* and *senkrecht*, respectively. The wave vectors and remaining fields are in the plane of incidence, with components parallel and perpendicular to the surface plane. We develop the topic by following the usual approach of getting the math right, then extracting the physics from the math. Some of what follows is a repeat of material in Sec. A, but is provided to better appreciate the entire picture.




We assume that our planar configuration is embedded in a transparent medium described by a real dielectric function $\varepsilon_1$ and illuminated by a plane wave of wave vector $\vec{k}_i$, as indicated in the diagram. If the wave is arriving at an angle of incidence $\theta$, then

$$\vec{k}_i = \hat{x}k_{ix} + \hat{z}k_{iz} = k_i(\hat{x}\sin\theta - \hat{z}\cos\theta)\,. \tag{13.46}$$

In Sec. B we found that if tangential $\vec{E}$ and $\vec{H}$ were to be continuous everywhere across an interface, then their phase projections $k_{jx}$ must be identical on both sides of an interface, and hence everywhere throughout the next layer. At deeper interfaces the process repeats, and hence $k_{jx} = k_{1x}$ everywhere throughout the configuration. Then since the dispersion equation $c^2k_j^2/\omega^2 = \varepsilon_j$ must also be satisfied, it follows that for any layer *j* the *z* component $k_{jz}$ of $\vec{k}_j$ must be given by

$$k_{jz} = \pm\sqrt{k_j^2 - k_1^2\sin^2\theta}\,, \tag{13.47}$$

or in dimensionless form

$$\frac{ck_{jz}}{\omega} = n_{jz} = \pm\sqrt{\varepsilon_j - \varepsilon_1\sin^2\theta}\,. \tag{13.48}$$

As usual, the $\pm$ sign refers to the direction of propagation, not to the radical. Except for the occurrence of $\vec{k}$ in the complex exponential, the normalization $c/\omega$ is preferred because it not only simplifies the notation but also recognizes that, because all wave vectors scale in the same way, the relevant frequency dependences lie in refractive indices and dielectric functions. It is always possible to transfer reversibly between any *k* or $k^2$ and the associated *n* or $\varepsilon$ with the $c/\omega$ or $(c/\omega)^2$ scaling factor.

For real dielectric functions, an immediate consequence of the equality of the tangential components is Snell's Law. If $k_2$ is real, then an analogous angle $\theta_2$ can be defined in medium 2, and $k_1\sin\theta_1 = k_2\sin\theta_2$, or in refractive-index terms

$$n_1\sin\theta_1 = n_2\sin\theta_2\,. \tag{13.49}$$

If $\varepsilon_2$ is complex, then the real part of $\vec{k}_2$ defines planes of equal phase, as with Snell's Law, but the imaginary part defines planes of constant attenuation. Thus in an absorbing material Snell's Law divides into two parts.

We now consider the implications that follow from applying the dispersion-equation constraint. One consequence is total internal reflection. If $\varepsilon_2$ is real and $\varepsilon_1\sin^2\theta > \varepsilon_2$, then the quantity under the radical in Eq. (13.48) is negative, and $k_{2z}$ is pure imaginary. The wave cannot propagate in medium 2 but instead decays exponentially away from the interface. With the $90^\circ$ phase shift between $\vec{E}$ and $\vec{H}$, the time average of the component of the Poynting vector normal to the surface is zero. This occurs when a wave propagating in a more optically dense medium arrives at an interface with a less optically dense material when the angle of incidence is too large. As a result, the view of the external world is constrained to a cone of angle $\sin\theta < n_2/n_1$. Anyone who has opened his or her eyes underwater has experienced this phenomenon; in the water/air case the critical angle is $\theta_c = \sin^{-1}(1/1.33) \simeq 49^\circ$. An analogous situation occurs in light-

emitting diodes: extracting radiation from a medium with $n \sim 3$ to an outside world of $n = 1$ is a challenge that must be met with textured surfaces. This is also the condition necessary for the existence of interface plasmons, as discussed in detail in the next section.

An interesting example of total "internal" reflection occurs when X-rays arrive at an interface at near-tangential incidence. At X-ray frequencies the dielectric functions of most materials are slightly less than 1. As noted in Ch. 7, this follows because at these energies electrons with shallower binding energies can be considered free and those at deeper energies ignored. The calculation is done for Si exposed to 1 keV X-rays in a homework assignment. Total external reflection is used for routing X-rays in beam lines at synchrotrons as well as other sources of "hard" radiation.

We next consider how best to describe waves propagating at non-normal incidence. Once this is determined, expressions for TE and TM reflectances and transmittances follow directly. $\vec{E} = E\hat{y}$ is the clear choice for representing TE modes, because in general $\vec{H}$ has two components in the plane of incidence, and it is obviously better to work with one component than two. Given $\vec{E}$, then $\vec{H}$ is calculated using Eq. (13.12). The TM case is the inverse: the efficient approach here is to cast everything in terms of $\vec{H} = H\hat{y}$ and calculate $\vec{E}$ according to Eq. (13.16). Then in both cases each wave is described entirely by a single component.

As with normal incidence, calculating non-normal-incidence reflectances and transmittances is a bookkeeping problem. The sensible approach here is to construct a table, then apply the boundary conditions. For the two-phase model, the TE table with $\mu_1 = \mu_2 = 1$ is

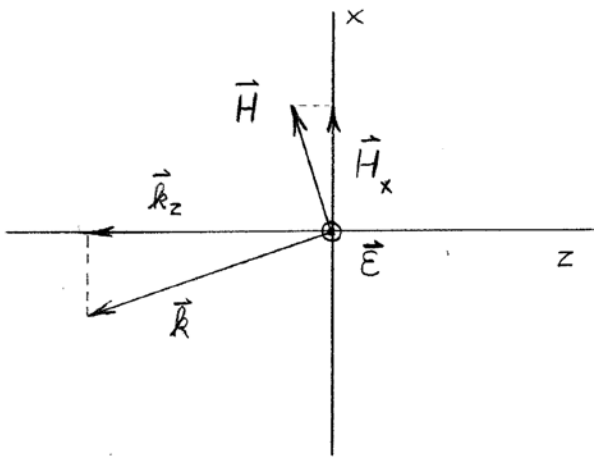
| | Wave vector | Electric field | Magnetic field |
|---|---|---|---|
| Incident: | $k_{1x}\hat{x} - k_{1z}\hat{z}$ | $E_{1i}\hat{y}$ | $E_{1i}(\hat{x}\, n_{1z} + \hat{z}\, n_{1x})$; |
| Reflected: | $k_{1x}\hat{x} + k_{1z}\hat{z}$ | $E_{1r}\hat{y}$ | $E_{1r}(-\hat{x}\, n_{1z} + \hat{z}\, n_{1x})$ |
| Transmitted: | $k_{2x}\hat{x} - k_{2z}\hat{z}$ | $E_{2t}\hat{y}$ | $E_{2t}(\hat{x}\, n_{2z} + \hat{z}\, n_{1x})$; |
| Back-reflected: | $k_{2x}\hat{x} + k_{2z}\hat{z}$ | $E_{2b}\hat{y}$ | $E_{2b}(-\hat{x}\, n_{2z} + \hat{z}\, n_{1x})$; |

