Chapter 13: Plane waves (2): Physics at planar interfaces (06 Mar 2022).

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 λ 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.

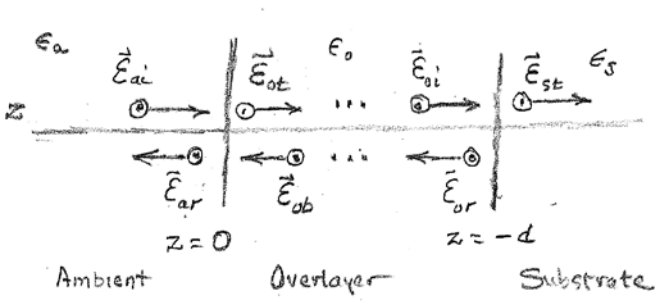
Chapter 12 is built on Fourier analysis. Chapter 13 is built on matrix algebra. To simplify matters, we assume throughout that all dielectric functions are scalars, in which case the solutions of the wave equation are transverse plane waves with wave vectors satisfying standard dispersion relations. In this case the coefficients describing the incident and reflected waves on one side of an interface can be written as a (2×1) field vector that is connected to a similar (2×1) field vector describing the transmitted and back-reflected waves on the other side of the interface through a 2×2 reflection matrix.

Another 2×2 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 basics, then get the math right, explain why we do what we do, then extract the physics.

B. Configurations.

It's worth laying out the configuration first, because an overview of this provides a framework for discussion and hence a better idea of what is being done. For simplicity the overview considers only normal incidence. This allows definitions to be established with minimal complications.



The configuration shown in the figure is the so-called 3-phase (ambient-overlayer-substrate) model, which is the minimum configuration that includes everything that needs to be defined. The ambient on the left is transparent with a dielectric function ε_a, the an overlayer of thickness *d* and dielectric function ε_o is in the middle, and a substrate with dielectric function ε_s is on the right. By assumption all dielectric functions treated here are scalar, and all permeabilities are equal to 1. The positive *z* direction is to the left because we in calculations we represent this configuration as a matrix product involving reflection/transmission and propagation matrices, placing the boundary condition on the right. This ordering conforms to that of standard matrix multiplication.

For a given polarization, each region has two transverse modes. Although $\vec{E}$ can be oriented in any direction perpendicular to $\hat{z}$, we select the $\hat{y}$ direction because this is consistent with the non-normal-incidence reflection convention discussed below. The normal modes of the ambient are then the incident and reflected fields $\vec{E}_{ai}(z,t) = E_{ai}\hat{y}e^{-ik_a z - i\omega t}$ and $\vec{E}_{ar}(z,t) = E_{ar}\hat{y}e^{ik_a z - i\omega t}$, respectively. The normal modes of the overlayer at the ambient/overlayer interface are termed transmitted and back-reflected, $\vec{E}_{ot}(z,t) = E_{ot}\hat{y}e^{-ik_o z - i\omega t}$ and $\vec{E}_{ob}(z,t) = E_{ob}\hat{y}e^{ik_o z - i\omega t}$, respectively. Note that after propagation these become the incident and reflected modes at the second interface. The substrate contains only one normal mode $\vec{E}_{st}(z,t) = E_{st}\hat{y}e^{ik_x z - i\omega t}$, since by definition there is no back-reflected mode in the substrate. This is also the boundary condition.

For mathematical operation these waves are represented by their coefficients arranged in two-vector form, i.e., $\begin{pmatrix} E_{ai} \\ E_{ar} \end{pmatrix}$ and $\begin{pmatrix} E_{ot} \\ E_{ob} \end{pmatrix}$ at the first interface, and $\begin{pmatrix} E_{oi} \\ E_{or} \end{pmatrix}$ and $\begin{pmatrix} E_{st} \\ 0 \end{pmatrix}$ at the second. These are connected by (2×2) reflection/transmission and propagation

matrices, see below. The mathematics is similar to that used in Jones-matrix calculations, although the meanings are different. The above structure is generic and can be extended to as many overlayers as necessary simply by including more matrices.

