1

Quantum Behavior

Note: This chapter is almost exactly the same as Chapter 37 of Volume I.

1-1 Atomic mechanics

"Quantum mechanics" is the description of the behavior of matter and light in all its details and, in particular, of the happenings on an atomic scale. Things on a very small scale behave like nothing that you have any direct experience about. They do not behave like waves, they do not behave like particles, they do not behave like clouds, or billiard balls, or weights on springs, or like anything that you have ever seen.

Newton thought that light was made up of particles, but then it was discovered that it behaves like a wave. Later, however (in the beginning of the twentieth century), it was found that light did indeed sometimes behave like a particle. Historically, the electron, for example, was thought to behave like a particle, and then it was found that in many respects it behaved like a wave. So it really behaves like neither. Now we have given up. We say: "It is like *neither*."

There is one lucky break, however—electrons behave just like light. The quantum behavior of atomic objects (electrons, protons, neutrons, photons, and so on) is the same for all, they are all "particle waves," or whatever you want to call them. So what we learn about the properties of electrons (which we shall use for our examples) will apply also to all "particles," including photons of light.

The gradual accumulation of information about atomic and small-scale behavior during the first quarter of this century, which gave some indications about how small things do behave, produced an increasing confusion which was finally resolved in 1926 and 1927 by Schrödinger, Heisenberg, and Born. They finally obtained a consistent description of the behavior of matter on a small scale. We take up the main features of that description in this chapter.

Because atomic behavior is so unlike ordinary experience, it is very difficult to get used to, and it appears peculiar and mysterious to everyone—both to the novice and to the experienced physicist. Even the experts do not understand it the way they would like to, and it is perfectly reasonable that they should not, because all of direct, human experience and of human intuition applies to large objects. We know how large objects will act, but things on a small scale just do not act that way. So we have to learn about them in a sort of abstract or imaginative fashion and not by connection with our direct experience.

In this chapter we shall tackle immediately the basic element of the mysterious behavior in its most strange form. We choose to examine a phenomenon which is impossible, *absolutely* impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the *only* mystery. We cannot make the mystery go away by "explaining" how it works. We will just *tell* you how it works. In telling you how it works we will have told you about the basic peculiarities of all quantum mechanics.

1-2 An experiment with bullets

To try to understand the quantum behavior of electrons, we shall compare and contrast their behavior, in a particular experimental setup, with the more familiar behavior of particles like bullets, and with the behavior of waves like water waves. We consider first the behavior of bullets in the experimental setup shown diagrammatically in Fig. 1-1. We have a machine gun that shoots a stream of bullets. It is not a very good gun, in that it sprays the bullets (randomly) over a fairly large angular spread, as indicated in the figure. In front of the gun we have

of I_1 and I_2. We say that there is "interference" of the two waves. At some places (where the curve I_{12} has its maxima) the waves are "in phase" and the wave peaks add together to give a large amplitude and, therefore, a large intensity. We say that the two waves are "interfering constructively" at such places. There will be such constructive interference wherever the distance from the detector to one hole is a whole number of wavelengths larger (or shorter) than the distance from the detector to the other hole.

a wall (made of armor plate) that has in it two holes just about big enough to let a bullet through. Beyond the wall is a backstop (say a thick wall of wood) which will "absorb" the bullets when they hit it. In front of the wall we have an object which we shall call a "detector" of bullets. It might be a box containing sand. Any bullet that enters the detector will be stopped and accumulated. When we wish, we can empty the box and count the number of bullets that have been caught. The detector can be moved back and forth (in what we will call the x-direction). With this apparatus, we can find out experimentally the answer to the question: "What is the probability that a bullet which passes through the holes in the wall will arrive at the backstop at the distance x from the center?" First, you should realize that we should talk about probability, because we cannot say definitely

At those places where the two waves arrive at the detector with a phase difference of π (where they are "out of phase") the resulting wave motion at the detector will be the difference of the two amplitudes. The waves "interfere destructively," and we get a low value for the wave intensity. We expect such low values wherever the distance between hole 1 and the detector is different from the distance between hole 2 and the detector by an odd number of half-wavelengths. The low values of I_{12} in Fig. 1–2 correspond to the places where the two waves interfere destructively.

You will remember that the quantitative relationship between I_1, I_2, and I_{12} can be expressed in the following way: The instantaneous height of the water wave at the detector for the wave from hole 1 can be written as (the real part of) $h_1e^{i\omega t}$, where the "amplitude" h_1 is, in general, a complex number. The intensity is proportional to the mean squared height or, when we use the complex numbers, to the absolute value squared $|h_1|^2$. Similarly, for hole 2 the height is $h_2e^{i\omega t}$ and the intensity is proportional to $|h_2|^2$. When both holes are open, the wave heights add to give the height$(h_1 + h_2)e^{i\omega t}$ and the intensity $|h_1 + h_2|^2$. Omitting the constant of proportionality for our present purposes, the proper relations for *interfering waves* are

$$I_1 = |h_1|^2, \qquad I_2 = |h_2|^2, \qquad I_{12} = |h_1 + h_2|^2. \tag{1.2}$$

You will notice that the result is quite different from that obtained with bullets (Eq. 1–1). If we expand $|h_1 + h_2|^2$ we see that

$$|h_1 + h_2|^2 = |h_1|^2 + |h_2|^2 + 2|h_1||h_2| \cos \delta, \tag{1.3}$$

where δ is the phase difference between h_1 and h_2. In terms of the intensities, we could write

$$I_{12} = I_1 + I_2 + 2\sqrt{I_1I_2} \cos \delta. \tag{1.4}$$

The last term in (1.4) is the "interference term." So much for water waves. The intensity can have any value, and it shows interference.

1–4 An experiment with electrons

Now we imagine a similar experiment with electrons. It is shown diagrammatically in Fig. 1–3. We make an electron gun which consists of a tungsten wire heated by an electric current and surrounded by a metal box with a hole in it. If the wire is at a negative voltage with respect to the box, electrons emitted by the wire will be accelerated toward the walls and some will pass through the hole. All the electrons which come out of the gun will have (nearly) the same energy. In front of the gun is again a wall (just a thin metal plate) with two holes in it. Beyond the wall is another plate which will serve as a "backstop." In front of the backstop we place a movable detector. The detector might be a geiger counter or, perhaps better, an electron multiplier, which is connected to a loudspeaker.

We should say right away that you should not try to set up this experiment (as you could have done with the two we have already described). This experiment

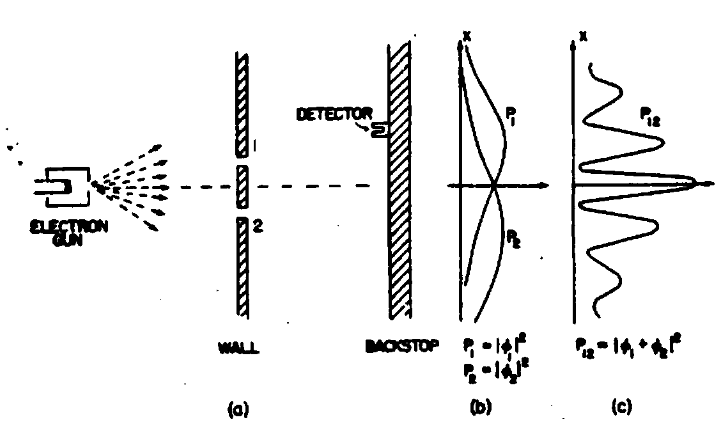

Fig. 1–3. Interference experiment with electrons.

has never been done in just this way. The trouble is that the apparatus would have to be made on an impossibly small scale to show the effects we are interested in. We are doing a "thought experiment," which we have chosen because it is easy to think about. We know the results that *would* be obtained because there *are* many experiments that have been done, in which the scale and the proportions have been chosen to show the effects we shall describe.

The first thing we notice with our electron experiment is that we hear sharp "clicks" from the detector (that is, from the loudspeaker). And all "clicks" are the same. There are *no* "half-clicks."

We would also notice that the "clicks" come very erratically. Something like: click . . . . . click-click . . . click . . . . click . . . . click-click . . . . . . click . . . , etc., just as you have, no doubt, heard a geiger counter operating. If we count the clicks which arrive in a sufficiently long time—say for many minutes—and then count again for another equal period, we find that the two numbers are very nearly the same. So we can speak of the *average rate* at which the clicks are heard (so-and-so-many clicks per minute on the average).

As we move the detector around, the *rate* at which the clicks appear is faster or slower, but the size (loudness) of each click is always the same. If we lower the temperature of the wire in the gun, the rate of clicking slows down, but still each click sounds the same. We would notice also that if we put two separate detectors at the backstop, one *or* the other would click, but never both at once. (Except that once in a while, if there were two clicks very close together in time, our ear might not sense the separation.) We conclude, therefore, that whatever arrives at the backstop arrives in "lumps." All the "lumps" are the same size: only whole "lumps" arrive, and they arrive one at a time at the backstop. We shall say: "Electrons always arrive in identical lumps."

Just as for our experiment with bullets, we can now proceed to find experimentally the answer to the question: "What is the relative probability that an electron 'lump' will arrive at the backstop at various distances x from the center?" As before, we obtain the relative probability by observing the rate of clicks, holding the operation of the gun constant. The probability that lumps will arrive at a particular x is proportional to the average rate of clicks at that x.

The result of our experiment is the interesting curve marked P_{12} in part (c) of Fig. 1–3. Yes! That is the way electrons go.

1–5 The interference of electron waves

Now let us try to analyze the curve of Fig. 1–3 to see whether we can understand the behavior of the electrons. The first thing we would say is that since they come in lumps, each lump, which we may as well call an electron, has come either through hole 1 or through hole 2. Let us write this in the form of a "Proposition":

Proposition A: Each electron *either* goes through hole 1 *or* it goes through hole 2.

Assuming Propositon A, all electrons that arrive at the backstop can be divided into two classes: (1) those that come through hole 1, and (2) those that come through hole 2. So our observed curve must be the sum of the effects of the electrons which come through hole 1 and the electrons which come through hole 2. Let us check this idea by experiment. First, we will make a measurement for those electrons that come through hole 1. We block off hole 2 and make our counts of the clicks from the detector. From the clicking rate, we get P_1. The result of the measurement is shown by the curve marked P_1 in part (b) of Fig. 1–3. The result seems quite reasonable. In a similar way, we measure P_2, the probability distribution for the electrons that come through hole 2. The result of this measurement is also drawn in the figure.

The result P_{12} obtained with *both* holes open is clearly not the sum of P_1 and P_2, the probabilities for each hole alone. In analogy with our water-wave experiment, we say: "There is interference."

$$\textit{For electrons:} \quad P_{12} \neq P_1 + P_2. \tag{1.5}$$

How can such an interference come about? Perhaps we should say: "Well, that means, presumably, that it is *not true* that the lumps go either through hole 1 or hole 2, because if they did, the probabilities should add. Perhaps they go in a more complicated way. They split in half and . . ." But no! They cannot, they always arrive in lumps . . . "Well, perhaps some of them go through 1, and then they go around through 2, and then around a few more times, or by some other complicated path . . . then by closing hole 2, we changed the chance that an electron that *started out* through hole 1 would finally get to the backstop . . ." But notice! There are some points at which very few electrons arrive when *both* holes are open, but which receive many electrons if we close one hole, so *closing* one hole *increased* the number from the other. Notice, however, that at the center of the pattern, P_{12} is more than twice as large as $P_1 + P_2$. It is as though closing one hole *decreased* the number of electrons which come through the other hole. It seems hard to explain *both* effects by proposing that the electrons travel in complicated paths.

It is all quite mysterious. And the more you look at it the more mysterious it seems. Many ideas have been concocted to try to explain the curve for P_{12} in terms of individual electrons going around in complicated ways through the holes. None of them has succeeded. None of them can get the right curve for P_{12} in terms of P_1 and P_2.

Yet, surprisingly enough, the *mathematics* for relating P_1 and P_2 to P_{12} is extremely simple. For P_{12} is just like the curve I_{12} of Fig. 1–2, and *that* was simple. What is going on at the backstop can be described by two complex numbers that we can call ϕ_1 and ϕ_2 (they are functions of x, of course). The absolute square of ϕ_1 gives the effect with only hole 1 open. That is, $P_1 = |\phi_1|^2$. The effect with only hole 2 open is given by ϕ_2 in the same way. That is, $P_2 = |\phi_2|^2$. And the combined effect of the two holes is just $P_{12} = |\phi_1 + \phi_2|^2$. The *mathematics* is the same as that we had for the water waves! (It is hard to see how one could get such a simple result from a complicated game of electrons going back and forth through the plate on some strange trajectory.)

We conclude the following: The electrons arrive in lumps, like particles, and the probability of arrival of these lumps is distributed like the distribution of intensity of a wave. It is in this sense that an electron behaves "sometimes like a particle and sometimes like a wave."

Incidentally, when we were dealing with classical waves we defined the intensity as the mean over time of the square of the wave amplitude, and we used complex numbers as a mathematical trick to simplify the analysis. But in quantum mechanics it turns out that the amplitudes *must* be represented by complex numbers. The real parts alone will not do. That is a technical point, for the moment, because the formulas look just the same.

Since the probability of arrival through both holes is given so simply, although it is not equal to $(P_1 + P_2)$, that is really all there is to say. But there are a large number of subtleties involved in the fact that nature does work this way. We would like to illustrate some of these subtleties for you now. First, since the number that arrives at a particular point is *not* equal to the number that arrives through 1 plus the number that arrives through 2, as we would have concluded from Proposition A, undoubtedly we should conclude that *Proposition A is false*. It is *not* true that the electrons go *either* through hole 1 or hole 2. But that conclusion can be tested by another experiment.