where the $n_{jz}$ follow from Eq. (13.41). The table is identical to that for normal incidence except that the $n_j$ are replaced by $n_{jz}$. Making the same substitutions in Eqs. (13.20a,b) we obtain (as a reminder, still for $\mu = 1$)

$$r_{12}^{TE} = \frac{n_{1z} - n_{2z}}{n_{1z} + n_{2z}}; \tag{13.50a}$$

$$t_{12}^{TE} = \frac{2n_{1z}}{n_{1z} + n_{2z}}. \tag{13.51b}$$

The impedance mismatch between the materials on the two sides of the interface is now defined by the normal components of the wave vectors, not the wave vectors themselves. It is evident that Eqs. (13.44) reduce correctly to normal-incidence expressions for $\theta = 0$. Had we retained $\mu$, all components $n_{jz}$ would be replaced by $n_{jz}/\mu_j$. The proof of this is left as a homework exercise. This replacement of $k_j$ with $k_{jz}$ also applies to the complex exponentials of the propagation matrix, because the phase of plane waves is given by $e^{i\vec{k}_j\cdot\vec{r}} = e^{ik_{jx}x + ik_{jz}z}$.




Calculations of reflectances and transmittances for the TM case follow similarly, except that in practical applications we are not allowed to set $\varepsilon_j = 1$. The above table then reads

| | Wave vector | Magnetic field | Electric field |
|---|---|---|---|
| Incident: | $k_{1x}\hat{x} - k_{1z}\hat{z}$ | $H_i\hat{y}$ | $(1/\varepsilon_1)H_i(-\hat{x}n_{1z} - \hat{z}n_{1x})$ |
| Reflected: | $k_{1x}\hat{x} + k_{1z}\hat{z}$ | $H_r\hat{y}$ | $(1/\varepsilon_1)H_r(\hat{x}n_{1z} - \hat{z}n_{1x})$ |
| Transmitted: | $k_{1x}\hat{x} - k_{1z}\hat{z}$ | $H_t\hat{y}$ | $(1/\varepsilon_2)H_t(-\hat{x}n_{2z} - \hat{z}n_{1x})$ |
| Back-reflected: | $k_{1x}\hat{x} + k_{1z}\hat{z}$ | $H_b\hat{y}$ | $(1/\varepsilon_2)H_b(\hat{x}n_{2z} - \hat{z}n_{1x})$ |

Comparing to the normal-incidence situation then drawing on Eqs. (13.20a,b), we have

$$r_{12}^{TM} = \frac{\dfrac{n_{1z}}{\varepsilon_1} - \dfrac{n_{2z}}{\varepsilon_2}}{\dfrac{n_{1z}}{\varepsilon_1} + \dfrac{n_{2z}}{\varepsilon_2}} = \frac{\varepsilon_2 n_{1z} - \varepsilon_1 n_{2z}}{\varepsilon_2 n_{1z} + \varepsilon_1 n_{2z}}; \tag{13.52a}$$

$$t_{12}^{TM} = \frac{2\dfrac{n_{1z}}{\varepsilon_1}}{\dfrac{n_{1z}}{\varepsilon_1} + \dfrac{n_{2z}}{\varepsilon_2}} = \frac{2\varepsilon_2 n_{1z}}{\varepsilon_2 n_{1z} + \varepsilon_1 n_{2z}}. \tag{13.52b}$$

These expressions are usually written in the form shown at the right in Eqs. (13.52).

These derivations apply to any interface, whether isolated or part of a multilayer stack. Therefore, all previous matrix relations, although derived for normal incidence, are valid for non-normal incidence with the substitution of the appropriate reflectances, that is, Eqs. (13.50) and (13.51) for the TE and TM cases, respectively.

G. Brewster's Angle and interface plasmons.

If $\mu = 1$, Brewster's Angle and interface plasmons are unique to TM polarization. Calling such widely different phenomena closely related appears to be a mistake, but the connection is established below.

We first consider Brewster's Angle $\theta = \theta_B$, the more familiar case. For transparent materials, $r_{12}^{TM}$ vanishes at $\theta_B$. Hence to determine $\theta_B$ we simply set the numerator of Eq. (13.52a) equal to zero:

$$\varepsilon_2 n_{1z} - \varepsilon_1 n_{2z} = 0\,, \tag{13.53}$$

and solve for $\theta_B$. The solution proceeds by moving the second term on the left to the right side of the equation and squaring the result to eliminate the radicals. We find

$$\begin{aligned} \varepsilon_2^2 n_{1z}^2 &= \varepsilon_1^2 n_{2z}^2 \\ &= \varepsilon_2^2(\varepsilon_1 - \varepsilon_1 \sin^2\theta_B) = \varepsilon_1^2(\varepsilon_2 - \varepsilon_1 \sin^2\theta_B)\,. \end{aligned} \tag{13.54}$$

Rearranging terms gives

$$\varepsilon_2\varepsilon_1(\varepsilon_2 - \varepsilon_1) = \varepsilon_1(\varepsilon_2^2 - \varepsilon_1^2)\sin^2\theta_B\,. \tag{13.55}$$

Cancelling common factors leads to

$$\sin^2\theta_B = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2}\,, \tag{13.56a}$$

so evidently

$$\cos^2\theta_B = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2}\,. \tag{13.56b}$$

Taking the ratio of Eqs. (13.56a) and (13.56b), we obtain the well-known result

$$\frac{\sin\theta_B}{\cos\theta_B} = \tan\theta_B = \frac{n_2}{n_1}\,. \tag{13.57}$$

Left as a homework assignment is the proof that at $\theta = \theta_B$, the reflected and transmitted wavevectors are at right angles to each other. This perpendicular orientation is no accident of mathematics: it is the atomic-scale explanation of why reflectance vanishes under this condition. When in Ch. 15 we treat reflectance as radiation arising from dipoles generated by the incoming wave, then reflectance vanishes because we are viewing the dipoles end-on. This is the Ewald-Oseen atomic-scale picture of the optical properties of materials.

