31

The Origin of the Refractive Index

31–1 The index of refraction

We have said before that light goes slower in water than in air, and slower, slightly, in air than in vacuum. This effect is described by the index of refraction n. Now we would like to understand how such a slower velocity could come about. In particular, we should try to see what the relation is to some physical assumptions, or statements, we made earlier, which were the following:

(a) That the total electric field in any physical circumstance can always be represented by the sum of the fields from all the charges in the universe.

(b) That the field from a single charge is given by its acceleration evaluated with a retardation at the speed c, *always* (for the *radiation* field).

But, for a piece of glass, you might think: "Oh, no, you should modify all this. You should say it is retarded at the speed c/n." That, however, is not right, and we have to understand why it is not.

It *is* approximately true that light or any electrical wave *does appear* to travel at the speed c/n through a material whose index of refraction is n, but the fields are still produced by the motions of *all* the charges—including the charges moving in the material—and with these basic contributions of the field travelling at the ultimate velocity c. Our problem is to understand how the *apparently* slower velocity comes about.

We shall try to understand the effect in a very simple case. A source which we shall call "the *external* source" is placed a large distance away from a thin plate of transparent material, say glass. We inquire about the field at a large distance on the opposite side of the plate. The situation is illustrated by the diagram of Fig. 31–1, where S and P are imagined to be very far away from the plate. According to the principles we have stated earlier, an electric field anywhere that is far from all moving charges is the (vector) sum of the fields produced by the external source (at S) *and* the fields produced by *each* of the charges in the plate of glass, *every one with its proper retardation at the velocity c.* Remember that the contribution of each charge is not changed by the presence of the other charges. These are our basic principles. The field at P can be written thus:

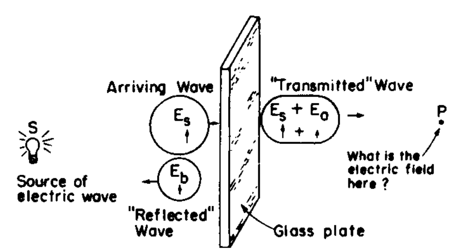

Fig. 31–1. Electric waves passing through a layer of transparent material.

$$\mathbf{E} = \sum_{\text{all charges}} \mathbf{E}_{\text{each charge}} \tag{31.1}$$

or

$$\mathbf{E} = \mathbf{E}_s + \sum_{\text{all other charges}} \mathbf{E}_{\text{each charge}}, \tag{31.2}$$

where $\mathbf{E}_s$ is the field due to the source alone and would be precisely the field at P *if there were no material present.* We expect the field at P to be different from $\mathbf{E}_s$ if there are any other moving charges.

Why should there be charges moving in the glass? We know that all material consists of atoms which contain electrons. When the electric field *of the source* acts on these atoms it drives the electrons up and down, because it exerts a force on the electrons. And moving electrons generate a field—they constitute new radiators. These new radiators are related to the source S, because they are driven by the field of the source. The total field is not just the field of the source S, but it is modified by the additional contribution from the other moving charges. This means that the field is not the same as the one which was there before the glass was there, but is modified, and it turns out that it is modified in such a way that the field inside the glass appears to be moving at a different speed. That is the idea which we would like to work out quantitatively.

Now this is, in the exact case, pretty complicated, because although we have said that all the other moving charges are driven by the source field, that is not quite true. If we think of a particular charge, it feels not only the source, but like anything else in the world, it feels *all* of the charges that are moving. It feels, in particular, the charges that are moving somewhere else in the glass. So the total field which is acting on a *particular charge* is a combination of the fields from the other charges, *whose motions depend on what this particular charge is doing!* You can see that it would take a complicated set of equations to get the complete and exact formula. It is so complicated that we postpone this problem until next year.

Instead we shall work out a very simple case in order to understand all the physical principles very clearly. We take a circumstance in which the effects from the other atoms is very small relative to the effects from the source. In other words, we take a material in which the total field is not modified very much by the motion of the other charges. That corresponds to a material in which the index of refraction is very close to 1, which will happen, for example, if the density of the atoms is very low. Our calculation will be valid for any case in which the index is for any reason very close to 1. In this way we shall avoid the complications of the most general, complete solution.

Incidentally, you should notice that there is another effect caused by the motion of the charges in the plate. These charges will also radiate waves back toward the source S. This backward-going field is the light we see reflected from the surfaces of transparent materials. It does not come from just the surface. The backward radiation comes from everywhere in the interior, but it turns out that the total effect is equivalent to a reflection from the surfaces. These reflection effects are beyond our approximation at the moment because we shall be limited to a calculation for a material with an index so close to 1 that very little light is reflected.

