Quantum Physics Notes

J D Cresser
Department of Physics
Macquarie University

31st August 2011

Preface

The world of our every-day experiences – the world of the not too big (compared to, say, a galaxy), and the not too small, (compared to something the size and mass of an atom), and where nothing moves too fast (compared to the speed of light) – is the world that is mostly directly accessible to our senses. This is the world usually more than adequately described by the theories of classical physics that dominated the nineteenth century: Newton's laws of motion, including his law of gravitation, Maxwell's equations for the electromagnetic field, and the three laws of thermodynamics. These classical theories are characterized by, amongst other things, the notion that there is a 'real' world out there, one that has an existence independent of ourselves, in which, for instance, objects have a definite position and momentum which we could measure to any degree of accuracy, limited only by our experimental ingenuity. According to this view, the universe is evolving in a way completely determined by these classical laws, so that if it were possible to measure the positions and momenta of all the constituent particles of the universe, and we knew all the forces that acted between the particles, then we could in principle predict to what ever degree of accuracy we desire, exactly how the universe (including ourselves) will evolve. Everything is predetermined – there is no such thing as free will, there is no room for chance. Anything apparently random only appears that way because of our ignorance of all the information that we would need to have to be able to make precise predictions.

This rather gloomy view of the nature of our world did not survive long into the twentieth century. It was the beginning of that century which saw the formulation of, not so much a new physical theory, but a new set of fundamental principles that provides a framework into which all physical theories must fit: quantum mechanics. To a greater or lesser extent all natural phenomena appear to be governed by the principles of quantum mechanics, so much so that this theory constitutes what is undoubtedly the most successful theory of modern physics. One of the crucial consequences of quantum mechanics was the realization that the world view implied by classical physics, as outlined above, was no longer tenable. Irreducible randomness was built into the laws of nature. The world is inherently probabilistic in that events can happen without a cause, a fact first stumbled on by Einstein, but never fully accepted by him. But more than that, quantum mechanics admits the possibility of an interconnectedness or an 'entanglement' between physical systems, even those possibly separated by vast distances, that has no analogue in classical physics, and which plays havoc with our strongly held presumptions that there is an objectively real world 'out there'.

Quantum mechanics is often thought of as being the physics of the very small as seen through its successes in describing the structure and properties of atoms and molecules – the chemical properties of matter – the structure of atomic nuclei and the properties of elementary particles. But this is true only insofar as the fact that peculiarly quantum effects are most readily observed at the atomic level. In the everyday world that we usually experience, where the classical laws of Newton and Maxwell seem to be able to explain so much, it quickly becomes apparent that classical theory is unable to explain many things e.g. why a solid is 'solid', or why a hot object has the colour that it does. Beyond that, quantum mechanics is needed to explain radioactivity, how semiconducting devices – the backbone of modern high technology – work, the origin of superconductivity, what makes a laser do what it does .... Even on the very large scale, quantum effects leave their mark

in unexpected ways: the galaxies spread throughout the universe are believed to be macroscopic manifestations of microscopic quantum-induced inhomogeneities present shortly after the birth of the universe, when the universe itself was tinier than an atomic nucleus and almost wholly quantum mechanical. Indeed, the marriage of quantum mechanics – the physics of the very small – with general relativity – the physics of the very large – is believed by some to be the crucial step in formulating a general 'theory of everything' that will hopefully contain all the basic laws of nature in one package.

The impact of quantum mechanics on our view of the world and the natural laws that govern it, cannot be underestimated. But the subject is not entirely esoteric. Its consequences have been exploited in many ways that have an immediate impact on the quality of our lives. The economic impact of quantum mechanics cannot be ignored: it has been estimated that about 30% of the gross national product of the United States is based on inventions made possible by quantum mechanics. If anyone aims to have anything like a broad understanding of the sciences that underpin modern technology, as well as obtaining some insight into the modern view of the character of the physical world, then some knowledge and understanding of quantum mechanics is essential. In the broader community, the peculiar things that quantum mechanics says about the way the world works has meant that general interest books on quantum mechanics and related subjects continue to popular with laypersons. This is clear evidence that the community at large and not just the scientific and technological community are very interested in what quantum mechanics has to say. Note that even the term 'quantum' has entered the vernacular – it is the name of a car, a market research company, and a dishwasher amongst other things!! The phrase 'quantum jump' or 'quantum leap' is now in common usage, and incorrectly too: a quantum jump is usually understood to represent a substantial change whereas a quantum jump in its physics context is usually something that is very small.

The successes of quantum mechanics have been extraordinary. Following the principles of quantum mechanics, it is possible to provide an explanation of everything from the state of the universe immediately after the big bang, to the structure of DNA, to the colour of your socks. Yet for all of that, and in spite of the fact that the theory is now roughly 100 years old, if Planck's theory of black body radiation is taken as being the birth of quantum mechanics, it as true now as it was then that no one truly understands the theory, though in recent times, a greater awareness has developed of what quantum mechanics is all about: as well as being a physical theory, it is also a theory of information, that is, it is a theory concerning what information we can gain about the world about us – nature places limitations on what we can 'know' about the physical world, but it also gives us greater freedoms concerning what we can do with this 'quantum information' (as compared to what we could expect classically), as realized by recent developments in quantum computation, quantum teleportation, quantum cryptography and so on. For instance, hundreds of millions of dollars are being invested world-wide on research into quantum computing. Amongst other things, if quantum computing ever becomes realizable, then all security protocols used by banks, defense, and businesses can be cracked on the time scale on the order of months, or maybe a few years, a task that would take a modern classical computer 10^{10} years to achieve! On the other hand, quantum cryptography, an already functioning technology, offers us perfect security. It presents a means by which it is *always* possible to know if there is an eavesdropper listening in on what is supposed to be a secure communication channel. But even if the goal of building a quantum computer is never reached, trying to achieve it has meant an explosion in our understanding of the quantum information aspects of quantum mechanics, and which may perhaps one day finally lead us to a full understanding of quantum mechanics itself.

The Language of Quantum Mechanics

As mentioned above, quantum mechanics provides a framework into which all physical theories must fit. Thus any of the theories of physics, such as Maxwell's theory of the electromagnetic field,

or Newton's description of the mechanical properties of matter, or Einstein's general relativistic theory of gravity, or any other conceivable theory, must be constructed in a way that respects the edicts of quantum mechanics. This is clearly a very general task, and as such it is clear that quantum mechanics must refer to some deeply fundamental, common feature of all these theories. This common feature is the *information* that can be known about the physical state of a physical system. Of course, the theories of classical physics are built on the information gained about the physical world, but the difference here is that quantum mechanics provides a set of rules regarding the information that can be gained about the state of *any* physical system and how this information can be processed, that are quite distinct from those implicit in classical physics. These rules tell us, amongst other things, that it is possible to have exact information about *some* physical properties of a system, but everything else is subject to the laws of probability.

To describe the quantum properties of any physical system, a new mathematical language is required as compared to that of classical mechanics. At its heart quantum mechanics is a mathematically abstract subject expressed in terms of the language of complex linear vector spaces — in other words, linear algebra. In fact, it was in this form that quantum mechanics was first worked out, by Werner Heisenberg, in the 1920s who showed how to represent the physically observable properties of systems in terms of matrices. But not long after, a second version of quantum mechanics appeared, that due to Erwin Schrödinger. Instead of being expressed in terms of matrices and vectors, it was written down in the terms of waves propagating through space and time (at least for a single particle system). These waves were represented by the so-called wave function $\Psi(x,t)$, and the equation that determined the wave function in any given circumstance was known as the Schrödinger equation.

This version of the quantum theory was, and still is, called 'wave mechanics'. It is fully equivalent to Heisenberg's version, but because it is expressed in terms of the then more familiar mathematical language of functions and wave equations, and as it was usually far easier to solve Schrödinger's equation than it was to work with (and understand) Heisenberg's version, it rapidly became 'the way' of doing quantum mechanics, and stayed that way for most of the rest of the 20th century. Its most usual application, built around the wave function Ψ and the interpretation of $|\Psi|^2$ as giving the probability of finding a particle in some region in space, is to describing the structure of matter at the atomic level where the positions of the particles is important, such as in the distribution in space of electrons and nuclei in atomic, molecular and solid state physics. But quantum mechanics is much more than the mechanics of the wave function, and its applicability goes way beyond atomic, molecular or solid state theory. There is an underlying, more general theory of which wave mechanics is but one mathematical manifestation or representation. In a sense wave mechanics is one step removed from this deeper theory in that the latter highlights the informational interpretation of quantum mechanics. The language of this more general theory is the language of vector spaces, of state vectors and linear superpositions of states, of Hermitean operators and observables, of eigenvalues and eigenvectors, of time evolution operators, and so on. As the subject has matured in the latter decades of the 20th century and into the 21st century, and with the development of the 'quantum information' interpretation of quantum mechanics, more and more the tendency is to move away from wave mechanics to the more abstract linear algebra version, chiefly expressed in the notation due to Dirac. It is this more general view of quantum mechanics that is presented in these notes.

The starting point is a look at what distinguishes quantum mechanics from classical mechanics, followed by a quick review of the history of quantum mechanics, with the aim of summarizing the essence of the wave mechanical point of view. A study is then made of the one experiment that is supposed to embody all of the mystery of quantum mechanics – the double slit interference experiment. A closer analysis of this experiment also leads to the introduction of a new notation – the Dirac notation – along with a new interpretation in terms of vectors in a Hilbert space. Subsequently, working with this general way of presenting quantum mechanics, the physical content of

the theory is developed.

The overall approach adopted here is one of inductive reasoning, that is the subject is developed by a process of trying to see what might work, or what meaning might be given to a certain mathematical or physical result or observation, and then testing the proposal against the scientific evidence. The procedure is not a totally logical one, but the result *is* a logical edifice that is only logical after the fact, i.e. the justification of what is proposed is based purely on its ability to agree with what is known about the physical world.

Contents

Chapter 1

Introduction

THERE are three fundamental theories on which modern physics is built: the theory of relativity, statistical mechanics/thermodynamics, and quantum mechanics. Each one has forced upon us the need to consider the possibility that the character of the physical world, as we perceive it and understand it on a day to day basis, may be far different from what we take for granted.

Already, the theory of special relativity, through the mere fact that nothing can ever be observed to travel faster than the speed of light, has forced us to reconsider the nature of space and time – that there is no absolute space, nor is time 'like a uniformly flowing river'. The concept of 'now' or 'the present' is not absolute, something that everyone can agree on – each person has their own private 'now'. The theory of general relativity then tells us that space and time are curved, that the universe ought to be expanding from an initial singularity (the big bang), and will possibly continue expanding until the sky, everywhere, is uniformly cold and dark.

Statistical mechanics/thermodynamics gives us the concept of entropy and the second law: the entropy of a closed system can never decrease. First introduced in thermodynamics – the study of matter in bulk, and in equilibrium – it is an aid, amongst other things, in understanding the 'direction in time' in which natural processes happen. We remember the past, not the future, even though the laws of physics do not make a distinction between the two temporal directions 'into the past' and 'into the future'. All physical processes have what we perceive as the 'right' way for them to occur – if we see something happening 'the wrong way round' it looks very odd indeed: eggs are often observed to break, but never seen to reassemble themselves. The sense of uni-directionality of events defines for us an 'arrow of time'. But what is entropy? Statistical mechanics – which attempts to explain the properties of matter in bulk in terms of the aggregate behaviour of the vast numbers of atoms that make up matter – stepped in and told us that this quantity, entropy, is not a substance in any sense. Rather, it is a measure of the degree of disorder that a physical system can possess, and that the natural direction in which systems evolve is in the direction such that, overall, entropy never decreases. Amongst other things, this appears to have the consequence that the universe, as it ages, could evolve into a state of maximum disorder in which the universe is a cold, uniform, amorphous blob – the so-called heat death of the universe.

So what does quantum mechanics do for us? What treasured view of the world is turned upside down by the edicts of this theory? It appears that quantum mechanics delivers to us a world view in which

- There is a loss of certainty – unavoidable, unremovable randomness pervades the physical world. Einstein was very dissatisfied with this, as expressed in his well-known statement: "God does not play dice with the universe." It even appears that the very process of making an observation can affect the subject of this observation in an uncontrollably random way (even if no physical contact is made with the object under observation!).

- Physical systems appear to behave as if they are doing a number of mutually exclusive things simultaneously. For instance an electron fired at a wall with two holes in it can appear to behave as if it goes through both holes simultaneously.
- Widely separated physical systems can behave as if they are entangled by what Einstein termed some 'spooky action at a distance' so that they are correlated in ways that appear to defy either the laws of probability or the rules of special relativity.

It is this last property of quantum mechanics that leads us to the conclusion that there are some aspects of the physical world that cannot be said to be objectively 'real'. For instance, in the game known as *The Shell Game*, a pea is hidden under one of three cups, which have been shuffled around by the purveyor of the game, so that bystanders lose track of which cup the pea is under. Now suppose you are a bystander, and you are asked to guess which cup the pea is under. You might be lucky and guess which cup first time round, but you might have to have another attempt to find the cup under which the pea is hidden. But whatever happens, when you do find the pea, you implicitly believe that the pea was under that cup all along. But is it possible that the pea *really* wasn't at any one of the possible positions at all, and the sheer process of looking to see which cup the pea is under, which amounts to measuring the position of the pea, 'forces' it to be in the position where it is ultimately observed to be? Was the pea 'really' there beforehand? Quantum mechanics says that, just maybe, it wasn't there all along! Einstein had a comment or two about this as well. He once asked a fellow physicist (Pascual Jordan): "Do you believe the moon exists only when you look at it?"

The above three points are all clearly in defiance of our classical view of the world, based on the theories of classical physics, which goes had-in-hand with a particular view of the world sometimes referred to as objective realism.

