

Beats

48-1 Adding two waves

Some time ago we discussed in considerable detail the properties of light waves and their interference—that is, the effects of the superposition of two waves from different sources. In all these analyses we assumed that the frequencies of the sources were all the same. In this chapter we shall discuss some of the phenomena which result from the interference of two sources which have *different* frequencies.

It is easy to guess what is going to happen. Proceeding in the same way as we have done previously, suppose we have two equal oscillating sources of the same frequency whose phases are so adjusted, say, that the signals arrive in phase at some point P . At that point, if it is light, the light is very strong; if it is sound, it is very loud; or if it is electrons, many of them arrive. On the other hand, if the arriving signals were 180° out of phase, we would get no signal at P , because the net amplitude there is then a minimum. Now suppose that someone twists the “phase knob” of one of the sources and changes the phase at P back and forth, say, first making it 0° and then 180° , and so on. Of course, we would then find variations in the net signal strength. Now we also see that if the phase of one source is slowly changing relative to that of the other in a gradual, uniform manner, starting at zero, going up to ten, twenty, thirty, forty degrees, and so on, then what we would measure at P would be a series of strong and weak “pulsations,” because when the phase shifts through 360° the amplitude returns to a maximum. Of course, to say that one source is shifting its phase relative to another at a uniform rate is the same as saying that the number of oscillations per second is slightly different for the two.

So we know the answer: if we have two sources at slightly different frequencies we should find, as a net result, an oscillation with a slowly pulsating intensity. That is all there really is to the subject!

It is very easy to formulate this result mathematically also. Suppose, for example, that we have two waves, and that we do not worry for the moment about all the spatial relations, but simply analyze what arrives at P . From one source, let us say, we would have $\cos \omega_1 t$, and from the other source, $\cos \omega_2 t$, where the two ω 's are not exactly the same. Of course the amplitudes may not be the same, either, but we can solve the general problem later; let us first take the case where the amplitudes are equal. Then the total amplitude at P is the sum of these two cosines. If we plot the amplitudes of the waves against the time, as in Fig. 48-1, we see that

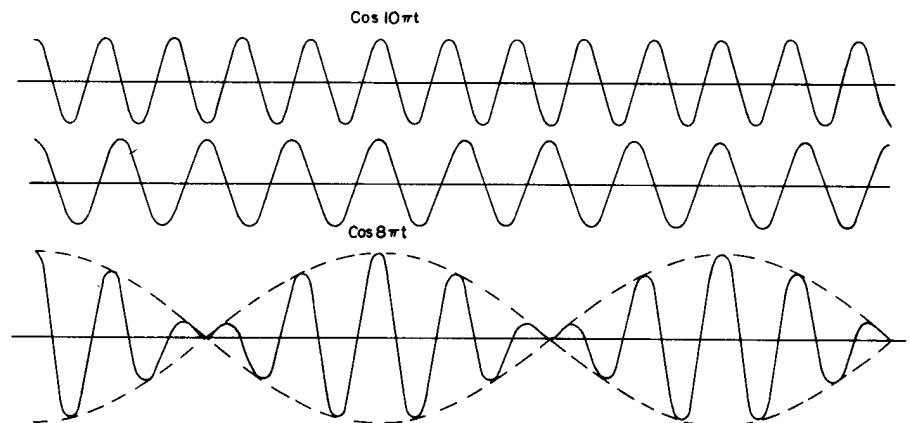


Fig. 48-1. The superposition of two cosine waves with frequencies in the ratio 8:10. The precise repetition of the pattern within each “beat” is not typical of the general case.

48-1 Adding two waves

48-2 Beat notes and modulation

48-3 Side bands

48-4 Localized wave trains

48-5 Probability amplitudes for particles

48-6 Waves in three dimensions

48-7 Normal modes

where the crests coincide we get a strong wave, and where a trough and crest coincide we get practically zero, and then when the crests coincide again we get a strong wave again.

Mathematically, we need only to add two cosines and rearrange the result somehow. There exist a number of useful relations among cosines which are not difficult to derive. Of course we know that

$$e^{i(a+b)} = e^{ia}e^{ib}, \quad (48.1)$$

and that e^{ia} has a real part, $\cos a$, and an imaginary part, $i \sin a$. If we take the real part of $e^{i(a+b)}$, we get $\cos(a+b)$. If we multiply out:

$$e^{ia}e^{ib} = (\cos a + i \sin a)(\cos b + i \sin b),$$

we get $\cos a \cos b - \sin a \sin b$, plus some imaginary parts. But we now need only the real part, so we have

$$\cos(a+b) = \cos a \cos b - \sin a \sin b. \quad (48.2)$$

Now if we change the sign of b , since the cosine does not change sign while the sine does, the same equation, for negative b , is