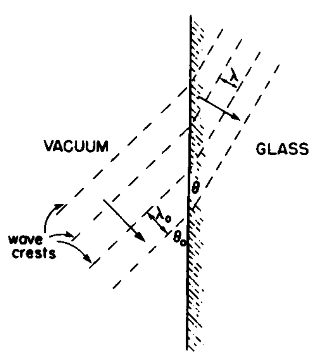

Fig. 31–2. Relation between refraction and velocity change.

Before we proceed with our study of how the index of refraction comes about, we should understand that all that is required to understand refraction is to understand why the apparent wave *velocity* is different in different materials. The *bending* of light rays comes about just *because* the effective speed of the waves is different in the materials. To remind you how that comes about we have drawn in Fig. 31–2 several successive crests of an electric wave which arrives from a vacuum onto the surface of a block of glass. The arrow perpendicular to the wave crests indicates the direction of travel of the wave. Now all oscillations in the wave must have the same *frequency*. (We have seen that driven oscillations have the same frequency as the driving source.) This means, also, that the wave crests for the waves on both sides of the surface must have the *same spacing along the surface* because they must travel together, so that a charge sitting at the boundary will feel only one frequency. The *shortest* distance between crests of the wave, however, is the wavelength which is the velocity divided by the frequency. On the vacuum side it is $\lambda_0 = 2\pi c/\omega$, and on the other side it is $\lambda = 2\pi v/\omega$ or $2\pi c/\omega n$, if $v = c/n$ is the velocity of the wave. From the figure we can see that the only way for the waves to "fit" properly at the boundary is for the waves in the material to be travelling at a different angle with respect to the surface. From the geometry of the figure you can see that for a "fit" we must have $\lambda_0/\sin\theta_0 = \lambda/\sin\theta$, or $\sin\theta_0/\sin\theta = n$, which is Snell's law. We shall, for the rest of our discussion, consider only why light has an effective speed of c/n in material of index n, and no longer worry, in this chapter, about the bending of the light direction.

We go back now to the situation shown in Fig. 31–1. We see that what we have to do is to calculate the field produced at P by all the oscillating charges in the glass plate. We shall call this part of the field E_a, and it is just the sum written as the second term in Eq. (31.2). When we add it to the term E_s, due to the source, we will have the total field at P.

This is probably the most complicated thing that we are going to do this year, but it is complicated only in that there are many pieces that have to be put together; each piece, however, is very simple. Unlike other derivations where we say, "Forget the derivation, just look at the answer!," in this case we do not need the answer so much as the derivation. In other words, the thing to understand now is the physical machinery for the production of the index.

To see where we are going, let us first find out what the "correction field" E_a would have to be if the total field at P is going to look like radiation from the source that is slowed down while passing through the thin plate. If the plate had no effect on it, the field of a wave travelling to the right (along the z-axis) would be

$$E_s = E_0 \cos \omega(t - z/c) \tag{31.3}$$

or, using the exponential notation,

$$E_s = E_0 e^{i\omega(t-z/c)}. \tag{31.4}$$

Now what would happen if the wave travelled more slowly in going through the plate? Let us call the thickness of the plate Δz. If the plate were not there the wave would travel the distance Δz in the time $\Delta z/c$. But if it appears to travel at the speed c/n then it should take the longer time $n\,\Delta z/c$ or the *additional* time $\Delta t = (n - 1)\,\Delta z/c$. After that it would continue to travel at the speed c again. We can take into account the extra delay in getting through the plate by replacing t in Eq. (31.4) by $(t - \Delta t)$ or by $[t - (n - 1)\,\Delta z/c]$. So the wave after insertion of the plate should be written

$$E_{\text{after plate}} = E_0 e^{i\omega[t-(n-1)\Delta z/c-z/c]}. \tag{31.5}$$

We can also write this equation as

$$E_{\text{after plate}} = e^{-i\omega(n-1)\Delta z/c} E_0 e^{i\omega(t-z/c)}, \tag{31.6}$$

which says that the wave after the plate is obtained from the wave which could exist without the plate, i, e., from E_s, by multiplying by the factor $e^{-i\omega(n-1)\Delta z/c}$. Now we know that multiplying an oscillating function like $e^{i\omega t}$ by a factor $e^{i\theta}$ just says that we change the phase of the oscillation by the angle θ, which is, of course, what the extra delay in passing through the thickness Δz has done. It has retarded the phase by the amount $\omega(n - 1)\,\Delta z/c$ (retarded, because of the minus sign in the exponent).