We next consider the interface plasmon (surface plasmon polariton). This is an elementary excitation of the configuration. As usual, elementary excitations are identified by setting denominators equal to zero, which we recall is equivalent to expecting an output for no input. Setting the denominator of either of Eqs. (13.52) equal to zero requires

$$\varepsilon_2 n_{1z} + \varepsilon_1 n_{2z} = 0 . \tag{13.58}$$

Following the same procedure used to obtain Brewster's Angle, we find

$$\varepsilon_2^2(\varepsilon_1 - \varepsilon_1 \sin^2\theta) = \varepsilon_1^2(\varepsilon_2 - \varepsilon_1 \sin^2\theta) . \tag{13.59}$$

But this is exactly Eq. (13.54). Consequently, it must also have the same solution, or

$$\sin^2\theta = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2}; \quad \cos^2\theta = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2} . \tag{13.49a,b}$$

This is odd - did we simply rederive Brewster's Angle? Yes – but look closely at Eq. (13.52a). If $\varepsilon_2$ is positive, then the numerator can vanish (the Brewster-Angle condition), but the denominator cannot. Thus there can be no elementary excitation in this case. If $\varepsilon_2$ is negative, then the denominator can vanish. This is the plasmon condition. Hence $\varepsilon_2$ has the opposite sign in the two cases: material #2 exhibits dielectric behavior for Brewster's Angle and metallic behavior for plasmons. The sign matters, but it disappeared in the squaring operation.

However, there is more to the picture than just setting $\varepsilon_2 < 0$. While in this case the numerator of Eq. (13.49a) is negative, if $0 \geq \varepsilon_2 \geq -\varepsilon_1$ then the denominator is positive, so $\sin^2\theta$ is negative. Recalling that $\sin\theta$ is shorthand for $k_{1x}/k_1$ and that $k_1$ is real, this says that $k_{1x}$ is pure imaginary. Because the wave is proportional to $e^{ik_{1x}x + ik_{1z}z - i\omega t}$, the wave must therefore either increase or decrease exponentially with increasing *x*, which means that it cannot be normalized. Hence this is a second range within this range a steady-state excitation cannot exist.

Accordingly, the requirement $\varepsilon_2 < 0$ must be replaced by the stronger requirement

$$\varepsilon_s < -\varepsilon_a , \tag{13.60}$$

in which case both numerator and denominator are negative, and $\sin^2\theta = (k_{1x}/k_1)^2$ is once again positive. In this case it is $\cos\theta = k_{1z}/k_1$ that is pure imaginary. Hence the excitation decays on both sides of the interface. Consequently, Eq. (13.53) becomes the defining relation for the existence of the interface plasmon, as we will see more clearly below.

Because we don't normally encounter situations where $\sin\theta$ is greater than 1, these equations may be more acceptable if we write the above expressions in terms of the components of the wave vector of the incident wave in the transparent medium. In this case Eq. (13.49a) is

$$\frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2} = \sin^2\theta = \frac{k_{1x}^2}{k_1^2} = \frac{k_x^2}{k_1^2} , \tag{13.61}$$

since $k_x$ is common to all layers. Put this way, the existence of the interface plasmon requires the component of the wave vector parallel to the surface on the ambient side to

exceed the value that can propagate in the ambient medium. We describe two ways to accomplish this below. Because $c^2 k_1^2 / \omega^2 = \varepsilon_1$, the above can also be written as

$$\frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 + \varepsilon_2} = \frac{c^2 k_x^2}{\omega^2}. \tag{13.62}$$

We return to this expression below.

The fields associated with the plasmon are easily determined. Start with

$$\vec{H}_1 = \hat{y} H_o e^{ik_x x} e^{-k_{1z} z}, \; z \geq 0, \tag{13.63a}$$

$$\vec{H}_2 = \hat{y} H_o e^{ik_x x} e^{k_{2z} z}, \; z \leq 0, \tag{13.63b}$$

where

$$k_{1z}^2 = \frac{\omega^2}{c^2} \varepsilon_1 - k_x^2 < 0; \tag{13.64a}$$

$$k_{2z}^2 = \frac{\omega^2}{c^2} \varepsilon_2 - k_x^2. \tag{13.64b}$$

Since $\varepsilon_2$ is already negative, the normal mode is seen to decay exponentially in both directions from the interface. The sign is selected to ensure attenuation with increasing distance.

We next check the result for consistency. The boundary condition on tangential $\vec{H}$ is already built into Eqs. (13.63). The boundary condition for tangential $\vec{E}$ follows from the Faraday-Maxwell Equation. We find that

$$\vec{E}_1 = \frac{c}{\omega \varepsilon_1} H_o \left( -\hat{z} k_x + i \hat{x} k_{1z} \right) e^{ik_x x} e^{-k_{1z} z}, \tag{13.65a}$$

$$\vec{E}_2 = \frac{c}{\omega \varepsilon_2} H_o \left( -\hat{z} k_x - i \hat{x} k_{2z} \right) e^{ik_x x} e^{k_{2z} z}. \tag{13.65b}$$

As a cross-check, Eqs. (13.65) show that the normal components of $\vec{D}$ are continuous across the interface. The continuity condition on tangential $\vec{E}$ is therefore

$$\frac{k_{1z}}{\varepsilon_1} H_o = -\frac{k_{2z}}{\varepsilon_2} H_o = \frac{H_o}{\varepsilon_1} \sqrt{\frac{\omega^2}{c^2} \varepsilon_1 - k_x^2} = -\frac{H_o}{\varepsilon_2} \sqrt{\frac{\omega^2}{c^2} \varepsilon_2 - k_x^2}\, s. \tag{13.66}$$

As a second cross-check, it is seen that Eq. (13.66) is precisely the condition for which the denominator of $r_{sa}^{TM}$ vanishes, closing the loop.

The instantaneous Poynting vector on the ambient ( $j = 1$ ) side follows as

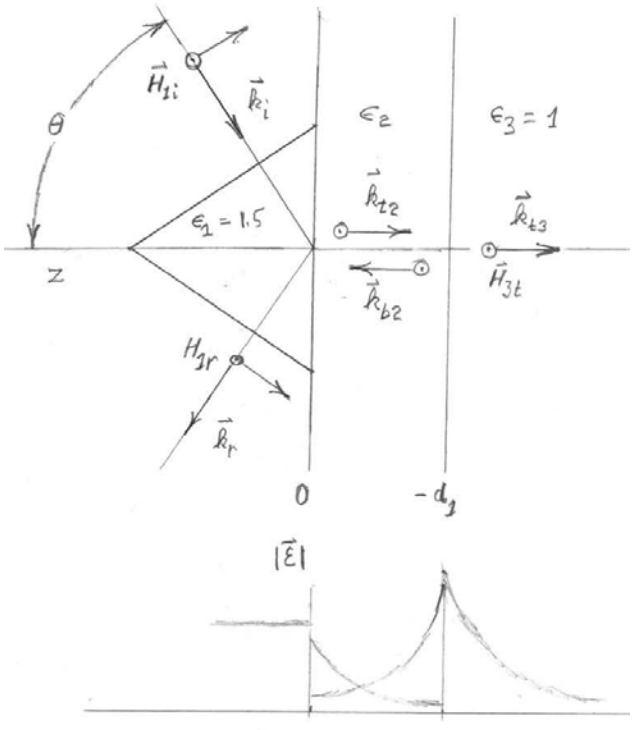
$$\vec{S}_1(x,z,t) = \frac{c^2 H_o^2}{4\pi\omega\varepsilon_1} e^{-2k_{1z} z} \left( \hat{x} \cos^2(k_x x - \omega t) + \hat{z} \sin(k_x x - \omega t) \cos(k_x x - \omega t) \right), \tag{13.67}$$

with a similar expression for the substrate. Hence energy can flow parallel to the interface, but the net energy flow perpendicular to the interface is zero, as expected.