1.1 Classical Physics

Before we look at what quantum mechanics has to say about how we are to understand the natural world, it is useful to have a look at what the classical physics perspective is on this. According to classical physics, by which we mean pre-quantum physics, it is essentially taken for granted that there is an 'objectively real world' out there, one whose properties, and whose very existence, is totally indifferent to whether or not we exist. These ideas of classical physics are not tied to any one person – it appears to be the world-view of Galileo, Newton, Laplace, Einstein and many other scientists and thinkers – and in all likelihood reflects an intuitive understanding of reality, at least in the Western world. This view of classical physics can be referred to as 'objective reality'.

The equations of the theories of classical physics, which include Newtonian mechanics, Maxwell's theory of the electromagnetic field and Einstein's theory of general relativity, are then presumed to describe what is 'really happening' with a physical system. For example, it is assumed that every particle has a definite position and velocity and that the solution to Newton's equations for a particle in motion is a perfect representation of what the particle is 'actually doing'.

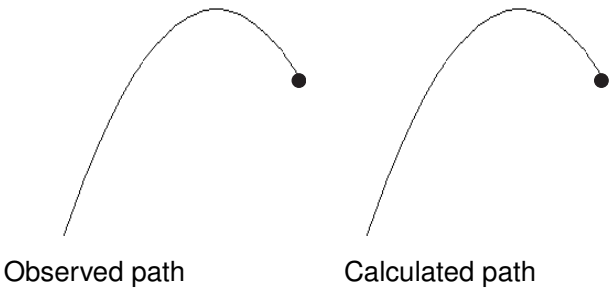


Figure 1.1: Comparison of observed and calculated paths of a tennis ball according to classical physics

Within this view of reality, we can speak about a particle moving through space, such as a tennis ball flying through the air, as if it has, at any time, a definite position and velocity. Moreover, it

would have that definite position and velocity whether or not there was anyone or anything monitoring its behaviour. After all, these are properties of the tennis ball, not something attributable to our measurement efforts. Well, that is the classical way of looking at things. It is then up to us to decide whether or not we want to measure this pre-existing position and velocity. They both have definite values at any instant in time, but it is totally a function of our experimental ingenuity whether or not we can measure these values, and the level of precision to which we can measure them. There is an implicit belief that by refining our experiments — e.g. by measuring to to the 100th decimal place, then the 1000th, then the 10000th — we are getting closer and closer to the values of the position and velocity that the particle 'really' has. There is no law of physics, at least according to classical physics, that says that we definitely cannot determine these values to as many decimal places as we desire – the only limitation is, once again, our experimental ingenuity. We can also, in principle, calculate, with unlimited accuracy, the future behaviour of any physical system by solving Newton's equations, Maxwell's equations and so on. In practice, there are limits to accuracy of measurement and/or calculation, but in principle there are no such limits.

1.1.1 Classical Randomness and Ignorance of Information

Of course, we recognize, for a macroscopic object, that we cannot hope to measure all the positions and velocities of all the particles making such an object. In the instance of a litre of air in a bottle at room temperature, there are something like 10^{26} particles whizzing around in the bottle, colliding with one another and with the walls of the bottle. There is no way of ever being able to measure the position and velocities of each one of these gas particles at some instant in time. But that does not stop us from believing that each particle does in fact possess a definite position and velocity at each instant. It is just too difficult to get at the information.

Likewise, we are unable to predict the motion of a pollen grain suspended in a liquid: Brownian motion (random walk) of pollen grain due to collisions with molecules of liquid. According to classical physics, the information is 'really there' – we just can't get at it.

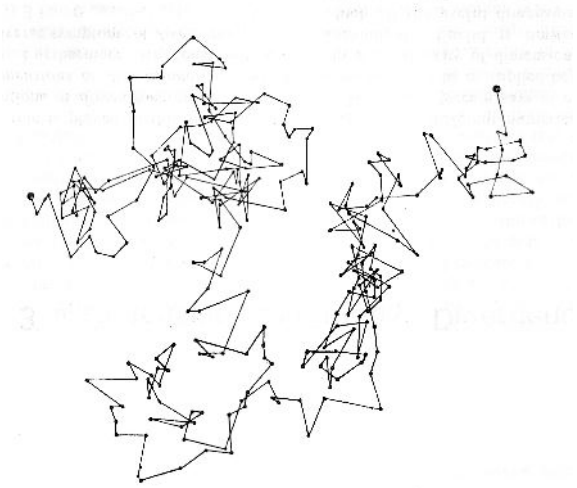


Figure 1.2: Random walk of a pollen grain suspended in a liquid.

Random behaviour only *appears* random because we do not have enough information to describe it exactly. It is not really random because we believe that if we could repeat an experiment under *exactly* identical conditions we ought to get the same result every time, and hence the outcome of the experiment would be perfectly predictable.

In the end, we accept a certain level of ignorance about the possible information that we could, in principle, have about the gas. Because of this, we cannot hope to make accurate predictions about what the future behaviour of the gas is going to be. We compensate for this ignorance by using statistical methods to work out the chances of the gas particles behaving in various possible ways. For instance, it is possible to show that the chances of all the gas particles spontaneously rushing to one end of the bottle is something like 1 in $10^{10^{26}}$ – appallingly unlikely.

The use of statistical methods to deal with a situation involving ignorance of complete information is reminiscent of what a punter betting on a horse race has to do. In the absence of complete information about each of the horses in the race, the state of mind of the jockeys, the state of the track, what the weather is going to do in the next half hour and any of a myriad other possible

influences on the outcome of the race, the best that any punter can do is assign odds on each horse winning according to what information is at hand, and bet accordingly. If, on the other hand, the punter knew *everything* beforehand, the outcome of the race is totally foreordained in the mind of the punter, so (s)he could make a bet that was guaranteed to win.

According to classical physics, the situation is the same when it comes to, for instance, the evolution of the whole universe. If we knew at some instant all the positions and all the velocities of all the particles making up the universe, and all the forces that can act between these particles, then we ought to be able to calculate the entire future history of the universe. Even if we cannot carry out such a calculation, the sheer fact that, in principle, it could be done, tells us that the future of the universe is already ordained. This prospect was first proposed by the mathematical physicist Pierre-Simon Laplace (1749-1827) and is hence known as Laplacian determinism, and in some sense represents the classical view of the world taken to its most extreme limits. So there is no such thing, in classical physics, as true randomness. Any uncertainty we experience is purely a consequence of our ignorance – things only appear random because we do not have enough information to make precise predictions. Nevertheless, behind the scenes, everything is evolving in an entirely preordained way – everything is deterministic, there is no such thing as making a decision, free will is merely an illusion!!!

1.2 Quantum Physics

The classical world-view works fine at the everyday (macroscopic) level – much of modern engineering relies on this – but there are things at the macroscopic level that cannot be understood using classical physics, these including the colour of a heated object, the existence of solid objects .... So where does classical physics come unstuck?

Non-classical behaviour is most readily observed for microscopic systems – atoms and molecules, but is in fact present at all scales. The sort of behaviour exhibited by microscopic systems that are indicators of a failure of classical physics are

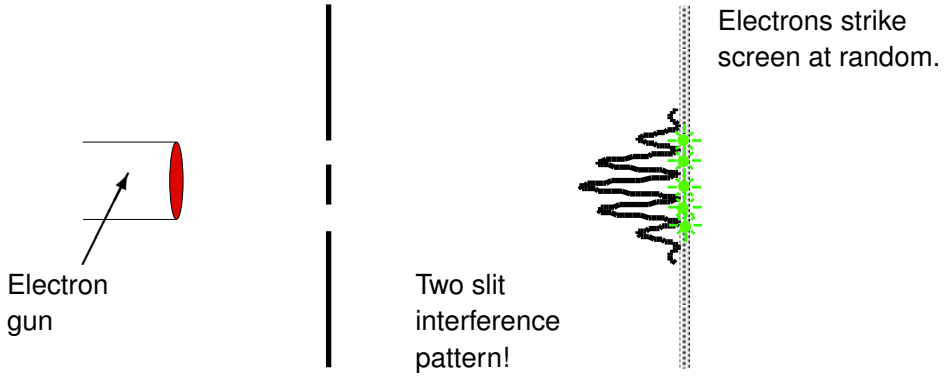
- Intrinsic Randomness
- Interference phenomena (e.g. particles acting like waves)
- Entanglement

Intrinsic Randomness It is impossible to prepare *any* physical system in such a way that *all* its physical attributes are precisely specified at the same time – e.g. we cannot pin down both the position *and* the momentum of a particle at the same time. If we trap a particle in a tiny box, thereby giving us a precise idea of its position, and then measure its velocity, we find, after many repetitions of the experiment, that the velocity of the particle always varies in a random fashion from one measurement to the next. For instance, for an electron trapped in a box 1 micron in size, the velocity of the electron can be measured to vary by at least ± 50 ms^{-1}. Refinement of the experiment *cannot* result in this randomness being reduced — it can never be removed, and making the box even tinier just makes the situation worse. More generally, it is found that for *any* experiment repeated under exactly identical conditions there will always be some physical quantity, some physical property of the systems making up the experiment, which, when measured, will always yield randomly varying results from one run of the experiment to the next. This is *not* because we do a lousy job when setting up the experiment or carrying out the measurement. The randomness is *irreducible*: it *cannot* be totally removed by improvement in experimental technique.

What this is essentially telling us is that nature places limits on how much information we can gather about any physical system. We apparently cannot know with precision as much about

a system as we thought we could according to classical physics. This tempts us to ask if this missing information is still there, but merely inaccessible to us for some reason. For instance, does a particle whose position is known also have a precise momentum (or velocity), but we simply cannot measure its value? It appears that in fact this information is not missing – it is not there in the first place. Thus the randomness that is seen to occur is not a reflection of our ignorance of some information. It is not randomness that can be resolved and made deterministic by digging deeper to get at missing information – it is apparently 'uncaused' random behaviour.

Interference Microscopic physical systems can behave as if they are doing mutually exclusive things at the same time. The best known example of this is the famous two slit experiment in which electrons are fired, one at a time, at a screen in which there are two narrow slits. The electrons are observed to strike an observation screen placed beyond the screen with the slits. What is expected is that the electrons will strike this second screen in regions immediately opposite the two slits. What is observed is that the electrons arriving at this observation screen tend to arrive in preferred locations that are found to have all the characteristics of a wave-like interference pattern, i.e. the pattern formed as would be observed if it were waves (e.g. light waves) being directed towards the slits.



The detailed nature of the interference pattern is determined by the separation of the slits: increasing this separation produces a finer interference pattern. This seems to suggest that an electron, which, being a particle, can only go through one slit or the other, somehow has 'knowledge' of the position of the other slit. If it did not have that information, then it is hard to see how the electron could arrive on the observation screen in such a manner as to produce a pattern whose features are directly determined by the slit separation! And yet, if the slit through which each electron passes is observed in some fashion, the interference pattern disappears – the electrons strike the screen at positions directly opposite the slits! The uncomfortable conclusion that is forced on us is that if the path of the electron is not observed then, in some sense, it passes through both slits much as waves do, and ultimately falls on the observation screen in such a way as to produce an interference pattern, once again, much as waves do.

This propensity for quantum system to behave as if they can be two places at once, or more generally in different states at the same time, is termed 'the superposition of states' and is a singular property of quantum systems that leads to the formulation of a mathematical description based on the ideas of vector spaces.

Entanglement Suppose for reasons known only to yourself that while sitting in a hotel room in Sydney looking at a pair of shoes that you really regret buying, you decided to send one of the pair to a friend in Brisbane, and the other to a friend in Melbourne, without observing which shoe went where. It would not come as a surprise to hear that if the friend in Melbourne discovered that the shoe they received was a left shoe, then the shoe that made it to Brisbane was a right shoe,

and vice versa. If this strange habit of splitting up perfectly good pairs of shoes and sending one at random to Brisbane and the other to Melbourne were repeated many times, then while it is not possible to predict for sure what the friend in, say Brisbane, will observe on receipt of a shoe, it is nevertheless always the case that the results observed in Brisbane and Melbourne were always perfectly correlated – a left shoe paired off with a right shoe.

Similar experiments can be undertaken with atomic particles, though it is the spins of pairs of particles that are paired off: each is spinning in exactly the opposite fashion to the other, so that the total angular momentum is zero. Measurements are then made of the spin of each particle when it arrives in Brisbane, or in Melbourne. Here it is not so simple as measuring whether or not the spins are equal and opposite, i.e. it goes beyond the simple example of left or right shoe, but the idea is nevertheless to measure the correlations between the spins of the particles. As was shown by John Bell, it is possible for the spinning particles to be prepared in states for which the correlation between these measured spin values is *greater* than what classical physics permits. The systems are in an 'entangled state', a quantum state that has no classical analogue. This is a conclusion that is experimentally testable via Bell's inequalities, and has been overwhelmingly confirmed. Amongst other things it seems to suggest the two systems are 'communicating' instantaneously, i.e. faster than the speed of light which is inconsistent with Einstein's theory of relativity. As it turns out, it can be shown that there is no faster-than-light communication at play here. But it can be argued that this result forces us to the conclusion that physical systems acquire some (maybe all?) properties only through the act of observation, e.g. a particle does not 'really' have a specific position until it is measured.

The sorts of quantum mechanical behaviour seen in the three instances discussed above are believed to be common to all physical systems. So what is quantum mechanics? It is saying something about *all* physical systems. Quantum mechanics is not a physical theory specific to a limited range of physical systems i.e. it is not a theory that applies only to atoms and molecules and the like. It is a meta-theory. At its heart, quantum mechanics is a set of fundamental principles that *constrain the form of physical theories themselves*, whether it be a theory describing the mechanical properties of matter as given by Newton's laws of motion, or describing the properties of the electromagnetic field, as contained in Maxwell's equations or any other conceivable theory. Another example of a meta-theory is relativity — both special and general — which places strict conditions on the properties of space and time. In other words, space and time must be treated in all (fundamental) physical theories in a way that is consistent with the edicts of relativity.

