

## Chapter 1

### Some History and the Mysteries of Superposition

#### 1.1 Some Brief Historical Background: Particles and Waves

At the end of the 1800's, physics was taken to deal with two fundamental types of things, matter and radiation.<sup>1</sup> Matter was believed to be constituted by atoms, particles obeying Newton's laws of classical mechanics, and radiation to be constituted by electromagnetic waves obeying Maxwell's laws.

The cornerstone of the Newtonian version of classical mechanics is Newton's Equation of Motion, which says that the force  $F$  applied to a particle is equal to the particle's mass  $m$  times its acceleration  $a$ . In one dimension (that is, if we consider only motions and forces on a straight line),

$$ma = F. \tag{1.1.1}$$

Since motion is change of position during time, velocity (how quickly the change of position occurs) is the rate of change of position. Mathematically, this is expressed using derivatives, so that

$$v = \frac{d}{dt}x, \tag{1.1.2}$$

---

<sup>1</sup> For a readable account of the basic transition from classical to quantum physics and relativity and the controversies that ensued, see Sachs, M., (1998) and Kragh, H., (1999). What follows is a bit whiggish but sufficient to give a historical reference frame.

where  $v$  is velocity.<sup>2</sup> In other words, velocity is obtained by taking the (first) derivative of position (represented by  $x$ ) with respect to time (represented by  $t$ ).

As velocity is the rate of change of position, so acceleration  $a$  is the rate of change of velocity; in other words, acceleration determines how quickly velocity changes.

Hence,

$$a = \frac{d}{dt} v. \quad (1.1.3)$$

By plugging (1.1.2) into (1.1.3), we obtain

$$a = \frac{d}{dt} \left( \frac{d}{dt} x \right), \quad (1.1.4)$$

which is usually expressed as

$$a = \frac{d^2}{dt^2} x. \quad (1.1.5)$$

In other words, acceleration is the rate of change of the rate of change of position. By plugging (1.1.5) into (1.1.1), we get

$$m \frac{d^2}{dt^2} x = F, \quad (1.1.6)$$

the standard formulation of the Equation of Motion. If we define the momentum  $p$  of a particle as

$$p = mv, \quad (1.1.7)$$

we can rewrite (1.1.6) as

$$\frac{d}{dt} p = F, \quad (1.1.8)$$

---

<sup>2</sup> For a brief introduction to derivatives (which is not really needed at this stage), see appendix 1.

which says that the force applied on the particle is equal to the change of momentum.

In principle, any mechanical problem can be solved as follows. First, by a study of the situation one determines the force  $F$  acting on the particle and plugs it into (1.1.6). Then, one solves (1.1.6). For mathematical reasons that need not concern us, it is possible to obtain a solution that fits the particle only by providing two extra pieces of information, namely, the position and velocity of the particle at a time  $t$ . Finally, one ends up with an equation that gives the position of the particle at *any* time, namely, the two pieces of information that completely determine the mechanical status of the particle.<sup>3</sup>

Maxwell's equations, which determine the behavior of electromagnetic waves, are mathematically too complex for us to tackle. However, we can understand the basic ideas associated with waves by considering mechanical waves. A wave occurs when a disturbance propagates from a location to another; for example, if one plucks a taut string once, one produces a pulse that travels along the string while retaining its shape. Obviously, the pulse travels because each particle in the string transmits its motion to the next one. All mechanical waves propagate through the motion of the particles in a medium such as air, water, or a string. Depending on the motion of the particles, a wave is transverse (the particles oscillate up and down perpendicularly to the motion of the

---

<sup>3</sup> We have assumed throughout that the mass of the particle does not change. In addition, in practice determining the force to be plugged into (1.1.6), or solving (1.1.6) may be very hard or downright impossible. Then, one uses other methods, often involving conservation laws such as the laws of conservation of mechanical energy or of momentum.

wave) or longitudinal (the particles oscillate back and forth along the motion of the wave). For example, ordinary water waves or waves produced by plucking a string are transversal. By contrast, pressure waves in a tube caused by pushing a piston back and forth, or sound waves in the air are longitudinal.