We have said earlier that the plate should *add* a field E_a to the original field $E_s = E_0 e^{i\omega(t-z/c)}$, but we have found instead that the effect of the plate is to *multiply* the field by a factor which shifts its phase. However, that is really all right because we can get the same result by adding a suitable complex number. It is particularly easy to find the right number to add in the case that Δz is small, for you will remember that if x is a small number then e^x is nearly equal to $(1 + x)$. We can write, therefore,

$$e^{-i\omega(n-1)\Delta z/c} = 1 - i\omega(n - 1)\,\Delta z/c. \tag{31.7}$$

Using this equality in Eq. (31.6), we have

$$E_{\text{after plate}} = \underbrace{E_0 e^{i\omega(t-z/c)}}_{E_s} - \underbrace{\frac{i\omega(n - 1)\,\Delta z}{c} E_0 e^{i\omega(t-z/c)}}_{E_a}. \tag{31.8}$$

The first term is just the field from the source, and the second term must just be equal to E_a, the field produced to the right of the plate by the oscillating charges of the plate—expressed here in terms of the index of refraction n, and depending, of course, on the strength of the wave from the source.

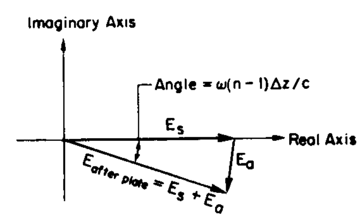

Fig. 31–3. Diagram for the transmitted wave at a particular t and z.

What we have been doing is easily visualized if we look at the complex number diagram in Fig. 31–3. We first draw the number E_s (we chose some values for z and t so that E_s comes out horizontal, but this is not necessary). The delay due to slowing down in the plate would delay the phase of this number, that is, it would rotate E_s through a negative angle. But this is equivalent to adding the small vector E_a at roughly right angles to E_s. But that is just what the factor $-i$ means in the second term of Eq. (31.8). It says that if E_s is real, then E_a is negative imaginary or that, in general, E_s and E_a make a right angle.

31-2 The field due to the material

We now have to ask: Is the field E_a obtained in the second term of Eq. (31.8) the kind we would expect from oscillating charges in the plate? If we can show that it is, we will then have calculated what the index n should be! [Since n is the only nonfundamental number in Eq. (31.8).] We turn now to calculating what field E_a the charges in the material will produce. (To help you keep track of the many symbols we have used up to now, and will be using in the rest of our calculation, we have put them all together in Table 31-1.)

Table 31-1

Symbols used in the calculations

E_s	=	field from the source
E_a	=	field produced by charges in the plate
Δz	=	thickness of the plate
z	=	perpendicular distance from the plate
n	=	index of refraction
ω	=	frequency (angular) of the radiation
N	=	number of charges per unit volume in the plate
η	=	number of charges per unit area of the plate
q_e	=	charge on an electron
m	=	mass of an electron
ω_0	=	resonant frequency of an electron bound in an atom

If the source S (of Fig. 31-1) is far off to the left, then the field E_s will have the same phase everywhere on the plate, so we can write that in the neighborhood of the plate

$$E_s = E_0 e^{i(\omega t - z/c)} \tag{31.9}$$

Right at the plate, where $z = 0$, we will have

$$E_s = E_0 e^{i\omega t} \text{ (at the plate)} \tag{31.10}$$

Each of the electrons in the atoms of the plate will feel this electric field and will be driven up and down (we assume the direction of E_0 is vertical) by the electric force qE. To find what motion we expect for the electrons, we will assume that the atoms are little oscillators, that is, that the electrons are fastened elastically to the atoms, which means that if a force is applied to an electron its displacement from its normal position will be proportional to the force.

You may think that this is a funny model of an atom if you have heard about electrons whirling around in orbits. But that is just an oversimplified picture. The correct picture of an atom, which is given by the theory of wave mechanics, says that, *so far as problems involving light are concerned*, the electrons behave as though they were held by springs. So we shall suppose that the electrons have a linear restoring force which, together with their mass m, makes them behave like little oscillators, with a resonant frequency ω_0. We have already studied such oscillators, and we know that the equation of their motion is written this way:

$$m\left(\frac{d^2x}{dt^2} + \omega_0^2 x\right) = F, \tag{31.11}$$

where F is the driving force.

For our problem, the driving force comes from the electric field of the wave from the source, so we should use

$$F = q_e E_s = q_e E_0 e^{i\omega t}, \tag{31.12}$$

where q_e is the electric charge on the electron and for E_s we use the expression $E_s = E_0 e^{i\omega t}$ from (31.10). Our equation of motion for the electron is then

$$m\left(\frac{d^2x}{dt^2} + \omega_0^2 x\right) = q_e E_0 e^{i\omega t}. \tag{31.13}$$

We have solved this equation before, and we know that the solution is

$$x = x_0 e^{i\omega t}, \tag{31.14}$$

where, by substituting in (31.13), we find that

$$x_0 = \frac{q_e E_0}{m(\omega_0^2 - \omega^2)}, \tag{31.15}$$

so that

$$x = \frac{q_e E_0}{m(\omega_0^2 - \omega^2)}\, e^{i\omega t}. \tag{31.16}$$

We have what we needed to know—the motion of the electrons in the plate. And it is the same for every electron, except that the mean position (the "zero" of the motion) is, of course, different for each electron.