$$\cos(a-b) = \cos a \cos b + \sin a \sin b. \quad (48.3)$$

If we add these two equations together, we lose the sines and we learn that the product of two cosines is half the cosine of the sum, plus half the cosine of the difference:

$$\cos a \cos b = \frac{1}{2} \cos(a+b) + \frac{1}{2} \cos(a-b). \quad (48.4)$$

Now we can also reverse the formula and find a formula for $\cos \alpha + \cos \beta$ if we simply let $\alpha = a+b$ and $\beta = a-b$. That is, $a = \frac{1}{2}(\alpha + \beta)$ and $b = \frac{1}{2}(\alpha - \beta)$, so that

$$\cos \alpha + \cos \beta = 2 \cos \frac{1}{2}(\alpha + \beta) \cos \frac{1}{2}(\alpha - \beta). \quad (48.5)$$

Now we can analyze our problem. The sum of $\cos \omega_1 t$ and $\cos \omega_2 t$ is

$$\cos \omega_1 t + \cos \omega_2 t = 2 \cos \frac{1}{2}(\omega_1 + \omega_2)t \cos \frac{1}{2}(\omega_1 - \omega_2)t. \quad (48.6)$$

Now let us suppose that the two frequencies are nearly the same, so that $\frac{1}{2}(\omega_1 + \omega_2)$ is the average frequency, and is more or less the same as either. But $\omega_1 - \omega_2$ is *much smaller* than ω_1 or ω_2 because, as we suppose, ω_1 and ω_2 are nearly equal. That means that we can represent the solution by saying that there is a high-frequency cosine wave more or less like the ones we started with, but that its "size" is slowly changing—its "size" is pulsating with a frequency which appears to be $\frac{1}{2}(\omega_1 - \omega_2)$. But is this the frequency at which the beats are heard? Although (48.6) says that the amplitude goes as $\cos \frac{1}{2}(\omega_1 - \omega_2)t$, what it is really telling us is that the high-frequency oscillations are contained between two opposed cosine curves (shown dotted in Fig. 48-1). On this basis one could say that the amplitude varies at the frequency $\frac{1}{2}(\omega_1 - \omega_2)$, but if we are talking about the *intensity* of the wave we must think of it as having twice this frequency. That is, the modulation of the amplitude, in the sense of the strength of its intensity, is at frequency $\omega_1 - \omega_2$, although the formula tells us that we multiply by a cosine wave at half that frequency. The technical basis for the difference is that the high frequency-wave has a little different phase relationship in the second half-cycle.

Ignoring this small complication, we may conclude that if we add two waves of frequency ω_1 and ω_2 , we will get a net resulting wave of average frequency $\frac{1}{2}(\omega_1 + \omega_2)$ which oscillates in strength with a frequency $\omega_1 - \omega_2$.

If the two amplitudes are different, we can do it all over again by multiplying the cosines by different amplitudes A_1 and A_2 , and do a lot of mathematics, rearranging, and so on, using equations like (48.2)–(48.5). However, there are other, easier ways of doing the same analysis. For example, we know that it is much

easier to work with exponentials than with sines and cosines and that we can represent $A_1 \cos \omega_1 t$ as the real part of $A_1 e^{i\omega_1 t}$. The other wave would similarly be the real part of $A_2 e^{i\omega_2 t}$. If we add the two, we get $A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t}$. If we then factor out the average frequency, we have

$$A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t} = e^{1/2i(\omega_1 + \omega_2)t} [A_1 e^{1/2i(\omega_1 - \omega_2)t} + A_2 e^{-1/2i(\omega_1 - \omega_2)t}]. \quad (48.7)$$

Again we have the high-frequency wave with a modulation at the lower frequency.

48-2 Beat notes and modulation

If we are now asked for the intensity of the wave of Eq. (48.7), we can either take the absolute square of the left side, or of the right side. Let us take the left side. The intensity then is

$$I = A_1^2 + A_2^2 + 2A_1 A_2 \cos(\omega_1 - \omega_2)t. \quad (48.8)$$

We see that the intensity swells and falls at a frequency $\omega_1 - \omega_2$, varying between the limits $(A_1 + A_2)^2$ and $(A_1 - A_2)^2$. If $A_1 \neq A_2$, the minimum intensity is not zero.

One more way to represent this idea is by means of a drawing, like Fig. 48-2. We draw a vector of length A_1 , rotating at a frequency ω_1 , to represent one of the waves in the complex plane. We draw another vector of length A_2 , going around at a frequency ω_2 , to represent the second wave. If the two frequencies are exactly equal, their resultant is of fixed length as it keeps revolving, and we get a definite, fixed intensity from the two. But if the frequencies are slightly different, the two complex vectors go around at different speeds. Figure 48-3 shows what the situation looks like relative to the vector $A_1 e^{i\omega_1 t}$. We see that A_2 is turning slowly away from A_1 , and so the amplitude that we get by adding the two is first strong, and then, as it opens out, when it gets to the 180° relative position the resultant gets particularly weak, and so on. As the vectors go around, the amplitude of the sum vector gets bigger and smaller, and the intensity thus pulsates. It is a relatively simple idea, and there are many different ways of representing the same thing.