A periodic wave is one that results from a displacement of the medium at regular intervals. For example, imagine a device that bobs up and down perpendicularly to the surface of a calm pond with frequency  $f$  (that is, in one second it bobs up and down  $f$  times) and therefore period  $T = 1/f$  (that is, it performs one whole bobbing cycle in  $T = 1/f$  seconds). Obviously, as the device moves up and down, it will displace the water molecules next to it and they in turn will transmit the motion to other molecules and so on. In short, all the molecules in the water will execute, successively, the same (harmonic) motion as the body, thus producing a train of concentric waves radiating away from the device. The wavelength  $\lambda$  is the distance between two successive crests (or troughs). Since a wave takes time  $T$  to move from the position occupied by a crest to that occupied by the next crest, it moves along the  $x$ -axis with velocity

$$v = \frac{\lambda}{T} = \lambda f. \quad (1.1.9)$$

When two waves encounter, they behave in remarkable ways. First, they do not change each other; rather, they go through each other: each wave propagates as if the others were not there in spite of the fact that they combine. Second, waves obey the principle of *superposition*: when two waves simultaneously arrive at a place, the resultant

wave is the (vector) sum of the two individual waves.<sup>4</sup> For example, imagine that a second device identical to the first is at work somewhere else in the pond and that a wave from the first device meets a wave from the second device. If the two waves are completely in phase (they meet crest to crest and trough to trough), their sum is a wave of twice the amplitude (the height) of the original waves, a case of total constructive interference. By contrast, if the two waves are totally out of phase (they meet crest to trough), the result is a flat surface, a case of total destructive interference. Obviously, when the waves have different amplitudes or are slightly out of phase neither constructive nor destructive interference will be so dramatic.

## 1.2 Light

Superposition played a crucial part in the debate on the nature of light, a source of controversy since the beginning of the scientific revolution, with the followers of Newton holding that light is a stream of corpuscles and those of Huygens that it is a wave.<sup>5</sup> Followers of Huygens could argue that two intersecting beams of light continue to propagate as if the other were not there, a fact easily explained in terms of wave but apparently incompatible with the corpuscular theory. Newtonians noted that waves can “bend” around corners (the phenomenon of diffraction): immediately after a water wave passes through a slit, it starts spreading out again. Since we hear around corners, sound is

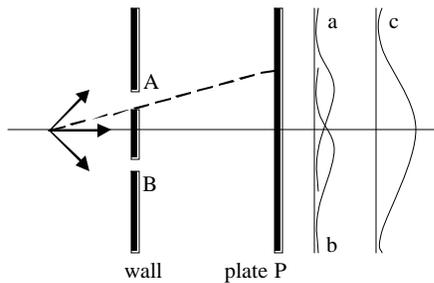
---

<sup>4</sup> This is not true in some cases, however. For example, when the waves are very intense, the principle may fail for reasons that need not concern us.

<sup>5</sup> Newton’s own position, however, is complex, and probably best understood as a mixture of wave and corpuscularian theories.

obviously a wave, but equally, since we do not see around corners, light must be a particle. Of course, one could argue that light does not go around corners because its wavelength is very small, but this is hardly a piece of evidence for the wave theory of light. Partially because of Newton's prestige, his followers won the day, but by the beginning of the 1800's the evidence that light is a wave became stronger and stronger.

Let us imagine a device that with great force randomly throws very hard marbles over a large angular spread against a barrier with two slits A and B just wide enough to let a marble pass and equidistant from the marble-thrower (Fig. 1).<sup>6</sup> Behind the barrier, at a distance much greater than that between the two slits, is a plate P capable of detecting where a marble hits.



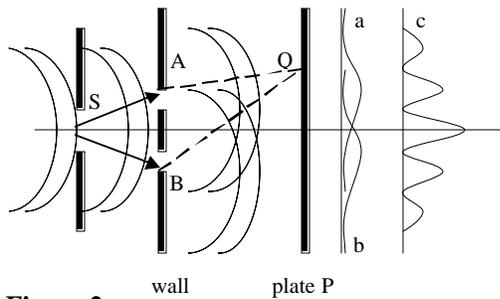
**Figure 1**

Suppose now that we cover slit A and use a grid of equal squares over P to determine the ratio between how many marbles have arrived in each square and the total number of hits all over P. In other words, suppose we determine the probability distribution of the hits of the marbles against P. What we get is a plot roughly like (a) with the maximum probability located about the point on P joined by a straight line to the marble-throwing device. If we cover slit A, open slit B, and repeat the experiment, we get a plot roughly

<sup>6</sup> See Feynman, R. P., Leighton, R. B., Sands, M., (1963), vol. III, sections 1.2 and 1.3.

like (b). If we open both slits, the probability distribution looks like figure (c), the sum of the two separate probability distributions.