1–6 Watching the electrons

We shall now try the following experiment. To our electron apparatus we add a very strong light source, placed behind the wall and between the two holes, as shown in Fig. 1–4. We know that electric charges scatter light. So when an

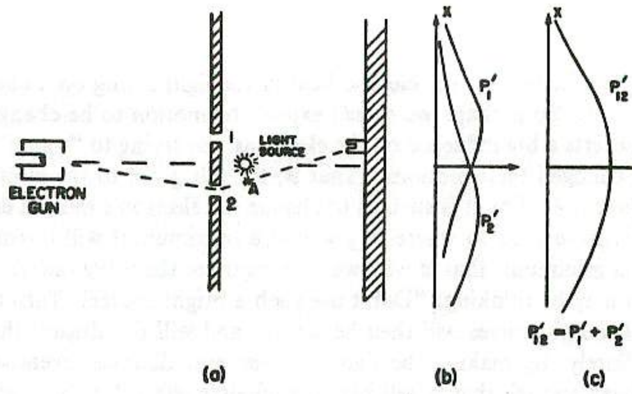

Fig. 1–4. A different electron experiment.

electron passes, however it does pass, on its way to the detector, it will scatter some light to our eye, and we can *see* where the electron goes. If, for instance, an electron were to take the path via hole 2 that is sketched in Fig. 1–4, we should see a flash of light coming from the vicinity of the place marked A in the figure. If an electron passes through hole 1, we would expect to see a flash from the vicinity of the upper hole. If it should happen that we get light from both places at the same time, because the electron divides in half . . . Let us just do the experiment!

Here is what we see: *every* time that we hear a "click" from our electron detector (at the backstop), we *also see* a flash of light *either* near hole 1 *or* near hole 2, but *never* both at once! And we observe the same result no matter where we put the detector. From this observation we conclude that when we look at the electrons we find that the electrons go either through one hole or the other. Experimentally, Proposition A is necessarily true.

What, then, is wrong with our argument *against* Proposition A? Why *isn't* P_{12} just equal to $P_1 + P_2$? Back to experiment! Let us keep track of the electrons and find out what they are doing. For each position (x-location) of the detector we will count the electrons that arrive and *also* keep track of which hole they went through, by watching for the flashes. We can keep track of things this way: whenever we hear a "click" we will put a count in Column 1 if we see the flash near hole 1, and if we see the flash near hole 2, we will record a count in Column 2. Every electron which arrives is recorded in one of two classes: those which come through 1 and those which come through 2. From the number recorded in Column 1 we get the probability P'_1 that an electron will arrive at the detector via hole 1; and from the number recorded in Column 2 we get P'_2, the probability that an electron will arrive at the detector via hole 2. If we now repeat such a measurement for many values of x, we get the curves for P'_1 and P'_2 shown in part (b) of Fig. 1–4.

Well, that is not too surprising! We get for P'_1 something quite similar to what we got before for P_1 by blocking off hole 2; and P'_2 is similar to what we got by blocking hole 1. So there is *not* any complicated business like going through both holes. When we watch them, the electrons come through just as we would expect them to come through. Whether the holes are closed or open, those which we see come through hole 1 are distributed in the same way whether hole 2 is open or closed.

But wait! What do we have *now* for the *total* probability, the probability that an electron will arrive at the detector by any route? We already have that information. We just pretend that we never looked at the light flashes, and we lump together the detector clicks which we have separated into the two columns. We *must* just *add* the numbers. For the probability that an electron will arrive at the backstop by passing through *either* hole, we do find $P'_{12} = P_1 + P_2$. That is, although we succeeded in watching which hole our electrons come through, we no longer get the old interference curve P_{12}, but a new one, P'_{12}, showing no interference! If we turn out the light P_{12} is restored.

We must conclude that *when we look at the electrons* the distribution of them on the screen is different than when we do not look. Perhaps it is turning on our light source that disturbs things? It must be that the electrons are very delicate, and the light, when it scatters off the electrons, gives them a jolt that changes their motion. We know that the electric field of the light acting on a charge will exert a force on it. So perhaps we *should* expect the motion to be changed. Anyway, the light exerts a big influence on the electrons. By trying to "watch" the electrons we have changed their motions. That is, the jolt given to the electron when the photon is scattered by it is such as to change the electron's motion enough so that if it *might* have gone to where P_{12} was at a maximum it will instead land where P_{12} was a minimum; that is why we no longer see the wavy interference effects.

You may be thinking: "Don't use such a bright source! Turn the brightness down! The light waves will then be weaker and will not disturb the electrons so much. Surely, by making the light dimmer and dimmer, eventually the wave will be weak enough that it will have a negligible effect." O.K. Let's try it. The first thing we observe is that the flashes of light scattered from the electrons as they pass by does *not* get weaker. *It is always the same-sized flash.* The only thing that happens as the light is made dimmer is that sometimes we hear a "click" from the detector but see *no flash at all.* The electron has gone by without being "seen." What we are observing is that light *also* acts like electrons, we *knew* that it was "wavy," but now we find that it is also "lumpy." It always arrives—or is scattered—in lumps that we call "photons." As we turn down the *intensity* of the light source we do not change the *size* of the photons, only the *rate* at which they are emitted. *That* explains why, when our source is dim, some electrons get by without being seen. There did not happen to be a photon around at the time the electron went through.

This is all a little discouraging. If it is true that whenever we "see" the electron we see the same-sized flash, then those electrons we see are *always* the disturbed ones. Let us try the experiment with a dim light anyway. Now whenever we hear a click in the detector we will keep a count in three columns: in Column (1) those electrons seen by hole 1, in Column (2) those electrons seen by hole 2, and in Column (3) those electrons not seen at all. When we work up our data (computing the probabilities) we find these results: Those "seen by hole 1" have a distribution like P'_1; those "seen by hole 2" have a distribution like P'_2 (so that those "seen by either hole 1 or 2" have a distribution like P'_{12}); and those "not seen at all" have a "wavy" distribution just like P_{12} of Fig. 1–3! *If the electrons are not seen, we have interference!*

That is understandable. When we do not see the electron, no photon disturbs it, and when we do see it, a photon has disturbed it. There is always the same amount of disturbance because the light photons all produce the same-sized effects and the effect of the photons being scattered is enough to smear out any interference effect.

Is there not *some* way we can see the electrons without disturbing them? We learned in an earlier chapter that the momentum carried by a "photon" is inversely proportional to its wavelength ($p = h/\lambda$). Certainly the jolt given to the electron when the photon is scattered toward our eye depends on the momentum that photon carries. Aha! If we want to disturb the electrons only slightly we should not have lowered the *intensity* of the light, we should have lowered its *frequency* (the same as increasing its wavelength). Let us use light of a redder color. We could even use infrared light, or radiowaves (like radar), and "see" where the electron went with the help of some equipment that can "see" light of these longer wavelengths. If we use "gentler" light perhaps we can avoid disturbing the electrons so much.

Let us try the experiment with longer waves. We shall keep repeating our experiment, each time with light of a longer wavelength. At first, nothing seems to change. The results are the same. Then a terrible thing happens. You remember that when we discussed the microscope we pointed out that, due to the *wave nature* of the light, there is a limitation on how close two spots can be and still be seen as two separate spots. This distance is of the order of the wavelength of light. So now, when we make the wavelength longer than the distance between our holes, we see a *big* fuzzy flash when the light is scattered by the electrons. We can no longer tell which hole the electron went through! We just know it went somewhere! And it is just with light of this color that we find that the jolts given to the electron are small enough so that P'_{12} begins to look like P_{12}—that we begin to get some interference effect. And it is only for wavelengths much longer than the separation of the two holes (when we have no chance at all of telling where the electron went) that the disturbance due to the light gets sufficiently small that we again get the curve P_{12} shown in Fig. 1-3.

In our experiment we find that it is impossible to arrange the light in such a way that one can tell which hole the electron went through, and at the same time not disturb the pattern. It was suggested by Heisenberg that the then new laws of nature could only be consistent if there were some basic limitation on our experimental capabilities not previously recognized. He proposed, as a general principle, his *uncertainty principle*, which we can state in terms of our experiment as follows: "It is impossible to design an apparatus to determine which hole the electron passes through, that will not at the same time disturb the electrons enough to destroy the interference pattern." If an apparatus is capable of determining which hole the electron goes through, it *cannot* be so delicate that it does not disturb the pattern in an essential way. No one has ever found (or even thought of) a way around the uncertainty principle. So we must assume that it describes a basic characteristic of nature.

The complete theory of quantum mechanics which we now use to describe atoms and, in fact, all matter, depends on the correctness of the uncertainty principle. Since quantum mechanics is such a successful theory, our belief in the uncertainty principle is reinforced. But if a way to "beat" the uncertainty principle were ever discovered, quantum mechanics would give inconsistent results and would have to be discarded as a valid theory of nature.

"Well," you say, "what about Proposition A? Is it true, or is it *not* true, that the electron either goes through hole 1 or it goes through hole 2?" The only answer that can be given is that we have found from experiment that there is a certain special way that we have to think in order that we do not get into inconsistencies. What we must say (to avoid making wrong predictions) is the following. If one looks at the holes or, more accurately, if one has a piece of apparatus which is capable of determining whether the electrons go through hole 1 or hole 2, then one *can* say that it goes either through hole 1 or hole 2. *But*, when one does *not* try to tell which way the electron goes, when there is nothing in the experiment to disturb the electrons, then one may *not* say that an electron goes either through hole 1 or hole 2. If one does say that, and starts to make any deductions from the statement, he will make errors in the analysis. This is the logical tightrope on which we must walk if we wish to describe nature successfully.

If the motion of all matter—as well as electrons—must be described in terms of waves, what about the bullets in our first experiment? Why didn't we see an interference pattern there? It turns out that for the bullets the wavelengths were so tiny that the interference patterns became very fine. So fine, in fact, that with any detector of finite size one could not distinguish the separate maxima and minima. What we saw was only a kind of average, which is the classical curve. In Fig. 1-5 we have tried to indicate schematically what happens with large-scale objects. Part (a) of the figure shows the probability distribution one might predict for bullets, using quantum mechanics. The rapid wiggles are supposed to represent the interference pattern one gets for waves of very short wavelength. Any physical detector, however, straddles several wiggles of the probability curve, so that the measurements show the smooth curve drawn in part (b) of the figure.

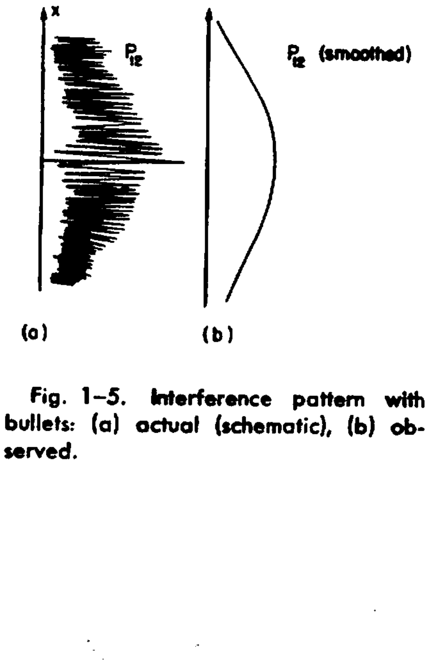

Fig. 1-5. Interference pattern with bullets: (a) actual (schematic), (b) observed.

1-7 First principles of quantum mechanics

We will now write a summary of the main conclusions of our experiments. We will, however, put the results in a form which makes them true for a general class of such experiments. We can write our summary more simply if we first define an "ideal experiment" as one in which there are no uncertain external influences, i.e., no jiggling or other things going on that we cannot take into account. We would be quite precise if we said: "An ideal experiment is one in which all of the initial and final conditions of the experiment are completely specified." What we will call "an event" is, in general, just a specific set of initial and final conditions. (For example: "an electron leaves the gun, arrives at the detector, and nothing else happens.") Now for our summary.

Summary

(1) The probability of an event in an ideal experiment is given by the square of the absolute value of a complex number ϕ which is called the probability amplitude:

$$\begin{aligned} P &= \text{probability}, \\ \phi &= \text{probability amplitude}, \\ P &= |\phi|^2. \end{aligned} \tag{1.6}$$

(2) When an event can occur in several alternative ways, the probability amplitude for the event is the sum of the probability amplitudes for each way considered separately. There is interference:

$$\begin{aligned} \phi &= \phi_1 + \phi_2, \\ P &= |\phi_1 + \phi_2|^2. \end{aligned} \tag{1.7}$$

(3) If an experiment is performed which is capable of determining whether one or another alternative is actually taken, the probability of the event is the sum of the probabilities for each alternative. The interference is lost:

$$P = P_1 + P_2. \tag{1.8}$$

One might still like to ask: "How does it work? What is the machinery behind the law?" No one has found any machinery behind the law. No one can "explain" any more than we have just "explained." No one will give you any deeper representation of the situation. We have no ideas about a more basic mechanism from which these results can be deduced.

We would like to emphasize a very important difference between classical and quantum mechanics. We have been talking about the probability that an electron will arrive in a given circumstance. We have implied that in our experimental arrangement (or even in the best possible one) it would be impossible to predict exactly what would happen. We can only predict the odds! This would mean, if it were true, that physics has given up on the problem of trying to predict exactly what will happen in a definite circumstance. Yes! physics *has* given up. *We do not know how to predict what would happen in a given circumstance,* and we believe now that it is impossible—that the only thing that can be predicted is the probability of different events. It must be recognized that this is a retrenchment in our earlier ideal of understanding nature. It may be a backward step, but no one has seen a way to avoid it.

We make now a few remarks on a suggestion that has sometimes been made to try to avoid the description we have given: "Perhaps the electron has some kind of internal works—some inner variables—that we do not yet know about. Perhaps that is why we cannot predict what will happen. If we could look more closely at the electron, we could be able to tell where it would end up." So far as we know, that is impossible. We would still be in difficulty. Suppose we were to assume that inside the electron there is some kind of machinery that determines where it is going to end up. That machine must *also* determine which hole it is going to go through on its way. But we must not forget that what is inside the electron should not be dependent on what *we* do, and in particular upon whether we open or close one of the holes. So if an electron, before it starts, has already made up its mind (a) which hole it is going to use, and (b) where it is going to land, we should find P_1 for those electrons that have chosen hole 1, P_2 for those that have chosen hole 2, *and necessarily* the sum $P_1 + P_2$ for those that arrive through the two holes. There seems to be no way around this. But we have verified experimentally that that is not the case. And no one has figured a way out of this puzzle. So at the present time we must limit ourselves to computing probabilities. We say "at the present time," but we suspect very strongly that it is something that will be with us forever—that it is impossible to beat that puzzle—that this is the way nature really *is*.