Now we are ready to find the field E_a that these atoms produce at the point P, because we have already worked out (at the end of Chapter 30) what field is produced by a sheet of charges that all move together. Referring back to Eq. (30.19), we see that the field E_a at P is just a negative constant times the velocity of the charges retarded in time the amount z/c. Differentiating x in Eq. (31.16) to get the velocity, and sticking in the retardation [or just putting x_0 from (31.15) into (30.18)] yields

$$E_a = -\frac{\eta q_e}{2\epsilon_0 c}\left[i\omega \frac{q_e E_0}{m(\omega_0^2 - \omega^2)}\, e^{i\omega(t - z/c)}\right]. \tag{31.17}$$

Just as we expected, the driven motion of the electrons produced an extra wave which travels to the right (that is what the factor $e^{i\omega(t-z/c)}$ says), and the amplitude of this wave is proportional to the number of atoms per unit area in the plate (the factor η) and also proportional to the strength of the source field (the factor E_0). Then there are some factors which depend on the atomic properties (q_e, m, and ω_0), as we should expect.

The most important thing, however, is that this formula (31.17) for E_a looks very much like the expression for E_a that we got in Eq. (31.8) by saying that the original wave was delayed in passing through a material with an index of refraction n. The two expressions will, in fact, be identical if

$$(n - 1)\,\Delta z = \frac{\eta q_e^2}{2\epsilon_0 m(\omega_0^2 - \omega^2)}. \tag{31.18}$$

Notice that both sides are proportional to Δz, since η, which is the number of atoms *per unit area*, is equal to $N\,\Delta Z$, where N is the number of atoms *per unit volume* of the plate. Substituting $N\,\Delta Z$ for η and cancelling the Δz, we get our main result, a formula for the index of refraction in terms of the properties of the atoms of the material—and of the frequency of the light:

$$n = 1 + \frac{N q_e^2}{2\epsilon_0 m(\omega_0^2 - \omega^2)}. \tag{31.19}$$

This equation gives the "explanation" of the index of refraction that we wished to obtain.

31-3 Dispersion

Notice that in the above process we have obtained something very interesting. For we have not only a number for the index of refraction which can be computed from the basic atomic quantities, but we have also learned how the index of refraction should vary with the frequency ω of the light. This is something we would never understand from the simple statement that "light travels slower in a transparent material." We still have the problem, of course, of knowing how many atoms per unit volume there are, and what is their natural frequency ω_0. We do not know this just yet, because it is different for every different material, and we cannot get a general theory of that now. Formulation of a general theory of the properties of different substances—their natural frequencies, and so on—is possible only with quantum atomic mechanics. Also, different materials have different properties and different indexes, so we cannot expect, anyway, to get a general formula for the index which will apply to all substances.

However, we shall discuss the formula we have obtained, in various possible circumstances. First of all, for most ordinary gases (for instance, for air, most colorless gases, hydrogen, helium, and so on) the natural frequencies of the electron oscillators correspond to ultraviolet light. These frequencies are higher than the frequencies of visible light, that is, ω_0 is much larger than ω of visible light, and to a first approximation, we can disregard ω^2 in comparison with ω_0^2. Then we find that the index is nearly constant. So for a gas, the index is nearly constant. This is also true for most other transparent substances, like glass. If we look at our expression a little more closely, however, we notice that as ω rises, taking a little bit more away from the denominator, the index also rises. So n rises slowly with frequency. The index is higher for blue light than for red light. That is the reason why a prism bends the light more in the blue than in the red.

The phenomenon that the index depends upon the frequency is called the phenomenon of *dispersion*, because it is the basis of the fact that light is "dispersed" by a prism into a spectrum. The equation for the index of refraction as a function of frequency is called a *dispersion equation*. So we have obtained a dispersion equation. (In the past few years "dispersion equations" have been finding a new use in the theory of elementary particles.)

Our dispersion equation suggests other interesting effects. If we have a natural frequency ω_0 which lies in the visible region, or if we measure the index of refraction of a material like glass in the ultraviolet, where ω gets near ω_0, we see that at frequencies very close to the natural frequency the index can get enormously large, because the denominator can go to zero. Next, suppose that ω is greater than ω_0. This would occur, for example, if we take a material like glass, say, and shine x-ray radiation on it. In fact, since many materials which are opaque to visible light, like graphite for instance, are transparent to x-rays, we can also talk about the index of refraction of carbon for x-rays. All the natural frequencies of the carbon atoms would be much lower than the frequency we are using in the x-rays, since x-ray radiation has a very high frequency. The index of refraction is that given by our dispersion equation if we set ω_0 equal to zero (we neglect ω_0^2 in comparison with ω^2).