If we repeat the experiment with light, a version of the two-slit interference experiment performed by Young in 1800, the results are quite different. The marble thrower is replaced with monochromatic light (light made up of waves of the same frequency) emerging from very narrow slit S. Behind the barrier, is a screen P (Fig. 2).



**Figure 2**

If we cover slit B, the plot for the light pattern on P is given by (a), and if we cover slit A the plot is given by (b). These two patterns are roughly similar to those of the marbles. However, if we open both slits, the plot is given by (c), which is *not* the sum of (a) and (b). In other words, what one observes on the screen is quite remarkable, namely, a pattern of alternate bright and dark lines, with the brightest line at the center of the screen, directly opposite the midpoint between A and B.

Although in the early 1800's the interpretation of these results was controversial, we can see that the light pattern on the screen is easily explained in terms of

superposition of waves.<sup>7</sup> Since all the light is monochromatic and comes from the same source, it emerges from S as a train of coherent waves, namely waves of identical frequency, wavelength, and in phase. Moreover, as A and B are equidistant from S, the waves emerging from them are in phase as well because they traveled the very same distance. However, they need not arrive at a point Q on the screen in phase because Q need not be equidistant from A and B (Fig. 2). Superposition takes over, creating the pattern of alternate bright and dark lines on the screen. By the end of the 1800's, the view that light is a wave was completely accepted, especially after Maxwell had managed to show that it is a type of electromagnetic wave.<sup>8</sup>

### 1.3 Photons

Classical physics, with its separation between particles and waves, was very successful; however several issues remained unresolved, two of which were directly relevant to the birth of quantum mechanics. The first issue centered on the black body radiation. Anyone who has ever fired clay pots knows that the inside of the very hot oven in which the firing occurs glows. Suppose now that we allow the oven to reach a certain temperature and then make a small hole in its side to study the radiation it emits. Crudely, we peep inside the oven. It turns out the color of the light emitted depends on

---

<sup>7</sup> 'Easily' is the operative word here. After Young's experiment, supporters of the corpuscularian view introduced a diffraction force that gave some account of the new results. However, the move was excessively ad hoc and added complexity to their theory.

<sup>8</sup> So, light is not a mechanical but an electromagnetic wave that can propagate in a vacuum, a feat impossible for any mechanical wave.

the temperature; we start with orange, then red, then yellow, and finally white. Why? Qualitatively, classical physics tried to explain the phenomenon by saying that as the oven gets hotter and hotter, the particles making up the oven vibrate at greater and greater frequency, with continuous transition from a frequency to another, producing corresponding waves in the electromagnetic field that we perceive as colored light. The problem was that the quantitative aspect of this view predicted that the oven would emit a blue glow at all temperatures.

In 1900, Plank solved the problem by making a radically new assumption: the energy exchange between matter and radiation takes place by discrete and indivisible quantities, *quanta* of energy. He showed that if we assume that the quantum of energy is proportional to the radiation frequency  $f$  so that  $E = hf$ , then as long as  $h$  is a constant of value

$$h = 6.626 \times 10^{-34} J \cdot s, \quad (1.3.1)$$

the theory could be made to agree with experience.<sup>9</sup> His work broke with the classical view that energy comes in a continuum and that consequently a particle can have any amount of energy; largely for this reason, it did not meet with much approval.

---

<sup>9</sup> Note that the value of  $h$  is unimaginably small. “ $J$ ” is the symbol for the joule, the unit of measurement of work. Work, in one dimension, is force times displacement, and therefore one joule is the work done by the force of one newton during the displacement of one meter:  $1J = 1N \cdot 1m$ . The newton ( $N$ ) is the unit of measure of force. One newton is the force necessary to impart the acceleration of one meter per second per second to a mass of one kilogram:  $1N = 1kg \cdot 1m/s^2$ .

However, in 1905 Einstein gave Plank's idea a radical twist in order to explain, among other phenomena, the photoelectric effect. When light falls on a metal plate, electrons are emitted from it. Qualitatively, this was not difficult to explain. Electrons were taken to be awash in the electromagnetic field of their atoms, and since light is just an electromagnetic wave it would trouble the waters of the electromagnetic field, as it were, thus causing some electrons to be bounced off. The problem was that the speed (actually, the kinetic energy) of these ejected electrons does not depend on the intensity of the light but only on its frequency (its color). Increasing the light intensity (troubling the magnetic waters more strongly, as it were) just increases the number of electrons bounced off, *not* their energy.