1-8 The uncertainty principle

This is the way Heisenberg stated the uncertainty principle originally: If you make the measurement on any object, and you can determine the x-component of its momentum with an uncertainty Δp, you cannot, at the same time, know its x-position more accurately than $\Delta x = h/\Delta p$, where h is a definite fixed number given by nature. It is called "Planck's constant," and is approximately 6.63×10^{-34} joule-seconds. The uncertainties in the position and momentum of a particle at any instant must have their product greater than Planck's constant. This is a special case of the uncertainty principle that was stated above more generally. The more general statement was that one cannot design equipment in any way to determine which of two alternatives is taken, without, at the same time, destroying the pattern of interference.

Let us show for one particular case that the kind of relation given by Heisenberg must be true in order to keep from getting into trouble. We imagine a modification of the experiment of Fig. 1-3, in which the wall with the holes consists of a plate mounted on rollers so that it can move freely up and down (in the x-direction), as shown in Fig. 1-6. By watching the motion of the plate carefully we can try to tell which hole an electron goes through. Imagine what happens when the detector is placed at $x = 0$. We would expect that an electron which passes through hole 1 must be deflected downward by the plate to reach the detector. Since the vertical component of the electron momentum is changed, the plate must recoil with an equal momentum in the opposite direction. The plate will get an upward kick. If the electron goes through the lower hole, the plate should feel a downward kick. It is clear that for every position of the detector, the momentum received by the plate will have a different value for a traversal via hole 1 than for a traversal via hole 2. So! Without disturbing the electrons *at all*, but just by watching the *plate*, we can tell which path the electron used.

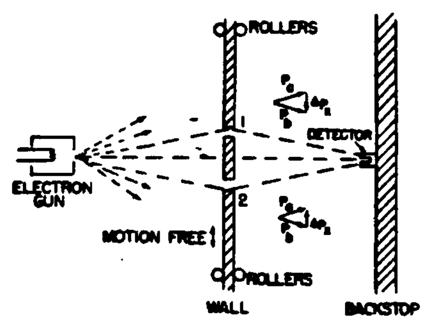

Fig. 1-6. An experiment in which the recoil of the wall is measured.

Now in order to do this it is necessary to know what the momentum of the screen is, before the electron goes through. So when we measure the momentum after the electron goes by, we can figure out how much the plate's momentum has changed. But remember, according to the uncertainty principle we cannot at the same time know the position of the plate with an arbitrary accuracy. But if we do not know exactly *where* the plate is, we cannot say precisely where the two holes are. They will be in a different place for every electron that goes through. This means that the center of our interference pattern will have a different location for each electron. The wiggles of the interference pattern will be smeared out. We shall show quantitatively in the next chapter that if we determine the momentum of the plate sufficiently accurately to determine from the recoil measurement which hole was used, then the uncertainty in the x-position of the plate will, according to the uncertainty principle, be enough to shift the pattern observed at the detector up and down in the x-direction about the distance from a maximum to its nearest minimum. Such a random shift is just enough to smear out the pattern so that no interference is observed.

The uncertainty principle "protects" quantum mechanics. Heisenberg recognized that if it were possible to measure the momentum and the position simultaneously with a greater accuracy, the quantum mechanics would collapse. So he proposed that it must be impossible. Then people sat down and tried to figure out ways of doing it, and nobody could figure out a way to measure the position and the momentum of anything—a screen, an electron, a billiard ball, anything—with any greater accuracy. Quantum mechanics maintains its perilous but still correct existence.

2

The Relation of Wave and Particle Viewpoints

Note: This chapter is almost exactly the same as Chapter 38 of Volume I.

2-1 Probability wave amplitudes

In this chapter we shall discuss the relationship of the wave and particle viewpoints. We already know, from the last chapter, that neither the wave viewpoint nor the particle viewpoint is correct. We would always like to present things accurately, or at least precisely enough that they will not have to be changed when we learn more—it may be extended, but it will not be changed! But when we try to talk about the wave picture or the particle picture, both are approximate, and both will change. Therefore what we learn in this chapter will not be accurate in a certain sense; we will deal with some half-intuitive arguments which will be made more precise later. But certain things will be changed a little bit when we interpret them correctly in quantum mechanics. We are doing this so that you can have some qualitative feeling for some quantum phenomena before we get into the mathematical details of quantum mechanics. Furthermore, all our experiences are with waves and with particles, and so it is rather handy to use the wave and particle ideas to get some understanding of what happens in given circumstances before we know the complete mathematics of the quantum-mechanical amplitudes. We shall try to indicate the weakest places as we go along, but most of it is very nearly correct—it is just a matter of interpretation.

First of all, we know that the new way of representing the world in quantum mechanics—the new framework—is to give an amplitude for every event that can occur, and if the event involves the reception of one particle, then we can give the amplitude to find that one particle at different places and at different times. The probability of finding the particle is then proportional to the absolute square of the amplitude. In general, the amplitude to find a particle in different places at different times varies with position and time.

In some special case it can be that the amplitude varies sinusoidally in space and time like $e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}$, where $\boldsymbol{r}$ is the vector position from some origin. (Do not forget that these amplitudes are complex numbers, not real numbers.) Such an amplitude varies according to a definite frequency ω and wave number $\boldsymbol{k}$. Then it turns out that this corresponds to a classical limiting situation where we would have believed that we have a particle whose energy E was known and is related to the frequency by

$$E = \hbar\omega, \tag{2.1}$$

and whose momentum $\boldsymbol{p}$ is also known and is related to the wave number by

$$\boldsymbol{p} = \hbar\boldsymbol{k}. \tag{2.2}$$

(The symbol $\hbar$ represents the number h divided by 2π; $\hbar = h/2\pi$.)

This means that the idea of a particle is limited. The idea of a particle—its location, its momentum, etc.—which we use so much, is in certain ways unsatisfactory. For instance, if an amplitude to find a particle at different places is given by $e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}$, whose absolute square is a constant, that would mean that the probability of finding a particle is the same at all points. That means we do not know *where* it is—it can be anywhere—there is a great uncertainty in its location.

On the other hand, if the position of a particle is more or less well known and we can predict it fairly accurately, then the probability of finding it in different places must be confined to a certain region, whose length we call Δx. Outside this region, the probability is zero. Now this probability is the absolute square of an amplitude, and if the absolute square is zero, the amplitude is also zero, so that we have a wave train whose length is Δx (Fig. 2-1), and the wavelength (the distance between nodes of the waves in the train) of that wave train is what corresponds to the particle momentum.

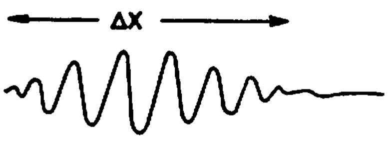

Fig. 2-1. A wave packet of length Δx.

Here we encounter a strange thing about waves; a very simple thing which has nothing to do with quantum mechanics strictly. It is something that anybody who works with waves, even if he knows no quantum mechanics, knows: namely, *we cannot define a unique wavelength for a short wave train.* Such a wave train does not *have* a definite wavelength; there is an indefiniteness in the wave number that is related to the finite length of the train, and thus there is an indefiniteness in the momentum.

2-2 Measurement of position and momentum

Let us consider two examples of this idea—to see the reason that there is an uncertainty in the position and/or the momentum, if quantum mechanics is right. We have also seen before that if there were not such a thing—if it were possible to measure the position and the momentum of anything simultaneously—we would have a paradox; it is fortunate that we do not have such a paradox, and the fact that such an uncertainty comes naturally from the wave picture shows that everything is mutually consistent.

Here is one example which shows the relationship between the position and the momentum in a circumstance that is easy to understand. Suppose we have a single slit, and particles are coming from very far away with a certain energy—so that they are all coming essentially horizontally (Fig. 2-2). We are going to concentrate on the vertical components of momentum. All of these particles have a certain horizontal momentum p_0, say, in a classical sense. So, in the classical sense, the vertical momentum p_y, before the particle goes through the hole, is definitely known. The particle is moving neither up nor down, because it came from a source that is far away—and so the vertical momentum is of course zero. But now let us suppose that it goes through a hole whose width is B. Then after it has come out through the hole, we know the position vertically—the y-position—with considerable accuracy—namely $\pm B$.† That is, the uncertainty in position, Δy, is of order B. Now we might also want to say, since we known the momentum is absolutely horizontal, that Δp_y is zero; but that is wrong. We *once* knew the momentum was horizontal, but we do not know it any more. Before the particles passed through the hole, we did not know their vertical positions. Now that we have found the vertical position by having the particle come through the hole, we have lost our information on the vertical momentum! Why? According to the wave theory, there is a spreading out, or diffraction, of the waves after they go through the slit, just as for light. Therefore there is a certain probability that particles coming out of the slit are not coming exactly straight. The pattern is spread out by the diffraction effect, and the angle of spread, which we can define as the angle of the first minimum, is a measure of the uncertainty in the final angle.

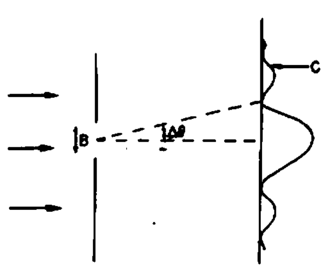

Fig. 2-2. Diffraction of particles passing through a slit.

How does the pattern become spread? To say it is spread means that there is some chance for the particle to be moving up or down, that is, to have a component of momentum up or down. We say *chance* and *particle* because we can detect this diffraction pattern with a particle counter, and when the counter receives the particle, say at C in Fig. 2-2, it receives the *entire* particle, so that, in a classical sense, the particle has a vertical momentum, in order to get from the slit up to C.

To get a rough idea of the spread of the momentum, the vertical momentum p_y has a spread which is equal to $p_0\,\Delta\theta$, where p_0 is the horizontal momentum. And how big is $\Delta\theta$ in the spread-out pattern? We know that the first minimum occurs at an angle $\Delta\theta$ such that the waves from one edge of the slit have to travel one wavelength farther than the waves from the other side—we worked that out before (Chapter 30 of Vol. I). Therefore $\Delta\theta$ is λ/B, and so Δp_y in this experiment is $p_0\lambda/B$. Note that if we make B smaller and make a more accurate measurement

† More precisely, the error in our knowledge of y is $\pm B/2$. But we are now only interested in the general idea, so we won't worry about factors of 2.

of the position of the particle, the diffraction pattern gets wider. So the narrower we make the slit, the wider the pattern gets, and the more is the likelihood that we would find that the particle has sidewise momentum. Thus the uncertainty in the vertical momentum is inversely proportional to the uncertainty of y. In fact, we see that the product of the two is equal to $p_0\lambda$. But λ is the wavelength and p_0 is the momentum, and in accordance with quantum mechanics, the wavelength times the momentum is Planck's constant h. So we obtain the rule that the uncertainties in the vertical momentum and in the vertical position have a product of the order h:

$$\Delta y\,\Delta p_y \approx h. \tag{2.3}$$

We cannot prepare a system in which we know the vertical position of a particle and can predict how it will move vertically with greater certainty than given by (2.3). That is, the uncertainty in the vertical momentum must exceed $h/\Delta y$, where Δy is the uncertainty in our knowledge of the position.

Sometimes people say quantum mechanics is all wrong. When the particle arrived from the left, its vertical momentum was zero. And now that it has gone through the slit, its position is known. Both position and momentum seem to be known with arbitrary accuracy. It is quite true that we can receive a particle, and on reception determine what its position is and what its momentum would have had to have been to have gotten there. That is true, but that is not what the uncertainty relation (2.3) refers to. Equation (2.3) refers to the *predictability* of a situation, not remarks about the *past*. It does no good to say "I knew what the momentum was before it went through the slit, and now I know the position," because now the momentum knowledge is lost. The fact that it went through the slit no longer permits us to predict the vertical momentum. We are talking about a predictive theory, not just measurements after the fact. So we must talk about what we can predict.

Now let us take the thing the other way around. Let us take another example of the same phenomenon, a little more quantitatively. In the previous example we measured the momentum by a classical method. Namely, we considered the direction and the velocity and the angles, etc., so we got the momentum by classical analysis. But since momentum is related to wave number, there exists in nature still another way to measure the momentum of a particle—photon or otherwise—which has no classical analog, because it uses Eq. (2.2). We measure the *wavelengths of the waves*. Let us try to measure momentum in this way.

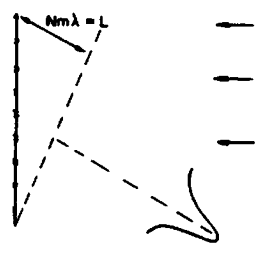

Fig. 2–3. Determination of momentum by using a diffraction grating.

Suppose we have a grating with a large number of lines (Fig. 2-3), and send a beam of particles at the grating. We have often discussed this problem: if the particles have a definite momentum, then we get a very sharp pattern in a certain direction, because of the interference. And we have also talked about how accurately we can determine that momentum, that is to say, what the resolving power of such a grating is. Rather than derive it again, we refer to Chapter 30 of Volume I, where we found that the relative uncertainty in the wavelength that can be measured with a given grating is $1/Nm$, where N is the number of lines on the grating and m is the order of the diffraction pattern. That is,

$$\Delta\lambda/\lambda = 1/Nm. \tag{2.4}$$

Now formula (2.4) can be rewritten as

$$\Delta\lambda/\lambda^2 = 1/Nm\lambda = 1/L, \tag{2.5}$$

where L is the distance shown in Fig. 2-3. This distance is the difference between the total distance that the particle or wave or whatever it is has to travel if it is reflected from the bottom of the grating, and the distance that it has to travel if it is reflected from the top of the grating. That is, the waves which form the diffraction pattern are waves which come from different parts of the grating. The first ones that arrive come from the bottom end of the grating, from the beginning of the wave train, and the rest of them come from later parts of the wave train, coming from different parts of the grating, until the last one finally arrives, and that involves a point in the wave train a distance L behind the first point. So in order that we shall have a sharp line in our spectrum corresponding to a definite momentum, with an uncertainty given by (2.4), we have to have a wave train of at least length L. If the wave train is too short, we are not using the entire grating. The waves which form the spectrum are being reflected from only a very short sector of the grating if the wave train is too short, and the grating will not work right—we will find a big angular spread. In order to get a narrower one, we need to use the whole grating, so that at least at some moment the whole wave train is scattering simultaneously from all parts of the grating. Thus the wave train must be of length L in order to have an uncertainty in the wavelength less than that given by (2.5). Incidentally,

$$\Delta\lambda/\lambda^2 = \Delta(1/\lambda) = \Delta k/2\pi. \tag{2.6}$$

Therefore

$$\Delta k = 2\pi/L, \tag{2.7}$$

where L is the length of the wave train.