To what aspect of all physical theories do the principles of quantum mechanics apply? The principles must apply to theories as diverse as Newton's Laws describing the mechanical properties of matter, Maxwell's equations describing the electromagnetic field, the laws of thermodynamics – what is the common feature? The answer lies in noting how a theory in physics is formulated.

1.3 Observation, Information and the Theories of Physics

Modern physical theories are not arrived at by pure thought (except, maybe, general relativity). The common feature of all physical theories is that they deal with the *information* that we can obtain about physical systems through experiment, or observation. For instance, Maxwell's equations for the electromagnetic field are little more than a succinct summary of the observed properties of electric and magnetic fields and any associated charges and currents. These equations were abstracted from the results of innumerable experiments performed over centuries, along with some clever interpolation on the part of Maxwell. Similar comments could be made about Newton's laws of motion, or thermodynamics. Data is collected, either by casual observation or controlled experiment on, for instance the motion of physical objects, or on the temperature, pressure, volume of solids, liquids, or gases and so on. Within this data, regularities are observed which are

best summarized as equations:

$$\mathbf{F} = m\mathbf{a} \qquad \text{— Newton's second law;}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{— One of Maxwell's equations (Faraday's law);}$$

$$PV = NkT \qquad \text{— Ideal gas law (not really a fundamental law)}$$

What these equations represent are relationships between *information* gained by *observation* of various physical systems and as such are a succinct way of summarizing the relationship between the data, or the information, collected about a physical system. The laws are expressed in a manner consistent with how we understand the world from the view point of classical physics in that the symbols replace precisely known or knowable values of the physical quantities they represent. There is no uncertainty or randomness as a consequence of our ignorance of information about a system implicit in any of these equations. Moreover, classical physics says that this information is a faithful representation of what is 'really' going on in the physical world. These might be called the 'classical laws of information' implicit in classical physics.

What these pre-quantum experimenters were not to know was that the information they were gathering was not refined enough to show that there were fundamental limitations to the accuracy with which they could measure physical properties. Moreover, there was some information that they might have taken for granted as being accessible, simply by trying hard enough, but which we now know could not have been obtained at all! There was in operation unsuspected laws of nature that placed constraints on the information that could be obtained about any physical system. In the absence in the data of any evidence of these laws of nature, the information that was gathered was ultimately organised into mathematical statements that constituted *classical* laws of physics: Maxwell's equations, or Newton's laws of motion. But in the late nineteenth century and on into the twentieth century, experimental evidence began to accrue that suggested that there was something seriously amiss with the classical laws of physics: the data could no longer be fitted to the equations, or, in other words, the theory could not explain the observed experimental results. The choice was clear: either modify the existing theories, or formulate new ones. It was the latter approach that succeeded. Ultimately, what was formulated was a new set of laws of nature, the laws of quantum mechanics, which were essentially a set of laws concerning the *information* that could be gained about the physical world.

These are *not* the same laws as implicit in classical physics. For instance, there are limits on the information that can be gained about a physical system. For instance, if in an experiment we measure the position x of a particle with an accuracy[1] of Δx, and then measure the momentum p of the particle we find that the result for p randomly varies from one run of the experiment to the next, spread over a range Δp. But there is still law here. Quantum mechanics tells us that

$$\Delta x \Delta p \geq \tfrac{1}{2}\hbar \text{ — the Heisenberg Uncertainty Relation}$$

Quantum mechanics also tells us how this information is processed e.g. as a system evolves in time (the Schrödinger equation) or what results might be obtained in in a randomly varying way in a measurement. Quantum mechanics is a theory of information, quantum information theory.

What are the consequences? First, it seems that we lose the apparent certainty and determinism of classical physics, this being replaced by uncertainty and randomness. This randomness is not due to our inadequacies as experimenters — it is built into the very fabric of the physical world. But on the positive side, these quantum laws mean that physical systems can do so much more within these restrictions. A particle with position or momentum uncertain by amounts Δx and Δp means we do not quite know where it is, or how fast it is going, and we can never know this. But

[1] Accuracy indicates closeness to the true value, precision is the repeatability or reproducibility of the measurement.

the particle can be doing a lot more things ‘behind the scenes’ as compared to a classical particle of precisely defined position and momentum. The result is infinitely richer physics — quantum physics.

Chapter 2

The Early History of Quantum Mechanics

In the early years of the twentieth century, Max Planck, Albert Einstein, Louis de Broglie, Neils Bohr, Werner Heisenberg, Erwin Schrödinger, Max Born, Paul Dirac and others created the theory now known as quantum mechanics. The theory was not developed in a strictly logical way – rather, a series of guesses inspired by profound physical insight and a thorough command of new mathematical methods was sewn together to create a theoretical edifice whose predictive power is such that quantum mechanics is considered the most successful theoretical physics construct of the human mind. Roughly speaking the history is as follows:

Planck's Black Body Theory (1900) One of the major challenges of theoretical physics towards the end of the nineteenth century was to derive an expression for the spectrum of the electromagnetic energy emitted by an object in thermal equilibrium at some temperature T. Such an object is known as a *black body*, so named because it absorbs light of any frequency falling on it. A black body also emits electromagnetic radiation, this being known as black body radiation, and it was a formula for the spectrum of this radiation that was being sort for. One popular candidate for the formula was Wein's law:

$$\mathcal{S}(f,T) = \alpha f^3 e^{-\beta f/T} \tag{2.1}$$

The quantity $\mathcal{S}(f,T)$, otherwise known as the *spectral distribution function*, is such that $\mathcal{S}(f,T)df$ is the energy contained in unit volume of electromagnetic radiation in thermal equilibrium at an absolute temperature T due to waves of frequency between f and $f+df$. The above expression for S was not so much derived from a more fundamental theory as quite simply guessed. It was a formula that worked well at high frequencies, but was found to fail when improved experimental techniques made it possible to measure $\mathcal{S}$ at lower (infrared) frequencies. There was another candidate for $\mathcal{S}$ which was derived using arguments from classical physics which lead to a formula for $\mathcal{S}(f,T)$ known as the Rayleigh-Jeans formula:

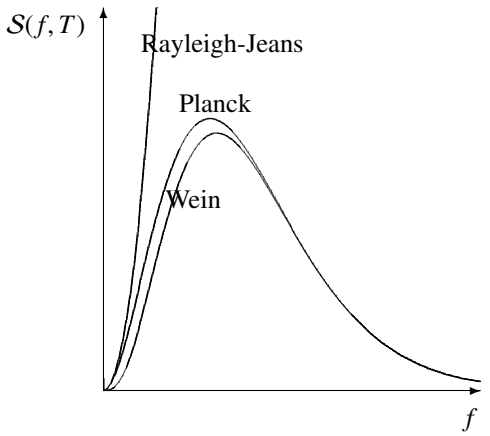


Figure 2.1: Rayleigh-Jeans (classical), Wein, and Planck spectral distributions.

$$\mathcal{S}(f,T) = \frac{8\pi f^2}{c^3} k_B T \tag{2.2}$$

where k_B is a constant known as Boltzmann's constant. This formula worked well at low frequencies, but suffered from a serious problem – it clearly increases without limit with increasing frequency – there is more and more energy in the electromagnetic field at higher and higher frequencies. This amounts to saying that an object at *any* temperature would radiate an infinite

amount of energy at infinitely high frequencies. This result, ultimately to become known as the 'ultra-violet catastrophe', is obviously incorrect, and indicates a deep flaw in classical physics.

In an attempt to understand the form of the spectrum of the electromagnetic radiation emitted by a black body, Planck proposed a formula which he obtained by looking for a formula that fitted Wein's law at high frequencies, and also fitted the new low frequency experimental results (which happen to be given by the Rayleigh-Jeans formula, though Planck was not aware of this). It was when he tried to provide a deeper explanation for the origin of this formula that he made an important discovery whose significance even he did not fully appreciate.

In this derivation, Planck proposed that the atoms making up the black body object absorbed and emitted light of frequency f in multiples of a fundamental unit of energy, or quantum of energy, $E = hf$. On the basis of this assumption, he was able to rederive the formula he had earlier guessed:

$$\mathcal{S}(f,T) = \frac{8\pi h f^3}{c^3}\frac{1}{\exp(hf/kT)-1}. \tag{2.3}$$

This curve did not diverge at high frequencies – there was no ultraviolet catastrophe. Moreover, by fitting this formula to experimental results, he was able to determine the value of the constant h, that is, $h = 6.6218 \times 10^{-34}$Joule-sec. This constant was soon recognized as a new fundamental constant of nature, and is now known as Planck's constant.

In later years, as quantum mechanics evolved, it was found that the ratio $h/2\pi$ arose time and again. As a consequence, Dirac introduced a new quantity $\hbar = h/2\pi$, pronounced 'h-bar', which is now the constant most commonly encountered. In terms of $\hbar$, Planck's formula for the quantum of energy becomes

$$E = hf = (h/2\pi)\,2\pi f = \hbar\omega \tag{2.4}$$

where ω is the angular frequency of the light wave.

Einstein's Light Quanta (1905) Although Planck believed that the rule for the absorption and emission of light in quanta applied only to black body radiation, and was a property of the atoms, rather than the radiation, Einstein saw it as a property of electromagnetic radiation, whether it was black body radiation or of any other origin. In particular, in his work on the photoelectric effect, he proposed that light of frequency ω was made up of quanta or 'packets' of energy $\hbar\omega$ which could be only absorbed or emitted in their entirety.

Bohr's Model of the Hydrogen Atom (1913) Bohr then made use of Einstein's ideas in an attempt to understand why hydrogen atoms do not self destruct, as they should according to the laws of classical electromagnetic theory. As implied by the Rutherford scattering experiments, a hydrogen atom consists of a positively charged nucleus (a proton) around which circulates a very light (relative to the proton mass) negatively charged particle, an electron. Classical electromagnetism says that as the electron is accelerating in its circular path, it should be radiating away energy in the form of electromagnetic waves, and do so on a time scale of $\sim 10^{-12}$ seconds, during which time the electron would spiral into the proton and the hydrogen atom would cease to exist. This obviously does not occur.

Bohr's solution was to propose that provided the electron circulates in orbits whose radii r satisfy an ad hoc rule, now known as a quantization condition, applied to the angular momentum L of the electron

$$L = mvr = n\hbar \tag{2.5}$$

where v is the speed of the electron and m its mass, and n a positive integer (now referred to as a *quantum number*), then these orbits would be *stable* – the hydrogen atom was said to be in a stationary state. He could give no physical reason why this should be the case, but on the basis of

this proposal he was able to show that the hydrogen atom could only have energies given by the formula

$$E_n = -\frac{ke^2}{2a_0}\frac{1}{n^2} \tag{2.6}$$

where $k = 1/4\pi\epsilon_0$ and

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} = 0.0529 \text{ nm} \tag{2.7}$$

is known as the Bohr radius, and roughly speaking gives an indication of the size of an atom as determined by the rules of quantum mechanics. Later we shall see how an argument based on the uncertainty principle gives a similar result.

The tie-in with Einstein's work came with the further proposal that the hydrogen atom emits or absorbs light quanta by 'jumping' between the energy levels, such that the frequency f of the photon emitted in a downward transition from the stationary state with quantum number n_i to another of lower energy with quantum number n_f would be

$$f = \frac{E_{n_i} - E_{n_f}}{h} = \frac{ke^2}{2a_0 h}\left[\frac{1}{n_f^2} - \frac{1}{n_i^2}\right]. \tag{2.8}$$

Einstein used these ideas of Bohr to rederive the black body spectrum result of Planck. In doing so, he set up the theory of emission and absorption of light quanta, including spontaneous (i.e. 'uncaused' emission) – the first intimation that there were processes occurring at the atomic level that were intrinsically probabilistic. This work also lead him to the conclusion that the light quanta were more than packets of energy, but carried momentum in a particular direction – the light quanta were, in fact, particles, subsequently named photons by the chemist Gilbert Lewis.

There was some success in extracting a general method, now known as the 'old' quantum theory, from Bohr's model of the hydrogen atom. But this theory, while quite successful for the hydrogen atom, was an utter failure when applied to even the next most complex atom, the helium atom. The ad hoc character of the assumptions on which it was based gave little clue to the nature of the underlying physics, nor was it a theory that could describe a dynamical system, i.e. one that was evolving in time. Its role seems to have been one of 'breaking the ice', freeing up the attitudes of researchers at that time to old paradigms, and opening up new ways of looking at the physics of the atomic world.

De Broglie's Hypothesis (1924) Inspired by Einstein's picture of light, a form of wave motion, as also behaving in some circumstances as if it was made up of particles, and inspired also by the success of the Bohr model of the hydrogen atom, de Broglie was lead, by purely aesthetic arguments to make a radical proposal: if light waves can behave under some circumstances like particles, then by symmetry it is reasonable to suppose that particles such as an electron (or a planet?) can behave like waves. More precisely, if light waves of frequency ω can behave like a collection of particles of energy $E = \hbar\omega$, then by symmetry, a massive particle of energy E, an electron say, should behave under some circumstances like a wave of frequency $\omega = E/\hbar$. But assigning a frequency to these waves is not the end of the story. A wave is also characterised by its wavelength, so it is also necessary to assign a wavelength to these 'matter waves'. For a particle of light, a photon, the wavelength of the associated wave is $\lambda = c/f$ where $f = \omega/2\pi$. So what is it for a massive particle? A possible formula for this wavelength can be obtained by looking a little further at the case of the photon. In Einstein's theory of relativity, a photon is recognized as a particle of zero rest mass, and as such the energy of a photon (moving freely in empty space) is related to its momentum p by $E = pc$. From this it follows that

$$E = \hbar\omega = \hbar\, 2\pi c/\lambda = pc \tag{2.9}$$

so that, since $\hbar = h/2\pi$

$$p = h/\lambda. \tag{2.10}$$

This equation then gave the wavelength of the photon in terms of its momentum, but it is also an expression that contains nothing that is specific to a photon. So de Broglie assumed that this relationship applied to all *free* particles, whether they were photons or electrons or anything else, and so arrived at the pair of equations

$$f = E/h \qquad \lambda = h/p \tag{2.11}$$

which gave the frequency and wavelength of the waves that were to be associated with a free particle of kinetic energy E and momentum p. Strictly speaking, the relativistic expressions for the momentum and energy of a particle of non-zero rest mass ought to be used in these formula, as these above formulae were derived by making use of results of special relativity. However, here we will be concerned solely with the non-relativistic limit, and so the non-relativistic expressions, $E = \frac{1}{2}mv^2$ and $p = mv$ will suffice[1].