H. Resonance (plasmon) singularities in reflectance.

Nominally slowly varying spectra of reflectance vs. energy can show "anomalous" decreases or dips at certain angles of incidence. Examples include the Wood's anomalies that are seen in gratings, and resonances that are observed in the internal-reflection spectra of thin metal films deposited on prisms with refractive indices significantly larger than 1. Both are generally attributed to plasmons. In the former case this is accurate, but in the latter case this is a boundary-condition effect at the metal-air interface rather than a stand-alone excitation. However, both provide excellent examples of representing relatively complex phenomena by simple limits to more elaborate equations, and extracting physics out of math.

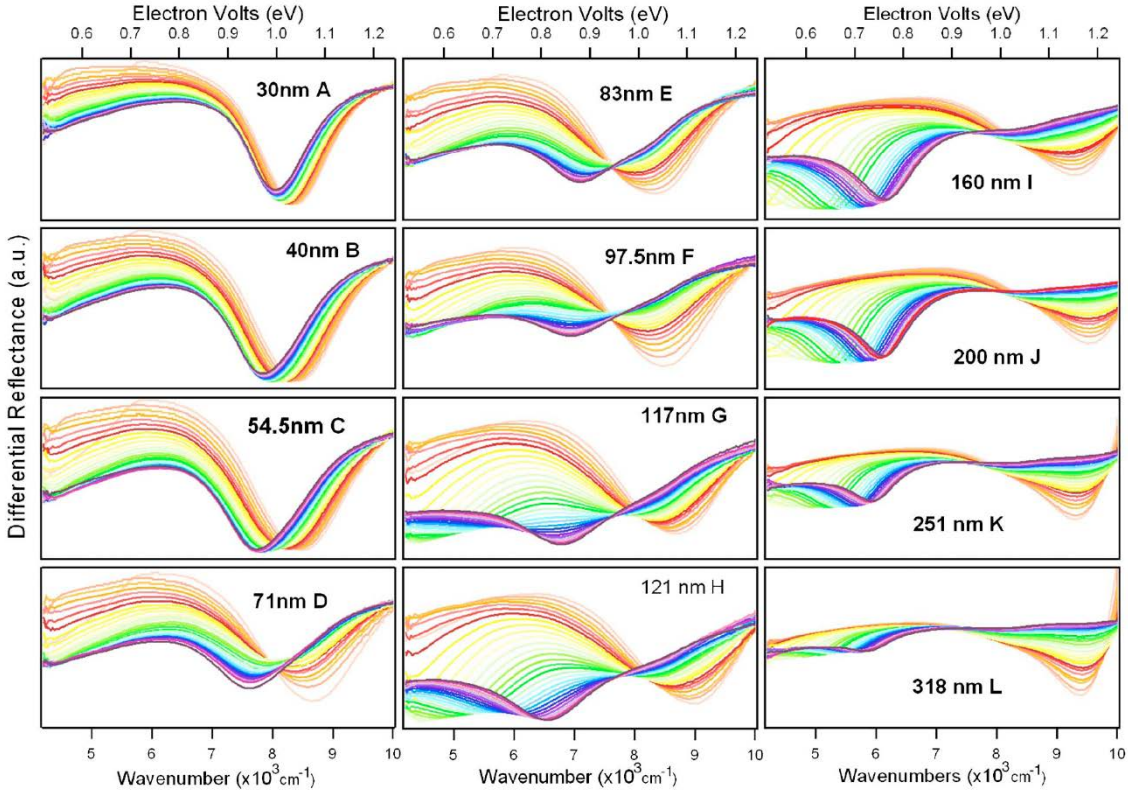
We consider the thin-metal-film case first, analyzing reflectance data obtained from the configuration shown at the top of the figure on the right. The data themselves are shown at the top of the next page. This is the Kretschmann configuration, commonly used to study interface plasmons. It consists of a thin metal film of dielectric function $\varepsilon_2$ and thickness $d_2$ deposited on a transparent prism of dielectric function $\varepsilon_1$. The entire assembly is located in air. A TM-polarized beam of photon energy $E = \hbar\omega$ is incident on the prism-metal interface at $z = 0$ at an angle of incidence $\theta$. The beam divides into a transmitted and reflected part. The presence of the metal-air interface at $z = -d_2$ generates a back-reflected beam that interacts with the other three beams at the prism-metal interface. The boundary condition is that only a transmitted beam is present for $z < -d_2$. The sketch at the bottom indicates magnitudes of fields associated with the configuration. With $\vec{k}$ purely imaginary in media 2 and 3, the amplitudes of the fields vary exponentially as shown.




In the configuration used to obtain the data on the next page, the prism is BK-7 glass, with $\varepsilon_1 = 2.3$. The "metal" film is a heavily doped semiconductor with an infrared dielectric function of

$$\varepsilon_2 = \varepsilon_2(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{(\omega + i\Gamma)^2}, \tag{13.68}$$

where $\varepsilon_\infty = 3.8$, $\hbar\omega_p = 2.19$ eV, and $\hbar\Gamma = 0.111$ eV. The term $\varepsilon_\infty$ in Eq. (13.61) describes the long-wavelength contribution of optical absorption at higher energies to $\varepsilon_1$. The dielectric function of air is of course $\varepsilon_3 = 1$.




Data illustrating various types of interface resonances. Details are given in the text. The figure is from C. Rhodes et al., J. Appl. Phys. 103, 093108 (2008).

The reflectance data $R_{123} = |r_{123}|^2$ shown in the figure were obtained from 0.5 to 1.25 eV. The figure is divided into panels corresponding to $R_{123}$ for samples of increasing ITO film thicknesses, ranging from 30 nm in panel A to 319 nm in panel L. Spectra from 0.5 to 1.25 eV are shown for angles of incidence $\theta$ in steps of 1 degree from 42° (orange) to 53° (violet). Rather than exhibiting a slowly varying change of $R_{123}$ with energy, as might be expected, the spectra show prominent dips at different thicknesses, slightly above 1 eV for the thinnest films, independent of thickness and angle of incidence, and a series of dips from about 0.6 to 0.75 eV most prominently for the 121 and 160 nm films, with energies increasing with angle of incidence. The first resonance dies out as the thickness increases, to be replaced by the second set. Because the reflected intensity and therefore reflected energy is reduced, these resonances evidently involve the activation of

channels that absorb energy. In some way they also involve interfaces, because for sufficiently thick films the features die out completely.