This means that if we have a wave train whose length is less than L, the uncertainty in the wave number must exceed $2\pi/L$. Or the uncertainty in a wave number times the length of the wave train—we will call that for a moment Δx—exceeds 2π. We call it Δx because that is the uncertainty in the location of the particle. If the wave train exists only in a finite length, then that is where we could find the particle, within an uncertainty Δx. Now this property of waves, that the length of the wave train times the uncertainty of the wave number associated with it is at least 2π, is a property that is known to everyone who studies them. It has nothing to do with quantum mechanics. It is simply that if we have a finite train, we cannot count the waves in it very precisely.

Let us try another way to see the reason for that. Suppose that we have a finite train of length L; then because of the way it has to decrease at the ends, as in Fig. 2–1, the number of waves in the length L is uncertain by something like ± 1. But the number of waves in L is $kL/2\pi$. Thus k is uncertain, and we again get the result (2.7), a property merely of waves. The same thing works whether the waves are in space and k is the number of radians per centimeter and L is the length of the train, or the waves are in time and ω is the number of oscillations per second and T is the "length" in time that the wave train comes in. That is, if we have a wave train lasting only for a certain finite time T, then the uncertainty in the frequency is given by

$$\Delta\omega = 2\pi/T. \tag{2.8}$$

We have tried to emphasize that these are properties of waves alone, and they are well known, for example, in the theory of sound.

The point is that in quantum mechanics we interpret the wave number as being a measure of the momentum of a particle, with the rule that $p = \hbar k$, so that relation (2.7) tells us that $\Delta p \approx h/\Delta x$. This, then, is a limitation of the classical idea of momentum. (Naturally, it has to be limited in some ways if we are going to represent particles by waves!) It is nice that we have found a rule that gives us some idea of when there is a failure of classical ideas.

2–3 Crystal diffraction

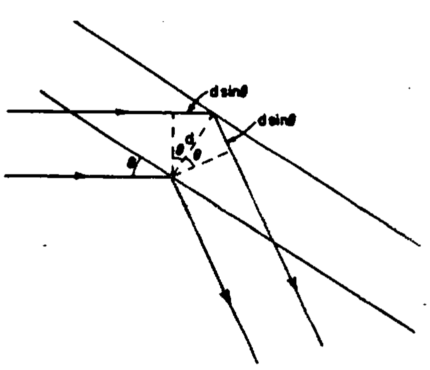

Fig. 2–4. Scattering of waves by crystal planes.

Next let us consider the reflection of particle waves from a crystal. A crystal is a thick thing which has a whole lot of similar atoms—we will include some complications later—in a nice array. The question is how to set the array so that we get a strong reflected maximum in a given direction for a given beam of, say, light (x-rays), electrons, neutrons, or anything else. In order to obtain a strong reflection, the scattering from all of the atoms must be in phase. There cannot be equal numbers in phase and out of phase, or the waves will cancel out. The way to arrange things is to find the regions of constant phase, as we have already explained; they are planes which make equal angles with the initial and final directions (Fig. 2–4).

If we consider two parallel planes, as in Fig. 2–4, the waves scattered from the two planes will be in phase, provided the difference in distance traveled by a wave front is an integral number of wavelengths. This difference can be seen to be $2d\sin\theta$, where d is the perpendicular distance between the planes. Thus the condition for coherent reflection is

$$2d\sin\theta = n\lambda \quad (n = 1, 2, \ldots). \tag{2.9}$$

If, for example, the crystal is such that the atoms happen to lie on planes obeying condition (2.9) with $n = 1$, then there will be a strong reflection. If, on the other hand, there are other atoms of the same nature (equal in density) halfway between, then the intermediate planes will also scatter equally strongly and will interfere with the others and produce no effect. So d in (2.9) must refer to *adjacent* planes; we cannot take a plane five layers farther back and use this formula!

As a matter of interest, actual crystals are not usually as simple as a single kind of atom repeated in a certain way. Instead, if we make a two-dimensional analog, they are much like wallpaper, in which there is some kind of figure which repeats all over the wallpaper. By "figure" we mean, in the case of atoms, some arrangement—calcium and a carbon and three oxygens, etc., for calcium carbonate, and so on—which may involve a relatively large number of atoms. But whatever it is, the figure is repeated in a pattern. This basic figure is called a *unit cell*.

The basic pattern of repetition defines what we call the *lattice type*; the lattice type can be immediately determined by looking at the reflections and seeing what their symmetry is. In other words, where we find any reflections *at all* determines the lattice type, but in order to determine what is in each of the elements of the lattice one must take into account the *intensity* of the scattering at the various directions. *Which* directions scatter depends on the type of lattice, but *how strongly* each scatters is determined by what is inside each unit cell, and in that way the structure of crystals is worked out.

Two photographs of x-ray diffraction patterns are shown in Figs. 2–5 and 2–6; they illustrate scattering from rock salt and myoglobin, respectively.

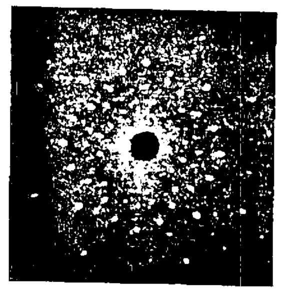

Fig. 2–5. The pattern produced by the diffraction of a beam of x-rays in a crystal of sodium chloride.

Fig. 2–6. The x-ray diffraction pattern of myoglobin.

Incidentally, an interesting thing happens if the spacings of the nearest planes are less than $\lambda/2$. In this case (2.9) has no solution for n. Thus if λ is bigger than twice the distance between adjacent planes, then there is no side diffraction pattern, and the light—or whatever it is—will go right through the material without bouncing off or getting lost. So in the case of light, where λ is much bigger than the spacing, of course it does go through and there is no pattern of reflection from the planes of the crystal.

This fact also has an interesting consequence in the case of piles which make neutrons (these are obviously particles, for anybody's money!). If we take these neutrons and let them into a long block of graphite, the neutrons diffuse and work their way along (Fig. 2–7). They diffuse because they are bounced by the atoms, but strictly, in the wave theory, they are bounced by the atoms because of diffraction from the crystal planes. It turns out that if we take a very long piece of graphite, the neutrons that come out the far end are all of long wavelength! In fact, if one plots the intensity as a function of wavelength, we get nothing except for wavelengths longer than a certain minimum (Fig. 2–8). In other words, we can get very slow neutrons that way. Only the slowest neutrons come through; they are not diffracted or scattered by the crystal planes of the graphite, but keep going right through like light through glass, and are not scattered out the sides. There are many other demonstrations of the reality of neutron waves and waves of other particles.

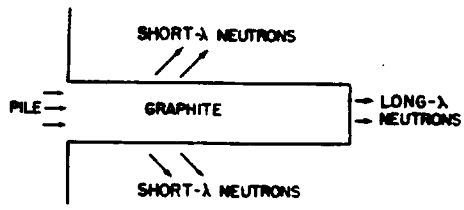

Fig. 2–7. Diffusion of pile neutrons through graphite block.

2–4 The size of an atom

We now consider another application of the uncertainty relation, Eq. (2.3). It must not be taken too seriously; the idea is right but the analysis is not very accurate. The idea has to do with the determination of the size of atoms, and the fact that, classically, the electrons would radiate light and spiral in until they settle down right on top of the nucleus. But that cannot be right quantum-mechanically because then we would know where each electron was and how fast it was moving.

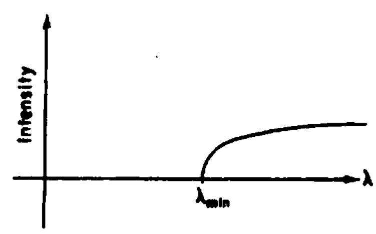

Fig. 2–8. Intensity of neutrons out of graphite rod as function of wavelength

Suppose we have a hydrogen atom, and measure the position of the electron; we must not be able to predict exactly where the electron will be, or the momentum spread will then turn out to be infinite. Every time we look at the electron, it is somewhere, but it has an amplitude to be in different places so there is a probability of it being found in different places. These places cannot all be at the nucleus; we shall suppose there is a spread in position of order a. That is, the distance of the electron from the nucleus is usually about a. We shall determine a by minimizing the total energy of the atom.

The spread in momentum is roughly h/a because of the uncertainty relation, so that if we try to measure the momentum of the electron in some manner, such as by scattering x-rays off it and looking for the Doppler effect from a moving scatterer, we would expect not to get zero every time—the electron is not standing still—but the momenta must be of the order $p \approx h/a$. Then the kinetic energy is roughly $\frac{1}{2}mv^2 = p^2/2m = h^2/2ma^2$. (In a sense, this is a kind of dimensional analysis to find out in what way the kinetic energy depends upon Planck's constant, upon m, and upon the size of the atom. We need not trust our answer to within factors like 2, π, etc. We have not even defined a very precisely.) Now the potential energy is minus e^2 over the distance from the center, say $-e^2/a$, where, as defined in Volume I, e^2 is the charge of an electron squared, divided by $4\pi\epsilon_0$. Now the point is that the potential energy is reduced if a gets smaller, but the smaller a is, the higher the momentum required, because of the uncertainty principle, and therefore the higher the kinetic energy. The total energy is

$$E = h^2/2ma^2 - e^2/a. \tag{2.10}$$

We do not know what a is, but we know that the atom is going to arrange itself to make some kind of compromise so that the energy is as little as possible. In order to minimize E, we differentiate with respect to a, set the derivative equal to zero, and solve for a. The derivative of E is

$$dE/da = -h^2/ma^3 + e^2/a^2, \tag{2.11}$$

and setting $dE/da = 0$ gives for a the value

$$\begin{aligned} a_0 = h^2/me^2 &= 0.528 \text{ angstrom} \\ &= 0.528 \times 10^{-10} \text{ meter.} \end{aligned} \tag{2.12}$$

This particular distance is called the *Bohr radius*, and we have thus learned that atomic dimensions are of the order of angstroms, which is right. This is pretty good—in fact, it is amazing, since until now we have had no basis for understanding the size of atoms! Atoms are completely impossible from the classical point of view, since the electrons would spiral into the nucleus.

Now if we put the value (2.12) for a_0 into (2.10) to find the energy, it comes out

$$E_0 = -e^2/2a_0 = -me^4/2h^2 = -13.6 \text{ ev.} \tag{2.13}$$

What does a negative energy mean? It means that the electron has less energy when it is in the atom than when it is free. It means it is bound. It means it takes energy to kick the electron out; it takes energy of the order of 13.6 ev to ionize a hydrogen atom. We have no reason to think that it is not two or three times this—or half of this—or $(1/\pi)$ times this, because we have used such a sloppy argument. However, we have cheated, we have used all the constants in such a way that it happens to come out the right number! This number, 13.6 electron volts, is called a Rydberg of energy; it is the ionization energy of hydrogen.

So we now understand why we do not fall through the floor. As we walk, our shoes with their masses of atoms push against the floor with *its* mass of atoms. In order to squash the atoms closer together, the electrons would be confined to a smaller space and, by the uncertainty principle, their momenta would have to be higher on the average, and that means high energy; the resistance to atomic compression is a quantum-mechanical effect and not a classical effect. Classically, we would expect that if we were to draw all the electrons and protons closer together, the energy would be reduced still further, and the best arrangement of positive and negative charges in classical physics is all on top of each other. This was well known in classical physics and was a puzzle because of the existence of the atom. Of course, the early scientists invented some ways out of the trouble—but never mind, we have the *right* way out, now!

Incidentally, although we have no reason to understand it at the moment, in a situation where there are many electrons it turns out that they try to keep away from each other. If one electron is occupying a certain space, then another does not occupy the same space. More precisely, there are two spin cases, so that two can sit on top of each other, one spinning one way and one the other way. But after that we cannot put any more there. We have to put others in another place, and that is the real reason that matter has strength. If we could put all the electrons in the same place, it would condense even more than it does. It is the fact that the electrons cannot all get on top of each other that makes tables and everything else solid.

Obviously, in order to understand the properties of matter, we will have to use quantum mechanics and not be satisfied with classical mechanics.

2-5 Energy levels

We have talked about the atom in its lowest possible energy condition, but it turns out that the electron can do other things. It can jiggle and wiggle in a more energetic manner, and so there are many different possible motions for the atom. According to quantum mechanics, in a stationary condition there can only be definite energies for an atom. We make a diagram (Fig. 2-9) in which we plot the energy vertically, and we make a horizontal line for each allowed value of the energy. When the electron is free, i.e., when its energy is positive, it can have any energy; it can be moving at any speed. But bound energies are not arbitrary. The atom must have one or another out of a set of allowed values, such as those in Fig. 2-9.