This work constituted de Broglie's PhD thesis. It was a pretty thin affair, a few pages long, and while it was looked upon with some scepticism by the thesis examiners, the power and elegance of his ideas and his results were immediately appreciated by Einstein, more reluctantly by others, and lead ultimately to the discovery of the wave equation by Schrödinger, and the development of wave mechanics as a theory describing the atomic world.

Experimentally, the first evidence of the wave nature of massive particles was seen by Davisson and Germer in 1926 when they fired a beam of electrons of known energy at a nickel crystal in which the nickel atoms are arranged in a regular array. Much to the surprise of the experimenters (who were not looking for any evidence of wave properties of electrons), the electrons reflected off the surface of the crystal to form an interference pattern. The characteristics of this pattern were entirely consistent with the electrons behaving as waves, with a wavelength given by the de Broglie formula, that were reflected by the periodic array of atoms in the crystal (which acted much like slits in a diffraction grating).

An immediate success of de Broglie's hypothesis was that it gave an explanation, of sorts, of the quantization condition $L = n\hbar$. If the electron circulating around the nucleus is associated with a wave of wavelength λ, then for the wave not to destructively interfere with itself, there must be a whole number of waves (see Fig. (2.2)) fitting into one circumference of the orbit, i.e.

$$n\lambda = 2\pi r. \tag{2.12}$$

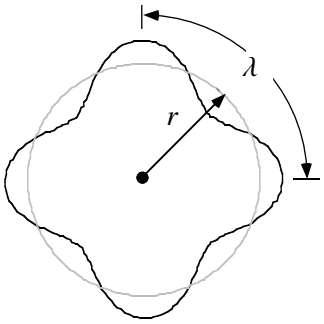


Figure 2.2: De Broglie wave for which four wavelengths λ fit into a circle of radius r.

Using the de Broglie relation $\lambda = h/p$ then gives $L = pr = n\hbar$ which is just Bohr's quantization condition. But now, given that particles can exhibit wave like properties, the natural question that arises is: what is doing the 'waving'? Further, as wave motion is usually describable in terms of some kind of wave equation, it is then also natural to ask what the wave equation is for these de Broglie waves. The latter question turned out to be much easier to answer than the first – these waves satisfy the famous Schrödinger wave equation. But what these waves are is still, largely speaking, an incompletely answered question: are they 'real' waves, as Schrödinger believed, in the sense that they represent some kind of physical vibration in the same way as water or sound or light waves, or are

[1]For a particle moving in the presence of a spatially varying potential, momentum is not constant so the wavelength of the waves will also be spatially dependent – much like the way the wavelength of light waves varies as the wave moves through a medium with a spatially dependent refractive index. In that case, the de Broglie recipe is insufficient, and a more general approach is needed – Schrödinger's equation.

they something more abstract, waves carrying information, as Einstein seemed to be the first to intimate. The latter is an interpretation that has been gaining in favour in recent times, a perspective that we can support somewhat by looking at what we can learn about a particle by studying the properties of these waves. It is this topic to which we now turn.

Chapter 3

The Wave Function

On the basis of the assumption that the de Broglie relations give the frequency and wavelength of some kind of wave to be associated with a particle, plus the assumption that it makes sense to add together waves of different frequencies, it is possible to learn a considerable amount about these waves without actually knowing beforehand what they represent. But studying different examples does provide some insight into what the ultimate interpretation is, the so-called Born interpretation, which is that these waves are 'probability waves' in the sense that the amplitude squared of the waves gives the probability of observing (or detecting, or finding – a number of different terms are used) the particle in some region in space. Hand-in-hand with this interpretation is the Heisenberg uncertainty principle which, historically, preceded the formulation of the probability interpretation. From this principle, it is possible to obtain a number of fundamental results even before the full machinery of wave mechanics is in place.

In this Chapter, some of the consequences of de Broglie's hypothesis of associating waves with particles are explored, leading to the concept of the wave function, and its probability interpretation.

3.1 The Harmonic Wave Function

On the basis of de Broglie's hypothesis, there is associated with a particle of energy E and momentum p, a wave of frequency f and wavelength λ given by the de Broglie relations Eq. (2.11). It is more usual to work in terms of the angular frequency $\omega = 2\pi f$ and wave number $k = 2\pi/\lambda$ so that the de Broglie relations become

$$\omega = E/\hbar \qquad k = p/\hbar. \tag{3.1}$$

With this in mind, and making use of what we already know about what the mathematical form is for a wave, we are in a position to make a reasonable guess at a mathematical expression for the wave associated with the particle. The possibilities include (in one dimension)

$$\Psi(x,t) = A\sin(kx-\omega t), \quad A\cos(kx-\omega t), \quad Ae^{i(kx-\omega t)}, \quad \ldots \tag{3.2}$$

At this stage, we have no idea what the quantity $\Psi(x,t)$ represents physically. It is given the name the *wave function*, and in this particular case we will use the term harmonic wave function to describe any trigonometric wave function of the kind listed above. As we will see later, in general it can take much more complicated forms than a simple single frequency wave, and is almost always a complex valued function. In fact, it turns out that the third possibility listed above is the appropriate wave function to associate with a free particle, but for the present we will work with real wave functions, if only because it gives us the possibility of visualizing their form while discussing their properties.

In order to gain an understanding of what a wave function might represent, we will turn things around briefly and look at what we can learn about a particle if we know what its wave function is. We are implicitly bypassing here any consideration of whether we can understand a wave function as being a physical wave in the same way that a sound wave, a water wave, or a light wave are physical waves, i.e. waves made of some kind of physical 'stuff'. Instead, we are going to look on a wave function as something that gives us information on the particle it is associated with. To this end, we will suppose that the particle has a wave function given by $\Psi(x,t) = A\cos(kx - \omega t)$. Then, given that the wave has angular frequency ω and wave number k, it is straightforward to calculate the wave velocity, that is, the phase velocity v_p of the wave, which is just the velocity of the wave crests. This phase velocity is given by

$$v_p = \frac{\omega}{k} = \frac{\hbar\omega}{\hbar k} = \frac{E}{p} = \frac{\frac{1}{2}mv^2}{mv} = \frac{1}{2}v. \tag{3.3}$$

Thus, given the frequency and wave number of a wave function, we can determine the speed of the particle from the phase velocity of its wave function, $v = 2v_p$. We could also try to learn from the wave function the position of the particle. However, the wave function above tells us nothing about where the particle is to be found in space. We can make this statement because this wave function is more or less the same everywhere. For sure, the wave function is not *exactly* the same everywhere, but any feature that we might decide as being an indicator of the position of the particle, say where the wave function is a maximum, or zero, will not do: the wave function is periodic, so any feature, such as where the wave function vanishes, reoccurs an infinite number of times, and there is no way to distinguish any one of these repetitions from any other, see Fig. (3.1).

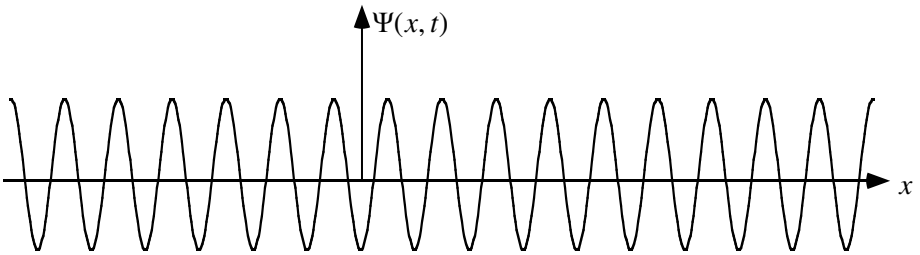


Figure 3.1: A wave function of constant amplitude and wavelength. The wave is the same everywhere and so there is no distinguishing feature that could indicate one possible position of the particle from any other.

Thus, this particular wave function gives no information on the whereabouts of the particle with which it is associated. So from a harmonic wave function it is possible to learn how fast a particle is moving, but not what the position is of the particle.

3.2 Wave Packets

From what was said above, a wave function constant throughout all space cannot give information on the position of the particle. This suggests that a wave function that did not have the same amplitude throughout all space might be a candidate for a giving such information. In fact, since what we mean by a particle is a physical object that is confined to a highly localized region in space, ideally a point, it would be intuitively appealing to be able to devise a wave function that is zero or nearly so everywhere in space except for one localized region. It is in fact possible to construct, from the harmonic wave functions, a wave function which has this property. To show how this is done, we first consider what happens if we combine together two harmonic waves whose wave numbers are very close together. The result is well-known: a 'beat note' is produced, i.e. periodically in space the waves add together in phase to produce a local maximum, while

midway in between the waves will be totally out of phase and hence will destructively interfere. This is illustrated in Fig. 3.2(a) where we have added together two waves $\cos(5x) + \cos(5.25x)$.

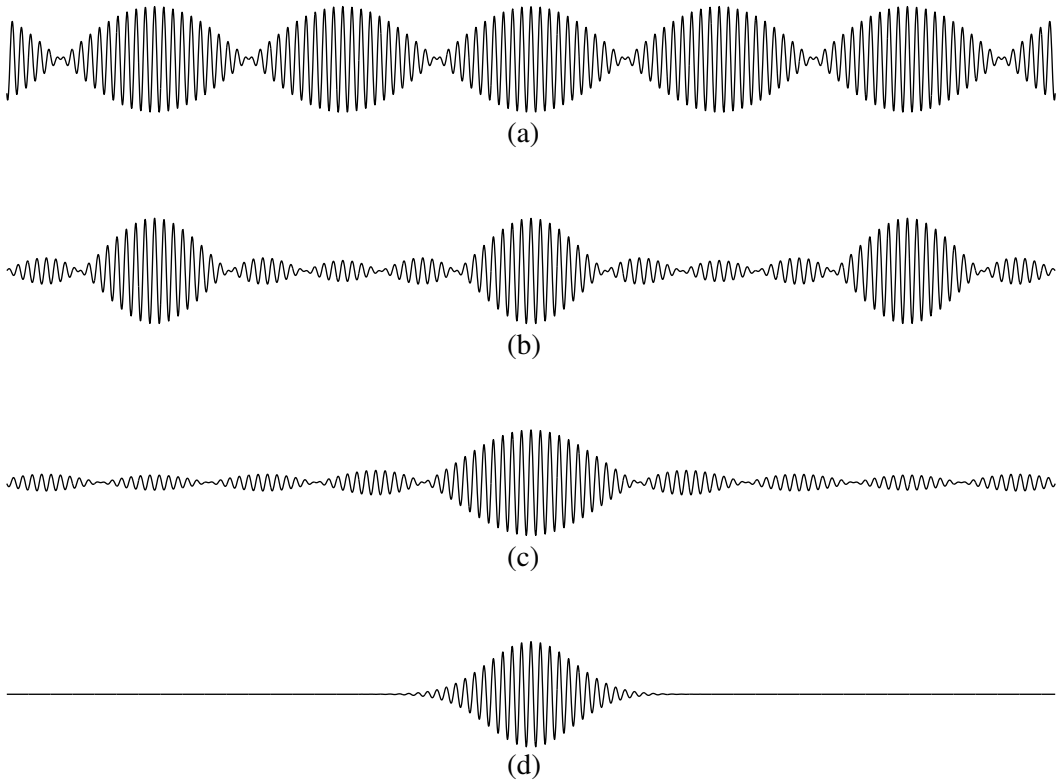


Figure 3.2: (a) Beat notes produced by adding together two cos waves: $\cos(5x) + \cos(5.25x)$.
(b) Combining five cos waves: $\cos(4.75x) + \cos(4.875x) + \cos(5x) + \cos(5.125x) + \cos(5.25x)$.
(c) Combining seven cos waves: $\cos(4.8125x) + \cos(4.875x) + \cos(4.9375x) + \cos(5x) + \cos(5.0625x) + \cos(5.125x) + \cos(5.1875x)$.
(d) An integral over a continuous range of wave numbers produces a single wave packet.

Now suppose we add five such waves together, as in Fig. 3.2(b). The result is that some beats turn out to be much stronger than the others. If we repeat this process by adding seven waves together, but now make them closer in wave number, we get Fig. 3.2(c), we find that most of the beat notes tend to become very small, with the strong beat notes occurring increasingly far apart. Mathematically, what we are doing here is taking a limit of a sum, and turning this sum into an integral. In the limit, we find that there is only one beat note – in effect, all the other beat notes become infinitely far away. This single isolated beat note is usually referred to as a wave packet.

