Chapter 4

The Two Slit Experiment

This experiment is said to illustrate the essential mystery of quantum mechanics. This mystery is embodied in the apparent ability of a system to exhibit properties which, from a classical physics point-of-view, are mutually contradictory. We have already touched on one such instance, which is that of a particle possessing wave-like properties, or a wave possessing particle-like properties, otherwise known as wave-particle duality. This property of physical systems must be mirrored in a mathematical language in terms of which the behaviour of such systems can be described and in some sense 'understood'. As we shall see in later chapters, the two slit experiment is a means by which we arrive at this new mathematical language.

But first, the experiment, which will be considered in three forms: performed with macroscopic particles, with waves, and with electrons. The first two experiments merely show what we expect to see based on our everyday experience. It is the third which displays the counterintuitive behaviour of microscopic systems – a peculiar combination of particle and wave like behaviour which cannot be understood in terms of the concepts of classical physics. The analysis of the two slit experiment presented below is more or less taken from Volume III of the Feynman Lectures in Physics.

4.1 An Experiment with Bullets

Imagine an experimental setup in which a machine gun is spraying bullets at a screen in which there are two narrow openings, or slits. Bullets that pass through the openings will then strike a further screen, the detection or observation screen, behind the first, and the point of impact of the bullets on this screen are noted.

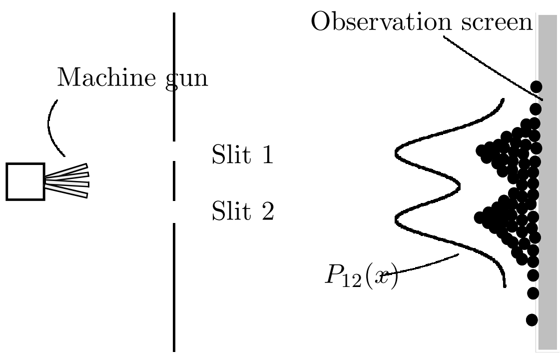

Figure 4.1: An erratic machine gun is firing bullets at a screen containing two small slits. The bullets accumulate on an observation screen, forming two small piles opposite each slit. The curve $P_{12}(x)$ represents the probability density of bullets landing at point x on the observation screen.

The first point to note is that the bullets arrive in 'lumps', (assuming indestructible bullets), i.e. every bullet that leaves the gun arrives as a whole somewhere on the detection screen. The second thing to note is that because the machine gun fires erratically in direction, then successive bullets will strike different parts of the detection screen. Our expectations are then that the bullets that make it through one or the other of the two slits will then strike the observation screen at points that, roughly speaking, will be directly aligned with the slits, and so will accumulate in two 'piles', as indicated in Fig. (4.1).

Now suppose that we perform this experiment with one slit closed. The result will then be a single pile opposite the open slit, see Fig. (4.2).

Figure 4.2: The result of firing bullets at the screen when only one slit is open. The curves $P_1(x)$ and $P_2(x)$ give the probability densities of a bullet passing through slit 1 or 2 respectively and striking ther screen at x.

The result obtained when both slits are opened would then, in some sense, be the result of simply adding together these two piles.

In order to quantify this last statement, we construct a histogram with which to specify the way the bullets spread themselves across the observation screen. We start by assuming that this screen is divided up into boxes of width δx, and then count the number of bullets that land in each box. Suppose that the number of bullets that make it to the observation screen is N, where N is a large number. If $\delta N(x)$ bullets land in the box occupying the range x to $x + \delta x$ then we can plot a histogram of $\delta N/N\delta x$, the fraction of all the bullets that arrive, per unit length, in each interval over the entire width of the screen. An illustrative example is given in Fig. (4.3) of the histogram obtained when $N = 133$ bullets strike the observation screen.

If the number of bullets is very large, and the width δx sufficiently small, then the histogram will define a smooth curve, $P(x)$ say. What this quantity $P(x)$ represents can be gained by considering

$$P(x)\delta x = \frac{\delta N}{N} \tag{4.1}$$

which is the fraction of all the bullets that reach the screen that end up in region x to $x + \delta x$. In other words, if N is very large, $P(x)\delta x$ approximates to the *probability* that any given bullet will arrive at the detection screen in the range x to $x + \delta x$. In Fig. (4.3) the approximate curve for $P(x)$ is also plotted.