The effect is very easy to observe experimentally. In the case of acoustics, we may arrange two loudspeakers driven by two separate oscillators, one for each loudspeaker, so that they each make a tone. We thus receive one note from one source and a different note from the other source. If we make the frequencies exactly the same, the resulting effect will have a definite strength at a given space location. If we then de-tune them a little bit, we hear some variations in the intensity. The farther they are de-tuned, the more rapid are the variations of sound. The ear has some trouble following variations more rapid than ten or so per second.

We may also see the effect on an oscilloscope which simply displays the sum of the currents to the two speakers. If the frequency of pulsing is relatively low, we simply see a sinusoidal wave train whose amplitude pulsates, but as we make the pulsations more rapid we see the kind of wave shown in Fig. 48-1. As we go to greater frequency differences, the "bumps" move closer together. Also, if the amplitudes are not equal and we make one signal stronger than the other, then we get a wave whose amplitude does not ever become zero, just as we expect. Everything works the way it should, both acoustically and electrically.

The opposite phenomenon occurs too! In radio transmission using so-called *amplitude modulation* (AM), the sound is broadcast by the radio station as follows: the radio transmitter has an AC electric oscillation which is at a very high frequency, for example 800 kilocycles per second, in the broadcast band. If this *carrier signal* is turned on, the radio station emits a wave which is of uniform amplitude at 800,000 oscillations a second. The way the "information" is transmitted, the useless kind of information about what kind of car to buy, is that when somebody talks into a microphone the amplitude of the carrier signal is changed in step with the vibrations of sound entering the microphone.

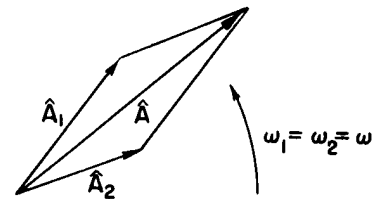


Fig. 48-2. The resultant of two complex vectors of equal frequency.

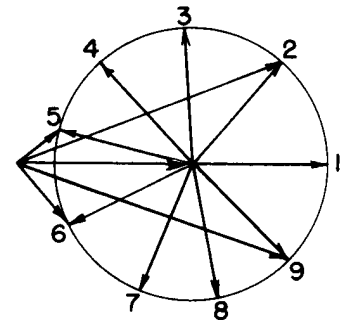


Fig. 48-3. The resultant of two complex vectors of unequal frequency, as seen in the rotating frame of reference of one vector. Nine successive positions of the slowly rotating vector are shown.

If we take as the simplest mathematical case the situation where a soprano is singing a perfect note, with perfect sinusoidal oscillations of her vocal cords, then we get a signal whose strength is alternating as shown in Fig. 48-4. The audiofrequency alternation is then recovered in the receiver; we get rid of the carrier wave and just look at the envelope which represents the oscillations of the vocal cords, or the sound of the singer. The loudspeaker then makes corresponding vibrations at the same frequency in the air, and the listener is then essentially unable to tell the difference, so they say. Because of a number of distortions and other subtle effects, it is, in fact, possible to tell whether we are listening to a radio or to a real soprano; otherwise the idea is as indicated above.

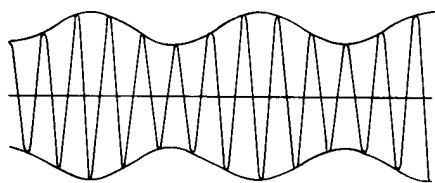


Fig. 48-4. A modulated carrier wave. In this schematic sketch, $\omega_c/\omega_m = 5$. In an actual radiowave, $\omega_c/\omega_m \sim 100$.

48-3 Side bands

Mathematically, the modulated wave described above would be expressed as

$$S = (1 + b \cos \omega_m t) \cos \omega_c t, \quad (48.9)$$

where ω_c represents the frequency of the carrier and ω_m is the frequency of the audio tone. Again we use all those theorems about the cosines, or we can use $e^{i\theta}$; it makes no difference—it is easier with $e^{i\theta}$, but it is the same thing. We then get

$$S = \cos \omega_c t + \frac{1}{2}b \cos (\omega_c + \omega_m)t + \frac{1}{2}b \cos (\omega_c - \omega_m)t. \quad (48.10)$$