Einstein proposed that light consists of a beam of photons, corpuscle-like packets of energy

$$E = hf, \quad (1.3.2)$$

where  $f$  represents the frequency of the light wave traveling at the speed of light. When a photon hits an electron in the metal, it is completely absorbed, transmitting its whole energy to the electron. In order to leave the metal, the electron must do a work equal to the energy  $W$  that binds it to its atom, thus bouncing off with a kinetic energy

$$\frac{1}{2}mv^2 = hf - W, \quad (1.3.3)$$

a prediction verified by experiment. An increase in light intensity without a change of its frequency (that is, of color) just amounted to an increase in the number, but not in the energy, of photons hitting the plate in a given time interval. Hence, the number, but not the energy, of emitted electrons increased. Einstein's idea was revolutionary in that it

endowed light, which everyone agreed to be a wave, with particle-like characteristics, thus seemingly giving it a dual nature of both particle and wave.

Before the early 1920's, few physicists accepted the existence of quanta or of photons. For example, in 1914 a group of influential German physicists, including Planck, recommended Einstein for membership of the Prussian Academy. However, they noted that the fact that Einstein "may sometimes have missed the target in his speculations, as, for example, in his hypothesis of light-quanta, cannot really be held too much against him, for it is not possible to introduce really new ideas even in the most exact sciences without sometimes taking a risk" (Quotation in Whitaker, A., (1996): 99). However, things changed in 1924.

When a beam of x-rays strikes matter, some of the radiation is scattered. While investigating the scattering of monochromatic x-rays from various materials, Compton noticed that some of the scattered radiation had lower energy (lower frequency, according to (1.3.2)) than before the scattering and that the change in wavelength depended on the angle of scattering: the closer the angle was to  $90^\circ$  the greater the change in wavelength. In spite of several attempts, Compton could not explain the effect that now carries his name in terms of classical electromagnetic theory, which predicted that the scattered wave should have the same wavelength as the incident wave. Eventually, he resorted to thinking of the incident radiation as a beam of photons and of the scattering process as a two-dimensional elastic collision between two particle-like things, a photon and an

electron at rest, and managed to explain the phenomenon, thus making it easier to accept that photons really exist.<sup>10</sup>

#### 1.4 Some More History: the Spectra of Elements

The second issue directly relevant to the birth of quantum mechanics had to do with the spectra of elements. Newton had observed that if sunlight passes through a prism it decomposes into a continuous spectrum of colors. However, in 1814 Fraunhofer, observing the spectrum with a magnifying glass, noted that it is crossed by narrow dark lines at specific wavelengths. Eventually, it became clear that each element such as hydrogen, oxygen, or gold when heated, produces a light that, when passed through a prism, shows a characteristic spectral pattern. Why? A good guess was provided by Maxwell's equations, according to which an accelerated electrical charge, such as the electron, discovered in 1897, emits electromagnetic waves, that is, light. The problem was to come up with an atomic model with electrons in motion that would fit the facts.

In 1911, on the basis of results obtained by bombarding atoms with alpha-particles (two protons and two neutrons), Rutherford proposed the idea that an atom is made up by a positively charged nucleus where most of the mass of the atom is located

---

<sup>10</sup> The formula Compton obtained is  $\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta_1)$ .  $\lambda' - \lambda$ , the shift in wavelength, depends on the photon's scattering angle  $\theta_1$ : the closer it is to  $90^\circ$ , the greater the shift in wavelength, and therefore the smaller the frequency of the scattered photon. The quantity  $h/mc$  is the Compton wavelength, whose value changes in relation to the mass  $m$  of the target particle ( $h/c$  is a constant).

plus negatively charged electrons orbiting it.<sup>11</sup> Qualitatively, then, one could say that the orbiting, and therefore accelerated, electrons produce the light of specific wavelength that compose the characteristic spectrum of each element. The problem was that since electrons emit radiation as they orbit the nucleus, they should lose energy and eventually spiral down into it while producing a continuous light spectrum. More generally, the model failed to explain the remarkable stability of atoms. For example, an atom of carbon remains an atom of carbon no matter what sort of chemical interactions it is involved in.