We need to look at this in a little more mathematical detail, so suppose we add together a large number of harmonic waves with wave numbers $k_1, k_2, k_3, \ldots$ all lying in the range:

$$\overline{k} - \Delta k \lesssim k_n \lesssim \overline{k} + \Delta k \tag{3.4}$$

around a value $\overline{k}$, i.e.

$$\begin{aligned}\Psi(x,t) =& A(k_1)\cos(k_1 x - \omega_1 t) + A(k_2)\cos(k_2 x - \omega_2 t) + \ldots \\ =& \sum_n A(k_n)\cos(k_n x - \omega_n t)\end{aligned} \tag{3.5}$$

where $A(k)$ is a function peaked about the value $\overline{k}$ with a full width at half maximum of $2\Delta k$. (There is no significance to be attached to the use of cos functions here – the idea is simply to illustrate a

point. We could equally well have used a sin function or indeed a complex exponential.) What is found is that in the limit in which the sum becomes an integral:

$$\Psi(x,t) = \int_{-\infty}^{+\infty} A(k)\cos(kx - \omega t)\,dk \tag{3.6}$$

all the waves interfere constructively to produce only a single beat note as illustrated in Fig. 3.2(d) above[1]. The wave function or wave packet so constructed is found to have essentially zero amplitude everywhere except for a *single* localized region in space, over a region of width $2\Delta x$, i.e. the wave function $\Psi(x,t)$ in this case takes the form of a single wave packet, see Fig. (3.3).

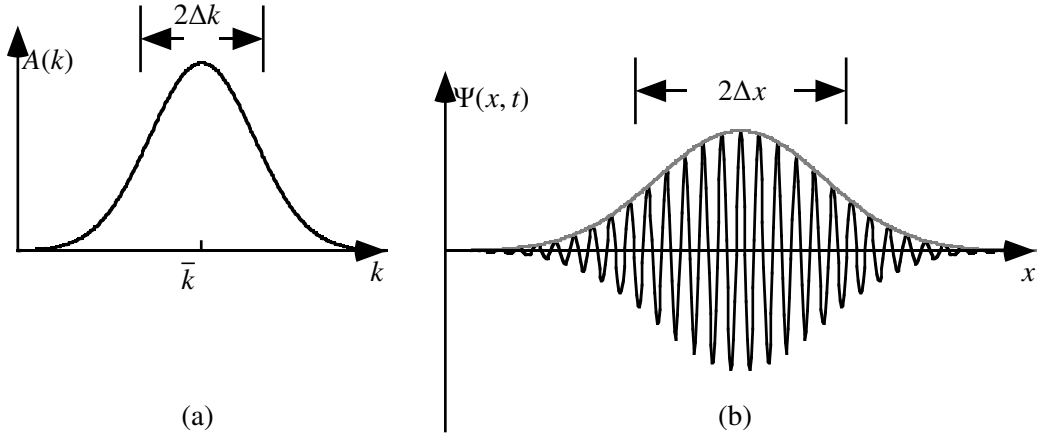


Figure 3.3: (a) The distribution of wave numbers k of harmonic waves contributing to the wave function $\Psi(x,t)$. This distribution is peaked about $\overline{k}$ with a width of $2\Delta k$. (b) The wave packet $\Psi(x,t)$ of width $2\Delta x$ resulting from the addition of the waves with distribution $A(k)$. The oscillatory part of the wave packet (the 'carrier wave') has wave number $\overline{k}$.

This wave packet is clearly particle-like in that its region of significant magnitude is confined to a localized region in space. Moreover, this wave packet is constructed out of a group of waves with an average wave number $\overline{k}$, and so these waves could be associated in some sense with a particle of momentum $\overline{p} = \hbar\overline{k}$. If this were true, then the wave packet would be expected to move with a velocity of $\overline{p}/m$. This is in fact found to be the case, as the following calculation shows.

Because a wave packet is made up of individual waves which themselves are moving, though not with the same speed, the wave packet itself will move (and spread as well). The speed with which the wave packet moves is given by its group velocity v_g:

$$v_g = \left(\frac{d\omega}{dk}\right)_{k=\overline{k}}. \tag{3.7}$$

This is the speed of the maximum of the wave packet i.e. it is the speed of the point on the wave packet where all the waves are in phase. Calculating the group velocity requires determining the relationship between ω to k, known as a dispersion relation. This dispersion relation is obtained from

$$E = \tfrac{1}{2}mv^2 = \frac{p^2}{2m}. \tag{3.8}$$

[1]In Fig. 3.2(d), the wave packet is formed from the integral

$$\Psi(x,0) = \frac{1}{4\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-((k-5)/4)^2}\cos(kx)\,dk.$$

Substituting in the de Broglie relations Eq. (2.11) gives

$$\hbar\omega = \frac{\hbar^2 k^2}{2m} \tag{3.9}$$

from which follows the dispersion relation

$$\omega = \frac{\hbar k^2}{2m}. \tag{3.10}$$

The group velocity of the wave packet is then

$$v_g = \left(\frac{d\omega}{dk}\right)_{k=\overline{k}} = \frac{\hbar\overline{k}}{m}. \tag{3.11}$$

Substituting $\overline{p} = \hbar\overline{k}$, this becomes $v_g = \overline{p}/m$. i.e. the packet is indeed moving with the velocity of a particle of momentum $\overline{p}$, as suspected. This is a result of some significance, i.e. we have constructed a wave function of the form of a wave packet which is particle-like in nature. But unfortunately this is done at a cost. We had to combine together harmonic wave functions $\cos(kx - \omega t)$ with a range of k values $2\Delta k$ to produce a wave packet which has a spread in space of size $2\Delta x$. The two ranges of k and x are not unrelated – their connection is embodied in an important result known as the Heisenberg Uncertainty Relation.

3.3 The Heisenberg Uncertainty Relation

The wave packet constructed in the previous section obviously has properties that are reminiscent of a particle, but it is not entirely particle-like — the wave function is non-zero over a region in space of size $2\Delta x$. In the absence of any better way of relating the wave function to the position of the atom, it is intuitively appealing to suppose that where $\Psi(x,t)$ has its greatest amplitude is where the particle is most likely to be found, i.e. the particle is to be found somewhere in a region of size $2\Delta x$. More than that, however, we have seen that to construct this wavepacket, harmonic waves having k values in the range $(\overline{k} - \Delta k, \overline{k} + \Delta k)$ were adding together. These ranges Δx and Δk are related by the bandwidth theorem, which applies when adding together harmonic waves, which tell us that

$$\Delta x \Delta k \gtrsim 1. \tag{3.12}$$

Using $p = \hbar k$, we have $\Delta p = \hbar \Delta k$ so that

$$\Delta x \Delta p \gtrsim \hbar. \tag{3.13}$$

A closer look at this result is warranted. A wave packet that has a significant amplitude within a region of size $2\Delta x$ was constructed from harmonic wave functions which represent a range of momenta $\overline{p} - \Delta p$ to $\overline{p} + \Delta p$. We can say then say that the particle is likely to be found somewhere in the region $2\Delta x$, and given that wave functions representing a range of possible momenta were used to form this wave packet, we could also say that the momentum of the particle will have a value in the range $\overline{p} - \Delta p$ to $\overline{p} + \Delta p$[2]. The quantities Δx and Δp are known as *uncertainties*, and the relation above Eq. (3.14) is known as the Heisenberg uncertainty relation for position and momentum.

All this is rather abstract. We do not actually 'see' a wave function accompanying its particle, so how are we to know how 'wide' the wave packet is, and hence what the uncertainty in position and momentum might be for a given particle, say an electron orbiting in an atomic nucleus, or the

[2]In fact, we can look on $A(k)$ as a wave function for k or, since $k = p/\hbar$ as effectively a wave function for momentum analogous to $\Psi(x,t)$ being a wave function for position.

nucleus itself, or an electron in a metal or ...? The answer to this question is intimately linked with what has been suggested by the use above of such phrases as 'where the particle is most likely to be found' and so on, words that are hinting at the fundamental role of randomness as an intrinsic property of quantum systems, and role of probability in providing a meaning for the wave function.

To get a flavour of what is meant here, we can suppose that we have a truly vast number of identical particles, say 10^{25}, all prepared one at a time in some experiment so that they all have associated with them the same wave packet. For half these particles, we measure their position at the same time, i.e. at, say, 10 sec after they emerge from the apparatus, and for the other half we measure their momentum. What we find is that the results for the position are not all the same: they are spread out randomly around some average value, and the range over which they are spread is most conveniently measured by the usual tool of statistics: the standard deviation. This standard deviation in position turns out to be just the uncertainty Δx we introduced above in a non-rigorous manner. Similarly, the results for the measurement of momentum for the other half are randomly scattered around some average value, and the spread around the average is given by the standard deviation once again. This standard deviation in momentum we identify with the uncertainty Δp introduced above.

With uncertainties defined as standard deviations of random results, it is possible to give a more precise statement of the uncertainty relation, which is:

$$\Delta x \Delta p \geq \tfrac{1}{2}\hbar \tag{3.14}$$

but we will mostly use the result Eq. (3.13). The detailed analysis is left to much later (See Chapter ∞).

The Heisenberg relation has an immediate interpretation. It tells us that we cannot determine, from knowledge of the wave function alone, the exact position and momentum of a particle at the same time. In the extreme case that $\Delta x = 0$, then the position uncertainty is zero, but Eq. (3.14) tells us that the uncertainty on the momentum is infinite, i.e. the momentum is entirely unknown. A similar statement applies if $\Delta p = 0$. In fact, this last possibility is the case for the example of a single harmonic wave function considered in Section 3.1. However, the uncertainty relation does not say that we cannot measure the position and the momentum at the same time. We certainly can, but we have to live with the fact that each time we repeat this simultaneous measurement of position and momentum on a collection of electrons all prepared such as to be associated with the same wave packet, the results that are obtained will vary randomly from one measurement to the next, i.e. the measurement results continue to carry with them uncertainty by virtue of the uncertainty relation.

This conclusion that it is impossible for a particle to have zero uncertainty in both position *and* momentum at the same time flies in the face of our intuition, namely our belief that a particle moving through space will at any instant have a definite position and momentum which we can, in principle, measure to arbitrary accuracy. We could then feel quite justified in arguing that our wave function idea is all very interesting, but that it is not a valid description of the physical world, or perhaps it is a perfectly fine concept but that it is incomplete, that there is information missing from the wave function. Perhaps there is a prescription still to be found that will enable us to complete the picture: retain the wave function but add something further that will then not forbid our being able to measure the position and the momentum of the particle precisely and at the same time. This, of course, amounts to saying that the wave function by itself does not give complete information on the state of the particle. Einstein fought vigorously for this position i.e. that the wave function was not a complete description of 'reality', and that there was somewhere, in some sense, a repository of missing information that will remove the incompleteness of the wave function — so-called 'hidden variables'. Unfortunately (for those who hold to his point

of view) evidence has mounted, particularly in the past few decades, that the wave function (or its analogues in the more general formulation of quantum mechanics) does indeed represent the full picture — the most that can ever be known about a particle (or more generally any system) is what can be learned from its wave function. This means that the difficulty encountered above concerning not being able to pinpoint the position and the momentum of a particle from knowledge of its wave function is not a reflection of any inadequacy on the part of experimentalists trying to measure these quantities, but is an irreducible property of the natural world. Nevertheless, at the macroscopic level the uncertainties mentioned above become so small as to be experimentally unmeasurable, so at this level the uncertainty relation has no apparent effect.

The limitations implied by the uncertainty relation as compared to classical physics may give the impression that something has been lost, that nature has prevented us, to an extent quantified by the uncertainty principle, from having complete information about the physical world. To someone wedded to the classical deterministic view of the the physical world (and Einstein would have to be counted as one such person), it appears to be the case that there is information that is hidden from us. This may then be seen as a cause for concern because it implies that we cannot, even in principle, make exact predictions about the behaviour of any physical system. However, the view can be taken that the opposite is true, that the uncertainty principle is an indicator of greater freedom. In a sense, the uncertainty relation means it is possible for a physical system to have a much broader range of possible physical properties consistent with the smaller amount of information that is available about its properties. This leads to a greater richness in the properties of the physical world than could ever be found within classical physics.

3.3.1 The Heisenberg microscope: the effect of measurement

The Heisenberg Uncertainty Relation is enormously general. It applies *without saying anything whatsoever about the nature of the particle, how it is prepared in an experiment, what it is doing, what it might be interacting with* .... It is clearly a profoundly significant physical result. But at its heart it is simply a mathematical statement about the properties of waves that flows from the assumed wave properties of matter plus some assumptions about the physical interpretation of these waves. There is little indication of what the *physics* might be that underlies it. One way to uncover what physics might be present is to study what takes place if we attempt to measure the position or the momentum of a particle. This is in fact the problem initially addressed by Heisenberg, and leads to a result that is superficially the same as 3.14, but, from a physics point of view, there is in fact a subtle difference between what Heisenberg was doing, and what Eq. (3.14) is saying.

Heisenberg's analysis was based on a *thought* experiment, i.e. an experiment that was not actually performed, but was instead analysed as a mental construct. From this experiment, it is possible to show, by taking account of the quantum nature of light and matter, that measuring the position of an electron results in an unavoidable, unpredictable change in its momentum. More than that, it is possible to show that if the particle's position were measured with ever increasing precision, the result was an ever greater disturbance of the particle's momentum. This is an outcome that is summarized mathematically by a formula essentially the same as Eq. (3.14).

In his thought experiment, Heisenberg considered what was involved in attempting to measure the position of a particle, an electron say, by shining light on the electron and observing the scattered light through a microscope. To analyse this measurement process, arguments are used which are a curious mixture of ideas from classical optics (the wave theory of light) and from the quantum theory of light (that light is made up of particles).

Classical optics enters the picture by virtue of the fact that in trying to measure the position of a *point* object using light, we must take into account the imprecision inherent in such a measurement. If a point object is *held fixed* (or is assumed to have infinite mass) and is illuminated by a steady beam of light, then the image of this object, as produced by a lens on a photographic image plate, is not itself a point — it is a diffraction pattern, a smeared out blob, brightest in the centre of the blob and becoming darker towards the edges (forming a so-called Airy disc). If there are two closely positioned point objects, then what is observed is two overlapping diffraction patterns. This overlap diminishes as the point objects are moved further apart until eventually, the edge of the central blob of one pattern will roughly coincide with the edge of the other blob, see Fig. (3.5).

