# 30

The Feynman Lectures on Physics
Addison Wesley, Reading 1963
Vol. 1

## Diffraction



### 30-1 The resultant amplitude due to *n* equal oscillators

This chapter is a direct continuation of the previous one, although the name has been changed from *Interference* to *Diffraction*. No one has ever been able to define the difference between interference and diffraction satisfactorily. It is just a question of usage, and there is no specific, important physical difference between them. The best we can do, roughly speaking, is to say that when there are only a few sources, say two, interfering, then the result is usually called interference, but if there is a large number of them, it seems that the word diffraction is more often used. So, we shall not worry about whether it is interference or diffraction, but continue directly from where we left off in the middle of the subject in the last chapter.

Thus we shall now discuss the situation where there are $n$ equally spaced oscillators, all of equal amplitude but different from one another in phase, either because they are driven differently in phase, or because we are looking at them at an angle such that there is a difference in time delay. For one reason or another, we have to add something like this:

$$R = A[\cos \omega t + \cos(\omega t + \phi) + \cos(\omega t + 2\phi) + \cdots + \cos(\omega t + (n-1)\phi)], \tag{30.1}$$

where $\phi$ is the phase difference between one oscillator and the next one, as seen in a particular direction. Specifically, $\phi = \alpha + 2\pi d \sin\theta/\lambda$. Now we must add all the terms together. We shall do this geometrically. The first one is of length $A$, and it has zero phase. The next is also of length $A$ and it has a phase equal to $\phi$. The next one is again of length $A$ and it has a phase equal to $2\phi$, and so on. So we are evidently going around an equiangular polygon with $n$ sides (Fig. 30-1).

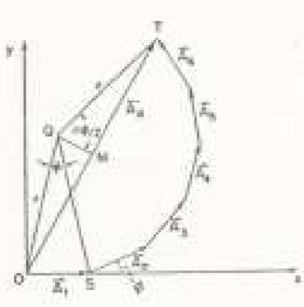


Fig. 30-1. The resultant amplitude of $n = 6$ equally spaced sources with net successive phase differences $\phi$.

Now the vertices, of course, all lie on a circle, and we can find the net amplitude most easily if we find the radius of that circle. Suppose that $Q$ is the center of the circle. Then we know that the angle $OQS$ is just a phase angle $\phi$. (This is because the radius $QS$ bears the same geometrical relation to $\mathbf{A}_2$ as $QO$ bears to $\mathbf{A}_1$, so they form an angle $\phi$ between them.) Therefore the radius $r$ must be such that $A = 2r \sin\phi/2$, which fixes $r$. But the large angle $OQT$ is equal to $n\phi$, and we thus find that $A_R = 2r\sin n\phi/2$. Combining these two results to eliminate $r$, we get

$$A_R = A\frac{\sin n\phi/2}{\sin \phi/2}. \tag{30.2}$$

The resultant intensity is thus

$$I = I_0 \frac{\sin^2 n\phi/2}{\sin^2 \phi/2}. \tag{30.3}$$

Now let us analyze this expression and study some of its consequences. In the first place, we can check it for $n = 1$. It checks: $I = I_0$. Next, we check it for $n = 2$: writing $\sin\phi = 2\sin\phi/2\cos\phi/2$, we find that $A_R = 2A\cos\phi/2$, which agrees with (29.12).

Now the idea that led us to consider the addition of several sources was that we might get a much stronger intensity in one direction than in another; that the nearby maxima which would have been present if there were only two sources will have gone down in strength. In order to see this effect, we plot the curve that comes from (30.3), taking $n$ to be enormously large and plotting the region near

$\phi = 0$. In the first place, if $\phi$ is exactly 0, we have 0/0, but if $\phi$ is infinitesimal, the ratio of the two sines squared is simply $n^2$, since the sine and the angle are approximately equal. Thus the intensity of the maximum of the curve is equal to $n^2$ times the intensity of one oscillator. That is easy to see, because if they are all in phase, then the little vectors have no relative angle and all $n$ of them add up so the amplitude is $n$ times, and the intensity $n^2$ times, stronger.

As the phase $\phi$ increases, the ratio of the two sines begins to fall off, and the first time it reaches zero is when $n\phi/2 = \pi$, because $\sin \pi = 0$. In other words, $\phi = 2\pi/n$ corresponds to the first minimum in the curve (Fig. 30–2). In terms of what is happening with the arrows in Fig. 30–1, the first minimum occurs when all the arrows come back to the starting point; that means that the total accumulated angle in all the arrows, the total phase difference between the first and last oscillator, must be $2\pi$ to complete the circle.

Now we go to the next maximum, and we want to see that it is really much smaller than the first one, as we had hoped. We shall not go precisely to the maximum position, because both the numerator and the denominator of (30.3) are variant, but $\sin \phi/2$ varies quite slowly compared with $\sin n\phi/2$ when $n$ is large, so when $\sin n\phi/2 = 1$ we are very close to the maximum. The next maximum of $\sin^2 n\phi/2$ comes at $n\phi/2 = 3\pi/2$, or $\phi = 3\pi/n$. This corresponds to the arrows having traversed the circle one and a half times. On putting $\phi = 3\pi/n$ into the formula to find the size of the maximum, we find that $\sin^2 3\pi/2 = 1$ in the numerator (because that is why we picked this angle), and in the denominator we have $\sin^2 3\pi/2n$. Now if $n$ is sufficiently large, then this angle is very small and the sine is equal to the angle; so for all practical purposes, we can put $\sin 3\pi/2n = 3\pi/2n$. Thus we find that the intensity at this maximum is $I = I_0(4n^2/9\pi^2)$. But $n^2 I_0$ was the maximum intensity, and so we have $4/9\pi^2$ times the maximum intensity, which is about 0.047, less than 5 percent, of the maximum intensity! Of course there are decreasing intensities farther out. So we have a very sharp central maximum with very weak subsidiary maxima on the sides.