The solution to these problems was provided by Niels Bohr, author of the first successful quantum theory. In 1913, he published a three part article in *Philosophical Magazine*, in the first installment of which he investigated the hydrogen atom, the simplest atom of all, made up of a proton “orbited” by an electron. In effect, he assumed that the angular momentum of the electron orbiting the nucleus in a circular path is *quantized*, more precisely, an integral multiple of  $h/2\pi$ , so that

$$mv_n r_n = n \frac{h}{2\pi}, \quad (1.4.1)$$

where  $v$  and  $m$  are the velocity and the mass of the orbiting electron,  $r$  the radius of the orbit, and  $n$  is a new quantity, the orbit’s quantum number. Bohr made two further assumptions. First, when the electrons are in permitted orbits, contrary to Maxwell’s

---

<sup>11</sup> The magnitudes are extraordinarily small: the nucleus’s dimension is about  $10^{-14}$  meters, while the atom’s is about  $10^{-10}$  meters, one Ångstrom (Å). The mass of the electron is very small:  $9.110 \times 10^{-31} kg$ . Its electrical charge,  $-e$  is a fundamental electrical charge in that all other particles have charges that are  $0, \pm 1e, \pm 2e, \dots$

laws they do not radiate energy at all, and consequently the atom is in a stationary energy state. This eliminates the spiraling down problem that afflicted Rutherford's model.

Second, the transition from an energy state  $E_i$  to another energy state  $E_j$  occurs through the emission or absorption of electromagnetic waves (light) of frequency  $f$  according to the equation

$$hf = E_j - E_i^{12} \quad (1.4.2)$$

Bohr's theory managed to explain the spectrum of hydrogen and a host of other phenomena. However, in spite of its successes, it was never fully satisfactory for it never really coped successfully with atoms more complex than hydrogen, and in some cases led to seriously wrong predictions. Even when in 1915 more powerful quantization techniques, developed by Wilson and Sommerfeld, allowed a very accurate analysis of the Hydrogen atom (the fine-structure of the Hydrogen atom), many of these problems did not go away. It became necessary to allow half-integral values for quantum numbers, and even so, the theory was unable to deal with relatively simple atoms such as helium. Nor was it possible to provide quantitative predictions of the probability of transition from a stationary state to another with the emission or the absorption of a photon. Even when in 1921-23 Bohr produced a new theory turning the orbits of the electrons into Keplerian ellipses, the predicted spectrum of helium, the simplest atom after hydrogen, failed to agree with observation. More fundamentally, Bohr's theory was a classical theory with quantization tacked on, a sort of unholy combination of contrasting views.

### **1.5 Heisenberg, Schrödinger, and Dirac**

---

<sup>12</sup> So, even if Bohr would not have liked it, one could say that the transition from  $E_i$  to  $E_j$  occurs through the absorption or emission of a photon.

The years 1925-26 saw the appearance of two new physically equivalent theories that form the core of contemporary quantum mechanics. The first was produced by Werner Heisenberg and quickly developed by him, Wolfgang Pauli, Pascual Jordan, Max Born, and Bohr. They constituted a close group of physicists in constant contact with each other and with common intellectual sources: Heisenberg and Pauli had been Sommerfeld's students, both worked with Born at Göttingen and with Bohr in Copenhagen. Heisenberg, like Pauli, held the empiricist view that the role of physics is to deal with properties that can be experimentally observed, and therefore was critical of Bohr's unobservable electronic orbits. Partly for this reason, while studying the light emitted by an atom hit by a light beam of a given frequency, instead of appealing to the unobservable frequencies of the electrons' orbits, he concentrated on the frequencies of emitted radiation.

In June 1925, while recuperating from a serious attack of hayfever on the island of Heligoland, he produced what amount to the first version of matrix mechanics. In fact, however, it was Born who realized that the mathematical items Heisenberg used are matrices, and in September of the same year he submitted, with Jordan, a paper in which Heisenberg's new mechanics was extended and explicitly presented in matrix form.<sup>13</sup> Later in the year, Heisenberg, Born, and Jordan co-authored a famous paper (the "three-

---

<sup>13</sup> Born, in a letter to Einstein of March 31, 1948 claims that "Heisenberg did not even know what a matrix was in those days (he was my assistant, that is how I know)" (Born, M., (ed.) (1971): 166). Perhaps, however, this remark should be taken with a grain of salt, since at the time Born was upset with Heisenberg, whom he considered still to be "...as pleasant and intelligent as ever, but noticeably 'Nazified'."

men” paper) giving an even more comprehensive account of the theory that, within a few months, was shown by Pauli to be able to deal with the hydrogen atom.