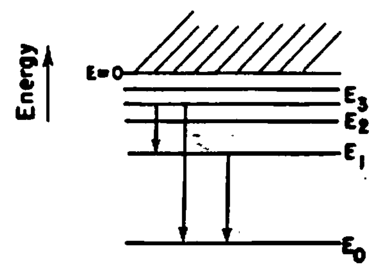

Fig. 2-9. Energy diagram for an atom, showing several possible transitions

Now let us call the allowed values of the energy E_0, E_1, E_2, E_3. If an atom is initially in one of these "excited states," E_1, E_2, *etc.*, it does not remain in that state forever. Sooner or later it drops to a lower state and radiates energy in the form of light. The frequency of the light that is emitted is determined by conservation of energy plus the quantum-mechanical understanding that the frequency of the light is related to the energy of the light by (2.1). Therefore the frequency of the light which is liberated in a transition from energy E_3 to energy E_1 (for example) is

$$\omega_{31} = (E_3 - E_1)/\hbar. \tag{2.14}$$

This, then, is a characteristic frequency of the atom and defines a spectral emission line. Another possible transition would be from E_3 to E_0. That would have a different frequency

$$\omega_{30} = (E_3 - E_0)/\hbar. \tag{2.15}$$

Another possibility is that if the atom were excited to the state E_1 it could drop to the ground state E_0, emitting a photon of frequency

$$\omega_{10} = (E_1 - E_0)/\hbar. \tag{2.16}$$

The reason we bring up three transitions is to point out an interesting relationship. It is easy to see from (2.14), (2.15), and (2.16) that

$$\omega_{30} = \omega_{31} + \omega_{10}. \tag{2.17}$$

In general, if we find two spectral lines, we shall expect to find another line at the sum of the frequencies (or the difference in the frequencies), and that all the lines can be understood by finding a series of levels such that every line corresponds to the difference in energy of some pair of levels. This remarkable coincidence in spectral frequencies was noted before quantum mechanics was discovered, and it is called the *Ritz combination principle*. This is again a mystery from the point of view of classical mechanics. Let us not belabor the point that classical mechanics is a failure in the atomic domain; we seem to have demonstrated that pretty well.

We have already talked about quantum mechanics as being represented by amplitudes which behave like waves, with certain frequencies and wave numbers. Let us observe how it comes about from the point of view of amplitudes that the atom has definite energy states. This is something we cannot understand from what has been said so far, but we are all familiar with the fact that confined waves have definite frequencies. For instance, if sound is confined to an organ pipe, or anything like that, then there is more than one way that the sound can vibrate, but for each such way there is a definite frequency. Thus an object in which the waves are confined has certain resonance frequencies. It is therefore a property of waves in a confined space—a subject which we will discuss in detail with formulas later on—that they exist only at definite frequencies. And since the general relation exists between frequencies of the amplitude and energy, we are not surprised to find definite energies associated with electrons bound in atoms.

2-6 Philosophical implications

Let us consider briefly some philosophical implications of quantum mechanics. As always, there are two aspects of the problem: one is the philosophical implication for physics, and the other is the extrapolation of philosophical matters to other fields. When philosophical ideas associated with science are dragged into another field, they are usually completely distorted. Therefore we shall confine our remarks as much as possible to physics itself.

First of all, the most interesting aspect is the idea of the uncertainty principle; making an observation affects the phenomenon. It has always been known that making observations affects a phenomenon, but the point is that the effect cannot be disregarded or minimized or decreased arbitrarily by rearranging the apparatus. When we look for a certain phenomenon we cannot help but disturb it in a certain minimum way, and *the disturbance is necessary for the consistency of the viewpoint.* The observer was sometimes important in prequantum physics, but only in a trivial sense. The problem has been raised: if a tree falls in a forest and there is nobody there to hear it, does it make a noise? A *real* tree falling in a *real* forest makes a sound, of course, even if nobody is there. Even if no one is present to hear it, there are other traces left. The sound will shake some leaves, and if we were careful enough we might find somewhere that some thorn had rubbed against a leaf and made a tiny scratch that could not be explained unless we assumed the leaf were vibrating. So in a certain sense we would have to admit that there is a sound made. We might ask: was there a *sensation* of sound? No, sensations have to do, presumably, with consciousness. And whether ants are conscious and whether there were ants in the forest, or whether the tree was conscious, we do not know. Let us leave the problem in that form.

Another thing that people have emphasized since quantum mechanics was developed is the idea that we should not speak about those things which we cannot measure. (Actually relativity theory also said this.) Unless a thing can be defined by measurement, it has no place in a theory. And since an accurate value of the momentum of a localized particle cannot be defined by measurement it therefore has no place in the theory. The idea that this is what was the matter with classical theory *is a false position*. It is a careless analysis of the situation. Just because we cannot *measure* position and momentum precisely does not *a priori* mean that we *cannot* talk about them. It only means that we *need* not talk about them. The situation in the sciences is this: A concept or an idea which cannot be measured or cannot be referred directly to experiment may or may not be useful. It need not exist in a theory. In other words, suppose we compare the classical theory of the world with the quantum theory of the world, and suppose that it is true experimentally that we can measure position and momentum only imprecisely. The question is whether the *ideas* of the exact position of a particle and the exact momentum of a particle are valid or not. The classical theory admits the ideas; the quantum theory does not. This does not in itself mean that classical physics is wrong. When the new quantum mechanics was discovered, the classical people—which included everybody except Heisenberg, Schrödinger, and Born—said: "Look, your theory is not any good because you cannot answer certain questions like: what is the exact position of a particle?, which hole does it go through?, and some others." Heisenberg's answer was: "I do not need to answer such questions because you cannot ask such a question experimentally." It is that we do not *have* to. Consider two theories (a) and (b); (a) contains an idea that cannot be checked directly but which is used in the analysis, and the other, (b), does not contain the idea. If they disagree in their predictions, one could not claim that (b) is false because it cannot explain this idea that is in (a), because that idea is one of the things that cannot be checked directly. It is always good to know which ideas cannot be checked directly, but it is not necessary to remove them all. It is not true that we can pursue science completely by using only those concepts which are directly subject to experiment.

In quantum mechanics itself there is a probability amplitude, there is a potential, and there are many constructs that we cannot measure directly. The basis of a science is its ability to *predict.* To predict means to tell what will happen in an experiment that has never been done. How can we do that? By assuming that we know what is there, independent of the experiment. We must extrapolate the experiments to a region where they have not been done. We must take our concepts and extend them to places where they have not yet been checked. If we do not do that, we have no prediction. So it was perfectly sensible for the classical physicists to go happily along and suppose that the position—which obviously means something for a baseball—meant something also for an electron. It was not stupidity. It was a sensible procedure. Today we say that the law of relativity is supposed to be true at all energies, but someday somebody may come along and say how stupid we were. We do not know where we are "stupid" until we "stick our neck out," and so the whole idea is to put our neck out. And the only way to find out that we are wrong is to find out *what* our predictions are. It is absolutely necessary to make constructs.

We have already made a few remarks about the indeterminacy of quantum mechanics. That is, that we are unable now to predict what will happen in physics in a given physical circumstance which is arranged as carefully as possible. If we have an atom that is in an excited state and so is going to emit a photon, we cannot say *when* it will emit the photon. It has a certain amplitude to emit the photon at any time, and we can predict only a probability for emission; we cannot predict the future exactly. This has given rise to all kinds of nonsense and questions on the meaning of freedom of will, and of the idea that the world is uncertain.

Of course we must emphasize that classical physics is also indeterminate, in a sense. It is usually thought that this indeterminacy, that we cannot predict the future, is an important quantum-mechanical thing, and this is said to explain the behavior of the mind, feelings of free will, etc. But if the world *were* classical—if the laws of mechanics were classical—it is not quite obvious that the mind would not feel more or less the same. It is true classically that if we knew the position and the velocity of every particle in the world, or in a box of gas, we could predict exactly what would happen. And therefore the classical world is deterministic. Suppose, however, that we have a finite accuracy and do not know *exactly* where just one atom is, say to one part in a billion. Then as it goes along it hits another atom, and because we did not know the position better than to one part in a billion, we find an even larger error in the position after the collision. And that is amplified, of course, in the next collision, so that if we start with only a tiny error it rapidly magnifies to a very great uncertainty. To give an example: if water falls over a dam, it splashes. If we stand nearby, every now and then a drop will land on our nose. This appears to be completely random, yet such a behavior would be predicted by purely classical laws. The exact position of all the drops depends upon the precise wigglings of the water before it goes over the dam. How? The tiniest irregularities are magnified in falling, so that we get complete randomness. Obviously, we cannot really predict the position of the drops unless we know the motion of the water *absolutely exactly*.

Speaking more precisely, given an arbitrary accuracy, no matter how precise, one can find a time long enough that we cannot make predictions valid for that long a time. Now the point is that this length of time is not very large. It is not that the time is millions of years if the accuracy is one part in a billion. The time goes, in fact, only logarithmically with the error, and it turns out that in only a very, very tiny time we lose all our information. If the accuracy is taken to be one part in billions and billions and billions—no matter how many billions we wish, provided we do stop somewhere—then we can find a time less than the time it took to state the accuracy—after which we can no longer predict what is going to happen! It is therefore not fair to say that from the apparent freedom and indeterminacy of the human mind, we should have realized that classical "deterministic" physics could not ever hope to understand it, and to welcome quantum mechanics as a release from a "completely mechanistic" universe. For already in classical mechanics there was indeterminability from a practical point of view.

3

Probability Amplitudes

3–1 The laws for combining amplitudes

When Schrödinger first discovered the correct laws of quantum mechanics, he wrote an equation which described the amplitude to find a particle in various places. This equation was very similar to the equations that were already known to classical physicists—equations that they had used in describing the motion of air in a sound wave, the transmission of light, and so on. So most of the time at the beginning of quantum mechanics was spent in solving this equation. But at the same time an understanding was being developed, particularly by Born and Dirac, of the basically new physical ideas behind quantum mechanics. As quantum mechanics developed further, it turned out that there were a large number of things which were not directly encompassed in the Schrödinger equation—such as the spin of the electron, and various relativistic phenomena. Traditionally, all courses in quantum mechanics have begun in the same way, retracing the path followed in the historical development of the subject. One first learns a great deal about classical mechanics so that he will be able to understand how to solve the Schrödinger equation. Then he spends a long time working out various solutions. Only after a detailed study of this equation does he get to the "advanced" subject of the electron's spin.

We had also originally considered that the right way to conclude these lectures on physics was to show how to solve the equations of classical physics in complicated situations—such as the description of sound waves in enclosed regions, modes of electromagnetic radiation in cylindrical cavities, and so on. That was the original plan for this course. However, we have decided to abandon that plan and to give instead an introduction to the quantum mechanics. We have come to the conclusion that what are usually called the advanced parts of quantum mechanics are, in fact, quite simple. The mathematics that is involved is particularly simple, involving simple algebraic operations and no differential equations or at most only very simple ones. The only problem is that we must jump the gap of no longer being able to describe the behavior *in detail* of particles in space. So this is what we are going to try to do: to tell you about what conventionally would be called the "advanced" parts of quantum mechanics. But they are, we assure you, by all odds the simplest parts—in a deep sense of the word—as well as the most basic parts. This is frankly a pedagogical experiment; it has never been done before, as far as we know.

In this subject we have, of course, the difficulty that the quantum mechanical behavior of things is quite strange. Nobody has an everyday experience to lean on to get a rough, intuitive idea of what will happen. So there are two ways of presenting the subject: We could either describe what can happen in a rather rough physical way, telling you more or less what happens without giving the precise laws of everything; or we could, on the other hand, give the precise laws in their abstract form. But, then because of the abstractions, you wouldn't know what they were all about, physically. The latter method is unsatisfactory because it is completely abstract, and the first way leaves an uncomfortable feeling because one doesn't know exactly what is true and what is false. We are not sure how to overcome this difficulty. You will notice, in fact, that Chapters 1 and 2 showed this problem. The first chapter was relatively precise; but the second chapter was a rough description of the characteristics of different phenomena. Here, we will try to find a happy medium between the two extremes.

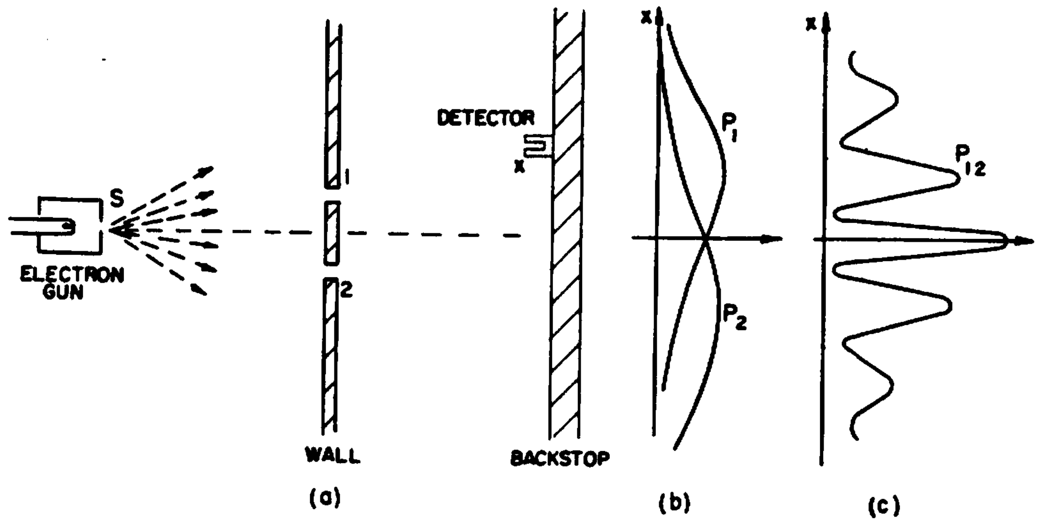

Fig 3–1. Interference experiment with electrons.

We will begin in this chapter by dealing with some general quantum mechanical ideas. Some of the statements will be quite precise, others only partially precise. It will be hard to tell you as we go along which is which, but by the time you have finished the rest of the book, you will understand in looking back which parts hold up and which parts were only explained roughly. The chapters which follow this one will not be so imprecise. In fact, one of the reasons we have tried carefully to be precise in the succeeding chapters is so that we can show you one of the most beautiful things about quantum mechanics—how much can be deduced from so little.