We can do the same in the two cases in which one or the other of the two slits are open. Thus, if slit 1 is open, then we get the curve $P_1(x)$ in Fig. (4.3(a)), while if only slit 2 is open, we get $P_2(x)$ such as that in Fig. (4.3(b)). What we are then saying is that if we leave both slits open, then the result will be just the sum of the two single slit curves, i.e.

$$P_{12}(x) = P_1(x) + P_2(x). \tag{4.2}$$

In other words, the *probability* of a bullet striking the screen in some region x to $x + \delta x$ when both slits are opened is just the sum of the probabilities of the bullet landing in region when one slit and then the other is closed. This is all perfectly consistent with what we understand about the properties and behaviour of macroscopic objects – they arrive in indestructible lumps, and the probability observed with two slits open is just the sum of the probabilities with each open individually.

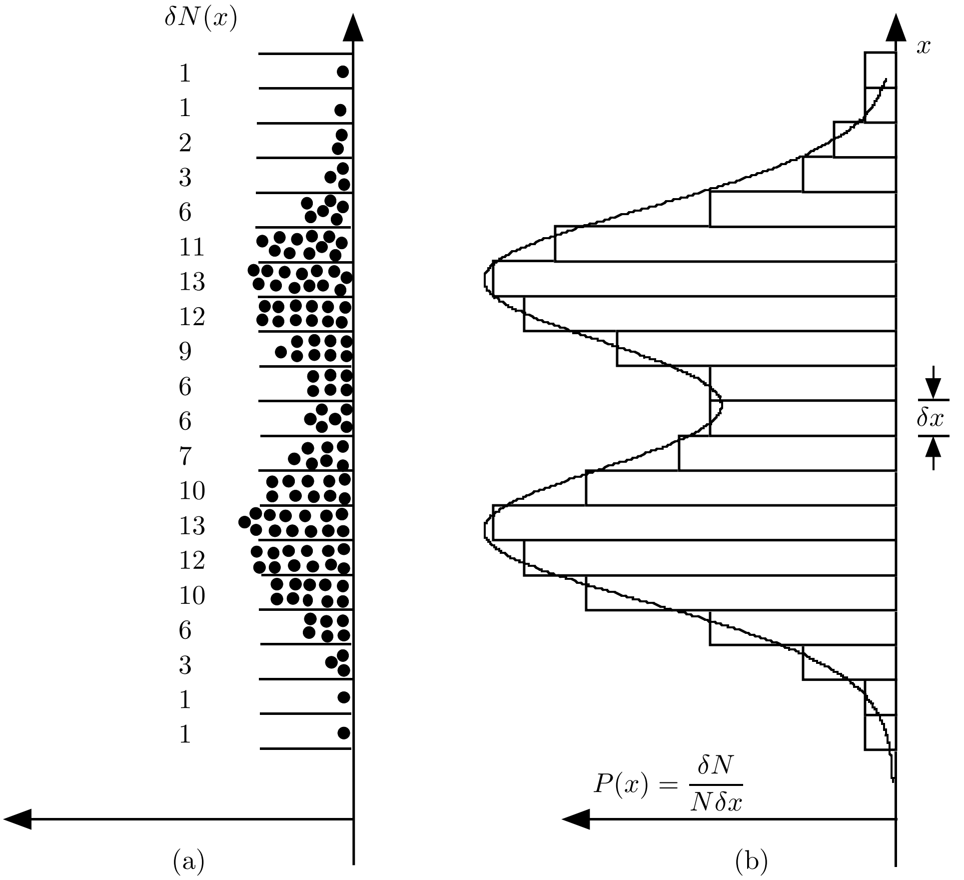

Figure 4.3: Bullets that have passed through the first screen collected in boxes all of the same size δx. (a) The number of bullets that land in each box is presented. There are $\delta N(x)$ bullets in box between x and $x + \delta x$. (b) A histogram is formed from the ratio $P(x) \approx \delta N/N\delta x$ where N is the total number of bullets in all the boxes.