The second theory was invented by Erwin Schrödinger. His work was stimulated by de Broglie, who in 1924 had assumed that as light (radiation) is particulate, so particles are wavy. Each particle of mass  $m$  and velocity  $v$  is associated to a wave of wavelength

$$\lambda = \frac{h}{mv} \tag{1.5.1}$$

and has energy

$$E = hf, \tag{1.5.2}$$

just as Einstein’s photons.

Although de Broglie’s work was ignored by most German physicists, it did impress Einstein and, through him, Schrödinger.<sup>14</sup> In late 1925, Schrödinger worked on a new wave theory based on de Broglie’s ideas and in the summer of the following year he came up with another type of mechanics based not on matrices, but on an equation whose solution is a wave function containing information about the particle under study.

---

<sup>14</sup> The relations between French and German physicists had been especially bad since the beginning of WWI. More generally, from 1919 to 1928 German scientists were not allowed to attend most international conferences. Einstein was one of the few exceptions because of his pacifism and of his holding a Swiss passport (he had double citizenship, German and Swiss). Even so, in 1922 he had to cancel a lecture at the Paris Academy of Sciences because thirty members had stated they would walk out as soon as he entered the room (Kragh, H., (1999): 143-48).

Although Schrödinger's early interpretation of the wave function proved problematic, later in the same year Born managed to provide for it the statistical interpretation that is now standard, while in the meantime Schrödinger showed that his mechanics and Heisenberg's are empirically equivalent.

Almost contemporaneously with Heisenberg's work, in the fall of 1925, Paul Dirac produced his own algebraic version of quantum mechanics, known as *q*-number algebra, which reproduced many of the results obtained by Born, Heisenberg, and Jordan by using Heisenberg's matrix mechanics. One year later, in late 1926, Dirac and Jordan independently developed a general formalism in which the states of quantum particles are represented by vectors in an abstract (purely mathematical) space. So, within the span of two years the central tenets and the basic formalism of quantum mechanics were developed. In a short time, quantum mechanics was extended to nuclear physics, when in 1928 George Gamow in Göttingen and Ronald Gurney and Edward Condon in Princeton independently managed to explain the process whereby radioactive elements produce alpha-particles.

Although, as we saw, the development of quantum mechanics was associated with the explanation of phenomena arising from the structure of atoms, the easiest way to begin to appreciate the strange nature of the quantum world is to consider two simple and well-known thought experiments.

## 1.6 A Double Slit Experiment with Quantum Particles

Let us see what happens if we perform the double-slit experiment with quantum particles, say electrons (Fig. 1).<sup>15</sup> When both slits are open, the probability distribution is identical to that for wave energy or light intensity, that is, the same distribution as in Young's double slit experiment. This intimates that electrons are akin to waves. However, electrons have a definite mass, and once an electron gets to the detection device, it does not behave like a wave but like a marble in the sense that it is not smeared all over the plate, but localized like a discrete lump. This intimates that electrons are akin to marbles.<sup>16</sup> So, perhaps the wave-like probability distribution is due to some sort of interaction among the lumpy, particle-like electrons bouncing off each other near the slits. However, the wave-like patterns persists even if we slow down the rate of electrons to such an extent that only one electron reaches the barrier at any given time, thus eliminating all possibility of electrons hitting each other. So, it looks as if electrons are particle-like after all but their probability distribution at the plate is wave-like.

Let us try, then, to see for any given particle-like electron whether it goes through slit A or slit B or, perhaps, splits into two parts, one going through A, the other through B, and then recomposes between the barrier and the plate, or it follows some other path.

---

<sup>15</sup> The experiment with electrons has been carried out and the results were those predicted by quantum mechanics (Greenstein, G., and Zajonc, A., (1997): 1). The same experiment has been performed using neutrons, helium and sodium atoms. Their classic description is in Feynman, R. P., Leighton, R. B., Sands, M., (1963), vol. III, sections 1.2 and 1.3.

<sup>16</sup> For what is worth, when imaged by using a scanning tunneling microscope, atoms even look lumpy, like marbles.

We can place detection devices near the slits, register what happens on a counter and look at the results. They record that every time an electron goes through the barrier it is through A or B and never through both. In addition, no electron is detected as arriving at the plate but by going through A or B.

