The double slit experiment and the essential mystery of quantum mechanics

(1) Read Feynman's Messenger Lecture about the double slit experiment and write a two page summary and explanation of it in your own words. Feynman's lecture is in my posted lecture notes for lecture 2.

(2) Read the two excerpts from the Feynman Lectures on Physics about the double slit experiment and write a two page summary and explanation about what they contain that his Messenger Lecture does not. The two excerpts are attached.

Read the notes about the double slit experiment that follow and then consider the double slit experiments described below. All of the experiments have

Two equal slits
With slit 1 open, 10,000 electrons per second
With slit 2 open, 10,000 electrons per second
Very narrow slits => 10,000 electrons per second across the entire screen

(3) Calculate and draw the intensity patterns on the screen for the three cases

only slit 1 open
only slit 2 open
both slits open

Next, consider adding a 50% efficient electron detector behind slit 1

(4) Calculate and draw the intensity patterns on the screen with both slits open for the two cases

When detector 1 sees the electron
When detector 1 does not see the electron

Finally, consider 50% efficient detectors behind both slits

(5) Calculate and draw the intensity patterns on the screen with both slits open for the four cases

When detector 1 sees the electron and detector 2 does not
When detector 2 sees the electron and detector 1 does not
When both detectors see the electron
When both detectors do not see the electron

(6) For each case above, show that electrons are conserved---i.e., show that the average number of electrons detected is equal to the number of electrons that pass thru the slits

Fig. 12-4. Experiment to show wave properties of electrons.

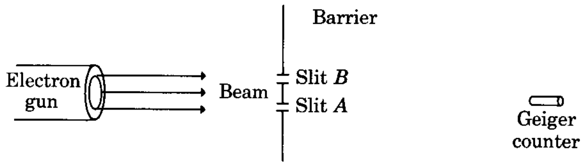
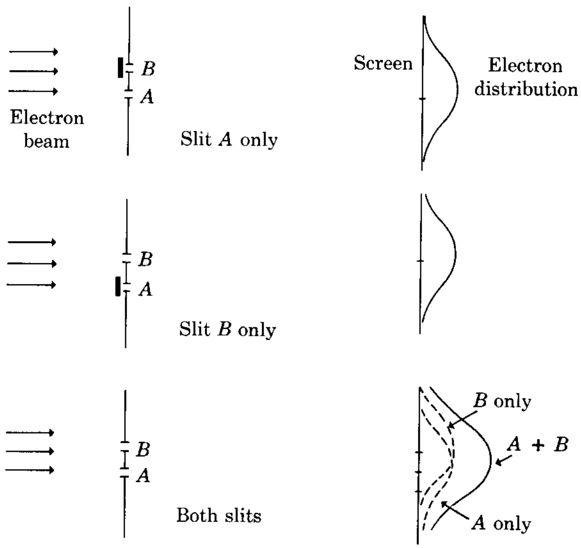


Fig. 12-5. Electron distributions according to classical physics.



Ans. 3: No. $\frac{(\text{KE})_2}{(\text{KE})_1} = \frac{2hf - \mathcal{W}}{hf - \mathcal{W}}$.

12-3 Wave-Particle Duality

Particles are waves are particles

Phenomena such as the photoelectric effect indicated that light must have the properties of particles along with its known wave nature. This behavior of light becomes paradoxical when one considers Young's double-slit experiment performed with a light source so faint that only one photon at a time can enter the apparatus. Photomultiplier tubes which utilize the photoelectric effect can detect the photons one at a time. Let us use photomultiplier tubes in place of the screen in Fig. 10-18. The distance from source to screen is typically about 100 cm, so each photon "lives" for

$$t = \frac{s}{v} = \frac{100 \text{ cm}}{3 \times 10^{10} \text{ cm/sec}} = 3 \times 10^{-9} \text{ sec}$$

If the time between "clicks" from the photomultiplier tubes is less than this, only one photon at a time reaches the double slit. This experiment can be performed (see the Educational Services Inc. film on double-slit interference) and the result is that the interference pattern remains exactly the same. How can *one* photon pass through *two* slits? One way to restate the question is, how can light have both particle and wave properties in the same experiment? This question is one of the most important in all of physics and we will try to answer it in the following paragraphs.

It is now known that this wave-particle relationship or "duality" applies to all particles and waves and is the basic principle of the modern quantum theory. At first it may sound quite farfetched to claim that material particles have a wave nature similar to that of photons. Before describing just what is meant by the wave associated with any particle, let us consider the idealized experiment of Fig. 12-4.

An electron gun shoots a beam of electrons at a barrier that has two slits, *A* and *B*. On the other side there is a Geiger counter that counts each individual electron that hits it. Suppose we count 100 electrons per minute coming from slit *A* (slit *B* is closed). The counting rate from slit *B* alone is also 100 counts per minute. If we open only slit *A* and then gradually open slit *B*, we would expect (according to common sense and everything we have ever learned) the counting rate

Fig. 12-6. Electron distribution according to quantum theory.