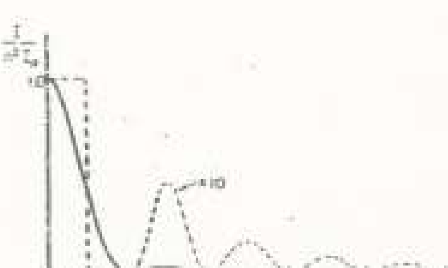

Fig. 30–2. The intensity as a function of phase angle for a large number of oscillators of equal strength.

It is possible to prove that the area of the whole curve, including all the little bumps, is equal to $2\pi n I_0$, or twice the area of the dotted rectangle in Fig. 30–2.

Now let us consider further how we may apply Eq. (30.3) in different circumstances, and try to understand what is happening. Let us consider our sources to be all on a line, as drawn in Fig. 30–3. There are $n$ of them, all spaced by a distance $d$, and we shall suppose that the intrinsic relative phase, one to the next, is $\alpha$. Then if we are observing in a given direction $\theta$ from the normal, there is an additional phase $2\pi d \sin \theta/\lambda$ because of the time delay between each successive two, which we talked about before. Thus

$$\begin{aligned} \phi &= \alpha + 2\pi d \sin \theta/\lambda \\ &= \alpha + kd \sin \theta. \end{aligned} \tag{30.4}$$

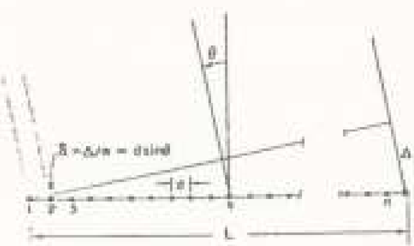

Fig. 30–3. A linear array of $n$ equal oscillators, driven with phases $\alpha_s = s\alpha$.

First, we shall take the case $\alpha = 0$. That is, all oscillators are in phase, and we want to know what the intensity is as a function of the angle $\theta$. In order to find out, we merely have to put $\phi = kd \sin \theta$ into formula (30.3) and see what happens. In the first place, there is a maximum when $\phi = 0$. That means that when all the oscillators are in phase there is a strong intensity in the direction $\theta = 0$. On the other hand, an interesting question is, where is the first minimum? That occurs when $\phi = 2\pi/n$. In other words, when $2\pi d \sin \theta/\lambda = 2\pi/n$, we get the first minumum of the curve. If we get rid of the $2\pi$'s so we can look at it a little better, it says that

$$nd \sin \theta = \lambda. \tag{30.5}$$

Now let us understand physically why we get a minimum at that position. $nd$ is the total length $L$ of the array. Referring to Fig. 30–3, we see that $nd \sin \theta = L \sin \theta = \Delta$. What (30.5) says is that when $\Delta$ is equal to *one wavelength*, we get a minimum. Now why do we get a minimum when $\Delta = \lambda$? Because the contributions of the various oscillators are then uniformly distributed in phase from 0° to

360°. The arrows (Fig. 30–1) are going around a whole circle—we are adding equal vectors in all directions, and such a sum is zero. So when we have an angle such that $\Delta = \lambda$, we get a minimum. That is the first minimum.

There is another important feature about formula (30.3), which is that if the angle $\phi$ is increased by any multiple of $2\pi$, it makes no difference to the formula. So we will get other strong maxima at $\phi = 2\pi, 4\pi, 6\pi$, and so forth. Near each of these great maxima the pattern of Fig. 30–2 is repeated. We may ask ourselves, what is the geometrical circumstance that leads to these other great maxima? The condition is that $\phi = 2\pi m$, where $m$ is any integer. That is, $2\pi d \sin \theta/\lambda = 2\pi m$. Dividing by $2\pi$, we see that

$$d \sin \theta = m\lambda. \tag{30.6}$$

This looks like the other formula, (30.5). No, that formula was $nd \sin \theta = \lambda$. The difference is that here we have to look at the *individual sources*, and when we say $d \sin \theta = m\lambda$, that means that we have an angle $\theta$ such that $\delta = m\lambda$. In other words, each source is now contributing a certain amount, and successive ones are out of phase by a whole multiple of 360°, and therefore are contributing *in phase*, because out of phase by 360° is the same as being in phase. So they all contribute in phase and produce just as good a maximum as the one for $m = 0$ that we discussed before. The subsidiary bumps, the whole shape of the pattern, is just like the one near $\phi = 0$, with exactly the same minima on each side, etc. Thus such an array will send beams in various directions—each beam having a strong central maximum and a certain number of weak "side lobes." The various strong beams are referred to as the zero-order beam, the first-order beam, etc., according to the value of $m$. $m$ is called the *order* of the beam.

We call attention to the fact that if $d$ is less than $\lambda$, Eq. (30.6) can have no solution except $m = 0$, so that if the spacing is too small there is only one possible beam, the zero-order one centered at $\theta = 0$. (Of course, there is also a beam in the opposite direction.) In order to get subsidiary great maxima, we must have the spacing $d$ of the array greater than one wavelength.

## 30–2 The diffraction grating

In technical work with antennas and wires it is possible to arrange that all the phases of the little oscillators, or antennas, are equal. The question is whether and how we can do a similar thing with light. We cannot at the present time literally make little optical-frequency radio stations and hook them up with infinitesimal wires and drive them all with a given phase. But there is a very easy way to do what amounts to the same thing.