We begin by discussing again the superposition of *probability amplitudes.* As an example we will refer to the experiment described in Chapter 1, and shown again here in Fig. 3–1. There is a source s of particles, say electrons; then there is a wall with two slits in it; after the wall, there is a detector located at some position x. We ask for the probability that a particle will be found at x. Our *first general principle* in quantum mechanics is that the *probability* that a particle will arrive at x, when let out at the source s, can be represented quantitatively by the absolute square of a complex number called a *probability amplitude*—in this case, the "amplitude that a particle from s will arrive at x." We will use such amplitudes so frequently that we will use a shorthand notation—invented by Dirac and generally used in quantum mechanics—to represent this idea. We write the probability amplitude this way:

$$\langle \text{Particle arrives at } x \mid \text{particle leaves } s \rangle. \tag{3.1}$$

In other words, the two brackets $\langle\ \rangle$ are a sign equivalent to "the amplitude that"; the expression at the *right* of the vertical line always gives the *starting* condition, and the one at the left, the *final* condition. Sometimes it will also be convenient to abbreviate still more and describe the initial and final conditions by single letters. For example, we may on occasion write the amplitude (3.1) as

$$\langle x \mid s \rangle. \tag{3.2}$$

We want to emphasize that such an amplitude is, of course, just a single number—a *complex* number.

We have already seen in the discussion of Chapter 1 that when there are two ways for the particle to reach the detector, the resulting probability is not the sum of the two probabilities, but must be written as the absolute square of the sum of two amplitudes. We had that the probability that an electron arrives at the detector when both paths are open is

$$P_{12} = |\phi_1 + \phi_2|^2. \tag{3.3}$$

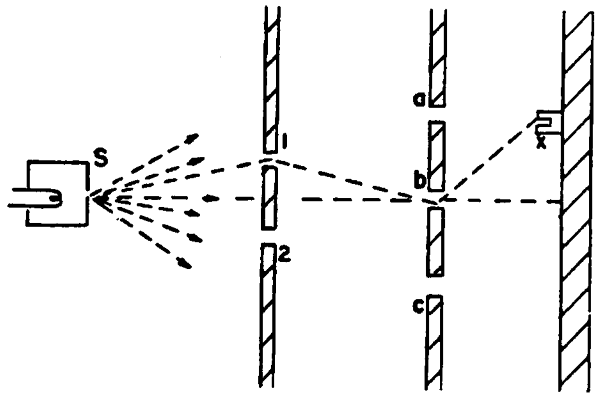

Fig. 3–2. A more complicated interference experiment.

We wish now to put this result in terms of our new notation. First, however, we want to state our *second general principle* of quantum mechanics: When a particle can reach a given state by two possible routes, the total amplitude for the process is the *sum of the amplitudes* for the two routes considered separately. In our new notation we write that

$$\langle x \mid s\rangle_{\text{both holes open}} = \langle x \mid s\rangle_{\text{through 1}} + \langle x \mid s\rangle_{\text{through 2}}. \tag{3.4}$$

Incidentally, we are going to suppose that the holes 1 and 2 are small enough that when we say an electron goes through the hole, we don't have to discuss which part of the hole. We could, of course, split each hole into pieces with a certain amplitude that the electron goes to the top of the hole and the bottom of the hole and so on. We will suppose that the hole is small enough so that we don't have to worry about this detail. That is part of the roughness involved; the matter can be made more precise, but we don't want to do so at this stage.

Now we want to write out in more detail what we can say about the amplitude for the process in which the electron reaches the detector at x by way of hole 1. We can do that by using our *third general principle*: When a particle goes by some particular route the amplitude for that route can be written as the *product* of the *amplitude* to go part way with the *amplitude* to go the rest of the way. For the setup of Fig. 3–1 the amplitude to go from s to x by way of hole 1 is equal to the amplitude to go from s to 1, multiplied by the amplitude to go from 1 to x.

$$\langle x \mid s\rangle_{\text{via 1}} = \langle x \mid 1\rangle\langle 1 \mid s\rangle. \tag{3.5}$$

Again this result is not completely precise. We should also include a factor for the amplitude that the electron will get through the hole at 1; but in the present case it is a simple hole, and we will take this factor to be unity.

You will note that Eq. (3.5) appears to be written in reverse order. It is to be read from right to left: The electron goes from s to 1 and then from 1 to x. In summary, if events occur in succession—that is, if you can analyze one of the routes of the particle by saying it does this, then it does this, then it does that—the resultant amplitude for that route is calculated by multiplying in succession the amplitude for each of the successive events. Using this law we can rewrite Eq. (3.4) as

$$\langle x \mid s\rangle_{\text{both}} = \langle x \mid 1\rangle\langle 1 \mid s\rangle + \langle x \mid 2\rangle\langle 2 \mid s\rangle.$$

Now we wish to show that just using these principles we can calculate a much more complicated problem like the one shown in Fig. 3–2. Here we have two walls, one with two holes, 1 and 2, and another which has three holes, a, b, and c. Behind the second wall there is a detector at x, and we want to know the amplitude for a particle to arrive there. Well, one way you can find this is by calculating the superposition, or interference, of the waves that go through; but you can also do it by saying that there are six possible routes and superposing an amplitude for each. The electron can go through hole 1, then through hole a, and then to x; or it could go through hole 1, then through hole b, and then to x; and so on. According to our second principle, the amplitudes for alternative routes add, so we should be able to write the amplitude from s to x as a sum of six separate amplitudes. On the other hand, using the third principle, each of these separate amplitudes can be written as a product of three amplitudes. For example, one of them is the amplitude for s to 1, times the amplitude for 1 to a, times the amplitude for a to x. Using our shorthand notation, we can write the complete amplitude to go from s to x as

$$\langle x \mid s\rangle = \langle x \mid a\rangle\langle a \mid 1\rangle\langle 1 \mid s\rangle + \langle x \mid b\rangle\langle b \mid 1\rangle\langle 1 \mid s\rangle + \cdots + \langle x \mid c\rangle\langle c \mid 2\rangle\langle 2 \mid s\rangle.$$

We can save writing by using the summation notation

$$\langle x \mid s\rangle = \sum_{\substack{i=1,2\\ \alpha=a,b,c}} \langle x \mid \alpha\rangle\langle \alpha \mid i\rangle\langle i \mid s\rangle. \tag{3.6}$$

In order to make any calculations using these methods, it is, naturally, necessary to know the amplitude to get from one place to another. We will give a rough idea of a typical amplitude. It leaves out certain things like the polarization of light or the spin of the electron, but aside from such features it is quite accurate. We give it so that you can solve problems involving various combinations of slits. Suppose a particle with a definite energy is going in empty space from a location $\boldsymbol{r}_1$ to a location $\boldsymbol{r}_2$. In other words, it is a free particle with no forces on it. Except for a numerical factor in front, the amplitude to go from $\boldsymbol{r}_1$ to $\boldsymbol{r}_2$ is

$$\langle \boldsymbol{r}_2 \mid \boldsymbol{r}_1\rangle = \frac{e^{i\boldsymbol{p}\cdot\boldsymbol{r}_{12}/\hbar}}{r_{12}}, \tag{3.7}$$

where $\boldsymbol{r}_{12} = \boldsymbol{r}_2 - \boldsymbol{r}_1$, and p is the momentum which is related to the energy E by the relativistic equation

$$p^2c^2 = E^2 - (m_0c^2)^2,$$

or the nonrelativistic equation

$$\frac{p^2}{2m} = \text{Kinetic energy.}$$

Equation (3.7) says in effect that the particle has wavelike properties, the amplitude propagating as a wave with a wave number equal to the momentum divided by $\hbar$.

In the most general case, the amplitude and the corresponding probability will also involve the time. For most of these initial discussions we will suppose that the source always emits the particles with a given energy so we will not need to worry about the time. But we could, in the general case, be interested in some other questions. Suppose that a particle is liberated at a certain place P at a certain time, and you would like to know the amplitude for it to arrive at some location, say $\boldsymbol{r}$, at some later time. This could be represented symbolically as the amplitude $\langle \boldsymbol{r}, t = t_1 \mid P, t = 0\rangle$. Clearly, this will depend upon both $\boldsymbol{r}$ and t. You will get different results if you put the detector in different places and measure at different times. This function of $\boldsymbol{r}$ and t, in general, satisfies a differential equation which is a wave equation. For example, in a nonrelativistic case it is the Schrödinger equation. One has then a wave equation analogous to the equation for electromagnetic waves or waves of sound in a gas. However, it must be emphasized that the wave function that satisfies the equation is not like a real wave in space; one cannot picture any kind of reality to this wave as one does for a sound wave.

Although one may be tempted to think in terms of "particle waves" when dealing with one particle, it is not a good idea, for if there are, say, two particles, the amplitude to find one at $\boldsymbol{r}_1$ and the other at $\boldsymbol{r}_2$ is not a simple wave in three-dimensional space, but depends on the *six* space variables $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. If we are, for example, dealing with two (or more) particles, we will need the following additional principle: Provided that the two particles do not interact, the amplitude that one particle will do one thing *and* the other one something else is the product of the two amplitudes that the two particles would do the two things separately. For example, if $\langle a \mid s_1\rangle$ is the amplitude for particle 1 to go from s_1 to a, and $\langle b \mid s_2\rangle$

is the amplitude for particle 2 to go from s_2 to b, the amplitude that *both* things will happen together is

$$\langle a \mid s_1\rangle\langle b \mid s_2\rangle.$$

There is one more point to emphasize. Suppose that we didn't know where the particles in Fig. 3-2 come from before arriving at holes 1 and 2 of the first wall. We can still make a prediction of what will happen beyond the wall (for example, the amplitude to arrive at x) provided that we are given two numbers: the amplitude to have arrived at 1 and the amplitude to have arrived at 2. In other words, because of the fact that the amplitude for successive events multiplies, as shown in Eq. (3.6), all you need to know to continue the analysis is two numbers—in this particular case $\langle 1 \mid s\rangle$ and $\langle 2 \mid s\rangle$. These two complex numbers are enough to predict all the future. That is what really makes quantum mechanics easy. It turns out that in later chapters we are going to do just such a thing when we specify a starting condition in terms of two (or a few) numbers. Of course, these numbers depend upon where the source is located and possibly other details about the apparatus, but given the two numbers, we do not need to know any more about such details.

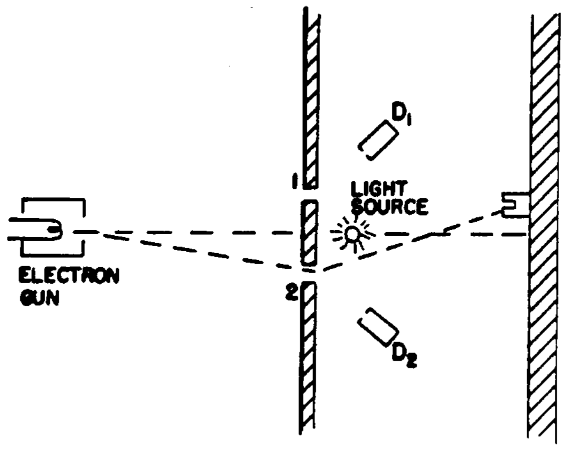

Fig. 3-3. An experiment to determine which hole the electron goes through.

3-2 The two-slit interference pattern

Now we would like to consider a matter which was discussed in some detail in Chapter 1. This time we will do it with the full glory of the amplitude idea to show you how it works out. We take the same experiment shown in Fig. 3-1, but now with the addition of a light source behind the two holes, as shown in Fig. 3-3. In Chapter 1, we discovered the following interesting result. If we looked behind slit 1 and saw a photon scattered from there, then the distribution obtained for the electrons at x in coincidence with these photons was the same as though slit 2 were closed. The total distribution for electrons that had been "seen" at either slit 1 or slit 2 was the sum of the separate distributions and was completely different from the distribution with the light turned off. This was true at least if we used light of short enough wavelength. If the wavelength was made longer so we could not be sure at which hole the scattering had occurred, the distribution became more like the one with the light turned off.

Let's examine what is happening by using our new notation and the principles of combining amplitudes. To simplify the writing, we can again let ϕ_1 stand for the amplitude that the electron will arrive at x by way of hole 1, that is,

$$\phi_1 = \langle x \mid 1\rangle\langle 1 \mid s\rangle.$$

Similarly, we'll let ϕ_2 stand for the amplitude that the electron gets to the detector by way of hole 2:

$$\phi_2 = \langle x \mid 2\rangle\langle 2 \mid s\rangle.$$

These are the amplitudes to go through the two holes and arrive at x if there is no light. Now if there is light, we ask ourselves the question: What is the amplitude for the process in which the electron starts at s and a photon is liberated by the light source L, ending with the electron at x and a photon seen behind slit 1? Suppose that we observe the photon behind slit 1 by means of a detector D_1, as shown in Fig. 3-3, and use a similar detector D_2 to count photons scattered behind hole 2. There will be an amplitude for a photon to arrive at D_1 and an electron at x, and also an amplitude for a photon to arrive at D_2 and an electron at x. Let's try to calculate them.

Although we don't have the correct mathematical formula for all the factors that go into this calculation, you will see the spirit of it in the following discussion. First, there is the amplitude $\langle 1 \mid s \rangle$ that an electron goes from the source to hole 1. Then we can suppose that there is a certain amplitude that while the electron is at hole 1 it scatters a photon into the detector D_1. Let us represent this amplitude by a. Then there is the amplitude $\langle x \mid 1 \rangle$ that the electron goes from slit 1 to the electron detector at x. The amplitude that the electron goes from s to x via slit 1 *and* scatters a photon into D_1 is then

$$\langle x \mid 1 \rangle \, a \, \langle 1 \mid s \rangle.$$

Or, in our previous notation, it is just $a\phi_1$.