We consider first the (2×2) reflection/transmission matrices. The constraints are that the tangential components of $\vec{E}$ and $\vec{H}$ must be continuous across an interface, a consequence of applying Stokes' Theorem to the Faraday-Maxwell and Ampère Equations, in the latter case with the assumption that no surface currents are present. You recall that the physics reason for doing the calculations in this way is to bypass any need to consider the interface itself, which is probably quite complicated on the atomic scale. The Stokes-Theorem approach casts the calculations in regions where everything is well defined. With $\vec{E}$ specified, $\vec{H}$ follows according to

$$\begin{aligned}\vec{H} &= \frac{c}{\omega}\vec{k}\times\vec{E}\,,\\ &= n\hat{k}\times\vec{E}\,,\end{aligned} \tag{13.1}$$

and thus has opposite signs for incident and reflected directions.

Taking $t=0$, using a local coordinate system where the interface is at $z=0$, and working through the math, we find

$$E_{1i}+E_{1r}=E_{2t}+E_{2b}\,; \tag{13.2a}$$

$$n_1E_{1i}-n_1E_{1r}=n_2E_{2t}-n_2E_{2b}\,. \tag{13.2b}$$

In Eqs. (13.2) we replace any reference to a particular material/interface with the generic 1 and 2, where 1 and 2 represent the incident and transmitted sides, respectively. The solution, written in matrix form, is

$$\begin{pmatrix}E_{1i}\\ E_{1r}\end{pmatrix}=\begin{pmatrix}\frac{1}{2}\left(1+\frac{n_2}{n_1}\right) & \frac{1}{2}\left(1-\frac{n_2}{n_1}\right)\\ \frac{1}{2}\left(1-\frac{n_2}{n_1}\right) & \frac{1}{2}\left(1+\frac{n_2}{n_1}\right)\end{pmatrix}\begin{pmatrix}E_{2t}\\ E_{2b}\end{pmatrix}. \tag{13.3}$$

We can simplify notation significantly by considering the expressions for reflectance and transmittance in the two-phase limit, where $E_b=0$. Then defining

$$r_{12}=\frac{E_r}{E_i}=\frac{n_1-n_2}{n_1+n_2}\,; \tag{13.4a}$$

$$t_{12}=\frac{2n_1}{n_1+n_2}\,; \tag{13.4b}$$

Eq. (13.3) can be written

$$\begin{pmatrix}E_{1i}\\ E_{1r}\end{pmatrix}=\frac{1}{t_{12}}\begin{pmatrix}1 & r_{12}\\ r_{12} & 1\end{pmatrix}\begin{pmatrix}E_{2t}\\ E_{2b}\end{pmatrix}. \tag{13.5}$$

In all cases the first index refers to the incident side, and the second to the transmitted side, hopefully making everything easier to remember.

The propagation matrix describes the phase difference between $\begin{pmatrix} E_{ot} \\ E_{ob} \end{pmatrix}$ and $\begin{pmatrix} E_{oi} \\ E_{or} \end{pmatrix}$ on opposite sides of the layer *o*. Two factors are involved: the direction of propagation, and the direction of multiplication. The way to remember the sign is to note that phase always accumulates in the propagation directions, but what means "propagation direction" depends on the perspective of the observer, which in these calculations is defined by the direction of matrix multiplication. With the boundary conditions on the right, the mode traveling toward the ambient accumulates phase while the mode traveling away from the ambient decrements phase. Hence the propagation matrix appropriate to the above order of matrix multiplication is

$$\begin{pmatrix} E_{ot} \\ E_{ob} \end{pmatrix} = \begin{pmatrix} e^{-ik_o d} & 0 \\ 0 & e^{ik_o d} \end{pmatrix} \begin{pmatrix} E_{oi} \\ E_{or} \end{pmatrix}. \tag{13.6}$$

Putting everything together, the three-phase model is described by

$$\begin{pmatrix} E_{1i} \\ E_{1r} \end{pmatrix} = \frac{1}{t_{12}} \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} e^{-ik_2 d} & 0 \\ 0 & e^{ik_2 d} \end{pmatrix} \frac{1}{t_{23}} \begin{pmatrix} 1 & r_{23} \\ r_{23} & 1 \end{pmatrix} \begin{pmatrix} E_{3t} \\ 0 \end{pmatrix}. \tag{13.7}$$

This matrix equation is easily solved for the reflectance $r_{123} = E_{1r}/E_{1i}$ of the three-layer stack. The closed-form result is the famous Airy Equation of Airy-function fame:

$$r_{123} = \frac{E_{1r}}{E_{1i}} = \frac{r_{12} + r_{23} e^{2ik_o d}}{1 + r_{12} r_{23} e^{2ik_o d}}. \tag{13.8}$$

Looking ahead to Sec. E, the non-normal-incidence version is identical except r_{12} and r_{23} are replaced by their non-normal-incidence equivalents, and k_o with its component normal to the surface. The general properties of reflectances described by the two-and three-phase models are given in the next section.