A similar situation would occur if we beam radiowaves (or light) on a gas of free electrons. In the upper atmosphere electrons are liberated from their atoms by ultraviolet light from the sun and they sit up there as free electrons. For free electrons $\omega_0 = 0$ (there is no elastic restoring force). Setting $\omega_0 = 0$ in our dispersion equation yields the correct formula for the index of refraction for radiowaves in the stratosphere, where N is now to represent the density of free electrons (number per unit volume) in the stratosphere. But let us look again at the equation, if we beam x-rays on matter, or radiowaves (or any electric waves) on free electrons the term $(\omega_0^2 - \omega^2)$ becomes *negative*, and we obtain the result that n is *less than one*. That means that the effective speed of the waves in the substance is *faster* than c! Can that be correct?

It is correct. In spite of the fact that it is said that you cannot send signals any faster than the speed of light, it is nevertheless true that the index of refraction of materials at a particular frequency can be either greater or less than 1. This just means that the *phase shift* which is produced by the scattered light can be either positive or negative. It can be shown, however, that the speed at which you can send a *signal* is not determined by the index at one frequency, but depends on what the index is at *many* frequencies. What the index tells us is the speed at which the *nodes* (or crests) of the wave travel. The *node* of a wave is not a signal by itself. In a perfect wave, which has no modulations of any kind, i.e., which is a steady oscillation, you cannot really say when it "starts," so you cannot use it for a timing signal. In order to send a *signal* you have to change the wave somehow, make a notch in it, make it a little bit fatter or thinner. That means that you have to have more than one frequency in the wave, and it can be shown that the speed at which *signals* travel is not dependent upon the index alone, but upon the way that the index changes with the frequency. This subject we must also delay (until Chapter 48). Then we will calculate for you the actual speed of *signals* through such a piece of glass, and you will see that it will not be faster than the speed of light, although the nodes, which are mathematical points, do travel faster than the speed of light.

Just to give a slight hint as to how that happens, you will note that the real difficulty has to do with the fact that the responses of the charges are opposite to the field, i.e., the sign has gotten reversed. Thus in our expression for x (Eq. 31.16) the displacement of the charge is in the direction opposite to the driving field, because $(\omega_0^2 - \omega^2)$ is negative for small ω_0. The formula says that when the electric field is pulling in one direction, the charge is moving in the opposite direction.

How does the charge happen to be going in the opposite direction? It certainly does not start off in the opposite direction when the field is first turned on. When the motion first starts there is a transient, which settles down after awhile, and only *then* is the phase of the oscillation of the charge opposite to the driving field. And it is then that the *phase* of the transmitted field can appear to be *advanced* with respect to the source wave. It is this *advance in phase* which is meant when we say that the "phase velocity" or velocity of the nodes is greater than c. In Fig. 31–4 we give a schematic idea of how the waves might look for a case where the wave is suddenly turned on (to make a signal). You will see from the diagram that the *signal* (i.e., the *start* of the wave) is *not* earlier for the wave which ends up with an advance in phase.

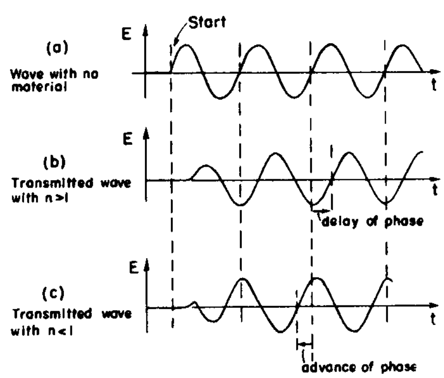

Fig. 31–4. Wave "signals."

Let us now look again at our dispersion equation. We should remark that our analysis of the refractive index gives a result that is somewhat simpler than you would actually find in nature. To be completely accurate we must add some refinements. First, we should expect that our model of the atomic oscillator should have some damping force (otherwise once started it would oscillate forever, and we do not expect that to happen). We have worked out before (Eq. 23.8) the motion of a damped oscillator and the result is that the denominator in Eq. (31.16), and therefore in (31.19), is changed from $(\omega_0^2 - \omega^2)$ to $(\omega_0^2 - \omega^2 + i\gamma\omega)$, where γ is the damping coefficient.