So, from another point of view, we can say that the output wave of the system consists of three waves added in superposition: first, the regular wave at the frequency ω_c , that is, at the carrier frequency, and then two new waves at two new frequencies. One is the carrier frequency plus the modulation frequency, and the other is the carrier frequency minus the modulation frequency. If, therefore, we make some kind of plot of the intensity being generated by the generator as a function of frequency, we would find a lot of intensity at the frequency of the carrier, naturally, but when a singer started to sing, we would suddenly also find intensity proportional to the strength of the singer, b^2 , at frequency $\omega_c + \omega_m$ and $\omega_c - \omega_m$, as shown in Fig. 48-5. These are called *side bands*; when there is a modulated signal from the transmitter, there are side bands. If there is more than one note at the same time, say ω_m and ω_m' , there are two instruments playing; or if there is any other complicated cosine wave, then, of course, we can see from the mathematics that we get some more waves that correspond to the frequencies $\omega_c \pm \omega_m'$.

Therefore, when there is a complicated modulation that can be represented as the sum of many cosines,* we find that the actual transmitter is transmitting over a range of frequencies, namely the carrier frequency plus or minus the maximum frequency that the modulation signal contains.

Although at first we might believe that a radio transmitter transmits only at the nominal frequency of the carrier, since there are big, superstable crystal oscillators in there, and everything is adjusted to be at precisely 800 kilocycles, the moment someone *announces* that they are at 800 kilocycles, he modulates the 800 kilocycles, and so they are no longer precisely at 800 kilocycles! Suppose that the amplifiers are so built that they are able to transmit over a good range of the ear's sensitivity (the ear can hear up to 20,000 cycles per second, but usually radio transmitters and receivers do not work beyond 10,000, so we do not hear the highest parts), then, when the man speaks, his voice may contain frequencies ranging up, say, to 10,000 cycles, so the transmitter is transmitting frequencies

* A slight side remark: In what circumstances can a curve be represented as a sum of a lot of cosines? *Answer:* In all ordinary circumstances, except for certain cases the mathematicians can dream up. Of course, the curve must have only one value at a given point, and it must not be a crazy curve which jumps an infinite number of times in an infinitesimal distance, or something like that. But aside from such restrictions any reasonable curve (one that a singer is going to be able to make by shaking her vocal cords) can always be compounded by adding cosine waves together.

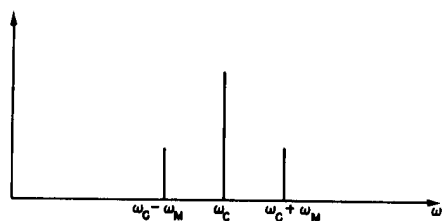


Fig. 48-5. The frequency spectrum of a carrier wave ω_c modulated by a single cosine wave ω_m .

which may range from 790 to 810 kilocycles per second. Now if there were another station at 795 kc/sec, there would be a lot of confusion. Also, if we made our receiver so sensitive that it picked up only 800, and did not pick up the 10 kilocycles on either side, we would not hear what the man was saying, because the information would be on these other frequencies! Therefore it is absolutely essential to keep the stations a certain distance apart, so that their side bands do not overlap and, also, the receiver must not be so selective that it does not permit reception of the side bands as well as of the main nominal frequency. In the case of sound, this problem does not really cause much trouble. We can hear over a ± 20 kc/sec range, and we have usually from 500 to 1500 kc/sec in the broadcast band, so there is plenty of room for lots of stations.

The television problem is more difficult. As the electron beam goes across the face of the picture tube, there are various little spots of light and dark. That "light" and "dark" is the "signal." Now ordinarily the beam scans over the whole picture, 500 lines, approximately, in a thirtieth of a second. Let us consider that the resolution of the picture vertically and horizontally is more or less the same, so that there are the same number of spots per inch along a scan line. We want to be able to distinguish dark from light, dark from light, dark from light, over, say, 500 lines. In order to be able to do this with cosine waves, the shortest wavelength needed thus corresponds to a wavelength, from maximum to maximum, of one 250th of the screen size. So we have $250 \times 500 \times 30$ pieces of information per second. The highest frequency that we are going to carry, therefore, is close to 4 megacycles per second. Actually, to keep the television stations apart, we have to use a little bit more than this, about 6 mc/sec; part of it is used to carry the sound signal, and other information. So, television channels are 6 megacycles per second wide. It certainly would not be possible to transmit tv on an 800 kc/sec carrier, since we cannot modulate at a higher frequency than the carrier.