4.2 An Experiment with Waves

Now repeat the experiment with waves. For definiteness, let us suppose that the waves are light waves of wavelength λ. The waves pass through the slits and then impinge on the screen where we measure the intensity of the waves as a function of position along the screen.

First perform this experiment with one of the slits open, the other closed. The resultant intensity distribution is then a curve which peaks behind the position of the open slit, much like the curve obtained in the experiment using bullets. Call it $I_1(x)$, which we know is just the square of the amplitude of the wave incident at x which originated from slit 1. If we deal only with the electric field, and let the amplitude[1] of the wave at x at time t be $E(x,t) = E(x)\exp(-i\omega t)$ in complex notation, then the intensity of the wave at x will be

$$I_1(x) = |E_1(x,t)|^2 = E_1(x)^2. \tag{4.3}$$

Close this slit and open the other. Again we get a curve which peaks behind the position of the open slit. Call it $I_2(x)$. These two outcomes are illustrated in Fig. (4.4)

Figure 4.4: The result of directing waves at a screen when only one slit is open. The curves $I_1(x)$ and $I_2(x)$ give the intensities of the waves passing through slit 1 or 2 respectively and reaching the screen at x. (They are just the central peak of a single slit diffraction pattern.)

Now open both slits. What results is a curve on the screen $I_{12}(x)$ which oscillates between maxima and minima – an interference pattern, as illustrated in Fig. (4.5). In fact, the theory of interference of waves tells us that

$$\begin{aligned} I_{12}(x) =& |E_1(x,t) + E_2(x,t)|^2 \\ =& I_1(x) + I_2(x) \\ &+ 2E_1E_2\cos\left(\frac{2\pi d\sin\theta}{\lambda}\right) \\ =& I_1(x) + I_2(x) \\ &+ 2\sqrt{I_1(x)I_2(x)}\cos\delta \end{aligned} \tag{4.4}$$

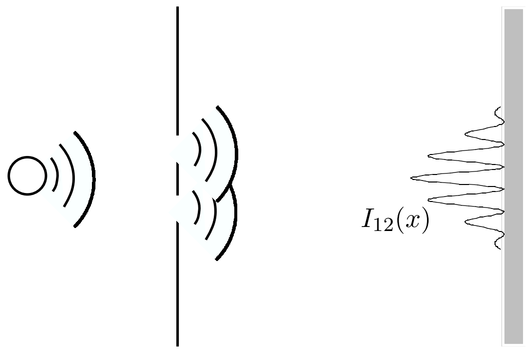

Figure 4.5: The usual two slit interference pattern.

where $\delta = 2\pi d\sin\theta/\lambda$ is the phase difference between the waves from the two slits arriving at point x on the screen at an angle θ to the straight through direction. This is certainly quite different from what was obtained with bullets where there was no interference term. Moreover, the detector does not register the arrival of individual lumps of wave energy: the intensity can have any value at all.

4.3 An Experiment with Electrons

We now repeat the experiment for a third time, but in this case we use electrons. Here we imagine that there is a beam of electrons incident normally on a screen with the two slits, with all the electrons having the same energy E and momentum p. The screen is a fluorescent screen, so that the arrival of each electron is registered as a flash of

[1]The word 'amplitude' is used here to represent the value of the wave at some point in time and space, and is *not* used to represent the maximum value of an oscillating wave.

light – the signature of the arrival of a *particle* on the screen. It might be worthwhile pointing out that the experiment to be described here was not actually performed until the very recent past, and even then not quite in the way described here. Nevertheless, the conclusions reached are what would be expected on the basis of what is now known about quantum mechanics from a multitude of other experiments. Thus, this largely hypothetical experiment (otherwise known as a thought experiment or gedanken experiment) serves to illustrate the kind of behaviour that quantum mechanics would produce, and in a way that can be used to establish the basic principles of the theory.