Suppose that we had a lot of parallel wires, equally spaced at a spacing $d$, and a radiofrequency source very far away, practically at infinity, which is generating an electric field which arrives at each one of the wires at the same phase (it is so far away that the time delay is the same for all of the wires). (One can work out cases with curved arrays, but let us take a plane one.) Then the external electric field will drive the electrons up and down in each wire. That is, the field which is coming from the original source will shake the electrons up and down, and in moving, these represent *new generators*. This phenomenon is called scattering: a light wave from some source can induce a motion of the electrons in a piece of material, and these motions generate their own waves. Therefore all that is necessary is to set up a lot of wires, equally spaced, drive them with a radiofrequency source far away, and we have the situation that we want, without a whole lot of special wiring. If the incidence is normal, the phases will be equal, and we will get exactly the circumstance we have been discussing. Therefore, if the wire spacing is greater than the wavelength, we will get a strong intensity of scattering in the normal direction, and in certain other directions given by (30.6).

*This can also be done with light!* Instead of wires, we use a flat piece of glass and make notches in it such that each of the notches scatters a little differently than the rest of the glass. If we then shine light on the glass, each one of the notches will represent a source, and if we space the lines very finely, but not closer than a wavelength (which is technically almost impossible anyway), then we would expect a miraculous phenomenon: the light not only will pass straight through, but there will also be a strong beam at a finite angle, depending on the spacing of the notches! Such objects have actually been made and are in common use—they are called *diffraction gratings.*

In one of its forms, a diffraction grating consists of nothing but a plane glass sheet, transparent and colorless, with scratches on it. There are often several hundred scratches to the millimeter, *very* carefully arranged so as to be equally spaced. The effect of such a grating can be seen by arranging a projector so as to throw a narrow, vertical line of light (the image of a slit) onto a screen. When we put the grating into the beam, with its scratches vertical, we see that the line is still there but, in addition, on each side we have *another* strong patch of light which is *colored.* This, of course, is the slit image spread out over a wide angular range, because the angle $\theta$ in (30.6) depends upon $\lambda$, and lights of different colors, as we know, correspond to different frequencies, and therefore different wavelengths. The longest visible wavelength is red, and since $d \sin \theta = \lambda$, that requires a larger $\theta$. And we do, in fact, find that red is at a greater angle out from the central image! There should also be a beam on the other side, and indeed we see one on the screen. Then, there might be another solution of (30.6) when $m = 2$. We do see that there is something vaguely there—very weak—and there are even other beams beyond.

We have just argued that all these beams ought to be of the same strength, but we see that they actually are not and, in fact, not even the first ones on the right and left are equal! The reason is that the grating has been carefully built to do just this. How? If the grating consists of very fine notches, infinitesimally wide, spaced evenly, then all the intensities would indeed be equal. But, as a matter of fact, although we have taken the simplest case, we could also have considered an array of *pairs* of antennas, in which each member of the pair has a certain strength and some relative phase. In this case, it is possible to get intensities which are different in the different orders. A grating is often made with little "sawtooth" cuts instead of little symmetrical notches. By carefully arranging the "sawteeth," more light may be sent into one particular order of spectrum than into the others. In a practical grating, we would like to have as much light as possible in one of the orders. This may seem a complicated point to bring in, but it is a very clever thing to do, because it makes the grating more useful.

So far, we have taken the case where all the phases of the sources are equal. But we also have a formula for $\phi$ when the phases differ from one to the next by an angle $\alpha$. That requires wiring up our antennas with a slight phase shift between each one. Can we do that with light? Yes, we can do it very easily, for suppose that there were a source of light at infinity, *at an angle* such that the light is coming in at an angle $\theta_{in}$, and let us say that we wish to discuss the scattered beam, which is leaving at an angle $\theta_{out}$. The $\theta_{out}$ is the same $\theta$ as we have had before, but the $\theta_{in}$ is merely a means for arranging that the phase of each source is different: the light coming from the distant driving source first hits one scratch, then the next, then the next, and so on, with a phase shift from one to the other, which, as we see, is $\alpha = -d \sin \theta_{in}/\lambda$. Therefore we have the formula for a grating in which light both comes in and goes out at an angle:

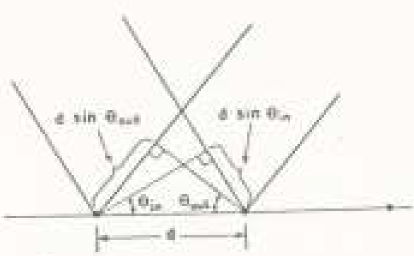


Fig. 30-4. The path difference for rays scattered from adjacent rulings of a grating is $d \sin \theta_{out} - d \sin \theta_{in}$.

$$\phi = 2\pi d \sin \theta_{out}/\lambda - 2\pi d \sin \theta_{in}/\lambda. \tag{30.7}$$

Let us try to find out where we get strong intensity in these circumstances. The condition for strong intensities is, of course, that $\phi$ should be a multiple of $2\pi$. There are several interesting points to be noted.

One case of rather great interest is that which corresponds to $m = 0$, where $d$ is less than $\lambda$; in fact, this is the only solution. In this case we see that $\sin \theta_{out} = \sin \theta_{in}$, which means that the light comes out in the *same direction* as the light which was exciting the grating. We might think that the light "goes right through." No, it is *different light* that we are talking about. The light that goes right through is from the original source; what we are talking about is the new light *which is* *generated by scattering*. It turns out that the scattered light is going in the same direction as the original light, in fact it can interfere with it—a feature which we will study later.

There is another solution for this same case. For a given $\theta_{in}$, $\theta_{out}$ may be the *supplement* of $\theta_{in}$. So not only do we get a beam in the same direction as the incoming beam but also one in another direction, which, if we consider it carefully, is such that the *angle of incidence is equal to the angle of scattering*. This we call the *reflected beam*.

So we begin to understand the basic machinery of reflection: the light that comes in generates motions of the atoms in the reflector, and the reflector then regenerates a *new wave*, and one of the solutions for the direction of scattering, the *only* solution if the spacing of the scatterers is small compared with one wavelength, is that the angle at which the light comes out is equal to the angle at which it comes in!