We need a second modification to take into account the fact that there are several resonant frequencies for a particular kind of atom. It is easy to fix up our dispersion equation by imagining that there are several different kinds of oscillators, but that each oscillator acts separately, and so we simply add the contributions of all the oscillators. Let us say that there are N_k electrons per unit of volume, whose natural frequency is ω_k and whose damping factor is γ_k. We would then have for our dispersion equation

$$n = 1 + \frac{q_e^2}{2\epsilon_0 m} \sum_k \frac{N_k}{\omega_k^2 - \omega^2 + i\gamma_k\omega}. \tag{31.20}$$

We have, finally, a complete expression which describes the index of refraction that is observed for many substances.* The index described by this formula varies with frequency roughly like the curve shown in Fig. 31–5.

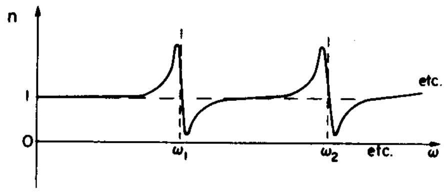

Fig. 31–5. The index of refraction as a function of frequency.

You will note that so long as ω is not too close to one of the resonant frequencies, the slope of the curve is positive. Such a positive slope is called "normal" dispersion (because it is clearly the most common occurrence). Very near the resonant frequencies, however, there is a small range of ω's for which the slope is negative. Such a negative slope is often referred to as "anomalous" (meaning abnormal) dispersion, because it seemed unusual when it was first observed, long before anyone even knew there were such things as electrons. From our point of view both slopes are quite "normal"!

31–4 Absorption

Perhaps you have noticed something a little strange about the last form (Eq. 31.20) we obtained for our dispersion equation. Because of the term $i\gamma$ we put in to take account of damping, the index of refraction is now a *complex number!* What does *that* mean? By working out what the real and imaginary parts of n are we could write

$$n = n' - in'', \tag{31.21}$$

where n' and n'' are real numbers. (We use the minus sign in front of the in'' because then n'' will turn out to be a positive number, as you can show for yourself.)

We can see what such a complex index means by going back to Eq. (31.6), which is the equation of the wave after it goes through a plate of material with an index n. If we put our complex n into this equation, and do some rearranging, we get

$$E_{\text{after plate}} = \underbrace{e^{-\omega n''\Delta z/c}}_{\text{A}} \underbrace{e^{-i\omega(n'-1)\Delta z/c}\, E_0 e^{i\omega(t-z/c)}}_{\text{B}} \tag{31.22}$$

The last factors, marked B in Eq. (31.22), are just the form we had before, and again describe a wave whose phase has been delayed by the angle $\omega(n' - 1)\,\Delta z/c$ in traversing the material. The first term (A) is new and is an exponential factor with a *real* exponent, because there were two i's that cancelled. Also, the exponent is negative, so the factor is a real number less than one. It describes a *decrease* in the magnitude of the field and, as we should expect, by an amount which is more the larger Δz is. As the wave goes through the material, it is weakened. The material is "absorbing" part of the wave. The wave comes out the other side with less energy. We should not be surprised at this, because the damping we put in for the oscillators is indeed a friction force and must be expected to cause a loss of energy. We see that the imaginary part n'' of a complex index of refraction represents an absorption (or "attenuation") of the wave. In fact, n'' is sometimes referred to as the "absorption index."

We may also point out that an imaginary part to the index n corresponds to bending the arrow E_a in Fig. 31–3 toward the origin. It is clear why the transmitted field is then decreased.

* Actually, although in quantum mechanics Eq. (31.20) is still valid, its interpretation is somewhat different. In quantum mechanics even an atom with one electron, like hydrogen, has several resonant frequencies. Therefore N_k is not really the number of electrons having the frequency ω_k, but is replaced instead by Nf_k, where N is the number of atoms per unit volume and f_k (called the oscillator strength) is a factor that tells how strongly the atom exhibits each of its resonant frequencies ω_k.

Normally, for instance as in glass, the absorption of light is very small. This is to be expected from our Eq. (31.20), because the imaginary part of the denominator, $i\gamma_k\omega$, is much smaller than the term $(\omega_k^2 - \omega^2)$. But if the light frequency ω is very close to ω_k then the resonance term $(\omega_k^2 - \omega^2)$ can become small compared with $i\gamma_k\omega$ and the index becomes almost completely imaginary. The absorption of the light becomes the dominant effect. It is just this effect that gives the dark lines in the spectrum of light which we receive from the sun. The light from the solar surface has passed through the sun's atmosphere (as well as the earth's), and the light has been strongly absorbed at the resonant frequencies of the atoms in the solar atmosphere.

The observation of such spectral lines in the sunlight allows us to tell the resonant frequencies of the atoms and hence the chemical composition of the sun's atmosphere. The same kind of observations tell us about the materials in the stars. From such measurements we know that the chemical elements in the sun and in the stars are the same as those we find on the earth.