The above results are easily generalized to multilayer laminar configurations, for example the broadband antireflection coatings used in high-end cameras, reducing these calculations to no more than bookkeeping. However, with few exceptions, determining the properties of multilayer stacks is a computer problem. But such calculations do have a built-in check: once the results for a configuration have been obtained, the computation should be reversed to ensure that E_{sb} is indeed zero, consistent with the boundary condition used to obtain them.

C. Reflectances in the two- and three-phase models; the Airy equation.

For the two-phase model we usually consider only power reflectance in an air ambient. In this situation it is useful to represent the (complex) index of refraction $n \to \tilde{n}$ as $\tilde{n} = n + i\kappa$, where *n* is the ordinary index of refraction and κ is the extinction coefficient. Then

$$R = |r|^2 = \left|\frac{1-n-i\kappa}{1+n+i\kappa}\right|^2 \tag{13.9a}$$

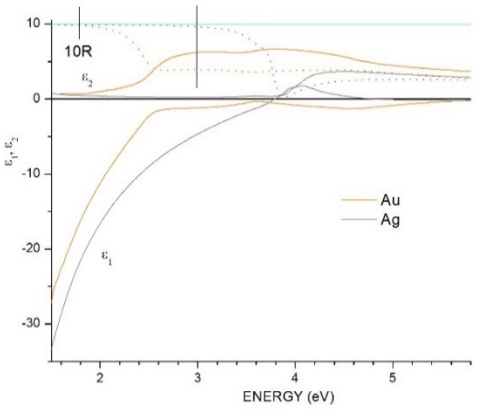
$$= \frac{(n-1)^2+\kappa^2}{(n+1)^2+\kappa^2}. \tag{13.9b}$$

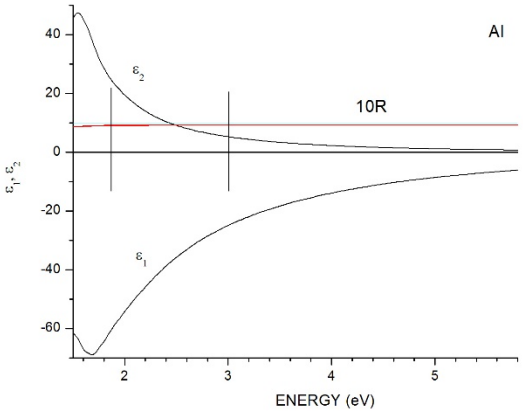
If the substrate is transparent, then $\kappa = 0$ and typically $n \sim 1.5$, so $R \sim 0.04$.

The more interesting case occurs with metals. For a Drude metal

$$\tilde{n} = n + i\kappa = \sqrt{1-\omega_p^2/\omega^2} \tag{13.10}$$

is pure imaginary for $\omega < \omega_p$. In this case $R = 1$. If there is no absorption loss the transition is abrupt, but in practice the rolloff is more graduate. In the noble metals this "plasma edge" occurs at about 650 nm for Cu, 590 nm for Au, and 330 nm for Ag (see figures.) Thus the reflectance is high at the red end of the spectrum for all metals but drops off at approximately yellow for Cu, green for Au, and far in the uv for Ag. This gives Cu and Au their characteristic colors, and the high energy of the plasma edge along with its relatively low loss in the visible range, makes Ag an excellent reflector (until it starts reacting with S.) Aluminum is also used as a reflector, but it exhibits significant loss at the red end of the visible spectrum, so its response is not as striking as Al. Transition metals show significantly more loss in the visible, and as a result have a characteristic grey appearance.