There is also some amplitude that an electron going through slit 2 will scatter a photon into counter D_1. You say, "That's impossible; how can it scatter into counter D_1 if it is only looking at hole 1?" If the wavelength is long enough, there are diffraction effects, and it is certainly possible. If the apparatus is built well and if we use photons of short wavelength, then the amplitude that a photon will be scattered into detector 1, from an electron at 2 is very small. But to keep the discussion general we want to take into account that there is always some such amplitude, which we will call b. Then the amplitude that an electron goes via slit 2 *and* scatters a photon into D_1 is

$$\langle x \mid 2 \rangle \, b \, \langle 2 \mid s \rangle = b\phi_2.$$

The amplitude to find the electron at x and the photon in D_1 is the sum of two terms, one for each possible path for the electron. Each term is in turn made up of two factors: first, that the electron went through a hole, and second, that the photon is scattered by such an electron into detector 1; we have

$$\left\langle \begin{matrix} \text{electron at } x \\ \text{photon at } D_1 \end{matrix} \middle| \begin{matrix} \text{electron from } s \\ \text{photon from } L \end{matrix} \right\rangle = a\phi_1 + b\phi_2. \tag{3.8}$$

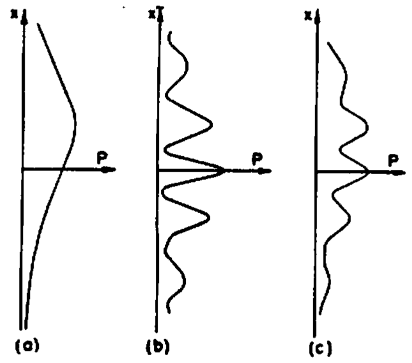

Fig. 3-4. The probability of counting an electron at x in coincidence with a photon at D in the experiment of Fig. 3-3: (a) for $b = 0$; (b) for $b = a$; (c) for $0 < b < a$.

We can get a similar expression when the photon is found in the other detector D_2. If we assume for simplicity that the system is symmetrical, then a is also the amplitude for a photon in D_2 when an electron passes through hole 2, and b is the amplitude for a photon in D_2 when the electron passes through hole 1. The corresponding total amplitude for a photon at D_2 and an electron at x is

$$\left\langle \begin{matrix} \text{electron at } x \\ \text{photon at } D_2 \end{matrix} \middle| \begin{matrix} \text{electron from } s \\ \text{photon from } L \end{matrix} \right\rangle = a\phi_2 + b\phi_1. \tag{3.9}$$

Now we are finished. We can easily calculate the probability for various situations. Suppose that we want to know with what probability we get a count in D_1 and an electron at x. That will be the absolute square of the amplitude given in Eq. (3.8), namely, just $|a\phi_1 + b\phi_2|^2$. Let's look more carefully at this expression. First of all, if b is zero—which is the way we would like to design the apparatus—then the answer is simply $|\phi_1|^2$ diminished in total amplitude by the factor $|a|^2$. This is the probability distribution that you would get if there were only one hole—as shown in the graph of Fig. 3-4(a). On the other hand, if the wavelength is very long, the scattering behind hole 2 into D_1 may be just about the same as for hole 1. Although there may be some phases involved in a and b, we can ask about a simple case in which the two phases are equal. If a is practically equal to b, then the total probability becomes $|\phi_1 + \phi_2|^2$ multiplied by $|a|^2$, since the common factor a can be taken out. This, however, is just the probability distribution we would have gotten without the photons at all. Therefore, in the case that the wavelength is very long—and the photon detection ineffective—you return to the original distribution curve which shows interference effects, as shown in Fig. 3–4(b). In the case that the detection is partially effective, there is an interference between a lot of ϕ_1 and a little of ϕ_2, and you will get an intermediate distribution such as is sketched in Fig. 3–4(c). Needless to say, if we look for coincidence counts of photons at D_2 and electrons at x, we will get the same kinds of results. If you remember the discussion in Chapter 1, you will see that these results give a quantitative description of what was described there.

Now we would like to emphasize an important point so that you will avoid a common error. Suppose that you only want the amplitude that the electron arrives at x, *regardless* of whether the photon was counted at D_1 or D_2. Should you add the amplitudes given in Eqs. (3.8) and (3.9)? No! You must *never add amplitudes for different and distinct final states.* Once the photon is accepted by one of the photon counters, we can always determine which alternative occurred if we want, without any further disturbance to the system. Each alternative has a probability completely independent of the other. To repeat, do not add amplitudes for different *final* conditions, where by "final" we mean at that moment the *probability* is desired—that is, when the experiment is "finished." You do add the amplitudes for the different *indistinguishable* alternatives inside the experiment, before the complete process is finished. At the end of the process you may say that you "don't want to look at the photon." That's your business, but you still do not add the amplitudes. Nature does not know what you are looking at, and she behaves the way she is going to behave whether you bother to take down the data or not. So here we must not add the amplitudes. We first square the amplitudes for all possible different final events and then sum. The correct result for an electron at x and a photon at either D_1 or D_2 is

$$\left|\left\langle \begin{matrix}\text{e at } x\\ \text{ph at } D_1\end{matrix} \middle| \begin{matrix}\text{e from } s\\ \text{ph from } L\end{matrix}\right\rangle\right|^2 + \left|\left\langle \begin{matrix}\text{e at } x\\ \text{ph at } D_2\end{matrix} \middle| \begin{matrix}\text{e from } s\\ \text{ph from } L\end{matrix}\right\rangle\right|^2 = |a\phi_1 + b\phi_2|^2 + |a\phi_2 + b\phi_1|^2. \tag{3.10}$$

3–3 Scattering from a crystal

Our next example is a phenomenon in which we have to analyze the interference of probability amplitudes somewhat carefully. We look at the process of the scattering of neutrons from a crystal. Suppose we have a crystal which has a lot of atoms with nuclei at their centers, arranged in a periodic array, and a neutron beam that comes from far away. We can label the various nuclei in the crystal by an index i, where i runs over the integers 1, 2, 3, ... N, with N equal to the total number of atoms. The problem is to calculate the probability of getting a neutron into a counter with the arrangement shown in Fig. 3–5. For any particular atom i, the amplitude that the neutron arrives at the counter C is the amplitude that the neutron gets from the source S to nucleus i, multiplied by the amplitude a that it gets scattered there, multiplied by the amplitude that it gets from i to the counter C. Let's write that down:

$$\langle \text{neutron at } C \mid \text{neutron from } S\rangle_{\text{via } i} = \langle C \mid i\rangle\, a\, \langle i \mid S\rangle. \tag{3.11}$$

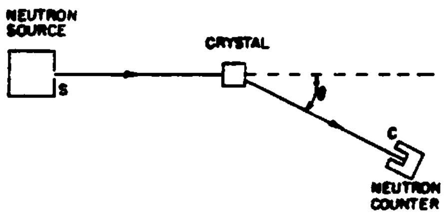

Fig. 3–5. Measuring the scattering of neutrons by a crystal.

In writing this equation we have assumed that the scattering amplitude a is the same for all atoms. We have here a large number of apparently indistinguishable routes. They are indistinguishable because a low-energy neutron is scattered from a nucleus without knocking the atom out of its place in the crystal—no "record" is left of the scattering. According to the earlier discussion, the total amplitude for a neutron at C involves a sum of Eq. (3.11) over all the atoms:

$$\langle \text{neutron at } C \mid \text{neutron from } S\rangle = \sum_{i=1}^{N} \langle C \mid i\rangle\, a\, \langle i \mid S\rangle. \tag{3.12}$$

Because we are adding amplitudes of scattering from atoms with different space positions, the amplitudes will have different phases giving the characteristic interference pattern that we have already analyzed in the case of the scattering of light from a grating.

The neutron intensity as a function of angle in such an experiment is indeed often found to show tremendous variations, with very sharp interference peaks and almost nothing in between—as shown in Fig. 3–6(a). However, for certain kinds of crystals it does not work this way, and there is—along with the interference peaks discussed above—a general background of scattering in all directions. We must try to understand the apparently mysterious reasons for this. Well, we have not considered one important property of the neutron. It has a spin of one-half, and so there are two conditions in which it can be: either spin "up" (say perpendicular to the page in Fig. 3–5) or spin "down." If the nuclei of the crystal have no spin, the neutron spin doesn't have any effect. But when the nuclei of the crystal also have a spin, say a spin of one-half, you will observe the background of smeared-out scattering described above. The explanation is as follows.

If the neutron has one direction of spin and the atomic nucleus has the same spin, then no change of spin can occur in the scattering process. If the neutron and atomic nucleus have opposite spin, then scattering can occur by two processes, one in which the spins are unchanged and another in which the spin directions are exchanged. This rule for no net change of the sum of the spins is analogous to our classical law of conservation of angular momentum. We can begin to understand the phenomenon if we assume that all the scattering nuclei are set up with spins in one direction. A neutron with the same spin will scatter with the expected sharp interference distribution. What about one with opposite spin? If it scatters without spin flip, then nothing is changed from the above; but if the two spins flip over in the scattering, we could, in principle, find out which nucleus had done the scattering, since it would be the only one with spin turned over. Well, if we can tell which atom did the scattering, what have the other atoms got to do with it? Nothing, of course. The scattering is exactly the same as that from a single atom.

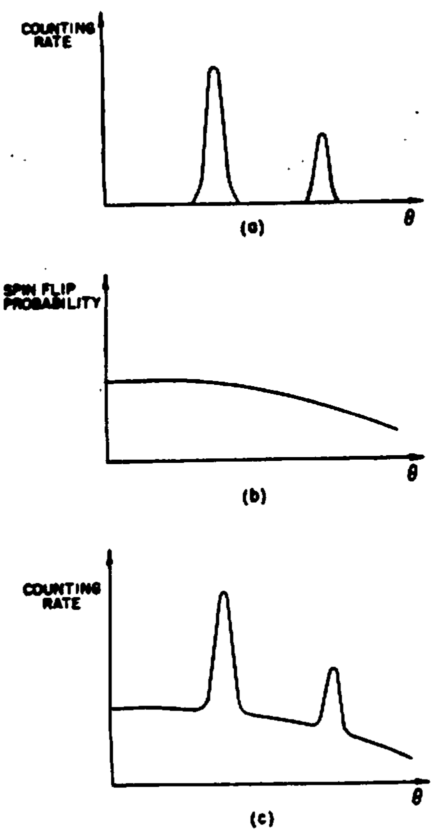

Fig. 3–6. The neutron counting rate as a function of angle: (a) for spin zero nuclei; (b) the probability of scattering with spin flip; (c) the observed counting rate with a spin one-half nucleus.

To include this effect, the mathematical formulation of Eq. (3.12) must be modified since we haven't described the states completely in that analysis. Let's start with all neutrons from the source having spin up and all the nuclei of the crystal having spin down. First, we would like the amplitude that at the counter the spin of the neutron is up *and* all spins of the crystal are still down. This is not different from our previous discussion. We will let a be the amplitude to scatter with no flip or spin. The amplitude for scattering from the ith atom is, of course,

$$\langle C_{\text{up}}, \text{crystal all down} \mid S_{\text{up}}, \text{crystal all down}\rangle = \langle C \mid i\rangle\, a\, \langle i \mid S\rangle.$$

Since all the atomic spins are still down, the various alternatives (different values of i) cannot be distinguished. There is clearly no way to tell which atom did the scattering. For this process, all the amplitudes interfere.

We have another case, however, where the spin of the detected neutron is down although it started from S with spin up. In the crystal, one of the spins must be changed to the up direction—let's say that of the kth atom. We will assume that there is the same scattering amplitude with spin flip for every atom, namely b. (In a real crystal there is the disagreeable possibility that the reversed spin moves to some other atom, but let's take the case of a crystal for which this probability is very low.) The scattering amplitude is then

$$\langle C_{\text{down}}, \text{nucleus } k \text{ up} \mid S_{\text{up}}, \text{crystal all down}\rangle = \langle C \mid k\rangle\, b\, \langle k \mid S\rangle. \tag{3.13}$$

If we ask for the probability of finding the neutron spin down and the kth nucleus spin up, it is equal to the absolute square of this amplitude, which is simply $|b|^2$ times $|\langle C \mid k\rangle\langle k \mid S\rangle|^2$. The second factor is almost independent of location in the crystal, and all phases have disappeared in taking the absolute square. The probability of scattering from *any nucleus* in the crystal with spin flip is now

$$|b|^2 \sum_{k=1}^{N} |\langle C \mid k\rangle\langle k \mid S\rangle|^2,$$

which will show a smooth distribution as in Fig. 3–6(b).

You may argue, "I don't care which atom is up." Perhaps you don't, but nature knows; and the probability is, in fact, what we gave above—there isn't any interference. On the other hand, if we ask for the probability that the spin is up at the detector and all the atoms still have spin down, then we must take the absolute square of

$$\sum_{i=1}^{N} \langle C \mid i\rangle\, a\, \langle i \mid S\rangle.$$

Since the terms in this sum have phases, they do interfere, and we get a sharp interference pattern. If we do an experiment in which we don't observe the spin of the detected neutron, then both kinds of events can occur; and the separate probabilities add. The total probability (or counting rate) as a function of angle then looks like the graph in Fig. 3–6(c).

Let's review the physics of this experiment. If you could, *in principle*, distinguish the alternative *final* states (even though you do not bother to do so), the total, final probability is obtained by calculating the *probability* for each state (not the amplitude) and then adding them together. If you *cannot* distinguish the final states *even in principle*, then the probability amplitudes must be summed before taking the absolute square to find the actual probability. The thing you should notice particularly is that if you were to try to represent the neutron by a wave alone, you would get the same kind of distribution for the scattering of a down-spinning neutron as for an up-spinning neutron. You would have to say that the "wave" would come from all the different atoms and interfere just as for the up-spinning one with the same wavelength. But we know that is not the way it works. So as we stated earlier, we must be careful not to attribute too much reality to the waves in space. They are useful for certain problems but not for all.

3–4 Identical particles

The next experiment we will describe is one which shows one of the beautiful consequences of quantum mechanics. It again involves a physical situation in which a thing can happen in two *indistinguishable* ways, so that there is an interference of amplitudes—as is *always* true in such circumstances. We are going to discuss the scattering, at relatively low energy, of nuclei on other nuclei. We start by thinking of α-particles (which, as you know, are helium nuclei) bombarding, say, oxygen. To make it easier for us to analyze the reaction, we will look at it in the center-of-mass system, in which the oxygen nucleus and the α-particle have their velocities in opposite directions before the collision and again in exactly opposite directions after the collision. See Fig. 3–7(a). (The magnitudes of the velocities are, of course, different, since the masses are different.) We will also suppose that there is conservation of energy and that the collision energy is low enough that neither particle is broken up or left in an excited state. The reason that the two particles deflect each other is, of course, that each particle carries a positive charge and, classically speaking, there is an electrical repulsion as they go by. The scattering will happen at different angles with different probabilities, and we would like to discuss something about the angle dependence of such scatterings. (It is possible, of course, to calculate this thing classically, and it is one of the most remarkable accidents of quantum mechanics that the answer to this problem comes out the same as it does classically. This is a curious point because it happens for no other force except the inverse square law—so it is indeed an accident.)