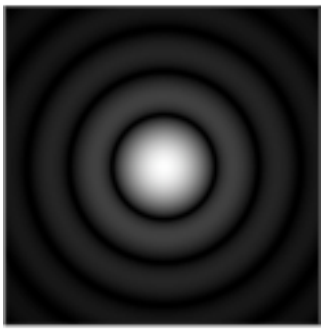


Figure 3.4: Airy disc diffraction pattern produced as image of a point object.

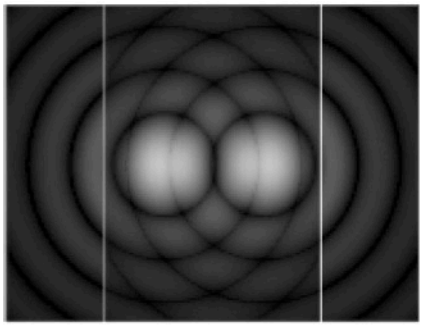


Figure 3.5: Airy disc diffraction pattern produced as image of a pair of point objects separated by distance d as give by Eqn. (3.15).

The separation d between the point objects for which this occurs can be shown to be given by

$$d = 2\frac{\lambda_l}{\sin\alpha} \tag{3.15}$$

where λ_l is the wavelength of the light, and 2α is the angle subtended at the lens by the object(s) (see Fig. (3.6)). This is a result that comes from the classical optics — it is essentially (apart from a factor of 2) the Rayleigh criterion for the resolution of a pair of images. For our purposes, we need to understand its meaning from a quantum mechanical perspective.

The quantum mechanical perspective arises because, according to quantum mechanics, a beam of light is to be viewed as a beam of individual particles, or photons. So, if we have a steady beam of light illuminating the fixed point object as before, then what we will observe on the photographic image plate is the formation of individual tiny spots, each associated with the arrival of a single photon. We would not see these individual photon arrivals with a normal every-day light source: the onslaught of photon arrivals is so overwhelming that all we see is the final familiar diffraction pattern. But these individual arrivals would be readily observed if the beam of light is weak, i.e. there is plenty of time between the arrival of one photon and the next. This gives rise to the question: we have individual particles (photons) striking the photographic plate, so where does the diffraction pattern come from? Answering this question goes to the heart of quantum mechanics. If we were to monitor where each photon strikes the image plate over a period of time, we find that the photons *strike at random*, more often in the centre, helping to build up the central bright region of the diffraction pattern, and more rarely towards the edges [3]. This probabilistic aspect of quantum mechanics we will study in depth in the following Chapter.

But what do we learn if we scatter just one photon off the point object? This photon will strike the image plate at some point, but we will have no way of knowing for sure if the point where

[3] In fact, the formation of a diffraction pattern, from which comes the Rayleigh criterion Eq. (3.15) is itself a consequence of the $\Delta x\Delta p \geq \frac{1}{2}\hbar$ form of the uncertainty relation applied to the photon making its way through the lens. The position at which the photon passes through the lens can be specified by an uncertainty $\Delta x \approx$ half the width of the lens, which means that the photon will acquire an uncertainty $\Delta p \approx h/\Delta x$ in a direction parallel to the lens. This momentum uncertainty means that photons will strike the photographic plate randomly over a region whose size is roughly the width of the central maximum of the diffraction pattern.

the photon arrives is a point near the centre of the diffraction pattern, or near the edge of such a pattern or somewhere in between — we cannot reconstruct the pattern from just one photon hit! But what we can say is that if one photon strikes the image plate, for example at the point c (on Fig. 3.6), then this point could be anywhere between two extreme possibilities. We could argue that the point object was sitting at a, and the photon has scattered to the far right of the diffraction pattern that would be built up by many photons being scattered from a point object at position a (labelled A in Fig. 3.6), or we could argue at the other extreme that the point object was at b, and the photon reaching c has simply landed at the far left hand edge of the diffraction pattern (labelled B in Fig. 3.6) associated with a point particle sitting at b. Or the point object could be somewhere in between a and b, in which case c would be within the central maximum of the associated diffraction pattern. The whole point is that the arrival of the photon at c is not enough for us to specify with certainty where the point object was positioned when it scattered the photon. If we let $2\delta x$ be the separation between a and b, then the best we can say, after detecting one photon only, is that the point object that scattered it was somewhere in the region between a and b, i.e., we can specify the position of the point object only to an accuracy of $\pm\delta x$. The argument that leads to Eq. (3.15) applies here, so we must put $d = 2\delta x$ in Eq. (3.15) and hence we have

$$\delta x = \frac{\lambda_l}{\sin\alpha}. \tag{3.16}$$

Note that the δx introduced here is to be understood as the resolution of the microscope. It is a property of the *apparatus* that we are using to measure the position of the fixed point object and so is not the same as the uncertainty Δx introduced earlier that appears in Eq. (3.14), that being a property of the wave function of the particle.

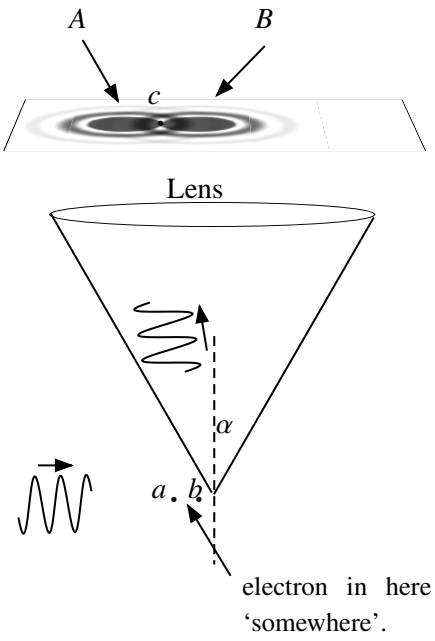


Figure 3.6: Diffraction images A and B corresponding to two extreme possible positions of a point object scattering a single photon that arrives at c. If one photon strikes the image plate, the point c where it arrives could be anywhere between two extreme possibilities: on the extreme right of the cenral maximum of a diffraction pattern (labelled A) built up by many photons scattered from a point object at position a, or on the extreme left of such a pattern (labelled B) built up by many photons being scattered from a point object at position b. With $2\delta x$ the separation beween a and b, the best we can say, after detecting one photon only, is that the object that scattered it was somewhere in the region between a and b, i.e., we can specify the position of the object only to an accuracy of $\pm\delta x$.

Now we turn to the experiment of interest, namely that of measuring the position of an electron presumably placed in the viewing range of the observing lens arrangement. In measuring its position we want to make sure that we disturb its position as little as possible. We can do this by using light whose intensity is as low as possible. Classically, this is not an issue: we can 'turn down' the light intensity to be as small as we like. But quantum mechanics gets in the way here. Given the quantum nature of light, the minimum intensity possible is that associated with the electron scattering only one photon. As we saw above, this single scattering event will enable us to

determine the position of the electron only to an accuracy given by Eq. (3.16). But this photon will have a momentum $p_l = h/\lambda_l$, and when it strikes the electron, it will be scattered, and the electron will recoil. It is reasonable to assume that the change in the wavelength of the light as a result of the scattering is negligibly small, but what cannot be neglected is the change in the direction of motion of the photon. If it is to be scattered in a direction so as to pass through the lens, and as we do not know the path that the photon follows through the lens — we only see where it arrives on the image plate — the best we can say is that it has two extreme possibilities defined by the edge of the cone of half angle α. Its momentum can therefore change by any amount up to $\pm p_l \sin\alpha$. Conservation of momentum then tells us that the momentum of the electron has consequently undergone a change of the same amount. In other words, the electron has now undergone a change in its momentum of an amount that could be as large as $\pm p_l \sin\alpha$. Just how big a change has taken place is not known as we do not know the path followed by the photon through the lens — we only know where the photon landed. So the momentum of the electron after the measurement has been *disturbed* by an unknown amount that could be as large as $\delta p = p_l \sin\alpha$. Once again, this quantity δp is not the same as the uncertainty Δp introduced earlier that appears in Eq. (3.14). Nevertheless we find that

$$\delta p \approx p_l \sin\alpha = \frac{h\sin\alpha}{\lambda_l} = \frac{h}{\delta x}, \tag{3.17}$$

using Eq. (3.16) and hence

$$\delta x \delta p \approx h \tag{3.18}$$

which apart from a factor $\sim 4\pi$, which can be neglected here given the imprecise way that we have defined δx and δp, is very similar to the uncertainty relation, $\Delta x \Delta p \geq \frac{1}{2}\hbar$!!!! In fact, to add to the confusion, the quantities δx and δp are also often referred to as 'uncertainties', but their meaning, and the meaning of Eq. (3.18) is not quite the same as Eq. (3.14).

Firstly, the derivation of Eq. (3.18) was explicitly based on the study of a measurement process. This is quite different from the derivation of the superficially identical relation, $\Delta x \Delta p \geq \frac{1}{2}\hbar$, derived by noting certain mathematical properties of the shape of a wave packet. Here, the 'uncertainty' δx is the resolution of the measuring apparatus, and δp is the disturbance in the momentum of the electron as a consequence of the physical effects of a measurement having been performed. In contrast, the uncertainties Δx and Δp, and the associated uncertainty relation was *not* derived by analysing some measurement processes — it simply states a property of wavepackets. The uncertainty Δp in momentum does not come about as a consequence of a measurement of the position of the particle, or vice versa.

Thus there are (at least) two 'versions' of Heisenberg's uncertainty relation. Which one is the more valid? Heisenberg's original version, $\delta x \delta p \approx h$, (the measurement-disturbance based 'δ version') played a very important role in the early development of quantum mechanics, but it has been recognized that the physical arguments used to arrive at the result are not strictly correct: the argument is neither fully correct classically or fully correct quantum mechanically. It can also be argued that the result follows from the use of the Rayleigh criterion, a *definition* based purely on experimental convenience, to derive a quantum mechanical result. On the other hand, the later formulation of the uncertainty relation, $\Delta x \Delta p \geq \frac{1}{2}\hbar$ (the statistical or 'Δ version'), in which the uncertainties in position and momentum are, in a sense, understood to be present at the same time for a particle, can be put on a sound physical and mathematical foundation, and is now viewed as being the more fundamental version. However, the close similarity of the two forms of the uncertainty relation suggests that this is more than just a coincidence. In fact, it is possible to show that in many circumstances, the measurement based 'δ version' does follow from the 'Δ version'. In each such case, the argument has to be tailored to suit the physics of the specific measurement procedure at hand, whether it be waves and optics as here, or masses on springs, or gravitational fields or whatever. The physical details of the measurement process can then be looked on as nature's way of guaranteeing that the electron indeed acquires an uncertainty

$\delta p \approx h/\delta x$ in its momentum if its position is measured with an uncertainty δx, while lurking in the background is the 'Δ version' of the uncertainty relation: in the example considered here, this describes how the uncertainty in the path followed by the photon through the lens leads to the formation of the diffraction pattern (see footnote 3 on p21). But the correspondence is not perfect — the two versions are not completely equivalent. No one has ever been able to show that Eq. (3.18) always follows from Eq. (3.14) for all and any measurement procedure, and for good reason. Einstein, Podolsky and Rosen showed that here are methods by which the position of a particle can be measured without physically interacting with the particle at all, so there is no prospect of the measurement disturbing the position of the particle.

Heisenberg's uncertainty principle has always been a source of both confusion and insight, not helped by Heisenberg's own shifting interpretation of his own work, and is still a topic that attracts significant research. The measurement-based 'δ version' has physical appeal as it seems to capture in an easily grasped fashion some of the peculiar predictions of quantum mechanics: measure the position of a particle to great accuracy, for instance, and you unavoidably thoroughly screw up its momentum. That performing an observation on a physical system can affect the system in an uncontrollably random fashion has been termed the 'observer effect', and is an aspect of quantum mechanics that has moved outside the purvey solely of quantum theory into other fields (such as sociology, for instance) involving the effects of making observations on other systems. But the statistical 'Δ version' is wholly quantum mechanical, and represents a significant constraint on the way physical systems can behave, and has remarkable predictive powers that belies the simplicity of its statement, as we will see in the next section.

3.3.2 The Size of an Atom

One important application of the uncertainty relation is to do with determining the size of atoms. Recall that classically atoms should not exist: the electrons must spiral into the nucleus, radiating away their excess energy as they do. However, if this were the case, then the situation would be arrived at in which the position and the momentum of the electrons would be known: stationary, and at the position of the nucleus. This is in conflict with the uncertainty principle, so it must be the case that the electron can spiral inward no further than an amount that is consistent with the uncertainty principle.

To see what the uncertainty principle does tell us about the behaviour of the electrons in an atom, consider as the simplest example a hydrogen atom. Here the electron is trapped in the Coulomb potential well due to the positive nucleus. We can then argue that if the electron cannot have a precisely defined position, then we can at least suppose that it is confined to a spherical (by symmetry) shell of radius a. Thus, the uncertainty Δx in x will be a, and similarly for the y and z positions. But, with the electron moving within this region, the x component of momentum, p_x, will, also by symmetry, swing between two equal and opposite values, p and $-p$ say, and hence p_x will have an uncertainty of $\Delta p_x \approx p$. By appealing to symmetry once again, the y and z components of momentum can be seen to have the same uncertainty.