The reflectance data in the above figure are described accurately by the Airy Equation. However, our goal is to understand the physical origin of these features, which is difficult to do with the full expression. Accordingly, we consider simplifications. The thin-film resonance may be the easiest, since to lowest order it appears to be independent of both film thickness and angle of incidence. This suggests that the loss mechanism may be due to a material property, although one that is somehow associated with interfaces. Taking "thin" as the cue, we expand the Airy Equation to first order in $(d/\lambda)$. The calculation is straightforward, and the result is

$$r_{123} = (r_{123})_o \left( 1 + \frac{4\pi i d n_1 \cos\theta}{\lambda} \frac{\varepsilon_3 - \varepsilon_2}{\varepsilon_3 - \varepsilon_1} \frac{1 - \left( \frac{1}{\varepsilon_2} + \frac{1}{\varepsilon_3} \right) \varepsilon_1 \sin^2\theta}{1 - \left( \frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_3} \right) \varepsilon_1 \sin^2\theta} \right), \tag{13.69}$$

an equation first derived Abelès in 1976.

In Eq. (13.69) both thickness and angle of incidence enter as scaling factors, consistent with observations. Hence we look for other possibilities, particularly vanishing denominators. Equation (13.68) shows that $\varepsilon_2$ essentially vanishes when $\omega^2 = \omega_p^2 / \varepsilon_\infty$. Substituting values given above, the resonance should occur at 1.12 eV. This is close enough; we have therefore identified the thin-film resonance as the bulk plasmon. The interfaces form a capacitor, allowing energy to be stored alternately in electric and magnetic form, first as an electric field resulting from charge accumulating at the interfaces, and second as a magnetic field from the current that results when the boundary charge is reversed. The bulk plasmon essentially amplifies the intrinsic loss of the metal layer, resulting in less energy in the reflected beam and therefore a decreased reflectance.

We now consider the resonances at higher energy. We can expect these to be related in some way to another vanishing denominator. This must involve the metal-air interface only, because it is not present until the interfaces are clearly separated and it disappears when the film becomes thick. Rather than examine the Airy Equation, we consider reflection at the metal-air boundary, noting the "substrate" condition that no back-reflected wave is present. The forward- and backward-propagating waves in the metal film adjacent to this interface are related to the transmitted wave by

$$\begin{pmatrix} E_{2i} \\ E_{2r} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left( 1 + \frac{n_{3z}/\varepsilon_3}{n_{2z}/\varepsilon_2} \right) & \frac{1}{2}\left( 1 - \frac{n_{3z}/\varepsilon_3}{n_{2z}/\varepsilon_2} \right) \\ \frac{1}{2}\left( 1 - \frac{n_{3z}/\varepsilon_3}{n_{2z}/\varepsilon_2} \right) & \frac{1}{2}\left( 1 + \frac{n_{3z}/\varepsilon_3}{n_{2z}/\varepsilon_2} \right) \end{pmatrix} \begin{pmatrix} E_{3t} \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{\varepsilon_3 n_{2z} + \varepsilon_2 n_{3z}}{\varepsilon_3 n_{2z}} \\ \dfrac{\varepsilon_3 n_{2z} - \varepsilon_2 n_{3z}}{\varepsilon_3 n_{2z}} \end{pmatrix} E_{3t} \,. \tag{13.70}$$

Thus

$$\frac{E_{2r}}{E_{2i}} = \frac{\varepsilon_3 n_{2z} - \varepsilon_2 n_{3z}}{\varepsilon_3 n_{2z} + \varepsilon_2 n_{3z}} \,. \tag{13.71}$$

The denominator is the same that defines the interface plasmon. Hence the boundary condition of no back-reflected beam arriving at the interface forces a potentially large increase in the magnitude of the beam back-reflected into the film. Because the decay is exponential in both directions, we identify this as the interface plasmon.

To confirm this interpretation quantitatively, we re-examine the equations leading to the plasmon. In this configuration the dielectric function of the prism is different from that of the dielectric (air) at the metal-air interface, so Eqs. (13.49) need to be re-examined as well. The equations appropriate to the present configuration are

$$\varepsilon_3^2(\varepsilon_2 - \varepsilon_1 \sin^2\theta) = \varepsilon_2^2(\varepsilon_3 - \varepsilon_1 \sin^2\theta) \,. \tag{13.72}$$

Following the same development we obtain

$$\frac{\varepsilon_2 \varepsilon_3}{\varepsilon_2 + \varepsilon_3} = \varepsilon_a \sin^2\theta \,, \tag{13.73}$$

where $\theta$ is the angle of incidence at the prism-metal interface, $\varepsilon_2(\omega)$ is given by Eq. (13.61), and $\varepsilon_3 = 1$.

The relevant variables $\omega$ and $\theta$ are connected via Eq. (13.61). Making the substitution we find

$$\varepsilon_1 \sin^2\theta = 2.3 \sin^2\theta = \frac{\varepsilon_\infty - \omega_p^2 / \omega^2}{1 + \varepsilon_\infty - \omega_p^2 / \omega^2} \,. \tag{13.74}$$

The broadening term $\Gamma$ is discarded. Although it is important for energy loss calculations, it isrelatively unimportant for the plasma energy itself. Solving for $(\hbar\omega)$ we find

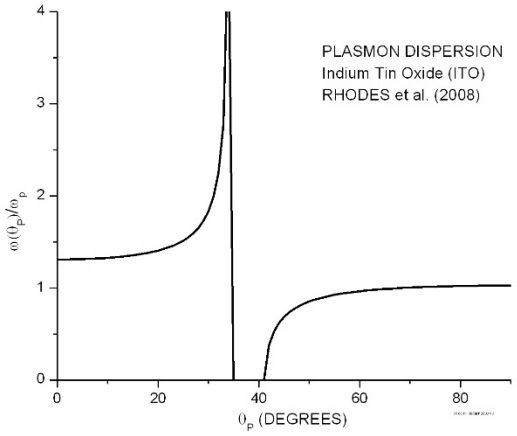
$$(\hbar\omega)^2 = (\hbar\omega_p)^2 \frac{2.3\sin^2\theta - 1}{1 + \varepsilon_\infty (2.3\sin^2\theta - 1)} \tag{13.75a}$$

$$= (2.19\,\mathrm{eV})^2 \frac{2.3\sin^2\theta - 1}{1 + 3.8(2.3\sin^2\theta - 1)} \,. \tag{13.75b}$$

The key point here is that the use of the prism has not only made confinement at the metal-air interface possible, it provides extra leverage for the $\sin^2\theta$ term, which appears twice in Eq. (13.75b).

The solution for $\hbar\omega$ is shown in the next figure. It has three ranges, which for the sample under investigation are defined by the denominator vanishing at $\theta = 34.5^\circ$ and the numerator vanishing at $\theta = 41.3^\circ$. For $\theta < 34.5^\circ$ we have already noted that the denominator of Eq. () cannot vanish, so in this range the interface plasmon does not exist. For $34.5^\circ < \theta < 41.3^\circ$, $\omega$ is purely imaginary, so the solution either grows or decreases exponentially in distance along the interface, in which case a steady-state solution also cannot exist. For $\theta > 41.3^\circ$ the denominator vanishes, confinement occurs in both directions away from the interface, so the interface plasmon exists.