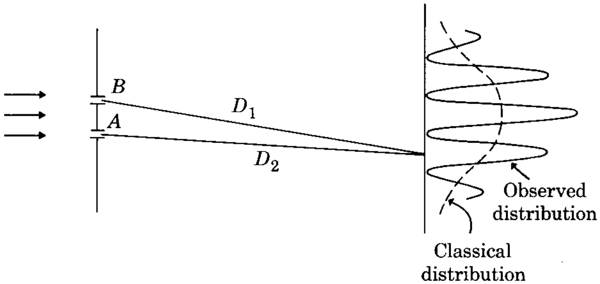
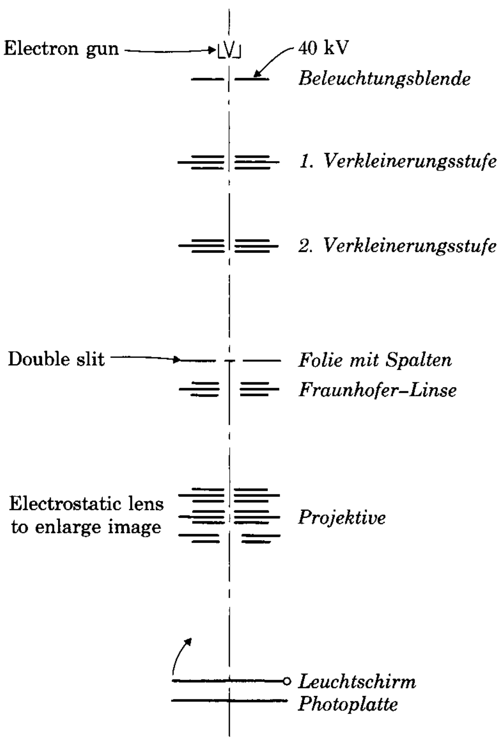


Fig. 12-7. Drawing from research paper by C. Jönsson in *Zeitschrift für Physik,* Vol. 161 (1961), showing his experimental arrangement for obtaining double-slit interference pattern of electrons.



to increase gradually from 100 to 200 counts per minute as slit B is opened. However, depending on the position of the counter, the true experimental result could be a gradual *decrease* from 100 to zero counts per minute! How can the act of opening slit B possibly influence those electrons that would have gone through slit A? Another violation of common sense is that a counter position can be found where the rate would increase from 100 to 400 counts per minute as slit B is opened. Then there would be twice as many electrons as obtained from the direct sum of the two separate contributions.

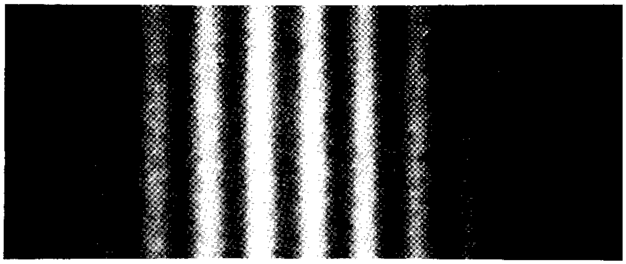
Figure 12-5 shows the expected classical electron distributions at the position of the counter as the counter is moved across the beam (electron intensity is plotted to the right in red). But if we actually do the experiment, we obtain a quite different pattern as shown in Fig. 12-6! Note that this experimental electron intensity pattern is of the same type as the double-slit interference pattern of light waves. If $D_1 - D_2 = N\lambda$, there would be an interference maximum; if $D_1 - D_2 = (N + \frac{1}{2})\lambda$, there would be a minimum of intensity. As we shall see in the next paragraph, electrons usually have wavelengths much smaller than that of visible light; hence it is difficult to perform Young's experiment with electrons. However, C. Jönsson in 1961 was able to obtain a genuine double-slit interference pattern of electrons on a photographic "screen." The experimental layout is shown schematically in Fig. 12-7 and the results are shown in Fig. 12-8*a*.

Each electron produces a black spot at the position where it hits the film. This photograph using a double-slit source of electrons is the result of thousands of electron impacts. For comparison, Fig. 12-8*b* shows a typical double-slit interference pattern using light.