At any rate, the television band starts at 54 megacycles. The first transmission channel, which is channel 2 (!), has a frequency range from 54 to 60 mc/sec, which is 6 mc/sec wide. "But," one might say, "we have just proved that there were side bands on both sides, and therefore it should be twice that wide." It turns out that the radio engineers are rather clever. If we analyze the modulation signal using not just cosine terms, but cosine and sine terms, to allow for phase differences, we then see that there is a definite, invariant relationship between the side band on the high-frequency side and the side band on the low-frequency side. What we mean is that there is no new information on that other side band. So what is done is to suppress one side band, and the receiver is wired inside such that the information which is missing is reconstituted by looking at the single side band and the carrier. Single side-band transmission is a clever scheme for decreasing the band widths needed to transmit information.

48-4 Localized wave trains

The next subject we shall discuss is the interference of waves in both space and time. Suppose that we have two waves travelling in space. We know, of course, that we can represent a wave travelling in space by $e^{i(\omega t - kx)}$. This might be, for example, the displacement in a sound wave. This is a solution of the wave equation provided that $\omega^2 = k^2 c^2$, where c is the speed of propagation of the wave. In this case we can write it as $e^{ik(x - ct)}$, which is of the general form $f(x - ct)$. Therefore this must be a wave which is travelling at this velocity, ω/k , and that is c and everything is all right.

Now we want to add two such waves together. Suppose we have a wave that is travelling with one frequency, and another wave travelling with another frequency. We leave to the reader to consider the case where the amplitudes are different; it makes no real difference. Thus we want to add $e^{i(\omega_1 t - k_1 x)} + e^{i(\omega_2 t - k_2 x)}$. We can add these by the same kind of mathematics we used when we added signal waves. Of course, if c is the same for both, this is easy, since it is the same as what we did before:

$$e^{i\omega_1(t-x/c)} + e^{i\omega_2(t-x/c)} = e^{i\omega_1 t'} + e^{i\omega_2 t'}, \quad (48.11)$$

except that $t' = t - x/c$ is the variable instead of t . So we get the same kind of modulations, naturally, but we see, of course, that those modulations are moving along with the wave. In other words, if we added two waves, but these waves were not just oscillating, but also moving in space, then the resultant wave would move along also, at the same speed.

Now we would like to generalize this to the case of waves in which the relationship between the frequency and the wave number k is not so simple. Example: material having an index of refraction. We have already studied the theory of the index of refraction in Chapter 31, where we found that we could write $k = n\omega/c$, where n is the index of refraction. As an interesting example, for x-rays we found that the index n is

$$n = 1 - \frac{Nq_e^2}{2\epsilon_0 m\omega^2}. \quad (48.12)$$

We actually derived a more complicated formula in Chapter 31, but this one is as good as any, as an example.

Incidentally, we know that even when ω and k are not linearly proportional, the ratio ω/k is certainly the speed of propagation for the particular frequency and wave number. We call this ratio the *phase velocity*; it is the speed at which the phase, or the nodes of a single wave, would move along:

$$v_p = \frac{\omega}{k}. \quad (48.13)$$

This phase velocity, for the case of x-rays in glass, is greater than the speed of light in vacuum (since n in 48.12 is less than 1), and that is a bit bothersome, because we do not think we can send signals faster than the speed of light!

What we are going to discuss now is the interference of two waves in which ω and k have a definite formula relating them. The above formula for n says that k is given as a definite function of ω . To be specific, in this particular problem, the formula for k in terms of ω is

$$k = \frac{\omega}{c} - \frac{a}{\omega c}, \quad (48.14)$$

where $a = Nq_e^2/2\epsilon_0 m$, a constant. At any rate, for each frequency there is a definite wave number, and we want to add two such waves together.

Let us do it just as we did in Eq. (48.7):

$$e^{i(\omega_1 t - k_1 x)} + e^{i(\omega_2 t - k_2 x)} = e^{1/2 i[(\omega_1 + \omega_2)t - (k_1 + k_2)x]} \times \{e^{1/2 i[(\omega_1 - \omega_2)t - (k_1 - k_2)x]} + e^{-1/2 i[(\omega_1 - \omega_2)t - (k_1 - k_2)x]}\}. \quad (48.15)$$

So we have a modulated wave again, a wave which travels with the mean frequency and the mean wave number, but whose strength is varying with a form which depends on the difference frequency and the difference wave number.

Now let us take the case that the difference between the two waves is relatively small. Let us suppose that we are adding two waves whose frequencies are nearly equal; then $(\omega_1 + \omega_2)/2$ is practically the same as either one of the ω 's, and similarly for $(k_1 + k_2)/2$. Thus the speed of the wave, the fast oscillations, the nodes, is still essentially ω/k . But look, the speed of propagation of the modulation is not the same! How much do we have to change x to account for a certain amount of t ? The speed of this modulation wave is the ratio

$$v_M = \frac{\omega_1 - \omega_2}{k_1 - k_2}. \quad (48.16)$$

The speed of modulation is sometimes called the *group velocity*. If we take the case that the difference in frequency is relatively small, and the difference in wave number is then also relatively small, then this expression approaches, in the limit,

$$v_g = \frac{d\omega}{dk}. \quad (48.17)$$

In other words, for the slowest modulation, the slowest beats, there is a definite speed at which they travel which is not the same as the phase speed of the waves—what a mysterious thing!