Let us suppose that the electron beam is made so weak that only one electron passes through the apparatus at a time. What we will observe on the screen will be individual point-flashes of light, and only one at a time as there is only one electron passing through the apparatus at a time. In other words, the electrons are arriving at the screen in the manner of particles, i.e. arriving in lumps. If we close first slit 2 and observe the result we see a localization of flashes in a region directly opposite slit 1. We can count up the number of flashes in a region of size δx to give the fraction of flashes that occur in the range x to $x+\delta x$, as in the case of the bullets. As there, we will call the result $P_1(x)$. Now do the same with slit 1 closed and slit 2 opened. The result is a distribution described by the curve $P_2(x)$. These two curves give, as in the case of the bullets, the probabilities of the electrons striking the screen when one or the other of the two slits are open. But, as in the case of the bullets, this randomness is not to be seen as all that unexpected – the electrons making their way from the source through the slits and then onto the screen would be expected to show evidence of some inconsistency in their behaviour which could be put down to, for instance, slight variations in the energy and direction of propagation of each electron as it leaves the source.

Now open both slits. What we notice now is that these flashes do not always occur at the same place – in fact they appear to occur randomly across the screen. But there is a pattern to this randomness. If the experiment is allowed to continue for a sufficiently long period of time, what is found is that there is an accumulation of flashes in some regions of the screen, and very few, or none, at other parts of the screen. Over a long enough observation time, the accumulation of detections, or flashes, forms an interference pattern, a characteristic of wave motion i.e. in contrast to what happens with bullets, we find that, for electrons, $P_{12}(x) \neq P_1(x) + P_2(x)$. In fact, we obtain a result of the form

$$P_{12}(x) = P_1(x) + P_2(x) + 2\sqrt{P_1(x)P_2(x)}\cos\delta \tag{4.5}$$

so we are forced to conclude that this is the result of the interference of two waves propagating from each of the slits. One feature of the waves, namely their wavelength, can be immediately determined from the separation between successive maxima of the interference pattern. It is found that $\delta = 2\pi d\sin\theta/\lambda$ where $\lambda = h/p$, and where p is the momentum of the incident electrons. Thus, these waves can be identified with the de Broglie waves introduced earlier, represented by the wave function $\Psi(x,t)$.

So what is going on here? If electrons are particles, like bullets, then we can make the proposal that they go *either* through slit 1 *or* through slit 2. The behaviour of the electrons going through slit 1 should not be affected by whether slit 2 is opened or closed. In other words, we have to expect that $P_{12}(x) = P_1(x) + P_2(x)$, but this not what is observed. So, if we want to retain the mental picture of electrons as particles, we must therefore conclude that the electrons pass through *both* slits in some way, because it is only by 'going through both slits' that there is any chance of an interference pattern forming. We could arrange this by supposing that the electrons split up in some way, but then they will have to subsequently recombine before striking the screen since all that is observed is single flashes of light. So what comes to mind is the idea of the electrons executing complicated paths that, perhaps, involve them looping back through each slit, which is scarcely believable. The question would have to be asked as to why the electrons execute such strange behaviour when there are a pair of slits present, but do not seem to when they are moving in free space. There is no way of understanding the double slit behaviour in terms of a particle picture only.

We may argue that one way of resolving the issue is to actually monitor the slits, and look to see when an electron passes through each slit. This could be done, for instance, by shining a light on each of the slits. If an electron goes through a slit, then it scatters some of this light, which can be observed with a microscope. We immediately know what slit the electron passed through, but unfortunately, as a consequence of gaining this knowledge, what is found is that the interference pattern disappears, and what is seen on the screen is the same result as for bullets. Thus, by monitoring an explicitly particle characteristic of the electron, i.e. where it is, the experiment yields the results that would be found with particles.

In fact, as far as we know *any* experiment that can be devised – either a real experiment or a gedanken experiment – that attempts to determine which slit the electron passes through always results in the disappearance of the interference pattern. In other words, if we *know* which slit the electron goes through, then the observed pattern of the screen is the same as found with bullets: no interference, with $P_{12}(x) = P_1(x) + P_2(x)$. Confirming that the electrons definitely go through one slit or another, which is a property that we expect particles to possess, then results in the electrons behaving as particles do. If we do not look, then in some sense each electron passes through both slits, resulting in the formation of an interference pattern, the signature of wave motion. Thus, the electrons behave either like particles or like waves, depending on what it is that is being observed, a dichotomy that is known as wave-particle duality.