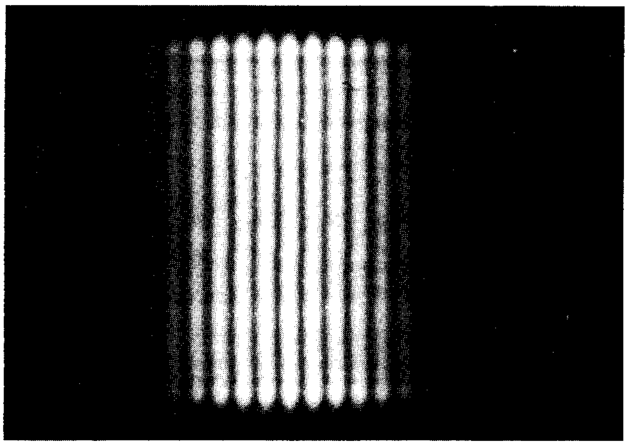
But how can electrons that we know to be particles of definite mass and charge at the same time be waves? Actually this possibility was first proposed by a student, Louis de Broglie, in his Ph. D. thesis in 1924. He proposed that all particles must have a wave nature in the same manner that light has a wave nature. *The physical interpretation of the wave-particle duality is that the intensity of the particle wave*

Fig. 12-8. (*a*) A double slit interference pattern of electrons Each grain in the photographic negative is produced by a single electron. For comparison, (*b*) is the double slit interference pattern of light shown in Chapter 10 (Fig. 10-15). Likewise, each grain in the negative is produced by a single photon. ((*a*) was made by Professor C. Jönsson at the University of Tubingen.)

(a)



(b)



at any given point is proportional to the probability of finding the particle at that point. This is what is meant by the wave-particle duality. The word duality is perhaps a poor choice. What is meant is that there is the definite relationship as italicized above between the particle characteristics and the wave characteristics of any particle (or wave). De Broglie proposed a quantitative relationship between the wavelength of the particle wave and the momentum of the particle:

The de Broglie relationship $$\lambda = \frac{h}{P} \quad (12\text{-}6)$$

for any particle of momentum P.

Example 4

Starting with the de Broglie relationship, derive the formula $W = hf$ for particles of zero rest mass (note that this formula holds only for particles of zero rest mass). Relativistically

$$P = Mv \quad \text{and} \quad M = \frac{W}{c^2}$$

Therefore

$$P = \frac{Wv}{c^2}$$

For a particle of zero rest mass $v = c$ and then $P = W/c$. Substituting in 12-6:

$$\lambda = \frac{h}{\frac{W}{c}} \quad \text{or} \quad W = h\frac{c}{\lambda} = hf$$

In Eq. 12-6, the wave nature, the left-hand side, is directly and intimately related to the particle nature, the right-hand side. The proportionality constant, h, is Planck's constant, which had previously been determined by phenomena such as black body radiation, the photoelectric effect, and the hydrogen spectrum.

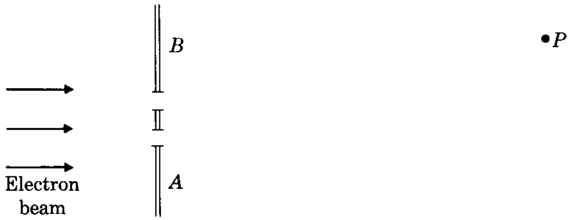
The wave-particle duality raises puzzling questions that require some further physical interpretation. Let us shoot only one electron at a time. Then according to this wave picture, each electron is represented by a wave train or wave packet that splits equally between the two slits. But we can put a geiger counter, cloud chamber, or other particle detector at slit A and observe that there is never in nature half of an electron. We either observe all of a particle, or else no particle at all. This is called the principle of indivisibility and is consistent with the hypothesis that the wave intensity at slit A is the probability of finding one whole electron at that position. Furthermore, if a detector is placed at slit A, the interference pattern smooths out and the classical result is then observed. To be detected, the electron must have an interaction with the detector. According to the quantum theory, we then have a new electron wave starting out from the point of interaction which produces just a single slit pattern. On the other hand, if an electron appears on the screen without being detected by the detector at slit A, we then know its wave must have gone through slit B only. Thus the presence of a detector changes the result from the interference pattern of Fig. 12-6 to the classical result of Fig. 12-5. Actually many physicists, including Einstein, have tried to contrive an experiment that would reveal the slit used by individual electrons without destroying the interference, but all such efforts have failed.

Just what is it that waves in an electron wave? We must give the same kind of answer we gave for photons. Electromagnetic waves travel freely through pure vacuum. In contrast to mechanical waves, no material of any kind is waving. Physicists use the symbol ψ for the amplitude of a particle wave. The intensity is the square of the absolute value of the amplitude or $|\psi|^2$. Hence $|\psi(x)|^2$ is proportional to the probability of finding the particle at position x. The wave amplitude $\psi(x)$ has no direct physical meaning, and in this sense nothing is waving. It is just that quantum mechanical problems are solved mathematically in the same way that water wave or other kinds of classical wave problems are solved. Classical waves and particle waves both obey the same kind of mathematical wave equation. However, in the case of classical waves, the wave amplitude is directly observable, whereas ψ is not (except in certain special cases for the photon). Another nonclassical characteristic of quantum mechanical waves is that even though the wave intensity is always a real positive number, sometimes the wave amplitude must be expressed as a complex number containing $\sqrt{-1}$. We will not deal with examples requiring complex

Q.4: Assume a thin-walled Geiger counter is placed behind slit A only. Whenever an electron goes through slit A it gives a "click" in the detector and passes through the Geiger counter tube to the screen. If the Geiger counter is turned off (but not removed), what will be the pattern on the screen?

numbers in this book. Now we shall illustrate the mathematics of particle waves by a few simple examples.