The group velocity is the derivative of ω with respect to k , and the phase velocity is ω/k .

Let us see if we can understand why. Consider two waves, again of slightly different wavelength, as in Fig. 48-1. They are out of phase, in phase, out of phase, and so on. Now these waves represent, really, the waves in space travelling with slightly different frequencies also. Now because the phase velocity, the velocity of the nodes of these two waves, is not precisely the same, something new happens. Suppose we ride along with one of the waves and look at the other one; if they both went at the same speed, then the other wave would stay right where it was relative to us, as we ride along on this crest. We ride on that crest and right opposite us we see a crest; if the two velocities are equal the crests stay on top of each other. But it is *not so* that the two velocities are really equal. There is only a small difference in frequency and therefore only a small difference in velocity, but because of that difference in velocity, as we ride along the other wave moves slowly forward, say, or behind, relative to our wave. So as time goes on, what happens to the node? If we move one wave train just a shade forward, the node moves forward (or backward) a considerable distance. That is, the sum of these two waves has an envelope, and as the waves travel along, the envelope rides on them at a different speed. The *group velocity* is the speed at which modulated signals would be transmitted.

If we made a signal, i.e., some kind of change in the wave that one could recognize when he listened to it, a kind of modulation, then that modulation would travel at the group velocity, provided that the modulations were relatively slow. (When they are fast, it is much more difficult to analyze.)

Now we may show (at long last), that the speed of propagation of x-rays in a block of carbon is *not* greater than the speed of light, although the phase velocity *is* greater than the speed of light. In order to do that, we must find $d\omega/dk$, which we get by differentiating (48.14): $dk/d\omega = 1/c + a/\omega^2c$. The group velocity, therefore, is the reciprocal of this, namely,

$$v_g = \frac{c}{1 + a/\omega^2}, \quad (48.18)$$

which is smaller than c ! So although the phases can travel faster than the speed of light, the modulation signals travel slower, and that is the resolution of the apparent paradox! Of course, if we have the simple case that $\omega = kc$, then $d\omega/dk$ is also c . So when all the phases have the same velocity, naturally the group has the same velocity.

48-5 Probability amplitudes for particles

Let us now consider one more example of the phase velocity which is extremely interesting. It has to do with quantum mechanics. We know that the amplitude to find a particle at a place can, in some circumstances, vary in space and time, let us say in one dimension, in this manner:

$$\psi = Ae^{i(\omega t - kx)}, \quad (48.19)$$

where ω is the frequency, which is related to the classical idea of the energy through $E = \hbar\omega$, and k is the wave number, which is related to the momentum through $p = \hbar k$. We would say the particle had a definite momentum p if the wave number were exactly k , that is, a perfect wave which goes on with the same amplitude everywhere. Equation (48.19) gives the amplitude, and if we take the absolute square, we get the relative probability for finding the particle as a function of position and time. This is a *constant*, which means that the probability is the same to find a particle anywhere. Now suppose, instead, that we have a situation where we know that the particle is more likely to be at one place than at another. We would

represent such a situation by a wave which has a maximum and dies out on either side (Fig. 48-6). (It is not quite the same as a wave like (48.1) which has a series of maxima, but it is possible, by adding several waves of nearly the same ω and k together, to get rid of all but one maximum.)

Now in those circumstances, since the square of (48.19) represents the chance of finding a particle somewhere, we know that at a given instant the particle is most likely to be near the center of the "lump," where the amplitude of the wave is maximum. If now we wait a few moments, the waves will move, and after some time the "lump" will be somewhere else. If we knew that the particle originally was situated somewhere, classically, we would *expect* that it would later be elsewhere as a matter of fact, because it has a *speed*, after all, and a momentum. The quantum theory, then, will go into the correct classical theory for the relationship of momentum, energy, and velocity only if the group velocity, the velocity of the modulation, is equal to the velocity that we would obtain classically for a particle of the same momentum.

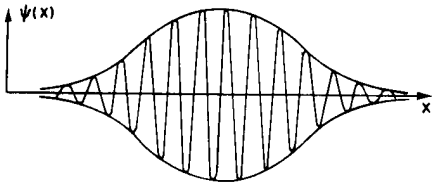


Fig. 48-6. A localized wave train.