Having established the energy dependence, we return to the data for the thicknesses 121 and 160 nm. The minimum shows a significant and monatomic increase in energy from 42 to 53 degrees. Referring to the figure, this is exactly what one should expect: there is a rise in the calculated value from about 0.6 to about 0.8 eV corresponding to this angular range in the figure. Note that the part giving this rise is the numerator, not the denominator.



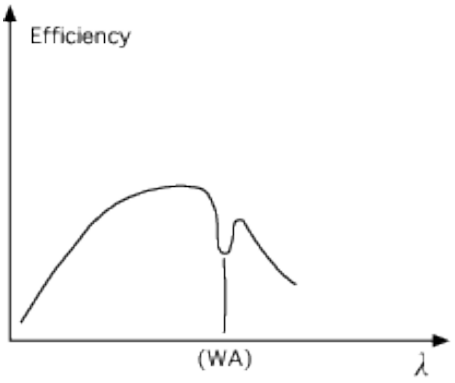

Thus all features in the data are identified. While evaluation of the full Airy Equation is necessary to prove that the theory describes the experiment, the physics is more clearly identified by reducing the full expression to simpler relations. Although we have discarded the loss term $\Gamma$, it should be noted that it is essential for describing the data. Without it, there is no loss mechanism in the calculation, and the dips would not appear.

Further insight about the physics of the interface plasmon is achieved by considering phase velocity $v_p = \omega/k$. The question as to what to use for *k* is answered by noting that the interface plasmon propagates along the interface. Hence the projection $k_x$ is appropriate. Given that $k_x = k_P \sin\theta_P$ and that $ck_P/\omega = n_P = \sqrt{\varepsilon_P}$, we find

$$v_p = \frac{\omega}{k_p \sin\theta} = \frac{c}{n_P \sin\theta} \tag{13.76}$$

independent of range. But this is exactly what we would expect from a system being driven by a wave arriving at an angle of incidence $\theta$ and observed along the interface. The result is independent of all details except $n_1$ and the angle of incidence of the driving field. It becomes infinite as normal incidence is approached.

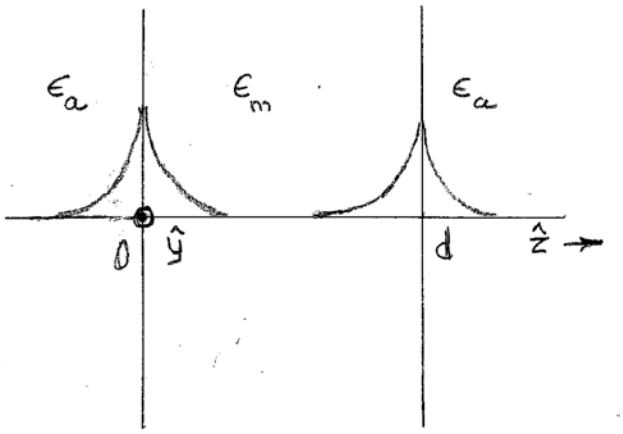
Although the Kretschmann configuration dominates current studies, interface plasmons were first observed as "Wood's anomalies" in the intensity of light diffracted from gratings. These are reductions in diffracted intensities that occur at wavelengths where a diffraction order emerges parallel to the grating, as indicated schematically in the diagram. In this situation the wave vector of the parallel order matches a component of the spatial Fourier transform of the grooves. The Fourier component is an integral multiple of the fundamental, and consequently can excite a surface plasmon and therefore open a new channel for dissipating energy.




The calculation is best cast in dipole-radiation form, which is covered in Ch. 15. These anomalies were first observed by Wood, but were not explained until much later. Because they are associated with the relief structure of the grating, these plasmons involve screening charge and hence are more nearly analogous to the structural plasmons that are discussed in the next section. The energies of the associated plasmons are given by the energies where the minima occur.

I. Coupled interface plasmons.

Additional physics of historical importance due to its misinterpretation as sub-wavelength imaging follows from the coupling of interface plasmons on either side of a free-standing metal film, as indicated in the diagram. The configuration combines in a single calculation several concepts with which we are already familiar. The asymmetric case of a thin film deposited on a substrate was originally solved by Raether. Although the configuration is more practical, it is also more difficult mathematically. Because the same points can be made working with a free-standing film, the asymmetric case is not worth the extra work. Some of the mathematical details will be left as a homework assignment.




Let a free-standing metal film of thickness $d_2$ and dielectric function $\varepsilon_2$ be immersed in a medium of dielectric function $\varepsilon_a$. We postulate the existence of interface plasmons on either side of the film, and seek to find the solution of the homogeneous propagation equations that satisfy the boundary conditions on continuity of tangential $\vec{H}$ and $\vec{E}$ across the two interfaces. As with the two-phase model, this can be achieved by setting the denominator of the Airy equation equal to zero. However, we use field-component approach below. Because the plasmon is a TM excitation, we start by making our usual

table, although it looks a bit different because the role of the incident wave is taken by the field in the region $z < 0$. We have

| Range: | Magnetic field: | |
|---|---|---|
| $z \le 0$: | $\vec{H}(z,t) = H_a\, \hat{y}\, e^{ik_x x + k_z z}$ | (13.77a) |
| $0 \le z \le d$: | $\vec{H}(z,t) = H_+\, \hat{y}\, e^{ik_x x + k_z z} + H_-\, \hat{y}\, e^{ik_x x - k_z z}$ | (13.77b) |
| $z \ge d$: | $\vec{H}(z,t) = H_b\, \hat{y}\, e^{ik_x x - k_z z}$, | (13.77c) |

where the time dependence $e^{-i\omega t}$ is suppressed. We get the tangential component of the electric field from the above using Ampère's Law:

$$i\vec{k} \times \vec{H} = \frac{-i\omega\varepsilon}{c}\vec{E}\,. \tag{13.78}$$

The result is

| Range: | Electric field (tangential component only): | |
|---|---|---|
| $z \le 0$: | $\frac{\omega}{ic}\vec{E}(z,t) = -\hat{x}\,\frac{k_{az}}{\varepsilon_a} H_a\, e^{k_{az} z}$; | (13.69a) |
| $0 \le z \le d$: | $\frac{\omega}{ic}\vec{E}(z,t) = -\hat{x}\,\frac{k_{mz}}{\varepsilon_m}\left(H_+\, e^{k_{mz} z} - H_-\, e^{-k_{mz} z}\right)$; | (13.79b) |
| $z \ge d$: | $\frac{\omega}{ic}\vec{E}(z,t) = \hat{x}\,\frac{k_{az}}{\varepsilon_a} H_b\, e^{k_{az} z}$; | (13.79c) |

where the common phase factor $e^{ik_x x - i\omega t}$ is ignored.