Fig. 12-9. Set up for Example 5.



Example 5

There is a Geiger counter at point P in Fig. 12-9. The part of the wave amplitude coming through slit A and reaching point P is $\psi_A = 2$ units, and through slit B is $\psi_B = 6$ units. When only slit A is open, 100 electrons per second are observed at point P.

(a) How many electrons per second are observed when only slit B is open?

(b) Assuming a constructive interference, how many electrons per second are observed when both slits are open?

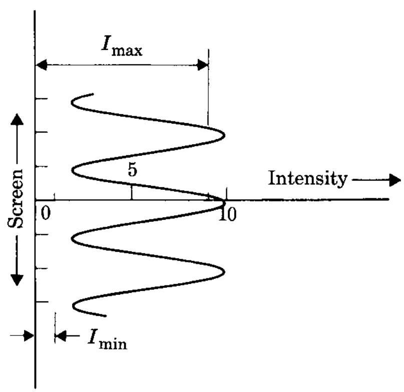
(c) Assuming a destructive interference, how many electrons per second are observed when both slits are open?

Answer: We are told that the particle wave intensity $\psi_A{}^2 = 4$ corresponds to 100 electrons per second. Hence $\psi_B{}^2 = 36$ will correspond to nine times as many particles, or 900 electrons per second.

For part (b) the total wave amplitude is $\psi = \psi_A + \psi_B$, or $\psi = 8$. Since $\psi^2 = 64$ is sixteen times $\psi_A{}^2$, there will be 1600 electrons per second.

For part (c) ψ_A and ψ_B must be of opposite sign to give a destructive interference. Hence $\psi = \psi_A + \psi_B = 2 - 6 = -4$. Now $\psi^2 = 16$ which is four times $\psi_A{}^2$. This corresponds to 400 electrons per second.

Fig. 12-10. Intensity pattern of electrons hitting screen in Example 6.



Example 6

What will be the intensity pattern of a double-slit interference experiment if slit B is four times as wide as slit A?

Answer: Four times as many electrons can get through slit B, hence the intensity $\psi_B{}^2 = 4\psi_A{}^2$, or $\psi_B = 2\psi_A$. The total intensity observed on a screen at a maximum is proportional to $(\psi_A + \psi_B)^2$, or

$$I_{max} = (\psi_A + 2\psi_A)^2 = 9\psi_A{}^2$$

At an intensity minimum, we use a minus sign giving

$$I_{min} = (\psi_A - 2\psi_A)^2 = \psi_A{}^2$$

The ratio $\dfrac{I_{max}}{I_{min}} = \dfrac{9}{1}$

The intensity pattern would appear as in Fig. 12-10.

Example 7

Now we go back again to two slits of equal sizes; however, we put a small, thin detector behind slit A in an attempt to determine which slit each electron goes through. Suppose, however, that the detector is not perfect; that is, it is so thin that an electron has one chance out of four of passing through it without interacting. What now is the interference pattern?

Ans. 4: One still observes the classical pattern on the screen. The electrons are still interacting in the Geiger tube whether or not we bother to record those interactions.

Answer: Let us call ψ_A the part of the beam going through slit A that is undetected and can therefore interfere with the electron waves going through slit B. Since only $\frac{1}{4}$ get through the detector $\psi_A{}^2 = \frac{1}{4}\psi_B{}^2$. Let us denote the intensity of the detected electrons at slit A by $\psi_{A'}{}^2$ where $\psi_{A'}{}^2 = \frac{3}{4}\psi_B{}^2$ and cannot interfere. We must treat $\psi_{A'}$ as a new localized source of electron waves which have no fixed phase relation with respect to ψ_B. In such cases of independent (also called incoherent) sources of particles one must add intensities, not amplitudes. The grand total intensity is then the sum of the two intensities:

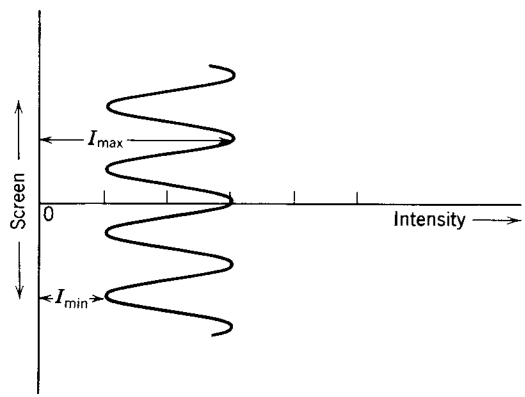
$$I = (\psi_A + \psi_B)^2 + \psi_{A'}{}^2$$

Now substitute $\frac{1}{2}\psi_B$ for ψ_A and $\frac{3}{4}\psi_B{}^2$ for $\psi_{A'}{}^2$

The result is $$\frac{I_{\max}}{I_{\min}} = \frac{(\frac{1}{2}\psi_B + \psi_B)^2 + \frac{3}{4}\psi_B{}^2}{(\frac{1}{2}\psi_B - \psi_B)^2 + \frac{3}{4}\psi_B{}^2} = \frac{\frac{9}{4}\psi_B{}^2 + \frac{3}{4}\psi_B{}^2}{\frac{1}{4}\psi_B{}^2 + \frac{3}{4}\psi_B{}^2} = \frac{3}{1}$$

The resulting interference pattern is shown in Fig. 12-11. If the detector were removed the intensity minima would be zero and one would then have a pure double-slit interference pattern.