However, something funny happens once we start the observation (once the detection devices near the slits are recording): the probability distribution at the plate changes from the wave-like pattern to the marble pattern (c) in figure 1. Perhaps, the measurement disturbs the electrons that are measured. To avoid this possibility, let us remove the A-slit detector and keep only the B-slit detector going. Hence, when the detector registers nothing, but plate P does, an electron that presumably has not been disturbed by any measurement has gone through slit A. It turns out that the distribution pattern is still marble-like. That is, any sort of measurement or detection, even apparently innocuous, changes the distribution from wave-like to particle-like. All of this is very puzzling and beyond the capacity of classical physics to explain or predict.

### 1.7 A Spin-half Experiment

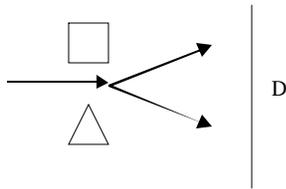
As a rigid mass moving in a loop or spinning on itself has mechanical angular momentum, so an electrically charged particle moving in a loop (for example, an electron 'orbiting' its atom's nucleus) or 'spinning' on itself (a bit like the Earth spinning on its axis) has magnetic momentum. We have, then, an orbital magnetic momentum and a spin magnetic momentum. If we shoot a particle with magnetic angular momentum through an inhomogeneous magnetic field, it will be deflected in proportion to its angular momentum. The Stern-Gerlach apparatus is constituted by a magnet that creates an

Formatted

Formatted

Formatted

inhomogeneous magnetic field and a detector D registering the deflections of the trajectories of particles shot through the field (Fig. 3).



**Figure 3**

If we shoot neutral atoms through the Stern-Gerlach apparatus, their deflections cannot be explained simply by appealing to the orbital magnetic momentums of their electrons but they can be explained by assuming that electrons have spin (or something producing the same results as spin) as well. As an aid to the imagination, then, we can then think of the motions of an electron in an atom as *roughly* analogous to the revolution of the Earth about the Sun and to the spinning of the Earth on its axis.

For reasons that will become clear later, we can properly work only with one component of the spin of an electron at a time; in other words, either we work with  $S_x$ , the component of the electron's spin in the  $x$ -direction, or  $S_z$ , or  $S_y$ , where  $x$ ,  $y$ , and  $z$  are the three mutually perpendicular Cartesian axes. It turns out that every time we measure the magnitude of a component of an electron's spin we obtain only one of two results,

namely,  $+\frac{\hbar}{2}$  or  $-\frac{\hbar}{2}$ .<sup>17</sup> Suppose now that we perform a measurement of  $S_z$  on many

electrons and set aside a thousand for which we have obtained  $S_z = \frac{\hbar}{2}$ . Let us call this

---

<sup>17</sup>  $\hbar$  (h-bar) is Plank's original constant divided by  $2\pi$ .

group of electrons “the  $z$ -spin-up ensemble,” or more simply “the  $|\uparrow_z\rangle$  ensemble,” and its members “ $|\uparrow_z\rangle$  electrons.” Not surprisingly, if we measure  $S_z$  again on the  $|\uparrow_z\rangle$  electrons, we obtain  $S_z = \frac{\hbar}{2}$  all the times. However, if we measure  $S_x$ , we get  $S_x = \frac{\hbar}{2}$  50% of the times and  $S_x = -\frac{\hbar}{2}$  50% of the times. Now let us make the electrons for which we got the former result be members of the  $|\uparrow_x\rangle$  ensemble, and those for which we got the latter result be members of the  $|\downarrow_x\rangle$  ( $x$ -spin-down) ensemble. Curiously, if we take the 500  $|\uparrow_x\rangle$  electrons and we measure  $S_z$ , we get  $S_z = \frac{\hbar}{2}$  50% of the times (that is, 250 times) and  $S_z = -\frac{\hbar}{2}$  50% of the times. We obtain exactly the same outcome if we use the 500  $|\downarrow_x\rangle$  electrons instead of the  $|\uparrow_x\rangle$  electrons. Hence, we expect that if we put together the  $|\downarrow_x\rangle$  and the  $|\uparrow_x\rangle$  electrons and measure  $S_z$ , we obtain  $S_z = \frac{\hbar}{2}$  50% of the times (that is, 500 times) and  $S_z = -\frac{\hbar}{2}$  50% of the times. Indeed, when we perform the measurement that is exactly what we get.