The probability of scattering in different directions can be measured by an experiment as shown in Fig. 3–7(a). The counter at position 1 could be designed to detect only α-particles; the counter at position 2 could be designed to detect

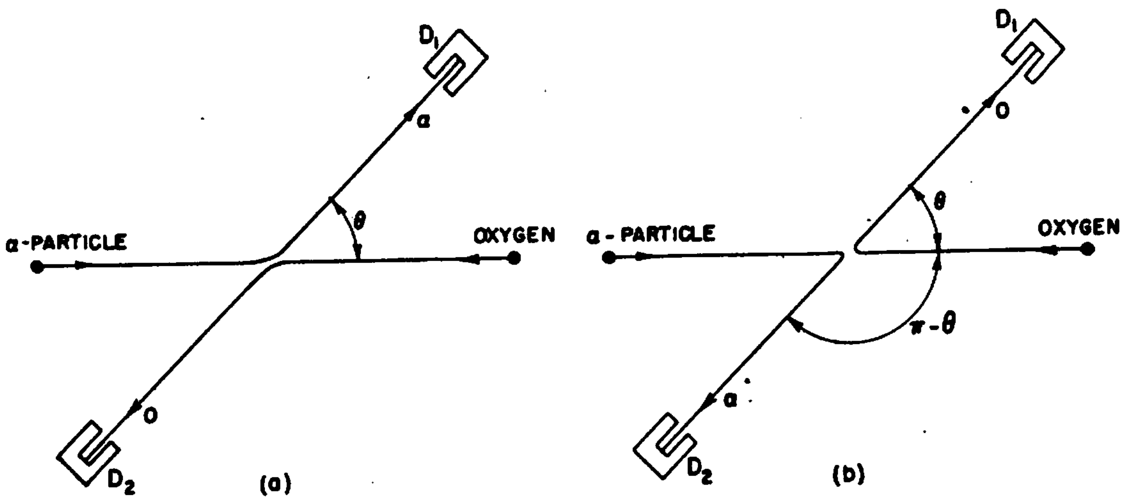

Fig. 3–7. The scattering of α-particles from oxygen nuclei, as seen in the center-of-mass system.

only oxygen—just as a check. (In the laboratory system the detectors would not be opposite; but in the CM system they are.) Our experiment consists in measuring the probability of scattering in various directions. Let's call $f(\theta)$ the amplitude to scatter into the counters when they are at the angle θ; then $|f(\theta)|^2$ will be our experimentally determined probability.

Now we could set up another experiment in which our counters would respond to *either* the α-particle *or* the oxygen nucleus. Then we have to work out what happens when we do not bother to distinguish which particles are counted. Of course, if we are to get an oxygen in the position θ, there must be an α-particle on the opposite side at the angle $(\pi - \theta)$, as shown in Fig. 3–7(b). So if $f(\theta)$ is the amplitude for α-scattering through the angle θ, then $f(\pi - \theta)$ is the amplitude for oxygen scattering through the angle θ.† Thus, the probability for having *some* particle in the detector at position 1 is:

$$\text{Probability of } \textit{some} \text{ particle in } D_1 = |f(\theta)|^2 + |f(\pi - \theta)|^2. \qquad (3.14)$$

Note that the two states are distinguishable in principle. Even though in this experiment we *do not* distinguish them, *we could*. According to the earlier discussion, then, we must add the probabilities, not the amplitudes.

The result given above is correct for a variety of target nuclei—for α-particles on oxygen, on carbon, on beryllium, on hydrogen. *But it is wrong for* α-particles on α-particles. For the one case in which both particles are exactly the same, the experimental data disagree with the prediction of (3.14). For example, the scattering probability at 90° is exactly twice what the above theory predicts and has nothing to do with the particles being "helium" nuclei. If the target is He^3, but the projectiles are α-particles (He^4), then there is agreement. Only when the target is He^4—so its nuclei are identical with the incoming α-particle—does the scattering vary in a peculiar way with angle.

Perhaps you can already see the explanation. There are two ways to get an α-particle into the counter: by scattering the bombarding α-particle at an angle θ, or by scattering it at an angle of $(\pi - \theta)$. How can we tell whether the bombarding particle or the target particle entered the counter? The answer is that we cannot. In the case of α-particles with α-particles there are two alternatives that cannot be distinguished. Here, we must let the probability *amplitudes* interfere by addition,

† In general, a scattering direction should, of course, be described by two angles, the polar angle ϕ, as well as the azimuthal angle θ. We would then say that an oxygen nucleus at (θ, ϕ) means that the α-particle is at $(\pi - \theta, \phi + \pi)$. However, for Coulomb scattering (and for many other cases), the scattering amplitude is independent of ϕ. Then the amplitude to get an oxygen at θ is the same as the amplitude to get the α-particle at $(\pi - \theta)$.

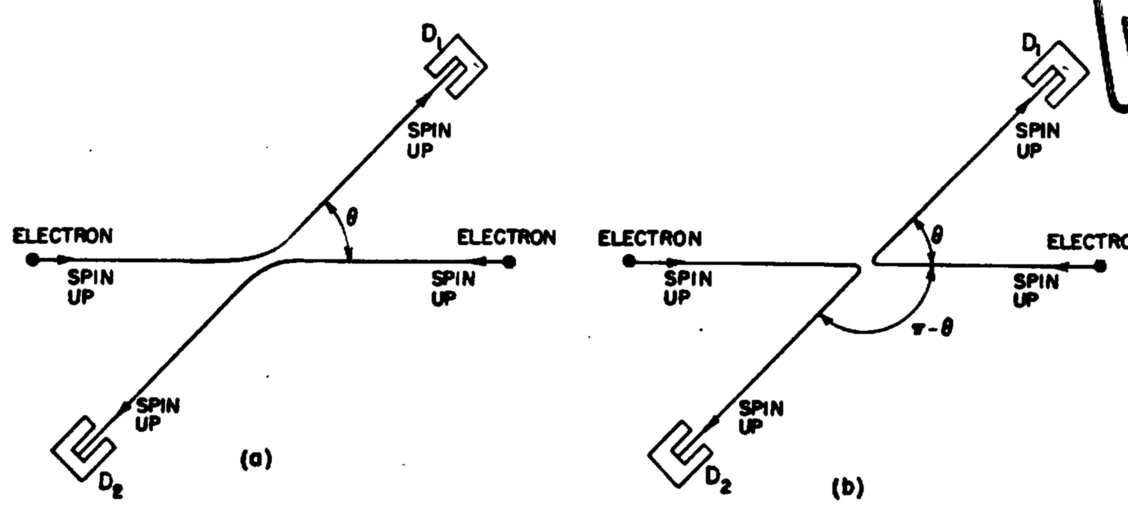

Fig. 3–8. The scattering of electrons on electrons. If the incoming electrons have parallel spins, the processes (a) and (b) are indistinguishable.

and the probability of finding an α-particle in the counter is the square of their sum:

$$\text{Probability of an } \alpha\text{-particle at } D_1 = |f(\theta) + f(\pi - \theta)|^2. \qquad (3.15)$$

This is quite a different result than that in Eq. (3.14). We can take an angle of $\pi/2$ as an example, because it is easy to figure out. For $\theta = \pi/2$, we obviously have $f(\theta) = f(\pi - \theta)$, so the probability in Eq. (3.15) becomes $|f(\pi/2) + f(\pi/2)|^2 = 4|f(\pi/2)|^2$.

On the other hand, if they did not interfere, the result of Eq. (3.14) gives only $2|f(\pi/2)|^2$. So there is twice as much scattering at 90° as we might have expected. Of course, at other angles the results will also be different. And so you have the unusual result that when particles are identical, a certain new thing happens that doesn't happen when particles can be distinguished. In the mathematical description you must add the amplitudes for alternative process in which the two particles simply exchange roles and there is an interference.

An even more perplexing thing happens when we do the same kind of experiment by scattering electrons on electrons, or protons on protons. Neither of the above results is then correct! For these particles, we must invoke still a new rule, a most peculiar rule, which is the following: When you have a situation in which the identity of the electron which is arriving at a point is exchanged with another one, the new amplitude interferes with the old one with an *opposite phase*. It is interference all right, but with a minus sign. In the case of α-particles, when you exchange the α-particle entering the detector, the interfering amplitudes interfere with the positive sign. *In the case of electrons, the interfering amplitudes for exchange interfere with a negative sign.* Except for another detail to be discussed below, the proper equation for electrons in an experiment like the one shown in Fig. 3–8 is

$$\text{Probability of e at } D_1 = |f(\theta) - f(\pi - \theta)|^2. \qquad (3.16)$$

The above statement must be qualified, because we have not considered the spin of the electron (α-particles have no spin). The electron spin may be considered to be either "up" or "down" with respect to the plane of the scattering. If the energy of the experiment is low enough, the magnetic forces due to the currents will be small and the spin will not be affected. We will assume that this is the case for the present analysis, so that there is no chance that the spins are changed during the collision. Whatever spin the electron has, it carries along with it. Now you see there are many possibilities. The bombarding and target particles can have both spins up, both down, or opposite spins. If both spins are up, as in Fig. 3–8 (or if both spins are down), the same will be true of the recoil particles and the *amplitude* for the process is the *difference* of the amplitudes for the two possibilities

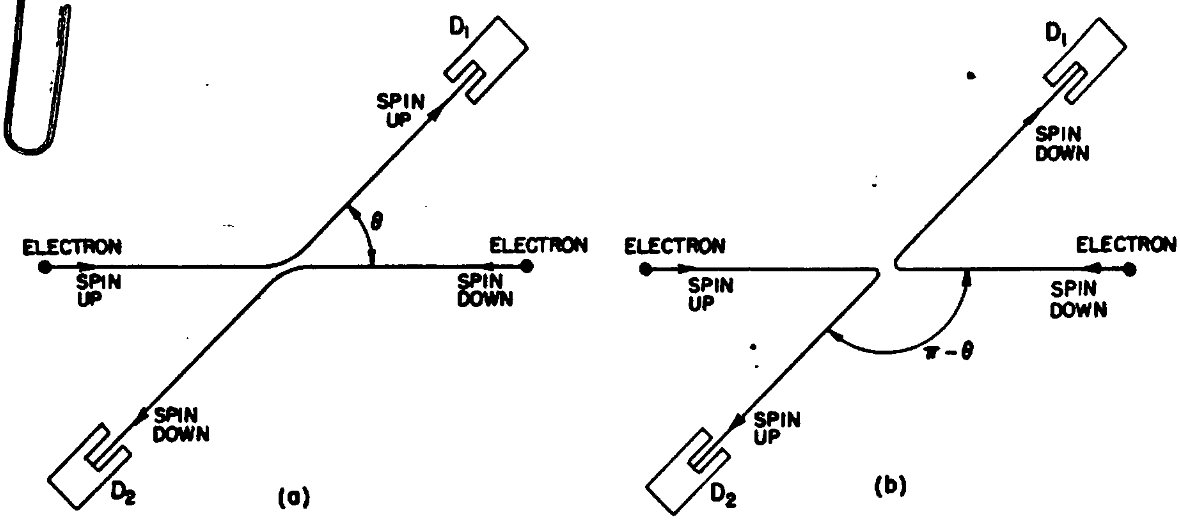

Fig. 3–9. The scattering of electrons with antiparallel spins.

shown in Fig. 3–8(a) and (b). The *probability* of detecting an electron in D_1 is then given by Eq. (3.16).

Suppose, however, the "bombarding" spin is up and the "target" spin is down. The electron entering counter 1 can have spin up or spin down, and by measuring this spin we can tell whether it came from the bombarding beam or from the target. The two possibilities are shown in Fig. 3–9(a) and (b); they are distinguishable in principle, and hence there will be no interference—merely an addition of the two probabilities. The same argument holds if both of the original spins are reversed—that is, if the left-hand spin is down and the right-hand spin is up.

Now if we take our electrons at random—as from a tungsten filament in which the electrons are completely unpolarized—then the odds are fifty-fifty that any particular electron comes out with spin up or spin down. If we don't bother to measure the spin of the electrons at any point in the experiment, we have what we call an unpolarized experiment. The results for this experiment are best calculated by listing all of the various possibilities as we have done in Table 3–1. A separate *probability* is computed for each distinguishable alternative. The total probability is then the sum of all the separate probabilities. Note that for unpolarized beams the result for $\theta = \pi/2$ is one-half that of the classical result with independent particles. The behavior of identical particles has many interesting consequences; we will discuss them in greater detail in the next chapter.

Table 3–1

Scattering of unpolarized spin one-half particles

Fraction of cases	Spin of particle 1	Spin of particle 2	Spin at D_1	Spin at D_2	Probability
$\frac{1}{4}$	up	up	up	up	$\lvert f(\theta) - f(\pi - \theta)\rvert^2$
$\frac{1}{4}$	down	down	down	down	$\lvert f(\theta) - f(\pi - \theta)\rvert^2$
$\frac{1}{4}$	up	down	up	down	$\lvert f(\theta)\rvert^2$
			down	up	$\lvert f(\pi - \theta)\rvert^2$
$\frac{1}{4}$	down	up	up	down	$\lvert f(\pi - \theta)\rvert^2$
			down	up	$\lvert f(\theta)\rvert^2$
Total probability $= \frac{1}{2}\lvert f(\theta) - f(\pi - \theta)\rvert^2 + \frac{1}{2}\lvert f(\theta)\rvert^2 + \frac{1}{2}\lvert f(\pi - \theta)\rvert^2$					