Fig. 12-11. Intensity pattern of electrons hitting screen in Example 7.



Q.5: Two light beams strike a screen. There are 3 times as many photons per sec in beam 1 as in beam 2. What is the ratio of the electric fields in the two beams?

12-4 Electron Diffraction

An accident

De Broglie's hypothesis was first verified by the experimental observation of electron diffraction in 1927 by two American physicists, C. J. Davisson and L. H. Germer. It is interesting that in this experiment, as in others that were of extreme importance to physics, the great discovery was "accidental." Davisson and Germer were not looking for electron diffraction. In fact, in the early stages of their experiment, they had never even heard of electron diffraction. In 1926, Davisson took some of his preliminary data to an international conference in Oxford, England. European physicists suggested to him that his results might be interpreted as electron diffraction rather than the classical electron scattering that he had been studying. Just a few months later, Davisson and Germer obtained data that conclusively demonstrated the wave nature of electrons and gave Planck's constant to an accuracy of about 1%. They scattered low-energy electrons off the surface of a single metallic crystal. The regular rows of atoms at the surface act as the lines of a very fine diffraction grating. The electron wavelength is determined by knowing the atomic spacing.

EXAMPLE 1

EQUAL SLITS

With Slit 1 open 100 photons/second => A_1=10

With Slit 2 open 100 photons/second => A_2=10

HOW MANY WHEN 1 AND 2 CONSTRUCTIVE?

$A = A_1 + A_2 = 20$

A^2 = 400 photons/second

HOW MANY WHEN 1 AND 2 DESTRUCTIVE?

$A = A_1 + A_2 = 0$

A^2 = 0 photons/second

EXAMPLE 2

$A_1=2$

$A_2=6$

With slit 1 open 100 photons/second

HOW MANY WITH ONLY SLIT 2 OPEN?

$(A_1)^2 = 4$

$4*(25 \text{ photons/sec}) = 100 \text{ photons/second}$

$(A_2)^2 = 36$

$36*(25 \text{ photons/sec}) = 900 \text{ photons/second}$

HOW MANY WHEN 1 AND 2 CONSTRUCTIVE?

$A = A_1 + A_2 = 8$

$A^2 = 64$

$64*(25 \text{ photons/sec}) = 1600 \text{ photons/second}$

HOW MANY WHEN 1 AND 2 DESTRUCTIVE?

$A = A_2 - A_1 = 4$

$A^2 = 16$

$16*(25 \text{ photons/sec}) = 400 \text{ photons/second}$

EXAMPLE 3

put a partial detector behind slit 1
for a total of 100 photons/second thru slit 1
50 photons/second are not detected => $A_1 = 7.07$
50 photons/second are detected => $A_{1d} = 7.07$

no partial detector behind slit 2 => $A_2 = 10$
100 photons/second not detected thru slit 2

FOR THE PHOTONS NOT DETECTED:
HOW MANY CONSTRUCTIVE?
$A = A_1 + A_2 = 7.07 + 10$
$A^2 = 291.42$ **photons/second**
HOW MANY DESTRUCTIVE?
$A = A_1 + A_2 = 10 - 7.07$
$A^2 = 8.58$ **photons/second**

FOR THE PHOTONS DETECTED:
NO INTERFERENCE
$(A_{1d})^2 = 50$ **photons/second**

EXAMPLE 4

equal slits
put a partial detector behind slit 1
3/4 of the photons/second are not detected A_1
1/4 of the photons/second are detected A_{1d}

$(A_{1d})^2 = 1/4\ (A_2)^2 \Rightarrow A_{1d} = 1/2\ A_2$

$(A_1)^2 = 3/4\ (A_2)^2 \Rightarrow A_1 = (3/4)^{0.5}\ A_2$

The total intensity is proportional to

$I \sim (A_1 + A_2)^2 + (A_{1d})^2$

The Intensity Contrast = I_{max}/I_{min} is given by

$((1/2)*A_2 + A_2)^2 + 3/4\ (A_2)^2$ divided by

$((1/2)*A_2 - A_2)^2 + 3/4\ (A_2)^2$

So, the Intensity Contrast = 3/1

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 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 where any particular bullet will go. A bullet which happens to hit one of the holes may bounce off the edges of the hole, and may end up anywhere at all. By "probability" we mean the chance that the bullet will arrive at the detector, which we can measure by counting the number which arrive at the detector in a certain time and then taking the ratio of this number to the *total* number that hit the backstop during that time. Or, if we assume that the gun always shoots at the same rate during the measurements, the probability we want is just proportional to the number that reach the detector in some standard time interval.