The Airy Equation, Eq. (13.8), describes a wide range of phenomena, from antireflection and decorative coatings through oil slicks to soap bubbles. The antireflection calculation is straightforward. Given n_s and n_a, what are the values of n_o and d_o that minimize r_{soa}? Considering n_a, n_o, and n_s to be real and $n_a < n_o < n_s$, then both terms in the numerator of Eq. (13.8) are negative. Thus if the numerator is to vanish, then $e^{2ik_o d_o} = -1$. The second requirement is $r_{oa} = r_{so}$. This leads to $n_o = \sqrt{n_a n_s}$. Finally,

$$d_o = d_{o\nu} = \frac{1}{2k_o}(\pi + 2\pi\nu) = \frac{\lambda_a}{4n_o}(1+2\nu), \tag{13.11a}$$

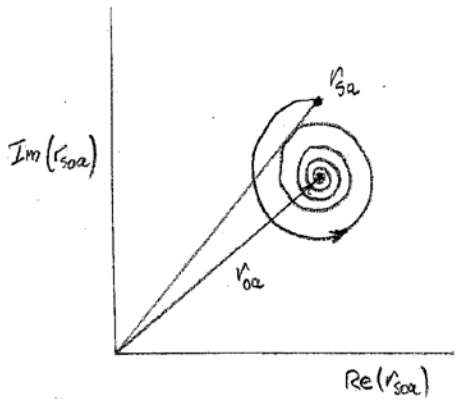
$$= \frac{\lambda_o}{4}, \tag{13.11b}$$

where ν is an integer and $\lambda_o = \lambda_a / n_o$ is the wavelength of light in the overlayer, written in terms of the wavelength in the ambient assuming $n_a = 1$. The choice $\nu = 0$ is made to make the change of phase with d_o to be as slow as possible, and hence make the antireflection coating effective over the largest wavelength range as possible. The maximum reflectance condition occurs for $e^{2ik_o d} = 1$, with a reflectance the same as if the overlayer were not present. This can be verified by direct calculation.

In informal situations the condition $\nu = 0$ 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 ($n_o \sim 1.5$) and water ($n_s = 1.33$) is small. Similar effects are seen in soap bubbles, where the difference between $n_a = n_s = 1$ and $n_o = 1.33$ 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 $d_o = 0$, $e^{ik_o d} \to 1$ 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 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_{oa}, 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 $d_o = 0$ 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}. 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* increases this length increases and decreases, leading to the well-known interference oscillations.

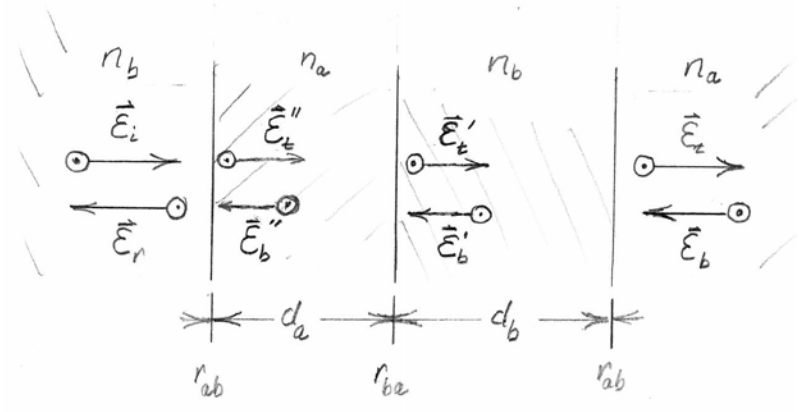
The exponential-spiral limit has also been used in a "virtual-interface" approach to assess compositions of materials being deposited in real time. In this application the substrate is treated as an unknown, to be determined if the rate of deposition is known. The result has been used to control compositions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

D. Multilayer stacks: laser mirrors and one-dimensional photonic crystals.

Configurations with more than one layer are generally computer problems, but what are currently termed one-dimensional photonic crystals are exceptions. Because they are repeating two-layer pairs, the configurations can be solved analytically. The simplest of these 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 found to be 100%.