By the uncertainty principle $\Delta p_x \Delta x \approx \hbar$, (and similarly for the other two components), the uncertainty in the x component of momentum will then be $\Delta p_x \approx \hbar/a$, and hence $p \approx \hbar/a$. The kinetic energy of the particle will then be

$$T = \frac{p^2}{2m} \approx \frac{\hbar^2}{2ma^2} \tag{3.19}$$

so including the Coulomb potential energy, the total energy of the particle will be

$$E \approx \frac{\hbar^2}{2ma^2} - \frac{e^2}{4\pi\epsilon_0 a}. \tag{3.20}$$

The lowest possible energy of the atom is then obtained by simple differential calculus. Thus, taking the derivative of E with respect to a and equating this to zero and solving for a gives

$$a \approx \frac{4\pi\epsilon_0\hbar^2}{me^2} \approx 0.5 \text{ nm} \tag{3.21}$$

and the minimum energy

$$E_{\min} \approx -\tfrac{1}{2}\frac{me^4}{(4\pi\epsilon_0)^2\hbar^2} \tag{3.22}$$

$$\approx -13.6 \text{ eV}. \tag{3.23}$$

The above values for atomic size and atomic energies are what are observed in practice. The uncertainty relation has yielded considerable information on atomic structure without knowing all that much about what a wave function is supposed to represent! The exactness of the above result is somewhat fortuitous, but the principle is nevertheless correct: the uncertainty principle demands that there be a minimum size to an atom. If a hydrogen atom has an energy above this minimum, it is free to radiate away energy by emission of electromagnetic energy (light) until it reaches this minimum. Beyond that, it cannot radiate any more energy. Classical EM theory says that it should, but it does not. The conclusion is that there must also be something amiss with classical EM theory, which in fact turns out to be the case: the EM field too must treated quantum mechanically. When this is done, there is consistency between the demands of quantum EM theory and the quantum structure of atoms – an atom in its lowest energy level (the ground state) cannot, in fact, radiate – the ground state of an atom is stable.

Another important situation for which the uncertainty principle gives a surprising amount of information is that of the harmonic oscillator.

3.3.3 The Minimum Energy of a Simple Harmonic Oscillator

By using Heisenberg's uncertainty principle in the form $\Delta x\Delta p \approx \hbar$, it is also possible to estimate the lowest possible energy level (ground state) of a simple harmonic oscillator. The simple harmonic oscillator potential is given by

$$U = \tfrac{1}{2}m\omega^2x^2 \tag{3.24}$$

where m is the mass of the oscillator and ω is its natural frequency of oscillation. This is a particularly important example as the simple harmonic oscillator potential is found to arise in a wide variety of circumstances such as an electron trapped in a well between two nuclei, or the oscillations of a linear molecule, or indeed in a manner far removed from the image of an oscillator as a mechanical object, the lowest energy of a single mode quantum mechanical electromagnetic field.

We start by assuming that in the lowest energy level, the oscillations of the particle have an amplitude of a, so that the oscillations swing between $-a$ and a. We further assume that the momentum of the particle can vary between p and $-p$. Consequently, we can assign an uncertainty $\Delta x = a$ in the position of the particle, and an uncertainty $\Delta p = p$ in the momentum of the particle. These two uncertainties will be related by the uncertainty relation

$$\Delta x\Delta p \approx \hbar \tag{3.25}$$

from which we conclude that

$$p \approx \hbar/a. \tag{3.26}$$

The total energy of the oscillator is

$$E = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2 \tag{3.27}$$

so that roughly, if a is the amplitude of the oscillation, and $p \approx \hbar/a$ is the maximum momentum of the particle then

$$E \approx \tfrac{1}{2}\left(\frac{1}{2m}\frac{\hbar^2}{a^2} + \tfrac{1}{2}m\omega^2 a^2\right) \tag{3.28}$$

where the extra factor of $\frac{1}{2}$ is included to take account of the fact that the kinetic and potential energy terms are each their maximum possible values.

The minimum value of E can be found using differential calculus i.e.

$$\frac{dE}{da} = \tfrac{1}{2}\left(-\frac{1}{m}\frac{\hbar^2}{a^3} + m\omega^2 a\right) = 0. \tag{3.29}$$

Solving for a gives

$$a^2 = \frac{\hbar}{m\omega}. \tag{3.30}$$

Substituting this into the expression for E then gives for the minimum energy

$$E_{\min} \approx \tfrac{1}{2}\hbar\omega. \tag{3.31}$$

A more precise quantum mechanical calculation shows that this result is (fortuitously) exactly correct, i.e. the ground state of the harmonic oscillator has a non-zero energy of $\frac{1}{2}\hbar\omega$.

It was Heisenberg's discovery of the uncertainty relation, and various other real and imagined experiments that ultimately lead to a fundamental proposal (by Max Born) concerning the physical meaning of the wave function. We shall arrive at this interpretation by way of the famous two slit interference experiment.

Chapter 4

The Two Slit Experiment

THIS experiment is said to illustrate the essential mystery of quantum mechanics[1]. 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, in which the same physical system can exhibit under different circumstances, either particle or wave-like properties, otherwise known as wave-particle duality. This property of physical systems, otherwise known as 'the superposition of states', 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.

The experiment 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 which may or may not be covered. 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.

Suppose, in the first instance, that this experiment is carried out with only one slit opened, slit 1 say. A first point to note is that the bullets arrive in 'lumps', (assuming indestructible bullets), i.e. every bullet that leaves the gun (and makes it through the slits) arrives as a whole somewhere on the detection screen. Not surprisingly, what would be observed is the tendency for the bullets to strike the screen in a region somewhere immediately opposite the position of the open slit, but because the machine gun is firing erratically, we would expect that not all the bullets would strike the screen in exactly the same spot, but to strike the screen at random, though always in a region roughly opposite the opened slit. We can represent this experimental outcome by a curve $P_1(x)$ which is simply such that

$$P_1(x)\,\delta x = \text{ probability of a bullet landing in the range } (x, x+\delta x). \tag{4.1}$$

[1] Another property of quantum systems, known as 'entanglement', sometimes vies for this honour, but entanglement relies on this 'essential mystery' we are considering here

If we were to cover this slit and open the other, then what we would observe is the tendency for the bullets to strike the screen opposite this opened slit, producing a curve $P_2(x)$ similar to $P_1(x)$. These results are indicated in Fig. (4.1).

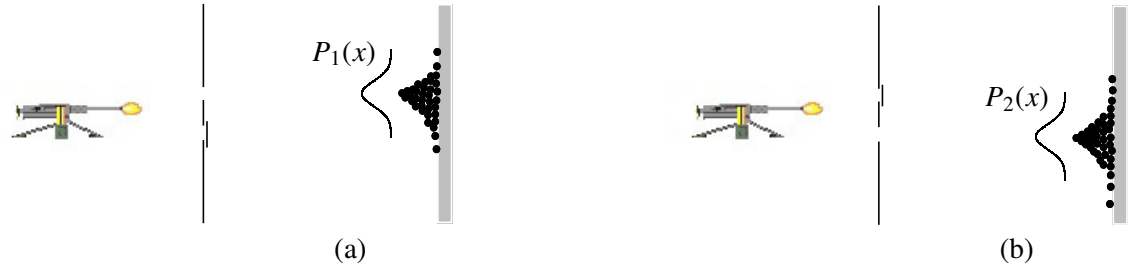


Figure 4.1: The result of firing bullets at the screen when only one slit is open. The curves $P_1(x)$ and $P_2(x)$ give the probability densities of a bullet passing through slit 1 or 2 respectively and striking the screen at x.

Finally, suppose that both slits are opened. We would then observe the bullets would sometimes come through slit 1 and sometimes through slit 2 – varying between the two possibilities in a random way – producing two piles behind each slit in a way that is simply the sum of the results that would be observed with one or the other slit opened, i.e.

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

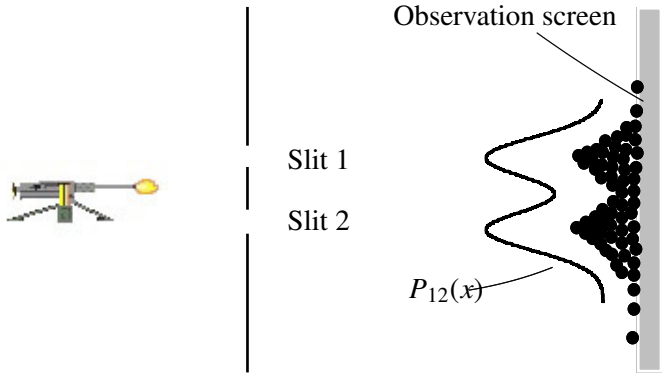


Figure 4.2: The result of firing bullets at the screen when both slits are open. The bullets accumulate on an observation screen, forming two small piles opposite each slit. The curve $P_{12}(x)$ represents the probability density of bullets landing at point x on the observation screen.

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 fired from the gun 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.

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

$$P(x)\,\delta x = \frac{\delta N}{N} \tag{4.3}$$

which is the fraction of all the bullets fired from the gun that end up landing on the screen 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 $x + \delta x$. An illustrative example is

given in Fig. (4.3) of the histogram obtained when 133 bullets strike the observation screen when both slits are open. In this figure the approximate curve for $P(x)$ also plotted, which in this case will be $P_{12}(x)$.

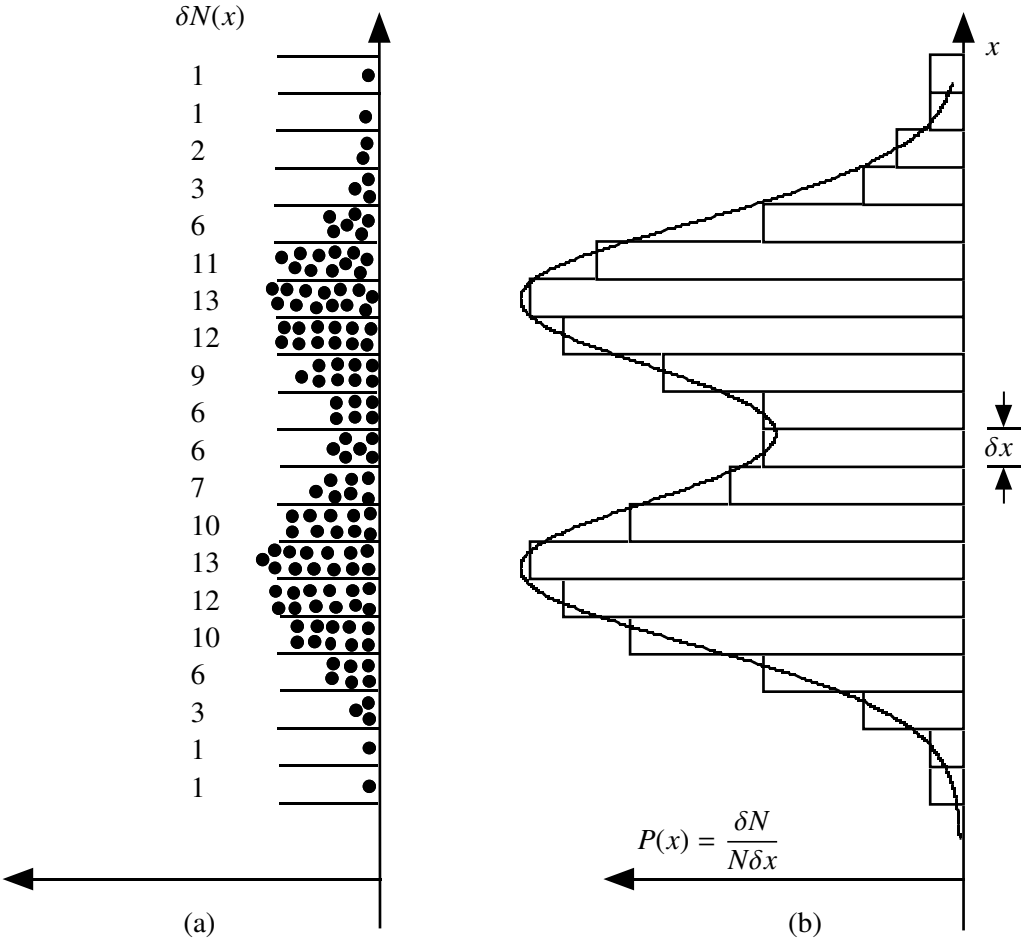


Figure 4.3: Bullets that have passed through the first screen collected in boxes all of the same size δx. (a) The number of bullets that land in each box is presented. There are $\delta N(x)$ bullets in the box between x and $x + \delta x$. (b) A histogram is formed from the ratio $P(x) \approx \delta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

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. (4.2(a)), while if only slit 2 is open, we get $P_2(x)$ such as that in Fig. (4.2(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.

$$P_{12}(x) = P_1(x) + P_2(x). \tag{4.4}$$

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.

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[2] 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

$$I_1(x) = |E_1(x,t)|^2 = E_1(x)^2. \tag{4.5}$$

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

$$\begin{aligned} I_{12}(x) =& |E_1(x,t) + E_2(x,t)|^2 \\ =& I_1(x) + I_2(x) + 2E_1E_2\cos\left(2\pi d\sin\theta/\lambda\right) \\ =& I_1(x) + I_2(x) + 2\sqrt{I_1(x)I_2(x)}\cos\delta \end{aligned} \tag{4.6}$$

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 waves arrive everywhere at the same time, and the intensity can have any value at all.

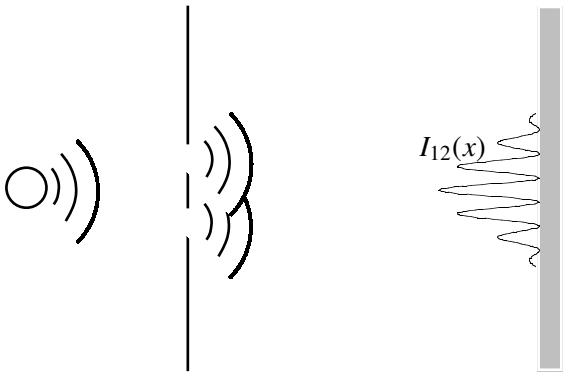


Figure 4.5: The usual two slit interference pattern.

[2]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.