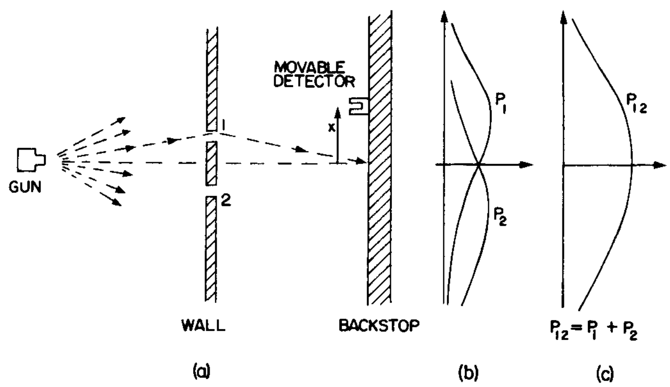


Fig. 1–1. Interference experiment with bullets.

For our present purposes we would like to imagine a somewhat idealized experiment in which the bullets are not real bullets, but are *indestructible* bullets—they cannot break in half. In our experiment we find that bullets always arrive in lumps, and when we find something in the detector, it is always one whole bullet. If the rate at which the machine gun fires is made very low, we find that at any given moment either nothing arrives, or one and only one—exactly one—bullet arrives at the backstop. Also, the size of the lump certainly does not depend on the rate of firing of the gun. We shall say: "Bullets *always* arrive in identical lumps." What we measure with our detector is the probability of arrival of a lump. And we measure the probability as a function of x. The result of such measurements with this apparatus (we have not yet done the experiment, so we are really imagining the result) are plotted in the graph drawn in part (c) of Fig. 1–1. In the graph we plot the probability to the right and x vertically, so that the x-scale fits the diagram of the apparatus. We call the probability P_{12} because the bullets may have come either through hole 1 or through hole 2. You will not be surprised that P_{12} is large near the middle of the graph but gets small if x is very large. You may wonder, however, why P_{12} has its maximum value at $x = 0$. We can understand this fact if we do our experiment again after covering up hole 2, and once more while covering up hole 1. When hole 2 is covered, bullets can pass only through hole 1, and we get the curve marked P_1 in part (b) of the figure. As you would expect, the maximum of P_1 occurs at the value of x which is on a straight line with the gun and hole 1. When hole 1 is closed, we get the symmetric curve P_2 drawn in the figure. P_2 is the probability distribution for bullets that pass through hole 2. Comparing parts (b) and (c) of Fig. 1–1, we find the important result that

$$P_{12} = P_1 + P_2. \tag{1.1}$$

The probabilities just add together. The effect with both holes open is the sum of the effects with each hole open alone. We shall call this result an observation of "*no interference*," for a reason that you will see later. So much for bullets. They come in lumps, and their probability of arrival shows no interference.

1–3 An experiment with waves

Now we wish to consider an experiment with water waves. The apparatus is shown diagrammatically in Fig. 1–2. We have a shallow trough of water. A small object labeled the "wave source" is jiggled up and down by a motor and makes circular waves. To the right of the source we have again a wall with two holes, and beyond that is a second wall, which, to keep things simple, is an "absorber," so that there is no reflection of the waves that arrive there. This can be done by building a gradual sand "beach." In front of the beach we place a detector which can be moved back and forth in the x-direction, as before. The detector is now a device which measures the "intensity" of the wave motion. You can imagine a gadget which measures the height of the wave motion, but whose scale is calibrated in proportion to the *square* of the actual height, so that the reading is proportional to the intensity of the wave. Our detector reads, then, in proportion to the *energy* being carried by the wave—or rather, the rate at which energy is carried to the detector.

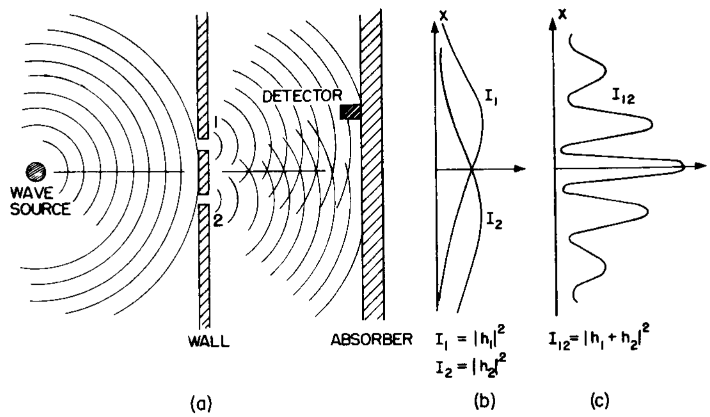


Fig. 1–2. Interference experiment with water waves.