Next, we discuss the special case when $d \to 0$. That is, we have just a solid piece of material, so to speak, but of finite length. In addition, we want the phase shift from one scatterer to the next to go to zero. In other words, we put more and more antennas between the other ones, so that each of the phase differences is getting smaller, but the number of antennas is increasing in such a way that the total phase difference, between one end of the line and the other, is constant. Let us see what happens to (30.3) if we keep the difference in phase $n\phi$ from one end to the other constant (say $n\phi = \Phi$), letting the number go to infinity and the phase shift $\phi$ of each one go to zero. But now $\phi$ is so small that $\sin \phi = \phi$, and if we also recognize $n^2 I_0$ as $I_m$, the maximum intensity at the center of the beam, we find

$$I = 4I_m \sin^2 \tfrac{1}{2}\Phi/\Phi^2. \tag{30.8}$$

This limiting case is what is shown in Fig. 30-2.

In such circumstances we find the same general kind of a picture as for finite spacing with $d > \lambda$; all the side lobes are practically the same as before, but there are no higher-order maxima. If the scatterers are all in phase, we get a maximum in the direction $\theta_{out} = 0$, and a minimum when the distance $\Delta$ is equal to $\lambda$, just as for finite $d$ and $n$. So we can even analyze a *continuous* distribution of scatterers or oscillators, by using integrals instead of summing.

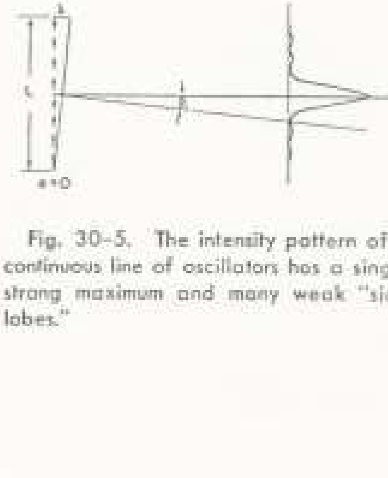

Fig. 30-5. The intensity pattern of a continuous line of oscillators has a single strong maximum and many weak "side lobes."

As an example, suppose there were a long line of oscillators, with the charge oscillating along the direction of the line (Fig. 30-5). From such an array the greatest intensity is perpendicular to the line. There is a little bit of intensity up and down from the equatorial plane, but it is very slight. With this result, we can handle a more complicated situation. Suppose we have a set of such lines, each producing a beam only in a plane perpendicular to the line. To find the intensity in various directions from a series of long wires, instead of infinitesimal wires, is the same problem as it was for infinitesimal wires, so long as we are in the central plane perpendicular to the wires; we just add the contribution from each of the long wires. That is why, although we actually analyzed only tiny antennas, we might as well have used a grating with long, narrow slots. Each of the long slots produces an effect only in its own direction, not up and down, but they are all set next to each other horizontally, so they produce interference that way.

Thus we can build up more complicated situations by having various distributions of scatterers in lines, planes, or in space. The first thing we did was to consider scatterers in a line, and we have just extended the analysis to strips; we can work it out by just doing the necessary summations, adding the contributions from the individual scatterers. The principle is always the same.

## 30-3 Resolving power of a grating

We are now in a position to understand a number of interesting phenomena. For example, consider the use of a grating for separating wavelengths. We noticed that the whole spectrum was spread out on the screen, so a grating can be used as an instrument for separating light into its different wavelengths. One of the interesting questions is: supposing that there were two sources of slightly different frequency, or slightly different wavelength, how close together in wavelength could they be such that the grating would be unable to tell that there were really two different wavelengths there? The red and the blue were clearly separated. But when one wave is red and the other is slightly redder, very close, how close can they be? This is called the *resolving power* of the grating, and one way of analyzing the problem is as follows. Suppose that for light of a certain color we happen to have the maximum of the diffracted beam occurring at a certain angle. If we vary the wavelength the phase $2\pi d \sin\theta/\lambda$ is different, so of course the maximum occurs at a different angle. That is why the red and blue are spread out. How different in angle must it be in order for us to be able to see it? If the two maxima are exactly on top of each other, of course we cannot see them. If the maximum of one is far enough away from the other, then we can see that there is a double bump in the distribution of light. In order to be able to just make out the double bump, the following simple criterion, called *Rayleigh's criterion*, is usually used (Fig. 30-6). It is that the first minimum from one bump should sit at the maximum of the other. Now it is very easy to calculate, when one minimum sits on the other maximum, how much the difference in wavelength is. The best way to do it is geometrically.

In order to have a maximum for wavelength $\lambda'$, the distance $\Delta$ (Fig. 30-3) must be $n\lambda'$, and if we are looking at the $m$th-order beam, it is $mn\lambda'$. In other words, $2\pi d\sin\theta/\lambda' = 2\pi m$, so $nd \sin\theta$, which is $\Delta$, is $\lambda'$ times $n$, or $mn\lambda'$. For the other beam, of wavelength $\lambda$, we want to have a *minimum* at this angle. That is, we want $\Delta$ to be exactly one wavelength $\lambda$ more than $mn\lambda$. That is, $\Delta = mn\lambda + \lambda = mn\lambda'$. Thus if $\lambda' = \lambda + \Delta\lambda$, we find

$$\Delta\lambda/\lambda = 1/mn. \tag{30.9}$$

Fig. 30-6. Illustration of the Rayleigh criterion. The maximum of one pattern falls on the first minimum of the other.