The reason that these relatively "exotic" mirrors are needed 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. (13.7). Consider the fundamental repeating unit from an interior segment of the stack, which is shown schematically in the figure. 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. (13.28) we find

$$\begin{pmatrix} E_t'' \\ E_b'' \end{pmatrix} = \begin{pmatrix} e^{-ik_a d_a} & 0 \\ 0 & e^{ik_a d_a} \end{pmatrix} \frac{1}{t_{ba}} \begin{pmatrix} 1 & r_{ba} \\ r_{ba} & 1 \end{pmatrix} \begin{pmatrix} e^{-ik_b d_b} & 0 \\ 0 & e^{ik_b d_b} \end{pmatrix} \frac{1}{t_{ab}} \begin{pmatrix} 1 & r_{ab} \\ r_{ab} & 1 \end{pmatrix} \begin{pmatrix} E_t \\ E_b \end{pmatrix}. \tag{13.12}$$

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

$$r_{ab} = \frac{n_b - n_a}{n_b + n_a}; \;\; t_{ab} = \frac{2n_b}{n_b + n_a}. \tag{13.13}$$

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

$$\begin{pmatrix} E_t'' \\ E_b'' \end{pmatrix} = -\frac{1}{2n_b n_a} \begin{pmatrix} n_b^2 + n_a^2 & n_b^2 - n_a^2 \\ n_b^2 - n_a^2 & n_b^2 + n_a^2 \end{pmatrix} \begin{pmatrix} E_t \\ E_b \end{pmatrix}. \tag{13.14}$$

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

$$\begin{pmatrix} E_t'' \\ E_b'' \end{pmatrix} = C_0 \begin{pmatrix} E_t \\ E_b \end{pmatrix}, \tag{13.15}$$

where C_0 is a constant to be determined. Substituting Eq. (13.15) into Eq. (13.14) yields

$$0 = -\frac{1}{2n_b n_a} \begin{pmatrix} (n_b^2 + n_a^2) + 2n_b n_a C_0 & (n_b^2 - n_a^2) \\ (n_b^2 - n_a^2) & (n_b^2 + n_a^2) + 2n_b n_a C_0 \end{pmatrix} \begin{pmatrix} E_t \\ E_b \end{pmatrix}. \tag{13.16}$$

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

$$2n_b n_a C_0 = -(n_b^2 + n_a^2) \pm (n_b^2 - n_a^2), \tag{13.17}$$

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

$$C_0^+ = -\frac{n_a}{n_b};\ \begin{pmatrix} E_t' \\ E_b' \end{pmatrix} = -\frac{n_a}{n_b}\begin{pmatrix} 1 \\ -1 \end{pmatrix} E_t, \tag{13.18a,b}$$

and for the lower sign

$$C_0^- = -\frac{n_b}{n_a};\ \begin{pmatrix} E_t' \\ E_b' \end{pmatrix} = -\frac{n_b}{n_a}\begin{pmatrix} 1 \\ 1 \end{pmatrix} E_t. \tag{13.19a,b}$$

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

$z \le 0$:

$$\vec{E}_-(z,t) = -2i\hat{y}E_t \sin(k_a z)e^{-i\omega t};\ \ \vec{H}_-(z,t) = 2n_a\hat{x}E_t \cos(k_a z)e^{-i\omega t}; \tag{13.20a,b}$$

$z \ge 0$:

$$\vec{E}_+(z,t) = -2i\hat{y}\frac{n_a}{n_b}E_t \sin(k_b z)e^{-i\omega t};\ \ \vec{H}_+(z,t) = 2n_a\hat{x}E_t \cos(k_b z)e^{-i\omega t}. \tag{13.20c,d}$$

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. (13.20), 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

$$\vec{E}(0^-,t) = \vec{E}(0^+,t) = -2i\hat{y}\frac{n_a}{n_b}e^{-i\omega t}; \tag{13.21a}$$

$$\vec{H}(0^-,t) = \vec{H}(0^+,t) = 0. \tag{13.21b}$$

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. (13.19). Here, $C_0 = -n_b/n_a$ and $E_b = E_t$. Repeating the same calculations yields

$z \le 0$:

$$\vec{E}_-(z,t) = -2i\hat{y}E_t \sin(k_a z)e^{-i\omega t}; \quad \vec{H}_-(z,t) = 2n_a \hat{x}E_t \cos(k_a z)e^{-i\omega t}; \tag{13.22a,b}$$