Suppose now that we construct a Stern-Gerlach device SGX (in effect, a Stern-Gerlach apparatus without the detector D) that sorts electrons shot through it by sending them on two different paths, A and B, and that by testing we can assure ourselves that when  $S_x$  is measured all the electrons on path A return  $S_x = \frac{\hbar}{2}$  and all the electrons on path B return  $S_x = -\frac{\hbar}{2}$ . Now, we shoot a  $|\uparrow_z\rangle$  ensemble through a SGX, after we have placed deflectors  $D_1$  and  $D_2$ , so that all the electrons on paths A and B eventually end up into a device R that recombines them, with the result that all the electrons stream out of R

on the same path C (Fig. 4).<sup>18</sup> We can satisfy ourselves that neither the deflectors nor the device R affect  $S_x$  by noticing that if we block path A, all the electrons on C return

$$S_x = -\frac{\hbar}{2} \text{ and if we block path B, they return } S_x = \frac{\hbar}{2}.$$

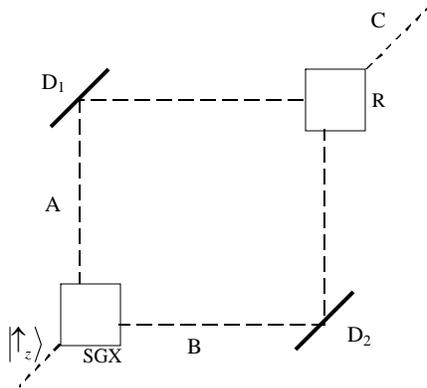


Figure 4

Suppose now that we perform some  $S_x$  measurements on the electrons on C without blocking either path A or path B. Not surprisingly, if we measure  $S_x$  we get  $S_x = \frac{\hbar}{2}$  50% of the times and  $S_x = -\frac{\hbar}{2}$  50% of the times. As we know from before, when measured for  $S_x$ , 50% of  $|\uparrow_z\rangle$  electrons return  $S_x = \frac{\hbar}{2}$  and 50% return  $S_x = -\frac{\hbar}{2}$ . Since the deflectors and R do not affect  $S_x$ , the result is exactly what we would expect.

However, if we measure  $S_z$ , we are in for a surprise, for *all* the electrons on path C return  $S_z = \frac{\hbar}{2}$ . This result is very strange. Since every electron on A, when measured,

<sup>18</sup> This standard example is discussed in Albert, D. Z., (1992) and Barrett, J., A., (1999).

returns  $S_x = \frac{\hbar}{2}$ , we expect the electrons on A to be  $|\uparrow_x\rangle$  electrons; similarly, since every electron on B, when measured, returns  $S_x = -\frac{\hbar}{2}$ , we expect the electrons on B to be  $|\downarrow_x\rangle$  electrons. As 50% of  $|\uparrow_x\rangle$  electrons return  $S_z = \frac{\hbar}{2}$  and 50% return  $S_z = -\frac{\hbar}{2}$ , and the same is true for  $|\downarrow_x\rangle$  electrons, we would expect only half of the electrons on C to return  $S_z = \frac{\hbar}{2}$ . It looks as if the electrons on A are not  $|\uparrow_x\rangle$  electrons and those on B are not  $|\downarrow_x\rangle$  electrons, although it is hard to see what else they could be since every time we measure  $S_x$  on an electron on A we obtain  $S_x = \frac{\hbar}{2}$  and every time we measure  $S_x$  on an electron on B we obtain  $S_x = -\frac{\hbar}{2}$ .

Suppose now that we check all the electrons on path A, verify that each returns  $S_x = \frac{\hbar}{2}$ , and then send them along A, while leaving the electrons on B unchecked. Strangely, 50% of the electrons on C will return  $S_z = \frac{\hbar}{2}$  and 50% will return  $S_z = -\frac{\hbar}{2}$ . Even more extraordinarily, we get the same result if we do not measure spin at all and only determine the position of electrons on A without disturbing those on B. Indeed, we get the same outcome if we block either path A or B. At this point we might think that the strange results are due to interactions among the electrons; however, this suggestion is unsatisfactory because we obtain exactly the same results even if at any given time there is only one electron in the whole apparatus. In short, if paths A and B are open and we do not check on the electrons on them, all the electrons on C return  $S_z = \frac{\hbar}{2}$ ; if we block a

path, or check on the electrons on any path (or both) in terms of position or spin, 50% of the electrons on C will return  $S_z = \frac{\hbar}{2}$  and 50% will return  $S_z = -\frac{\hbar}{2}$ .

In the next chapters, we shall learn enough mathematical machinery to understand how quantum mechanics deals with these bizarre results.