It is now necessary to demonstrate that this is, or is not, the case. According to the classical theory, the energy is related to the velocity through an equation like

$$E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}. \quad (48.20)$$

Similarly, the momentum is

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}}. \quad (48.21)$$

That is the classical theory, and as a consequence of the classical theory, by eliminating v , we can show that

$$E^2 - p^2c^2 = m^2c^4.$$

That is the four-dimensional grand result that we have talked and talked about, that $p_\mu p_\mu = m^2$; that is the relation between energy and momentum in the classical theory. Now that means, since these E 's and p 's are going to become ω 's and k 's, by substitution of $E = \hbar\omega$ and $p = \hbar k$, that for quantum mechanics it is necessary that

$$\frac{\hbar^2\omega^2}{c^2} - \hbar^2k^2 = m^2c^2. \quad (48.22)$$

This, then, is the relationship between the frequency and the wave number of a quantum-mechanical amplitude wave representing a particle of mass m . From this equation we can deduce that ω is

$$\omega = c\sqrt{k^2 + m^2c^2/\hbar^2}.$$

The phase velocity, ω/k , is here again faster than the speed of light!

Now let us look at the group velocity. The group velocity should be $d\omega/dk$, the speed at which the modulations move. We have to differentiate a square root, which is not very difficult. The derivative is

$$\frac{d\omega}{dk} = \frac{kc}{\sqrt{k^2 + m^2c^2/\hbar^2}}.$$

Now the square root is, after all, ω , so we could write this as $d\omega/dk = c^2k/\omega$. Further, k/ω is p/E , so

$$v_g = \frac{c^2p}{E}.$$

But from (48.20) and (48.21), $c^2p/E = v$, the velocity of the particle, according to classical mechanics. So we see that whereas the fundamental quantum-mechanical relationship $E = \hbar\omega$ and $p = \hbar k$, for the identification of ω and k with the classical E and p , only produces the equation $\omega^2 - k^2c^2 = m^2c^4/\hbar^2$, now we also understand the relationships (48.20) and (48.21) which connected E and p

to the velocity. Of course the group velocity must be the velocity of the particle if the interpretation is going to make any sense. If we think the particle is over here at one time, and then ten minutes later we think it is over there, as the quantum mechanics said, the distance traversed by the “lump,” divided by the time interval, must be, classically, the velocity of the particle.

48-6 Waves in three dimensions

We shall now bring our discussion of waves to a close with a few general remarks about the wave equation. These remarks are intended to give some view of the future—not that we can understand everything exactly just now, but rather to see what things are going to look like when we study waves a little more. First of all, the wave equation for sound in one dimension was

$$\frac{\partial^2 \chi}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2},$$

where c is the speed of whatever the wave is—in the case of sound, it is the sound speed; in the case of light, it is the speed of light. We showed that for a sound wave the displacements would propagate themselves at a certain speed. But the excess pressure also propagates at a certain speed, and so does the excess density. So we should expect that the pressure would satisfy the same equation, as indeed it does. We shall leave it to the reader to prove that it does. *Hint:* ρ_e is proportional to the rate of change of χ with respect to x . Therefore if we differentiate the wave equation with respect to x , we will immediately discover that $\partial\chi/\partial x$ satisfies the same equation. That is to say, ρ_e satisfies the same equation. But P_e is proportional to ρ_e , and therefore P_e does too. So the pressure, the displacements, everything, satisfy the same wave equation.

Usually one sees the wave equation for sound written in terms of pressure instead of in terms of displacement, because the pressure is a scalar and has no direction. But the displacement is a vector and has direction, and it is thus easier to analyze the pressure.

The next matter we discuss has to do with the wave equation in three dimensions. We know that the sound wave solution in one dimension is $e^{i(\omega t - kx)}$, with $\omega = kc_s$, but we also know that in three dimensions a wave would be represented by $e^{i(\omega t - k_x x - k_y y - k_z z)}$, where, in this case, $\omega^2 = k^2 c_s^2$, which is, of course, $(k_x^2 + k_y^2 + k_z^2) c_s^2$. Now what we want to do is to guess what the correct wave equation in three dimensions is. Naturally, for the case of sound this can be deduced by going through the same dynamic argument in three dimensions that we made in one dimension. But we shall not do that; instead we just write down what comes out: the equation for the pressure (or displacement, or anything) is

$$\frac{\partial^2 P_e}{\partial x^2} + \frac{\partial^2 P_e}{\partial y^2} + \frac{\partial^2 P_e}{\partial z^2} = \frac{1}{c_s^2} \frac{\partial^2 P_e}{\partial t^2}. \quad (48.23)$$

That this is true can be verified by substituting in $e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$. Clearly, every time we differentiate with respect to x , we multiply by $-ik_x$. If we differentiate twice, it is equivalent to multiplying by $-k_x^2$, so the first term would become $-k_x^2 P_e$, for that wave. Similarly, the second term becomes $-k_y^2 P_e$, and the third term becomes $-k_z^2 P_e$. On the right, we get $-(\omega^2/c_s^2) P_e$. Then, if we take away the P_e 's and change the sign, we see that the relationship between k and ω is the one that we want.