$z \geq 0$:

$$\vec{E}_+(z,t) = -2i\hat{y}\frac{n_a}{n_b}E_t \sin(k_b z)e^{-i\omega t}; \quad \vec{H}_+(z,t) = 2n_a \hat{x}E_t \cos(k_b z)e^{-i\omega t}. \tag{13.22c,d}$$

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

$$\vec{E}'(0^-,t) = \vec{E}'(0^+,t) = 0; \tag{13.23a}$$

$$\vec{H}'(0^-,t) = \vec{H}'(0^+,t) = -2i\hat{x}n_b E_t. \tag{13.23b}$$

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 must increase 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 $(n_b/n_a)^{\mathcal{V}}$ for $\mathcal{V}$ unit cells. To determine how this gain is achieved, write

$$\begin{pmatrix} E_t \\ E_b \end{pmatrix} = \begin{pmatrix} 1 \\ 1-\delta \end{pmatrix} E_t, \tag{13.24}$$

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

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

$$= -\frac{2n_b^2 - (n_b^2 - n_a^2)\delta}{2n_b n_a} \begin{pmatrix} 1 \\ 1 - \dfrac{2n_a^2}{n_b^2 + n_a^2}\delta \end{pmatrix} E_t. \tag{13.25}$$

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.

E. Non-normal-incidence reflection: basic physics.

Non-normal-incidence behavior follows the bookkeeping procedures outlined above, but many more physical phenomena are involved, so the topic is far more interesting. 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* that fills the half-space $z > 0$ adjacent to a substrate *s* that fills the half-space $z \le 0$. The permeabilities and dielectric functions of the ambient and substrate are μ_a, ε_a and μ_s, ε_s, respectively. We generally use the term "ambient" to describe 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 μ 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 $\mu \neq 1$, so retaining μ should not cause undue hardship. However, we do assume that the materials are isotropic, so ε and μ 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 θ_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 $\hat{y}$:

$$\vec{E}_i = \hat{y} E_i e^{-ik_{ix}x - ik_{iz}z - i\omega t}; \tag{13.26a}$$

$$\vec{E}_r = \hat{y} E_r e^{-ik_{rx}x + ik_{rz}z - i\omega t} \tag{13.26b}$$

$$\vec{E}_t = \hat{y} E_t e^{-ik_{tx}x - ik_{tz}z - i\omega t} \tag{13.26c}$$

$$\vec{E}_b = \hat{y} E_b e^{-ik_{bx}x + ik_{bz}z - i\omega t} \tag{13.26d}$$

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

$$\nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} + \frac{\mu\varepsilon}{c^2}\frac{\partial^2}{\partial t^2}\vec{E} = 0\,, \tag{13.27}$$

which follows by combining the Faraday-Maxwell and Ampère Equations. With $\vec{E}$, $\vec{H}$, and $\vec{k}$ mutually orthogonal, Eq. (13.27) reduces to the dispersion equation

$$\left(k^2 - \frac{\omega^2 \mu\varepsilon}{c^2}\right)\vec{E} = 0\,. \tag{13.28}$$

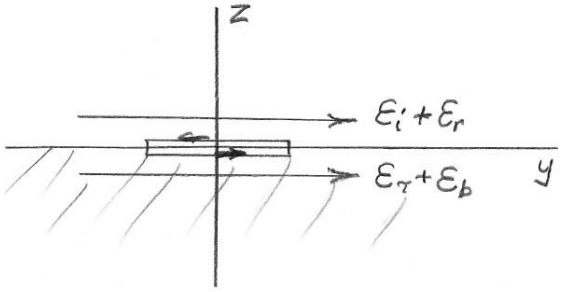
For medium *j* this can be written as

$$k_{xj}^2 + k_{zj}^2 = \frac{\omega^2 \mu_j \varepsilon_j}{c^2} \tag{13.29a}$$

or defining the complex refractive index n_j, as

$$\frac{ck_j}{\omega} = \sqrt{\mu_j \varepsilon_j} = n_j\,. \tag{13.29b}$$

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, μ, and ε 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 right. 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 Δy and a height Δz. Applying Stokes' Theorem to the Faraday-Maxwell Equation we obtain in the limit $\Delta z \to 0$

$$\int_S d^2 r\, \hat{x} \cdot \nabla \times \vec{E} = \oint_C \vec{E} \cdot d\vec{l} = \left(E_\tau + E_b - E_r - E_i\right)\Delta l \tag{13.30a}$$

$$= -\frac{1}{c}\int_S d^2r \frac{\partial \vec{B}}{\partial t} = 0 . \qquad (13.30b)$$

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

$$\hat{z}\times(\vec{E}_i + \vec{E}_r) = \hat{z}\times(\vec{E}_\tau + \vec{E}_b) . \qquad (13.31)$$

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. (13.31) is general, also applicable to the more complex modes encountered in crystal optics.