The ratio $\lambda/\Delta\lambda$ is called the *resolving power* of a grating; we see that it is equal to the total number of lines in the grating, times the order. It is not hard to prove that this formula is equivalent to the formula that the error in frequency is equal to the reciprocal time difference between extreme paths that are allowed to interfere:*

$$\Delta\nu = 1/T.$$

In fact, that is the best way to remember it, because the general formula works not only for gratings, but for any other instrument whatsoever, while the special formula (30.9) depends on the fact that we are using a grating.

### 30-4 The parabolic antenna

Now let us consider another problem in resolving power. This has to do with the antenna of a radio telescope, used for determining the position of radio sources in the sky, i.e., how large they are in angle. Of course if we use any old antenna and find signals, we would not know from what direction they came. We are very interested to know whether the source is in one place or another. One way we can find out is to lay out a whole series of equally spaced dipole wires on the Australian landscape. Then we take all the wires from these antennas and feed them into the same receiver, in such a way that all the delays in the feed lines are equal. Thus the receiver receives signals from all of the dipoles in phase. That is, it adds all the waves from every one of the dipoles in the same phase. Now what happens? If the source is directly above the array, at infinity or nearly so, then its radiowaves will excite all the antennas in the same phase, so they all feed the receiver together.

Now suppose that the radio source is at a slight angle $\theta$ from the vertical. Then the various antennas are receiving signals a little out of phase. The receiver adds all these out-of-phase signals together, and so we get nothing, if the angle

* In our case $T = \Delta/c = mn\lambda'/c$, where $c$ is the speed of light. The frequency $\nu = c/\lambda$, so $\Delta\nu = c\Delta\lambda/\lambda^2$.

$\theta$ is too big. How big may the angle be? *Answer:* we get zero if the angle $\Delta/L = \theta$ (Fig. 30-3) corresponds to a 360° phase shift, that is, if $\Delta$ is the wavelength $\lambda$. This is because the vector contributions form together a complete polygon with zero resultant. The smallest angle that can be resolved by an antenna array of length $L$ is $\theta = \lambda/L$. Notice that the receiving pattern of an antenna such as this is exactly the same as the intensity distribution we would get if we turned the receiver around and made it into a transmitter. This is an example of what is called a *reciprocity principle*. It turns out, in fact, to be generally true for any arrangement of antennas, angles, and so on, that if we first work out what the relative intensities would be in various directions if the receiver were a transmitter instead, then the relative directional sensitivity of a receiver with the same external wiring, the same array of antennas, is the same as the relative intensity of emission would be if it were a transmitter.

Some radio antennas are made in a different way. Instead of having a whole lot of dipoles in a long line, with a lot of feed wires, we may arrange them not in a line but in a curve, and put the receiver at a certain point where it can detect the scattered waves. This curve is cleverly designed so that if the radiowaves are coming down from above, and the wires scatter, making a new wave, the wires are so arranged that the scattered waves reach the receiver all at the same time (Fig. 26-12). In other words, the curve is a *parabola*, and when the source is exactly on its axis, we get a very strong intensity at the focus. In this case we understand very clearly what the resolving power of such an instrument is. The arranging of the antennas on a parabolic curve is not an essential point. It is only a convenient way to get all the signals to the same point with no relative delay and without feed wires. The angle such an instrument can resolve is still $\theta = \lambda/L$, where $L$ is the separation of the first and last antennas. It does not depend on the spacing of the antennas and they may be very close together or in fact be all one piece of metal. Now we are describing a telescope mirror, of course. We have found the resolving power of a telescope! (Sometimes the resolving power is written $\theta = 1.22\lambda/L$, where $L$ is the diameter of the telescope. The reason that it is not exactly $\lambda/L$ is this: when we worked out that $\theta = \lambda/L$, we assumed that all the lines of dipoles were equal in strength, but when we have a circular telescope, which is the way we usually arrange a telescope, not as much signal comes from the outside edges, because it is not like a square, where we get the same intensity all along a side. We get somewhat less because we are using only part of the telescope there; thus we can appreciate that the effective diameter is a little shorter than the true diameter, and that is what the 1.22 factor tells us. In any case, it seems a little pedantic to put such precision into the resolving power formula.*)

### 30-5 Colored films; crystals

The above, then, are some of the effects of interference obtained by adding the various waves. But there are a number of other examples, and even though we do not understand the fundamental mechanism yet, we will some day, and we can understand even now how the interference occurs. For example, when a light wave hits a surface of a material with an index $n$, let us say at normal incidence, some of the light is reflected. The *reason* for the reflection we are not in a position to understand right now; we shall discuss it later. But suppose we know that some of the light is reflected both on entering and leaving a refracting medium. Then, if we look at the reflection of a light source in a thin film, we see the sum of two waves; if the thicknesses are small enough, these two waves will produce an interference, either constructive or destructive, depending on the signs of the phases. It might be, for instance, that for red light, we get an enhanced reflection, but for

* This is because Rayleigh's criterion is a rough idea in the first place. It tells you where it begins to get very hard to tell whether the image was made by one or by two stars. Actually, if sufficiently careful measurements of the exact intensity distribution over the diffracted image spot can be made, the fact that two sources make the spot can be proved even if $\theta$ is less than $\lambda/L$.

blue light, which has a different wavelength, perhaps we get a destructively interfering reflection, so that we see a bright red reflection. If we change the thickness, i.e., if we look at another place where the film is thicker, it may be reversed, the red interfering and the blue not, so it is bright blue, or green, or yellow, or whatnot. So we see *colors* when we look at thin films and the colors change if we look at different angles, because we can appreciate that the timings are different at different angles. Thus we suddenly appreciate another hundred thousand situations involving the colors that we see on oil fims, soap bubbles, etc. at different angles. But the principle is all the same: we are only adding waves at different phases.