Evoking the boundary conditions at $z = 0$ we obtain the equations

$$H_a = H_+ + H_-; \tag{13.80a}$$

$$H_a = \eta H_+ - \eta H_-; \tag{13.80b}$$

where to save writing we have defined

$$\eta = \frac{\varepsilon_a k_{mz}}{\varepsilon_s k_{az}}\,. \tag{13.80c}$$

From the existence condition on the plasmon, that is, $\varepsilon_m < -\varepsilon_a$, we can guess that $\eta$ is negative and has a value reasonably close to $(-1)$. If we were working in the two-phase model $\eta$ would be exactly $(-1)$, because the defining condition for the plasmon is

$$\varepsilon_a k_{mz} + \varepsilon_m k_{az} = 0\,. \tag{13.81}$$

We can rewrite the above as

$$H_+ = \frac{\eta+1}{2\eta} H_a; \quad H_- = \frac{\eta-1}{2\eta} H_a. \tag{13.82}$$

This is consistent with expectations, because the "forward" wave $H_-$ should have an amplitude that is significantly larger than that of the "backward" wave $H_+$. This is the case for $\eta \cong -1$. The above can also be written

$$H_- = \frac{\eta-1}{\eta+1} H_+. \tag{13.83}$$

Considering next the boundary conditions at $z = d$, we obtain

$$H_+ e^{k_{sz}d} + H_- e^{-k_{sz}d} = H_b e^{-k_{az}d} = H_b'; \tag{13.84a}$$

$$\eta H_+ e^{k_{sz}d} - \eta H_- e^{-k_{sz}d} = -H_b e^{-k_{az}d} = -H_b'; \tag{13.84b}$$

where

$$H_b' = H_b e^{-k_{az}d}. \tag{13.84c}$$

These can be rewritten as

$$H_+ e^{k_{sz}d} = \frac{\eta-1}{2\eta} H_b' \tag{13.85a}$$

$$H_- e^{-k_{sz}d} = \frac{\eta+1}{2\eta} H_b'; \tag{13.85b}$$

This in turn gives

$$H_+ = H_- \frac{\eta-1}{\eta+1} e^{-2k_{mz}d}. \tag{13.86}$$

Recalling that we're looking for a solution of the homogeneous equations, we can now substitute an above expression for $H_-$, obtaining

$$H_+ = H_- \frac{\eta-1}{\eta+1} e^{-2k_{ms}d} \tag{13.87a}$$

$$= \left(\frac{\eta-1}{\eta+1}\right)^2 e^{-2k_{ms}d} H_+. \tag{13.87b}$$

This will have nontrivial solutions if

$$\left(\frac{\eta-1}{\eta+1}\right)^2 e^{-2k_{ms}d} = 1. \tag{13.88}$$

Accordingly, we find that

$$\frac{\eta+1}{\eta-1} = \pm e^{-k_{ms}d}. \tag{13.89}$$

Although we won't use the following relations, choosing the upper sign gives

$$\eta_+ = -\tanh\left(\frac{k_{sz}d}{2}\right), \tag{13.90a}$$

whereas the lower sign gives

$$\eta_- = -\coth\left(\frac{k_{sz}d}{2}\right). \tag{13.90b}$$

More useful are the fields themselves in terms of $H_a$. Doing the necessary algebra leads to

$$H_+ = \frac{\pm e^{-k_{sz}d}}{1 \pm e^{-k_{sz}d}} H_a \tag{13.91a}$$

$$H_- = \frac{1}{1 \pm e^{-k_{sz}d}} H_a \tag{13.91b}$$

$$H_b' = H_b e^{-k_{az}d} = \mp H_a. \tag{13.91c}$$

Thus the meaning of the two roots becomes clear: the upper sign corresponds to the antisymmetric combination, while the lower sign gives the symmetric combination. The interior fields reflect this by their spatial dependences. To proceed further it would be necessary to know the boundary conditions on the back side of the film. The signal on the back side can range from full amplitude to full amplitude inverted, with zero as an intermediate option.

In the limit that $d$ becomes very large, the solutions become independent, with $H_- = H_a$, and on the other side, $H_+ = H_b'$, in the latter case placing the origin at $z = d$. Because the purpose of solutions of homogeneous equations are to satisfy initial conditions, the division between symmetric and antisymmetric modes can only be established for specific cases.

Interface plasmons are an interesting topic historically. They were studied extensively in the 1960's and 1970's, although much of this early work was forgotten. Interest has revived for several reasons: first, plasmons involve light but are extremely localized, thereby offering the possibility of logic circuits operating at the speed of light in sizes much smaller than light wavelengths. Second, as resonances, these excitations amplify fields, improving observation capabilities particularly for nonlinear phenomena such as second-harmonic generation.

The start of the current interest in interface plasmons can be traced to a publication (Pendry PRL 2000) where plasmons coupled across a thin film were mistakenly interpreted as the abovementioned subwavelength imaging. The plasmon interpretation is straightforward: because the excitations on both sides are coherent, a plasmonic superposition forced on one side by e.g. structure in a metal overlayer appears on the

other side as well. The coupled-plasmon problem had already been solved a few years earlier and reported in an appendix of a comprehensive 136-page monograph "Surface Plasmons on Smooth and Rough Surfaces and on Gratings" by H. Raether, published by Springer-Verlag, Berlin, in 1988, two years after his death. This monograph covers essentially all aspects of plasmons.

J. Structural plasmons.

We encountered structural plasmons before, when we discussed the case of a dielectric sphere in a dielectric medium. Maxwell Garnett's 1904 solution of this configuration, intended to provide insight into the deep red color of cathedral windows containing dispersions of Au nanoparticles, is significant because it showed that optical properties of materials could be affected by their nanostructure (using current terminology). This meant that optical measurements could be used to obtain information about nanostructure. By 1935 the development of the "classical" part of the field was completed with the introduction of the Bruggeman effective- medium expression, although some advances continued to be made. The most recent advance, by Moharam et al., provides an exact connection between nanostructure and the properties of a reflected beam through a combination of Fourier analysis and scattering theory. This so-called rigorous coupled-wave analysis (RCWA) is widely used in integrated-circuits technology to maintain control of etching processes by the analysis of polarimetric (= ellipsometric + reflectometric) spectra obtained in-line during processing.

The Maxwell Garnett expression discussed in Ch. 7 can be generalized to

$$\frac{\varepsilon - \varepsilon_h}{\varepsilon + p\varepsilon_h} = f_a \frac{\varepsilon_a - \varepsilon_h}{\varepsilon_a + p\varepsilon_h} + f_b \frac{\varepsilon_b - \varepsilon_h}{\varepsilon_b + p\varepsilon_h} + \ldots, \tag{13.92}$$

where the depolarization factor $p$ ranges from zero to infinity. Special values are:

$p = 0$: laminar configuration, field perpendicular to layers, maximum screening;

$p = 1$: cylindrical (2D) geometry;

$p = 2$: spherical (3D) geometry;

$p \to \infty$: laminar configuration, field parallel to layers, no screening.