In the current plane-wave context, Eq. (13.31) imposes several restrictions, some of which are not immediately obvious. Evaluating Eqs. (13.26) at $z = 0$ explicitly and ignoring the common factor $e^{-i\omega t}$, Eq. (13.31) becomes

$$E_i e^{ik_{ix}x} + E_r e^{ik_{rx}x} = E_\tau e^{ik_{tx}x} + E_b e^{ik_{bx}x} . \qquad (13.32)$$

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

$$\hat{x}\cdot\vec{k}_r = k_{rx} = \hat{x}\cdot\vec{k}_t = k_{tx} = \hat{x}\cdot\vec{k}_b = k_{bx} = \hat{x}\cdot\vec{k}_i = k_{ix} = k_x . \qquad (13.33a\text{-}g)$$

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 $k_x = k_a \sin\theta_i$. 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. (13.29) and (13.33a). We have

$$k_x^2 + k_{iz}^2 = k_x^2 + k_{rz}^2 . \qquad (13.34)$$

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

$$k_{jz}^2 = k_j^2 - k_x^2 \qquad (13.35)$$

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

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

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

$c^2 k^2 / \omega^2 = \varepsilon$,

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

$$k_z^2 = k^2 - k_a^2 \sin^2 \theta$$

Which we usually write as

$$n_{jz} = \frac{ck_{jz}}{\omega} = \sqrt{\varepsilon_j - \varepsilon_a \sin^2 \theta}$$

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

$$\vec{k}_j \times \vec{E}_j = -\frac{\omega}{c} \mu_j \vec{H}_j . \tag{13.36}$$

Therefore

$$\vec{H}_j = -\sqrt{\frac{\varepsilon_j}{\mu_j}} \hat{k}_j \times \vec{E}_j . \tag{13.37}$$

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

$$\hat{z} \times (\vec{H}_i + \vec{H}_r) = \hat{z} \times (\vec{H}_\tau + \vec{H}_b) . \tag{13.38}$$

As with Eq. (13.31), Eq. (13.38) 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

$$\hat{z} \times (\vec{H}_i + \vec{H}_r) = \frac{4\pi}{c} \vec{K} , \tag{13.39}$$

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

$$\vec{S} = \frac{c}{4\pi}\vec{E}\times\vec{H}\,. \tag{13.40}$$

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

$$\vec{E}_j = \sqrt{\frac{\mu_j}{\varepsilon_j}}\hat{k}_j\times\vec{H}_j\,. \tag{13.41}$$

This expression is different from Eq. (13.37). 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 μ and ε, 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 normal component of $\vec{D}$ is $\hat{z}\cdot\vec{D} = \varepsilon\hat{z}\cdot\vec{E}$. Applying this relation to the j^{th} material, Eq. (13.16) shows that

$$\varepsilon_j\hat{z}\cdot\vec{E}_j = \varepsilon_j\sqrt{\frac{\mu_j}{\varepsilon_j}}\hat{z}\cdot\hat{k}_j\times\vec{H}_j = \sqrt{\mu_j\varepsilon_j}\left(\hat{k}_j\cdot\vec{H}_j\times\hat{z}\right). \tag{13.42}$$

Now $\vec{H}_j\times\hat{z}$ is the tangential component of $\vec{H}$, which by Eq. (13.38) is continuous across an interface. Also, $\vec{H}_j\times\hat{z}$ is tangential to the interface, so its scalar product with $\hat{k}_j$ extracts its tangential component. By Eq. (13.33), the tangential component of

$c\vec{k}_j/\omega = \sqrt{\mu_j \varepsilon_j}\hat{k}_j$ is the same everywhere in the configuration. Consequently, Eq. (13.17) 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. (13.37), and is left as a homework assignment. Thus overdetermination does not occur.