As another important application of diffraction, we may mention the following. We used a grating and we saw the diffracted image on the screen. If we had used monochromatic light, it would have been at a certain specific place. Then there were various higher-order images also. From the positions of the images, we could tell how far apart the lines on the grating were, if we knew the wavelength of the light. From the difference in intensity of the various images, we could find out the shape of the grating scratches, whether the grating was made of wires, sawtooth notches, or whatever, *without being able to see them*. This principle is used to discover the positions of the *atoms in a crystal*. The only complication is that a crystal is three-dimensional; it is a repeating three-dimensional array of atoms. We cannot use ordinary light, because we must use something whose wavelength is less than the space between the atoms or we get no effect; so we must use radiation of very short wavelength, i.e., x-rays. So, by shining x-rays into a crystal and by noticing how intense is the reflection in the various orders, we can determine the arrangement of the atoms inside without ever being able to see them with the eye! It is in this way that we know the arrangement of the atoms in various substances, which permitted us to draw those pictures in the first chapter, showing the arrangement of atoms in salt, and so on. We shall later come back to this subject and discuss it in more detail, and therefore we say no more about this most remarkable idea at present.

### 30-6 Diffraction by opaque screens

Now we come to a very interesting situation. Suppose that we have an opaque sheet with holes in it, and a light on one side of it. We want to know what the intensity is on the other side. What most people say is that the light shines through the holes, and produces an effect on the other side. It will turn out that one gets the right answer, to an excellent approximation, if he assumes that there are sources distributed with uniform density across the open holes, and that the phases of these sources are the same as they would have been if the opaque material were absent. Of course, actually there are *no* sources at the holes, in fact that is the only place that there are *certainly* no sources. Nevertheless, we get the correct diffraction patterns by considering the holes to be the only places that there *are* sources; that is a rather peculiar fact. We shall explain later why this is true, but for now let us just suppose that it is.

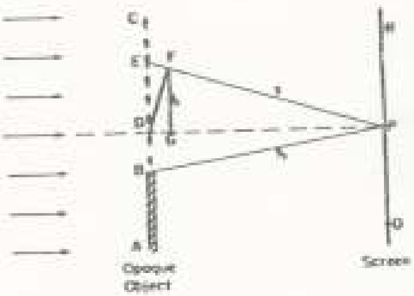


Fig. 30-7. A distant light source casts a shadow of an opaque object on a screen.

In the theory of diffraction there is another kind of diffraction that we should briefly discuss. It is usually not discussed in an elementary course as early as this, only because the mathematical formulas involved in adding these little vectors are a little elaborate. Otherwise it is exactly the same as we have been doing all along. All the interference phenomena are the same; there is nothing very much more advanced involved, only the circumstances are more complicated and it is harder to add the vectors together, that is all.

Suppose that we have light coming in from infinity, casting a shadow of an object. Figure 30-7 shows a screen on which the shadow of an object $AB$ is made by a light source very far away compared with one wavelength. Now we would expect that outside the shadow, the intensity is all bright, and inside it, it is all dark. As a matter of fact, if we plot the intensity as a function of position near the shadow edge, the intensity rises and then overshoots, and wobbles, and oscillates about in a very peculiar manner near this edge (Fig. 30-8). We now shall discuss the reason for this. If we use the theorem that we have not yet proved, then we can replace the actual problem by a set of effective sources uniformly distributed over the open space beyond the object.

We imagine a large number of very closely spaced antennas, and we want the intensity at some point $P$. That looks just like what we have been doing. Not quite; because our screen is not at infinity. We do not want the intensity at infinity, but at a finite point. To calculate the intensity at some particular place, we have to add the contributions from all the antennas. First there is an antenna at $D$, exactly opposite $P$; if we go up a little bit in angle, let us say a height $h$, then there is an increase in delay (there is also a change in amplitude because of the change in distance, but this is a very small effect if we are at all far away, and is much less important than the difference in the phases). Now the path difference $EP - DP$ is $h^2/2s$, so that the phase difference is proportional to the *square* of how far we go from $D$, while in our previous work $s$ was infinite, and the phase difference was *linearly* proportional to $h$. When the phases are linearly proportional, each vector adds at a constant angle to the next vector. What we now need is a curve which is made by adding a lot of infinitesimal vectors with the requirement that the angle they make shall increase, not linearly, but as the *square* of the length of the curve. To construct that curve involves slightly advanced mathematics, but we can always construct it by actually drawing the arrows and measuring the angles. In any case, we get the marvelous curve (called Cornu's spiral) shown in Fig. 30-8. Now how do we use this curve?

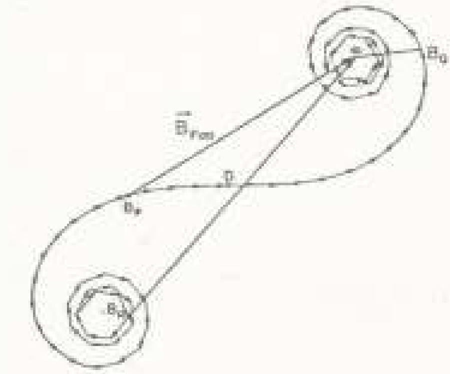


Fig. 30-8. The addition of amplitudes for many in-phase oscillators whose phase delays vary as the square of the distance from point D of the previous figure.

If we want the intensity, let us say, at point $P$, we add a lot of contributions of different phases from point $D$ on up to infinity, and from $D$ down only to point $B_P$. So we start at $B_P$ in Fig. 30-8, and draw a series of arrows of ever-increasing angle. Therefore the total contribution above point $B_P$ all goes along the spiraling curve. If we were to stop integrating at some place, then the total amplitude would be a vector from $B$ to that point; in this particular problem we are going to infinity, so the total answer is the vector $\mathbf{B}_{P\infty}$. Now the position on the curve which corresponds to point $B_P$ on the object depends upon where point $P$ is located, since point $D$, the inflection point, always corresponds to the position of point $P$. Thus, depending upon where $P$ is located above $B$, the beginning point will fall at various positions on the lower left part of the curve, and the resultant vector $\mathbf{B}_{P\infty}$ will have many maxima and minima (Fig. 30-9).