31–5 The energy carried by an electric wave

We have seen that the imaginary part of the index means absorption. We shall now use this knowledge to find out how much energy is carried by a light wave. We have given earlier an argument that the energy carried by light is proportional to $\overline{E^2}$, the time average of the square of the electric field in the wave. The decrease in E due to absorption must mean a loss of energy, which would go into some friction of the electrons and, we might guess, would end up as heat in the material.

If we consider the light arriving on a unit area, say one square centimeter, of our plate in Fig. 31–1, then we can write the following energy equation (if we assume that energy is conserved, as we *do!*):

$$\text{Energy in per sec} = \text{energy out per sec} + \text{work done per sec.} \tag{31.23}$$

For the first term we can write $\alpha\overline{E_s^2}$, where α is the as yet unknown constant of proportionality which relates the average value of E^2 to the energy being carried. For the second term we must include the part from the radiating atoms of the material, so we should use $\alpha\overline{(E_s + E_a)^2}$, or (evaluating the square) $\alpha(\overline{E_s^2} + 2\overline{E_sE_a} + \overline{E_a^2})$.

All of our calculations have been made for a thin layer of material whose index is not too far from 1, so that E_a would always be much less than E_s (just to make the calculations easier). In keeping with our approximations, we should, therefore, leave out the term $\overline{E_a^2}$, because it is much smaller than $\overline{E_sE_a}$. You may say: "Then you should leave out $\overline{E_sE_a}$ also, because *it* is much smaller than $\overline{E_s^2}$." It is true that $\overline{E_sE_a}$ is much smaller than $\overline{E_s^2}$, but we must keep $\overline{E_sE_a}$ or our approximation will be the one that would apply if we neglected the presence of the material completely! One way of checking that our calculations are consistent is to see that we always keep terms which are proportional to $N\,\Delta z$, the area density of atoms in the material, but we leave out terms which are proportional to $(N\,\Delta z)^2$ or any higher power of $N\,\Delta z$. Ours is what should be called a "low-density approximation."

In the same spirit, we might remark that our energy equation has neglected the energy in the reflected wave. But that is OK because this term, too, is proportional to $(N\,\Delta z)^2$, since the amplitude of the reflected wave is proportional to $N\,\Delta z$.

For the last term in Eq. (31.23) we wish to compute the rate at which the incoming wave is doing work on the electrons. We know that work is force times distance, so the *rate* of doing work (also called power) is the force times the velocity. It is really $\mathbf{F}\cdot\mathbf{V}$, but we do not need to worry about the dot product when the velocity and force are along the same direction as they are here (except for a possible minus sign). So for each atom we take $\overline{q_eE_sv}$ for the average rate of doing work. Since there are $N\,\Delta z$ atoms in a unit area, the last term in Eq. (31.23) should be $N\,\Delta z q_e\,\overline{E_s v}$. Our energy equation now looks like

$$\alpha\overline{E_s^2} = \alpha\overline{E_s^2} + 2\alpha\overline{E_s E_a} + N\,\Delta z\,q_e\,\overline{E_s v}. \tag{31.24}$$

The $\overline{E_s^2}$ terms cancel, and we have

$$2\alpha\overline{E_s E_a} = N\,\Delta z\,q_e\,\overline{E_s v}. \tag{31.25}$$

We now go back to Eq. (30.19), which tells us that for large z

$$E_a = \frac{N\,\Delta z q_e}{2\epsilon_0 c}\,v(\text{ret by } z/c) \tag{31.26}$$

(recalling that $\eta = N\,\Delta z$). Putting Eq. (31.26) into the left-hand side of (31.25), we get

$$2\alpha\,\frac{N\,\Delta z q_e}{2\epsilon_0 c}\,\overline{E_s(\text{at } z)\cdot v(\text{ret by } z/c)}.$$

However, E_s(at z) is E_s (at atoms) retarded by z/c. Since the average is independent of time, it is the same now as retarded by z/c, or is $\overline{E_s \text{ (at atom)} \cdot v}$, the same average that appears on the right-hand side of (31.25). The two sides are therefore equal if

$$\frac{\alpha}{\epsilon_0 c} = 1, \qquad \text{or} \qquad \alpha = \epsilon_0 c. \tag{31.27}$$

We have discovered that if energy is to be conserved, the energy carried in an electric wave per unit area and per unit time (or what we have called the *intensity*) must be given by $\epsilon_0 c\overline{E^2}$. If we call the intensity $\overline{S}$, we have

$$\overline{S} = \begin{Bmatrix} \text{intensity} \\ \text{or} \\ \text{energy/area/time} \end{Bmatrix} = \epsilon_0 c\overline{E^2}, \tag{31.28}$$

where the *bar* means the *time average*. We have a nice bonus result from our theory of the refractive index!