With our wave apparatus, the first thing to notice is that the intensity can have *any* size. If the source just moves a very small amount, then there is just a little bit of wave motion at the detector. When there is more motion at the source, there is more intensity at the detector. The intensity of the wave can have any value at all. We would *not* say that there was any "lumpiness" in the wave intensity.

Now let us measure the wave intensity for various values of x (keeping the wave source operating always in the same way). We get the interesting-looking curve marked I_{12} in part (c) of the figure.

We have already worked out how such patterns can come about when we studied the interference of electric waves in Volume I. In this case we would observe that the original wave is diffracted at the holes, and new circular waves spread out from each hole. If we cover one hole at a time and measure the intensity distribution at the absorber we find the rather simple intensity curves shown in part (b) of the figure. I_1 is the intensity of the wave from hole 1 (which we find by measuring when hole 2 is blocked off) and I_2 is the intensity of the wave from hole 2 (seen when hole 1 is blocked).

The intensity I_{12} observed when both holes are open is certainly *not* the sum 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.

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) $h_1 e^{i\omega t}$, 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 $h_2 e^{i\omega t}$ 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

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

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

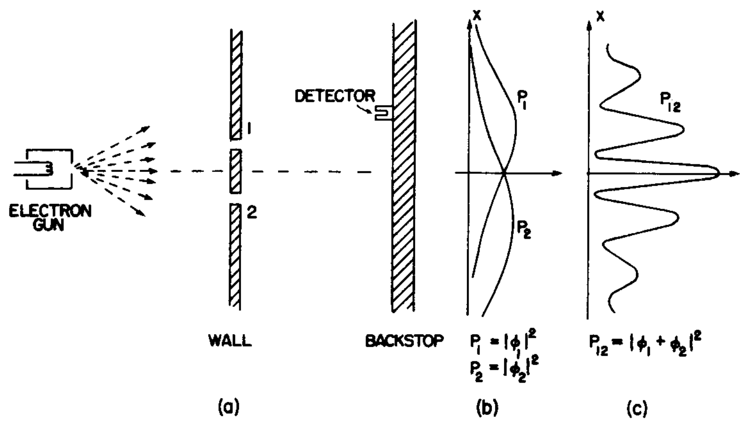


Fig. 1–3. Interference experiment with electrons.

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."

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

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

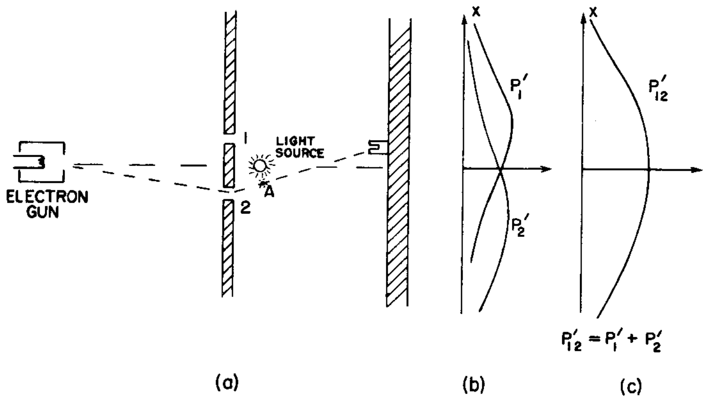


Fig. 1–4. A different electron experiment.

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.

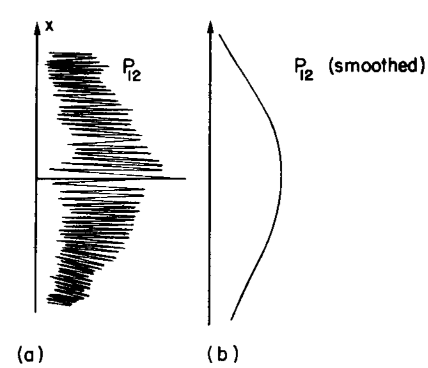


Fig. 1–5. Interference pattern with bullets: (a) actual (schematic), (b) observed.

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.

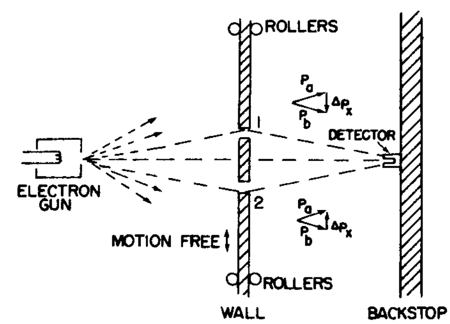


Fig. 1–6. An experiment in which the recoil of the wall is measured.

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.

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.