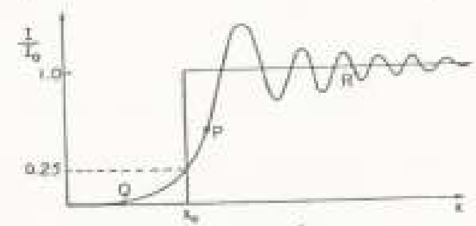


Fig. 30-9. The intensity near the edge of a shadow. The geometrical shadow edge is at $x_0$.

On the other hand, if we are at $Q$, on the other side of $P$, then we are using only one end of the spiral curve, and not the other end. In other words, we do not even start at $D$, but at $B_Q$, so on this side we get an intensity which continuously falls off as $Q$ goes farther into the shadow.

One point that we can immediately calculate with ease, to show that we really understand it, is the intensity exactly opposite the edge. The intensity here is 1/4 that of the incident light. Reason: Exactly at the edge (so the endpoint $B$ of the arrow is at $D$ in Fig. 30-8) we have half the curve that we would have had if we were far into the bright region. If our point $R$ is far into the light we go from one end of the curve to the other, that is, one full unit vector; but if we are at the edge of the shadow, we have only half the amplitude—1/4 the intensity.

In this chapter we have been finding the intensity produced in various directions from various distributions of sources. As a final example we shall derive a formula which we shall need for the next chapter on the theory of the index of refraction. Up to this point relative intensities have been sufficient for our purpose, but this time we shall find the complete formula for the field in the following situation.

## 30–7 The field of a plane of oscillating charges

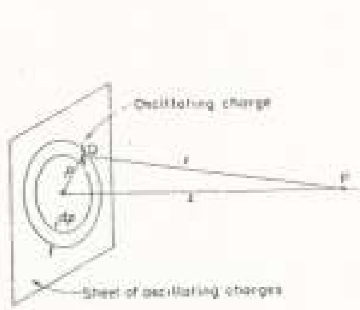


Fig. 30–10. Radiation field of a sheet of oscillating charges.

Suppose that we have a plane full of sources, all oscillating together, with their motion in the plane and all having the same amplitude and phase. What is the field at a finite, but very large, distance away from the plane? (We cannot get very close, of course, because we do not have the right formulas for the field close to the sources.) If we let the plane of the charges be the $XY$-plane, then we want to find the field at the point $P$ far out on the $Z$-axis (Fig. 30–10). We suppose that there are $\eta$ charges per unit area of the plane, and that each one of them has a charge $q$. All of the charges move with simple harmonic motion, with the same direction, amplitude, and phase. We let the motion of each charge, *with respect to its own average position*, be $x_0 \cos \omega t$. Or, using the complex notation and remembering that the real part represents the actual motion, the motion can be described by $x_0 e^{i\omega t}$.

Now we find the field at the point $P$ from all of the charges by finding the field there from each charge $q$, and then adding the contributions from all the charges. We know that the radiation field is proportional to the acceleration of the charge, which is $-\omega^2 x_0 e^{i\omega t}$ (and is the same for every charge). The electric field that we want at the point $P$ due to a charge at the point $Q$ is proportional to the acceleration of the charge $q$, but we have to remember that the field at the point $P$ at the instant $t$ is given by the acceleration of the charge at the earlier time $t' = t - r/c$, where $r/c$ is the time it takes the waves to travel the distance $r$ from $Q$ to $P$. Therefore the field at $P$ is proportional to

$$-\omega^2 x_0 e^{i\omega(t-r/c)}. \tag{30.10}$$

Using this value for the acceleration as seen from $P$ in our formula for the electric field at large distances from a radiating charge, we get

$$\begin{pmatrix}\text{Electric field at } P\\ \text{from charge at } Q\end{pmatrix} = \frac{q}{4\pi\epsilon_0 c^2}\,\frac{\omega^2 x_0 e^{i\omega(t-r/c)}}{r} \text{ (approx.).} \tag{30.11}$$

Now this formula is not quite right, because we should have used *not* the acceleration of the charge but *its component* perpendicular to the line $QP$. We shall suppose, however, that the point $P$ is so far away, compared with the distance of the point $Q$ from the axis (the distance $\rho$ in Fig. 30–9), for those charges that we need to take into account, that we can leave out the cosine factor (which would be nearly equal to 1 anyway).

To get the total field at $P$, we now add the effects of all the charges in the plane. We should, of course, make a *vector* sum. But since the direction of the electric field is nearly the same for all the charges, we may, in keeping with the approximation we have already made, just add the magnitudes of the fields. To our approximation the field at $P$ depends only on the distance $r$, so all charges at the same $r$ produce equal fields. So we add, first, the fields of those charges in a ring of width $d\rho$ and radius $\rho$. Then, by taking the integral over all $\rho$, we will obtain the total field.