The fact that any experiment that determines which slit the electron passes through always results in the disappearance of the interference pattern suggests that there is a deep physical principle at play here – a law of nature – that overrides any attempt to both watch where the particles are, and to observe the interference effects. The principle is the uncertainty principle. It can be argued that pinning down the position of an electron to be passing through a particular slit amounts to specifying its x position to within $\Delta x \approx d/2$, where d is the separation of the slits. But doing so implies that there is an uncertainty in the sideways momentum of the electron given by $\Delta p_x \approx 2\hbar/d$. Since the electrons have a total momentum p, this amounts to a change in direction through an angle

$$\Delta\theta = \frac{\Delta p_x}{p} \approx \frac{\lambda}{\pi d} \tag{4.6}$$

Since the angular separation between a minimum and a neighbouring maximum of the diffraction pattern is $\lambda/2d$, it is clear that the uncertainty in the sideways momentum arising from trying to observe through which slit the particle passes is enough to displace a particle from a maximum into a neighbouring minimum, washing out the interference pattern.

Using the uncertainty principle in this way does not supply a 'physical' explanation for why the interference pattern washes out, i.e. no physical mechanism appears to be at play, only the abstract requirements of the uncertainty principle. Nothing is said about how the position of the particle is pinned down to within an uncertainty Δx. If this information is provided, then usually a physical argument, of sorts, that mixes classical and quantum mechanical ideas can be provided that explains 'why' the interference pattern disappears.

For instance, in the case in which a light beam is shone on the slits, then an argument, due to Heisenberg, that shows why the interference pattern will be washed out. Essentially, the scattering of light by an electron also produces a recoil in the motion of the electron: it will be deflected from its path. If the light is of wavelength λ, then because the smallest 'package' that light comes in is a photon of momentum h/λ, a collision between the electron and the photon could result in a transfer of momentum to the electron of this amount. But, to pin down the position of the electron to better than the distance $d/2$, that is, half the separation between the slits we must have, from the classical theory of optical imaging, $\lambda < d/2$, otherwise we would not know which slit the electron went through. The collision between the electron and the photon could then result in a transfer of momentum to the electron of an amount $\Delta p \approx 2h/d$, and we arrive at the same situation (within a factor $\asymp 1$ as in the above analysis based directly on the uncertainty relation Eq. (4.6).

Other experimental ways of determining through which slit the electron passes will also give a result in agreement with expectations based on the uncertainty principle, i.e. that the interference pattern will be wiped out. However, the details of the way the physics conspires to produce this result will differ from one experiment to the other. It is the laws of quantum mechanics (from which the uncertainty principle follows) that tell us that the interference pattern must disappear if we measure particle properties of the electrons, and this is so *irrespective of the particular kind of physics involved in the measurement* – the individual physical effects that may be present in one experiment or another are subservient to the laws of quantum mechanics.

4.4 Probability Amplitudes

First, a summary of what has been seen so far. In the case of waves, we have seen that the total amplitude of the waves incident on the screen at the point x is given by

$$E(x,t) = E_1(x,t) + E_2(x,t) \tag{4.7}$$

where $E_1(x,t)$ and $E_2(x,t)$ are the waves arriving at the point x from slits 1 and 2 respectively. The intensity of the resultant interference pattern is then given by

$$\begin{aligned} I_{12}(x) &= |E(x,t)|^2 \\ &= |E_1(x,t) + E_2(x,t)|^2 \\ &= I_1(x) + I_2(x) + 2E_1E_2\cos\left(\frac{2\pi d\sin\theta}{\lambda}\right) \\ &= I_1(x) + I_2(x) + 2\sqrt{I_1(x)I_2(x)}\cos\delta \end{aligned} \tag{4.8}$$

where $\delta = 2\pi d\sin\theta/\lambda$ is the phase difference between the waves arriving at the point x from slits 1 and 2 respectively, at an angle θ to the straight through direction.

The point was then made that the probability density for an electron to arrive at the observation screen at point x had the same form, i.e. it was given by the same mathematical expression

$$P_{12}(x) = P_1(x) + P_2(x) + 2\sqrt{P_1(x)P_2(x)}\cos\delta. \tag{4.9}$$

so we were forced to conclude that this is the result of the interference of two waves propagating from each of the slits. Moreover, the wavelength of these waves was found to be given by $\lambda = h/p$, where p is the momentum of the incident electrons so that these waves can be identified with the de Broglie waves introduced earlier, represented by the wave function $\Psi(x,t)$.