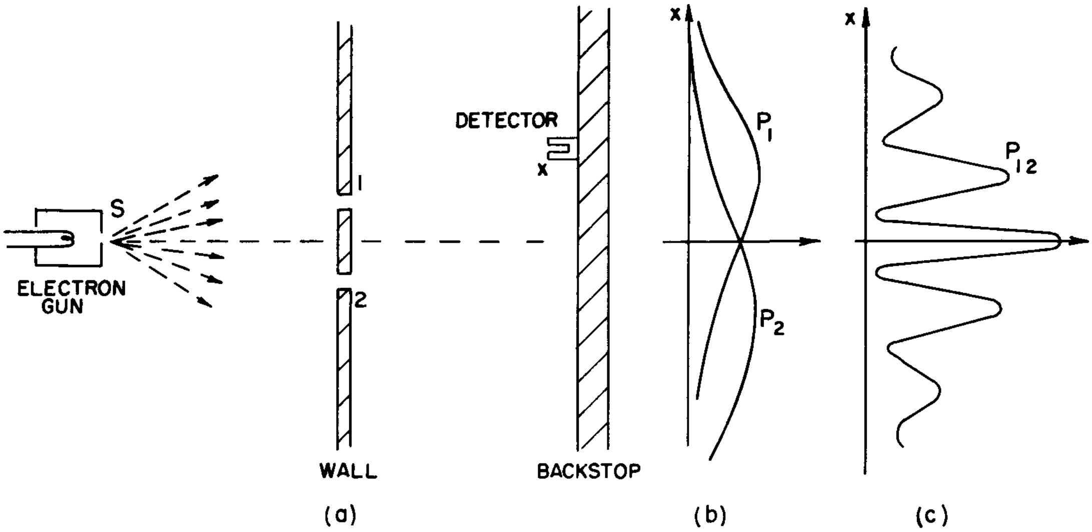


Fig 3–1. Interference experiment with electrons.

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

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

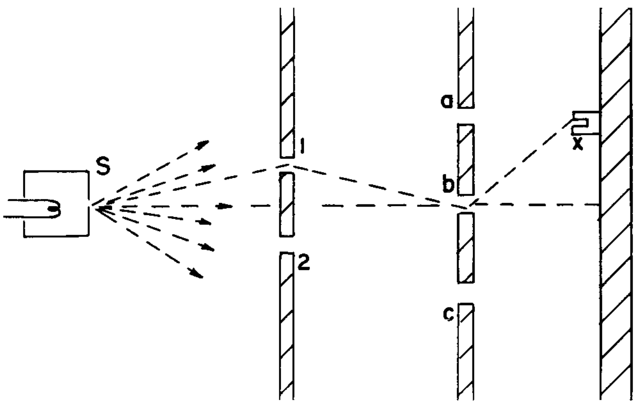


Fig. 3–2. A more complicated interference experiment.

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 $\boldsymbol{p}$ 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.

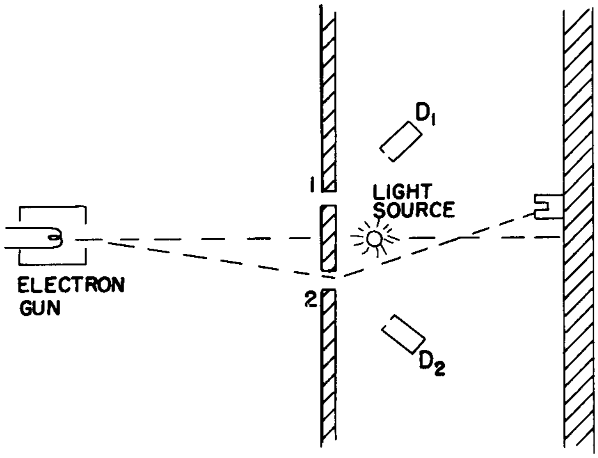


Fig. 3–3. An experiment to determine which hole the electron goes through.

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

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

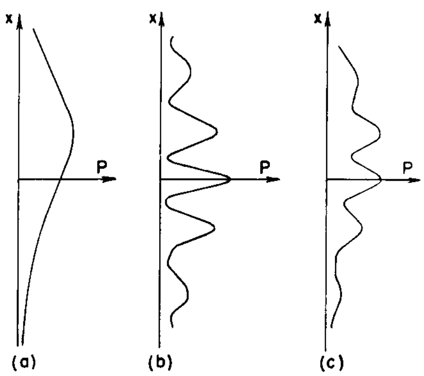


Fig. 3–4. The probability of counting an electron at x in coincidence with a photon at D in the experiment of Fig. 3 3: (a) for $b = 0$; (b) for $b = a$; (c) for $0 < b < a$.

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:

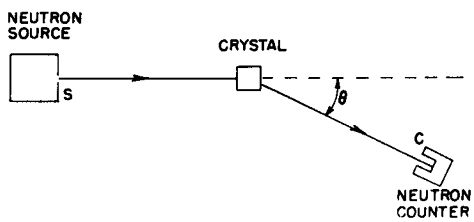


Fig. 3–5. Measuring the scattering of neutrons by a crystal.

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

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:

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