The number of charges in the ring is the product of the surface area of the ring, $2\pi\rho\, d\rho$, and $\eta$, the number of charges per unit area. We have, then,

$$\text{Total field at } P = \int \frac{q}{4\pi\epsilon_0 c^2}\,\frac{\omega^2 x_0 e^{i\omega(t-r/c)}}{r}\cdot \eta \cdot 2\pi\rho\, d\rho. \tag{30.12}$$

We wish to evaluate this integral from $\rho = 0$ to $\rho = \infty$. The variable $t$, of course, is to be held fixed while we do the integral, so the only varying quantities are $\rho$ and $r$. Leaving out all the constant factors, *including the factor* $e^{i\omega t}$, for the moment, the integral we wish is

$$\int_{\rho=0}^{\rho=\infty} \frac{e^{-i\omega r/c}}{r}\,\rho\, d\rho. \tag{30.13}$$

To do this integral we need to use the relation between $r$ and $\rho$:

$$r^2 = \rho^2 + z^2. \tag{30.14}$$

Since $z$ is independent of $\rho$, when we take the differential of this equation, we get

$$2r\,dr = 2\rho\,d\rho,$$

which is lucky, since in our integral we can replace $\rho\,d\rho$ by $r\,dr$ and the $r$ will cancel the one in the denominator. The integral we want is then the simpler one

$$\int_{r=z}^{r=\infty} e^{-i\omega r/c}\,dr. \tag{30.15}$$

To integrate an exponential is very easy. We divide by the coefficient of $r$ in the exponent and evaluate the exponential at the limits. But the limits of $r$ are not the same as the limits of $\rho$. When $\rho = 0$, we have $r = z$, so the limits of $r$ are $z$ to infinity. We get for the integral

$$-\frac{c}{i\omega}\left[e^{-i\infty} - e^{-(i\omega/c)z}\right], \tag{30.16}$$

where we have written $\infty$ for $(\omega/c)\infty$, since they both just mean a very large number!

Now $e^{-i\infty}$ is a mysterious quantity. Its real part, for example, is $\cos(-\infty)$, which, mathematically speaking, is completely indefinite (although we would expect it to be somewhere—or everywhere (?)—between $+1$ and $-1$!). But in a *physical* situation, it can mean something quite reasonable, and usually can just be taken to be zero. To see that this is so in our case, we go back to consider again the original integral (30.15).

We can understand (30.15) as a sum of many small complex numbers, each of magnitude $\Delta r$, and with the angle $\theta = -\omega r/c$ in the complex plane. We can try to evaluate the sum by a graphical method. In Fig. 30-11 we have drawn the first five pieces of the sum. Each segment of the curve has the length $\Delta r$ and is placed at the angle $\Delta\theta = -\omega\,\Delta r/c$ with respect to the preceding piece. The sum for these first five pieces is represented by the arrow from the starting point to the end of the fifth segment. As we continue to add pieces we shall trace out a polygon until we get back to the starting point (approximately) and then start around once more. Adding more pieces, we just go round and round, staying close to a circle whose radius is easily shown to be $c/\omega$. We can see now why the integral does not give a definite answer!

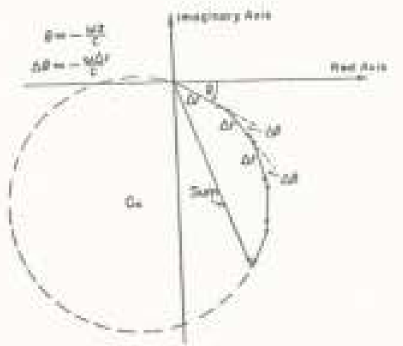


Fig. 30-11. Graphical solution of $\int_z^\infty e^{-i\omega r/c}dr$.

But now we have to go back to the *physics* of the situation. In any real situation the plane of charges *cannot* be infinite in extent, but must sometime stop. If it stopped suddenly, and was exactly circular in shape, our integral would have some value on the circle in Fig. 30-11. If, however, we let the number of charges in the plane gradually taper off at some large distance from the center (or else stop suddenly but in an irregular shape so for larger $\rho$ the entire ring of width $d\rho$ no longer contributes), then the coefficient $\eta$ in the exact integral would decrease toward zero. Since we are adding smaller pieces but still turning through the same angle, the graph of our integral would then become a curve which is a spiral. The spiral would eventually end up at the center of our original circle, as drawn in Fig. 30-12. The *physically* correct integral is the complex number $A$ in the figure represented by the interval from the starting point to the center of the circle, which is just equal to

$$\frac{c}{i\omega}e^{-i\omega z/c}, \tag{30.17}$$

as you can work out for yourself. This is the same result we would get from Eq. (30.16) if we set $e^{-i\infty} = 0$.

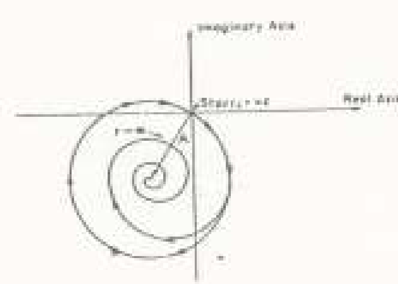


Fig. 30-12. Graphical solution of $\int_z^\infty \eta e^{-i\omega r/c}dr$.

(There is also another reason why the contribution to the integral tapers off for large values of $r$, and that is the factor we have omitted for the projection of the acceleration on the plane perpendicular to the line $PQ$.)

We are, of course, interested only in physical situations, so we will take $e^{-i\infty}$ equal to zero. Returning to our original formula (30.12) for the field and putting back all of the factors that go with the integral, we have the result

$$\text{Total field at } P = -\frac{\eta q}{2\epsilon_0 c}\, i\omega x_0 e^{i\omega(t - z/c)} \tag{30.18}$$

(remembering that $1/i = -i$).

It is interesting to note that $(i\omega x_0 e^{i\omega t})$ is just equal to the *velocity* of the charges, so that we can also write the equation for the field as

$$\text{Total field at } P = -\frac{\eta q}{2\epsilon_0 c}\,[\text{velocity of charges}]_{\text{at } t - z/c} \tag{30.19}$$

which is a little strange, because the retardation is just by the distance $z$, which is the shortest distance from $P$ to the plane of charges. But that is the way it comes out—fortunately a rather simple formula. (We may add, by the way, that although our derivation is valid only for distances far from the plane of oscillatory charges, it turns out that the formula (30.18) or (30.19) is correct at any distance $z$, even for $z < \lambda$.)