The remaining parameters are:

$\varepsilon =$ dielectric function of the composite material;

$\varepsilon_h =$ "host" dielectric function supporting the inclusions;

$\varepsilon_a, \varepsilon_b =$ dielectric functions of inclusions (more than two may be present);

$f_a, f_b =$ volume fractions of inclusions.

A main point here is that singularities (plasmons) occur whenever $\varepsilon_a = p\varepsilon_h$ etc. By choosing the proper materials for the inclusions and fabricating them into particular shapes, one can engineer a wide range of properties that do not exist in materials found in nature.

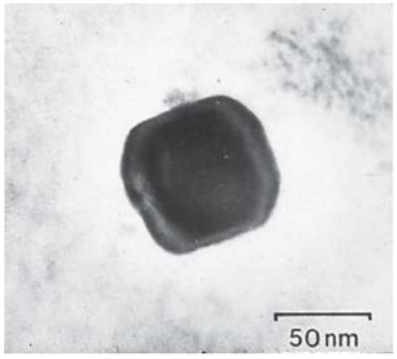
External illumination  Internal illumination

The Lycurgus Cup (right, Roman, ca. 400 a.d., now in the British Museum) is a good example. The glass contains inclusions of an Au-Ag alloy (see TEM micrograph on next page), so it exhibits structural-plasma resonances that affect its dielectric response. The resulting plasmonic effects are essentially the reverse of those seen in Au metal. The composite reflects well at short wavelengths and transmits well at long wavelengths. Hence it appears green under external illumination and red when illuminated internally.

K. Negative-index materials.

A related topic of current interest involves the so-called "negative-index" materials. In the ideal case, these are materials where $\mu = \varepsilon = -1$, so that the refractive index is $\tilde{n} = +\sqrt{\mu\varepsilon} = n + i\kappa = -1$. With $|n| = 1$, this hypothetical material is perfectly matched to an air ambient, and its reflectance is zero. Yet the normal component of its wave vector is reversed compared to that in "ordinary" material. This has given rise to a relatively large number of papers in the literature nominally explaining the behavior of negative-index materials, although in most cases these explanations consist of making assumptions, instead of going back to basics and building from there. We can do better.

**Figure 4**
*Transmission electron microscopy (TEM) image of a silver-gold alloy particle within the glass of the Lycurgus Cup [21]. © The Trustees of the British Museum.*




The history of this hypothetical material is itself interesting. Negative-index materials were first proposed by Veselago in a Russian journal in 1964. His work was completely ignored at the time: the article was not translated into English until 1968, at which point it was essentially forgotten. However, interest in these materials exploded in 2000 with Pendry's claim of subwavelength imaging. The hope was that negative-index materials would offer another route to this goal. They didn't, but the Veselago lens remains interesting and educational in its own right. Because fabrication techniques have now advanced to the stage where negative-index materials can be approximated, at least in the microwave region, the topic is no longer only of academic interest. The lesson to be

learned from this section is not necessarily how such materials behave, but how issues like this should be addressed.

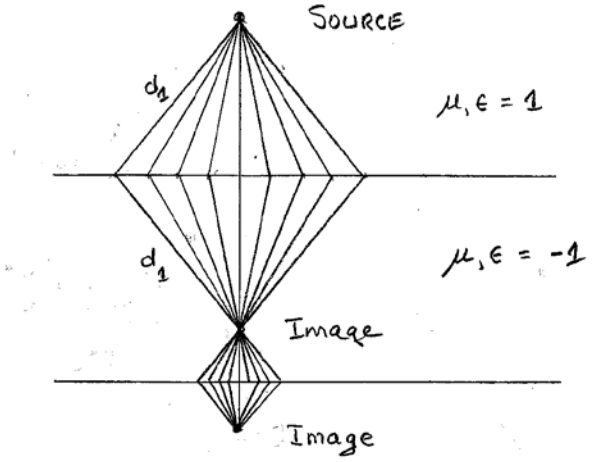
A second curious point concerns the evolution of the phase of the plane wave. With

$$k_s = -\frac{\omega}{c}, \tag{13.93}$$

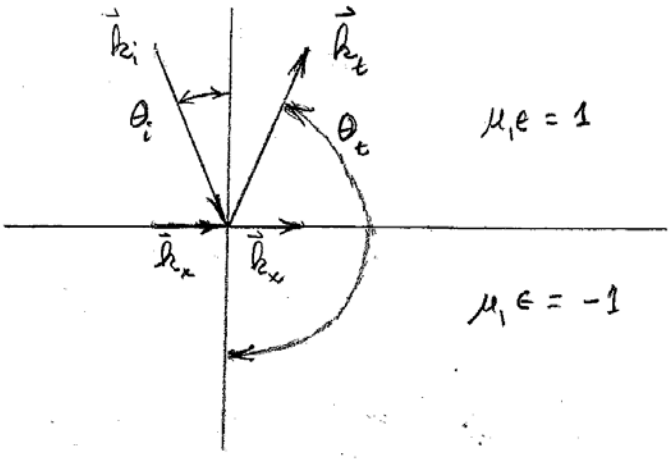
the plane wave is

$$\vec{E}_s(z,t) = \hat{y}E_i e^{+i|k_s|z - i\omega t}. \tag{13.94}$$

That is, propagation over a distance *d* in the negative-index substrate erases the phase change accumulated in traversing the same distance *d* in the ambient. This, plus Huygens' Principle, leads to a quick-and-dirty derivation of the Veselago lens: if the phase accumulated in any path in the ambient can be cancelled by traveling the same distance in the substrate ( $d_1$ for the leftmost ray in the diagram), then a point source a distance *d* above the interface will be mirror-imaged exactly at a distance *d* below the interface. As indicated in the diagram, this follows because the distance traveled by any ray leaving a point above the surface is the same as that needed to reach the mirror-image point below the surface. We have then realized a negative version of Snell's Law.




If the negative-index material is a slab that is thicker than the distance between the point and the surface of the slab, then an identical construction from the interior focus shows that the interior image is re-imaged to an exterior point the same distance from the second surface of the slab. However, this process is not imaging in the Abbé sense, because the rays do not arrive at the same time. Thus the Veselago lens is not a true lens, although under steady-state conditions it appears to function like one. Also, the magnification is restricted to 1. Modern metamaterial sheets with engineered structures can do significantly better – you'll



have them in your cell-phone camera soon.

The negative-index equivalent of Snell's Law is shown in the second diagram. With the *z* component of $\vec{k}_s$ reversed but the *x* component remaining the same, Snell's Law for the ideal negative-index material becomes

$$\sin\theta_i = \sin(\theta_t) = \sin(\pi - \theta_i) = \sin(\theta_i)\,, \tag{13.95}$$

so mathematically there is no apparent change. However, the wave vector in the substrate is identical to the wave vector of the reflected wave, if there actually were a reflected wave. Drawings in the literature tend reverse the sign of $\vec{k}_s$ to make it look more like the standard Snell's-Law diagram, but the physics then is incorrect.