Working backwards again, we cannot resist writing down the grand equation which corresponds to the dispersion equation (48.22) for quantum-mechanical waves. If ϕ represents the amplitude for finding a particle at position x, y, z , at the time t , then the great equation of quantum mechanics for free particles is this:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{m^2 c^2}{\hbar^2} \phi. \quad (48.24)$$

First of all, the relativity character of this expression is suggested by the appearance

of x , y , z and t in the nice combination relativity usually involves. Second, it is a wave equation which, if we try a plane wave, would produce as a consequence that $-k^2 + \omega^2/c^2 = m^2c^2/\hbar^2$, which is the right relationship for quantum mechanics. There is still another great thing contained in the wave equation: the fact that any superposition of waves is also a solution. So this equation contains all of the quantum mechanics and the relativity that we have been discussing so far, at least so long as it deals with a single particle in empty space with no external potentials or forces on it!

48-7 Normal modes

Now we turn to another example of the phenomenon of beats which is rather curious and a little different. Imagine two equal pendulums which have, between them, a rather weak spring connection. They are made as nearly as possible the same length. If we pull one aside and let go, it moves back and forth, and it pulls on the connecting spring as it moves back and forth, and so it really is a machine for generating a force which has the natural frequency of the other pendulum. Therefore, as a consequence of the theory of resonance, which we studied before, when we put a force on something at just the right frequency, it will drive it. So, sure enough, one pendulum moving back and forth drives the other. However, in this circumstance there is a new thing happening, because the total energy of the system is finite, so when one pendulum pours its energy into the other to drive it, it finds itself gradually losing energy, until, if the timing is just right along with the speed, it loses all its energy and is reduced to a stationary condition! Then, of course, it is the other pendulum ball that has all the energy and the first one which has none, and as time goes on we see that it works also in the opposite direction, and that the energy is passed back into the first ball; this is a very interesting and amusing phenomenon. We said, however, that this is related to the theory of beats, and we must now explain how we can analyze this motion from the point of view of the theory of beats.

We note that the motion of either of the two balls is an oscillation which has an amplitude which changes cyclically. Therefore the motion of one of the balls is presumably analyzable in a different way, in that it is the sum of two oscillations, present at the same time but having two slightly different frequencies. Therefore it ought to be possible to find two other motions in this system, and to claim that what we saw was a superposition of the two solutions, because this is of course a linear system. Indeed, it is easy to find two ways that we could start the motion, each one of which is a perfect, single-frequency motion—absolutely periodic. The motion that we started with before was not strictly periodic, since it did not last; soon one ball was passing energy to the other and so changing its amplitude; but there are ways of starting the motion so that nothing changes and, of course, as soon as we see it we understand why. For example, if we made both pendulums go together, then, since they are of the same length and the spring is not then doing anything, they will of course continue to swing like that for all time, assuming no friction and that everything is perfect. On the other hand, there is another possible motion which also has a definite frequency: that is, if we move the pendulums oppositely, pulling them aside exactly equal distances, then again they would be in absolutely periodic motion. We can appreciate that the spring just adds a little to the restoring force that the gravity supplies, that is all, and the system just keeps oscillating at a slightly higher frequency than in the first case. Why higher? Because the spring is pulling, in addition to the gravitation, and it makes the system a little “stiffer,” so that the frequency of this motion is just a shade higher than that of the other.

Thus this system has two ways in which it can oscillate with unchanging amplitude: it can either oscillate in a manner in which both pendulums go the same way and oscillate all the time at one frequency, or they could go in opposite directions at a slightly higher frequency.

Now the actual motion of the thing, because the system is linear, can be represented as a superposition of the two. (The subject of this chapter, remember,

is the effects of adding two motions with different frequencies.) So think what would happen if we combined these two solutions. If at $t = 0$ the two motions are started with equal amplitude and in the same phase, the sum of the two motions means that one ball, having been impressed one way by the first motion and the other way by the second motion, is at zero, while the other ball, having been displaced the same way in both motions, has a large amplitude. As time goes on, however, the two basic *motions* proceed independently, so the phase of one relative to the other is slowly shifting. That means, then, that after a sufficiently long time, when the time is enough that one motion could have gone “ $900\frac{1}{2}$ ” oscillations, while the other went only “900,” the relative phase would be just reversed with respect to what it was before. That is, the large-amplitude motion will have fallen to zero, and in the meantime, of course, the initially motionless ball will have attained full strength!

So we see that we could analyze this complicated motion either by the idea that there is a resonance and that one passes energy to the other, or else by the superposition of two constant-amplitude motions at two different frequencies.