Thus, we are proposing that incident on the observation screen is the de Broglie wave associated with each electron whose total amplitude at point x is given by

$$\Psi(x,t) = \Psi_1(x,t) + \Psi_2(x,t) \tag{4.10}$$

where $\Psi_1(x,t)$ and $\Psi_2(x,t)$ are the amplitudes at x of the waves emanating from slits 1 and 2 respectively. Further, since $P_{12}(x)\delta x$ is the probability of an electron being detected in the region $x, x+\delta x$, we are proposing that

$$|\Psi(x,t)|^2\delta x \propto \text{probability of observing an electron in } x, x+\delta x \tag{4.11}$$

so that we can interpret $\Psi(x,t)$ as a *probability amplitude*. This is the famous probability interpretation of the wave function first proposed by Born on the basis of his own observations of the outcomes of scattering experiments, as well as awareness of Einstein's own inclinations along these lines. Somewhat later, after proposing his uncertainty relation, Heisenberg made a similar proposal.

There are two other important features of this result that are worth taking note of:

- If the detection event can arise in two different ways (i.e. electron detected after having passed through either slit 1 or 2) and the two possibilities remain unobserved, then the total probability of detection is

$$P = |\Psi_1 + \Psi_2|^2 \tag{4.12}$$

  i.e. we add the amplitudes and then square the result.

- If the experiment contains a part that even *in principle* can yield information on which of the alternate paths were followed, then

$$P = P_1 + P_2 \tag{4.13}$$

  i.e. we add the probabilities associated with each path.

What this last point is saying, for example in the context of the two slit experiment, is that, as part of the experimental set-up, there is equipment that is monitoring through which slit the particle goes. Even if this equipment is automated, and simply records the result, say in some computer memory, and we do not even bother to look the results, the fact that they are still available means that we should add the probabilities. This last point can be understood if we view the process of observation of which path as introducing randomness in such a manner that the interference effects embodied in the $\cos\delta$ are smeared out. In other words, the $\cos\delta$ factor – which can range between plus and minus one – will average out to zero, leaving behind the sum of probability terms.

4.5 The Fundamental Nature of Quantum Probability

The fact that the results of the experiment performed with electrons yields outcomes which appear to vary in a random way from experiment to experiment at first appears to be identical to the sort of randomness that occurs in the experiment performed with the machine gun. In the latter case, the random behaviour can be explained by the fact that the machine gun is not a very well constructed device: it sprays bullets all over the place. This seems to suggest that simply by refining the equipment,the randomness can be reduced, in principle removing it all together if we are clever enough. At least, that is what classical physics would lead us to believe. Classical physics permits unlimited accuracy in the fixing the values of physical or dynamical quantities, and our failure to live up to this is simply a fault of inadequacies in our experimental technique.

However, the kind of randomness found in the case of the experiment performed with electrons is of a different kind. It is intrinsic to the physical system itself. We are unable to refine the experiment in such a way that we can know precisely what is going on. Any attempt to do so gives rise to unpredictable changes, via the uncertainty principle. Put another way, it is found that experiments on atomic scale systems (and possibly at macroscopic scales as well) performed under identical conditions, where everything is as precisely determined as possible, will always, in general, yield results that vary in a random way from one run of the experiment to the next. This randomness is irreducible, an intrinsic part of the physical nature of the universe.

Attempts to remove this randomness by proposing the existence of so-called 'classical hidden variables' have been made in the past. These variables are supposed to be classical in nature – we are simply unable to determine their values, or control them in any way, and hence give rise to the apparent random behaviour of physical systems. Experiments have been performed that test this idea, in which a certain inequality known as the Bell inequality, was tested. If these classical hidden variables did in fact exist, then the inequality would be satisfied. A number of experiments have yielded results that are clearly inconsistent with the inequality, so we are faced with having to accept that the physics of the natural world is intrinsically random at a fundamental level, and in a way that is *not* explainable classically, and that physical theories can do no more than predict the probabilities of the outcome of any measurement.