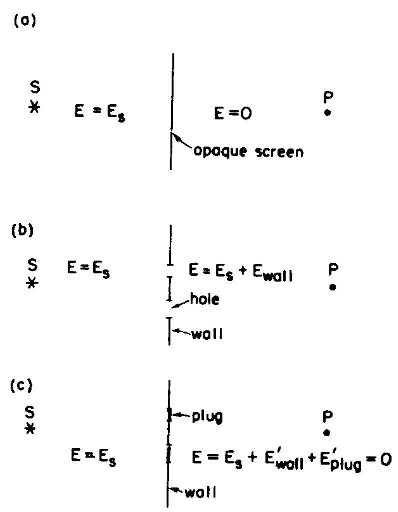

Fig. 31–6. Diffraction by a screen.

31–6 Diffraction of light by a screen

It is now a good time to take up a somewhat different matter which we can handle with the machinery of this chapter. In the last chapter we said that when you have an opaque screen and the light can come through some holes, the distribution of intensity—the diffraction pattern—could be obtained by imagining instead that the holes are replaced by sources (oscillators) uniformly distributed over the hole. In other words, the diffracted wave is the same as though the hole were a new source. We have to explain the reason for that, because the hole is, of course, just where there are *no* sources, where there are *no* accelerating charges.

Let us first ask: "What *is* an opaque screen?" Suppose we have a completely opaque screen between a source S and an observer at P, as in Fig. 31–6(a). If the screen is "opaque" there is no field at P. Why is there no field there? According to the basic principles we should obtain the field at P as the field E_s of the source delayed, plus the field from all the other charges around. But, as we have seen above, the charges in the screen will be set in motion by the field E_s, and these motions generate a new field which, if the screen is opaque, must *exactly cancel* the field E_s on the back side of the screen. You say: "What a miracle that it balances *exactly*! Suppose it was not exactly right!" If it were not exactly right (remember that this opaque screen has some thickness), the field toward the rear part of the screen would not be exactly zero. So, not being zero, it would set into motion some other charges in the material of the screen, and thus make a little more field, trying to get the total balanced out. So if we make the screen thick enough, there is no residual field, because there is enough opportunity to finally get the thing quieted down. In terms of our formulas above we would say that the screen has a large and imaginary index, so the wave is absorbed exponentially as it goes through. You know, of course, that a thin enough sheet of the most opaque material, even gold, is transparent.

Now let us see what happens with an opaque screen which has holes in it, as in Fig. 31–6(b). What do we expect for the field at P? The field at P can be represented as a sum of two parts—the field due to the source S plus the field due to the wall, i.e., due to the motions of the charges in the walls. We might expect the motions of the charges in the walls to be complicated, but we can find out *what fields they produce* in a rather simple way.

Suppose that we were to take the same screen, but plug up the holes, as indicated in part (c) of the figure. We imagine that the plugs are of exactly the same material as the wall. Mind you, the plugs go where the holes were in case (b). Now let us calculate the field at P. The field at P is certainly zero in case (c), but it is *also* equal to the field from the source plus the field due to all the motions of the atoms in the walls and in the plugs. We can write the following equations:

$$\textit{Case (b)}: \quad E_{\text{at }P} = E_s + E_{\text{wall}},$$
$$\textit{Case (c)}: \quad E'_{\text{at }P} = 0 = E_s + E'_{\text{wall}} + E'_{\text{plug}},$$

where the primes refer to the case where the plugs are in place, but E_s is, of course, the same in both cases. Now if we subtract the two equations, we get

$$E_{\text{at }P} = (E_{\text{wall}} - E'_{\text{wall}}) - E'_{\text{plug}}.$$

Now if the holes are not too small (say many wavelengths across), we would not expect the presence of the plugs to change the fields which arrive at the walls except possibly for a little bit around the edges of the holes. Neglecting this small effect, we can set $E_{\text{wall}} = E'_{\text{wall}}$ and obtain that

$$E_{\text{at }P} = -E'_{\text{plug}}.$$

We have the result that the field at P *when there are holes* in a screen (case *b*) is the same (except for sign) as the field that is produced by *that part* of a complete opaque wall which is *located where the holes are!* (The sign is not too interesting, since we are usually interested in intensity which is proportional to the square of the field.) It seems like an amazing backwards-forwards argument. It is, however, not only true (approximately for not too small holes), but useful, and is the justification for the usual theory of diffraction.

The field E'_{plug} is computed in any particular case by remembering that the motion of the charges *everywhere* in the screen is just that which will cancel out the field E_s on the back of the screen. Once we know these motions, we add the radiation fields at P due just to the charges in the plugs.

We remark again that this theory of diffraction is only approximate, and will be good only if the holes are not too small. For holes which are too small the E'_{plug} term will be small and then the difference between E'_{wall} and E_{wall} (which difference we have taken to be zero) may be comparable to or larger than the small E'_{plug} term, and our approximation will no longer be valid.