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 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

$$P_{12}(x) = P_1(x) + P_2(x) + 2\sqrt{P_1(x)P_2(x)}\cos\delta \tag{4.7}$$

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 it seems clear that the electrons go *either* through slit 1 *or* through slit 2, because that is what particles would do. The behaviour of the electrons going through slit 1 should then not be affected by whether slit 2 is opened or closed as those electrons would go nowhere near slit 2. In other words, we have to expect that $P_{12}(x) = P_1(x) + P_2(x)$, but this not what is observed. In fact, what is observed is impossible to understand on the basis of this argument: if only one slit is open, say slit 1, then we find electrons landing on the screen at points which, if we open slit 2, receive no electrons at all!

In other words, opening both slits and thereby providing an extra pathway by which the electrons reach the screen results in the number that arrive at some points actually decreasing.

It appears that we must abandon the idea that the particles go through one slit *or* the other. So if we want to retain the mental picture of electrons as particles, we must 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. After all, the interference term depends on d, the separation between the slits, so we must expect that the particles must 'know' how far apart the slits are in order for the positions that they strike the screen to depend on d, and they cannot 'know' this if each electron goes through only one slit. We could imagine that the electrons determine the separation between slits by supposing that they 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.

4.3.1 Monitoring the slits: the Feynman microscope

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. There are many many ways of doing this. One possibility, a variation on the Heisenberg microscope discussed in Section 3.3.1, was proposed by Feynman, and is known as the Feynman light microscope. The experiment consists of shining a light on the slits so that, if an electron goes through a slit, then it scatters some of this light, which can then be observed with a microscope. We immediately know what slit the electron passed through. Remarkably, 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.

The essence of the argument is exactly that presented in Section 3.3.1 in the discussion of the Heisenberg microscope experiment. The aim is to resolve the position of the electron to an accuracy of at least $\delta x \sim d/2$, but as a consequence, we find that the electron is given a random kick in momentum of magnitude $\delta p \sim 2h/d$, (see Eq. (3.17)).

Thus, an electron passing through, say, the upper slit, could be deflected by an amount up to an angle of $\delta\theta$ where

$$\delta\theta \sim \delta p/p \tag{4.8}$$

where $p = h/\lambda$ is the momentum of the electron and λ its de Broglie wavelength. Thus we find

$$\delta\theta \sim \frac{2h/d}{h/\lambda} = 2\lambda/d \tag{4.9}$$

i.e. the electron will be deflected through this angle, while at the same time, the photon will be seen as coming from the position of the upper slit.

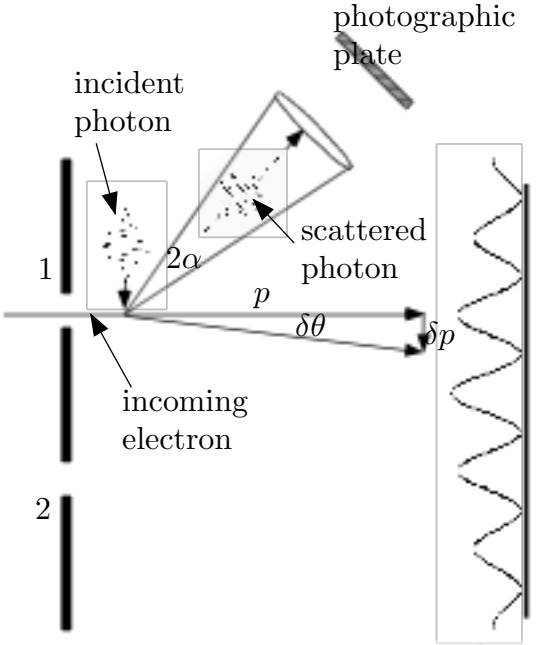


Figure 4.6: An electron with momentum p passing through slit 1 scatters a photon which is observed through a microscope. The electron gains momentum δp and is deflected from its original course.

Since the angular separation between a maximum of the interference pattern and a neighbouring minimum is $\lambda/2d$, the uncertainty in the deflection of an observed electron is at least enough to displace an electron from its path heading for a maximum of the interference pattern, to one where it will strike the screen at the position of a minimum. The overall result is to wipe out the interference pattern.

But we could ask why it is that the photons could be scattered through a range of angles, i.e. why do we say 'up to an angle of $\delta\theta$'? After all, the experiment could be set up so that every photon sent towards the slits has the same momentum every time, and similarly for every electron, so every scattering event should produce the same result, and the outcome would merely be the interference pattern displaced sideways. It is here where quantum randomness comes in. The amount by which the electron is deflected depends on what angle the photon has been scattered through, which amounts to knowing what part of the observation lens the photon goes through on its way to the photographic plate. But we cannot know this unless we 'look' to see where in the lens the photon passes. So there is an uncertainty in the position of the photon as it passes through the lens which, by the uncertainty relation, implies an uncertainty in the momentum of the photon in a direction parallel to the lens, so we cannot know with precision where the photon will land on the photographic plate. As a consequence, the position that each scattered photon will strike the photographic plate will vary randomly from one scattered photon to the next in much the same way as the position that the electron strikes the observation screen varies in a random fashion in the original (unmonitored) two slit experiment. The uncertainty as to which slit the electron has passed through has been transferred to the uncertainty about which part of the lens the scattered photon passes through.

So it would seem that by trying to determine through which slit the electron passes, the motion of the electron is always sufficiently disturbed so that no interference pattern is observed. As far as it is known, *any* experiment that can be devised – either a real experiment or a gedanken (i.e. a thought) experiment – that involves directly observing through which slit the electrons pass always results in a disturbance in the motion of the electron sufficient for the interference pattern to be wiped out. An 'explanation' of why this occurs, of which the above discussion is an example, could perhaps be constructed in each case, but in any such explanation, the details of the way the physics conspires to produce this result will differ from one experiment to the other [3]. These explanations often require a mixture of classical and quantum concepts, i.e. they have a foot in both camps and as such are not entirely satisfactory, but more thorough studies of, for instance, the Feynman microscope experiment using a fully quantum mechanical approach, confirm the results discussed here. So it may come as a surprise to find that there need not be any direct physical interaction between the electrons and the apparatus used to monitor the slits, so there is no place for the kind of quantum/classical explanation given above. It turns out that as long as there is *information* available regarding which slit the electron passes through, there is no interference, as the following example illustrates.

4.3.2 The Role of Information: The Quantum Eraser

Suppose we construct the experimental set-up illustrated in Fig. (4.7). The actual experiment is performed with photons, but we will stick with an 'electron version' here. In this experiment, the electron source sends out *pairs* of particles. One of the pair (sometimes called the 'signal' particle in the photon version of the experiment) heads towards the slits, while the other (the 'idler') heads

[3] Einstein, who did not believe quantum mechanics was a complete theory, played the game of trying to find an experimental setup that would bypass the uncertainty relation i.e. to know which slit an electron passes through AND to observe an interference pattern. Bohr answered all of Einstein's challenges, including, in one instance, using Einstein's own theory of general relativity to defeat a proposal of Einstein's concerning another form of the uncertainty relation, the time-energy uncertainty relation. It was at this point that Einstein abandoned the game, but not his attitude to quantum mechanics.

off at right angles. What we find here is that, if we send out many electrons in this fashion, then *no interference pattern* is built up on the observation screen, i.e. we get the result we expect if we are able to determine through which slit each electron passes. But this outcome is not as puzzling as it first might seem since, in fact, we actually do have this information: for each emitted electron there is a correlated idler particle whose direction of motion will tell us in what direction the electron is headed. Thus, in principle, we can determine which slit an electron will pass through by choosing to *detect* the correlated idler particle by turning on detectors a and b. Thus, if we detect a particle at a, then electron is heading for slit 1. However, we do not actually have to carry out this detection. The sheer fact that this information is available for us to access if we choose to do so is enough for the interference pattern not to appear.

But suppose we *wipe out* this information somehow. If we make use of a single detector that is capable of detecting either of the idler particles, then we cannot tell, if this detector registers a count, which idler particle it detected. In this case, the 'which path' information has been *erased*. If we erase this information for every electron emitted, there will appear an interference pattern even though *there has been no physical interaction with the electrons passing through the slits!!* This example suggests that provided there is information present as to which slit an electron passes through, even if we do not access this information, then there is no interference pattern. This is an example of quantum entanglement: the direction of the electron heading towards the slits is *entangled* with the direction of the 'information carrier' idler particle.

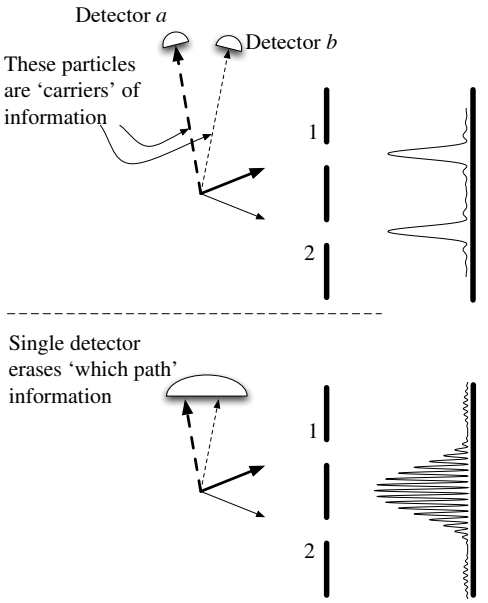


Figure 4.7: Illustrating the 'quantum eraser' experiment. In the first set-up, the information on 'which slit' is encoded in the auxiliary particle, and there is no interference. In the second, this information is erased and interference is observed.

This more general analysis shows that the common feature in all these cases is that the electron becomes 'entangled' with some other system, in such a manner that this other system carries with it information about the electron. In this example, entanglement is present right from the start of the experiment, while in the monitoring-by-microscope example, entanglement with the scattered photon has been actively established, with information on the position of the electron encoded in the photon. But that is all that we require, i.e. that the information be so encoded – we do not need to have some one consciously observe these photons for the interference pattern to disappear, it is merely sufficient that this information be available. It is up to use to choose to access it or not.

What is learned from this is that if information is in principle available on the position of the electron or, in other words, if we want to be anthropomorphic about all this, if we can in principle *know* which slit the electron goes through by availing ourselves of this information, even without interfering with the electron, then the observed pattern of the screen is the same as found with bullets: no interference, with $P_{12}(x) = P_1(x) + P_2(x)$. If this information is not available at all, then we are left with the uncomfortable thought that 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 information is extant, a dichotomy that is known as wave-particle duality.

4.3.3 Wave-particle duality

So it appears that if we gain information on which slit the electron passes through then we lose the interference pattern. Gaining information on which slit means we are identifying the electron as a particle — only particles go through one slit or the other. But once this identification is made, say an electron has been identified as passing through slit 1, then that electron behaves thereafter like a particle that has passed through one slit only, and so will contribute to the pattern normally produced when only slit 1 is open. And similarly for an electron identified as passing through slit 2. Thus, overall, no interference pattern is produced. If we do not determined which slit, the electrons behave as waves which are free to probe the presence of each slit, and so will give rise to an interference pattern.

This dichotomy, in which the same physical entity, an electron (or indeed, any particle), can be determined to be *either* a particle *or* a wave depending on what experiment is done to observe its properties, is known as wave-particle duality.

But in all instances, the inability to observe both the wave and particle properties – i.e. to know which slit a particle passes through AND for there to be an interference pattern – can be traced back to the essential role played by the uncertainty relation $\Delta x \Delta p \geq \frac{1}{2}\hbar$. As Feynman puts it, this relation 'protects' quantum mechanics. But note that this is NOT the measurement form of the uncertainty relation: there is no need for a disturbance to act on the particle as a consequence of a measurement, for quantum mechanics to be protected, as we have just seen.

The fact that *any* situation for which there is information on which slit the electron passes through always results in the non-appearance of the interference pattern suggests that there is a deep physical principle at play here – a law or laws of nature – that overrides, irrespective of the physical set-up, any attempt to have information on where a particle is, and to observe interference effects. The laws are the laws of quantum mechanics, and one of their consequences is the uncertainty principle, that is $\Delta x \Delta p \geq \frac{1}{2}\hbar$. If we want to pin down the slit through which the particle has past with, say, 97% certainty, then the consequent uncertainty Δx in the position of the particle can be estimated to be $\Delta x \approx 0.17d$, where d is the separation of the slits[4]. But doing so implies that there is an uncertainty in the sideways momentum of the electron given by $\Delta p \gtrsim \hbar/0.17d$. This amounts to a change in direction through an angle

$$\Delta\theta = \frac{\Delta p}{p} \gtrsim \frac{\lambda}{2d} \tag{4.10}$$

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 require us to invoke any 'physical mechanism' to explain why the interference pattern washes out; only the abstract requirements of the uncertainty principle are used. Nothing is said about how the position of the particle is pinned down to

[4]This estimate is arrived at by recognizing that the measurement is not perfect, i.e. we cannot be totally sure, when a photon arrives on the photographic plate, which slit it came from. So, if we assign a probability P that the particle is at the slit with position $d/2$ and a probability $1 - P$ that it is at the position of the slit at $-d/2$ based on the observed outcome of the measurement, then the mean position of the electron is now

$$\langle x \rangle = Pd/2 - (1 - P)d/2 = (P - \tfrac{1}{2})d$$

and the standard deviation of this outcome is

$$(\Delta x)^2 = P(d/2 - \langle x \rangle)^2 + (1 - P)(-d/2 - \langle x \rangle)^2 = P(1 - P)d$$

so $\Delta x = \sqrt{P(1 - P)}d = 0.17d$ for $P = 0.97$.

within an uncertainty Δx. If this information is provided, then a physical argument, of sorts, as in the case of Heisenberg's microscope, that mixes classical and quantum mechanical ideas might be possible that explains why the interference pattern disappears. However, the details of the way the physics conspires to produce this result will differ from one experiment to the other. And in any case, there might not be any physical interaction at all, as we have just seen. 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

$$E(x,t) = E_1(x,t) + E_2(x,t) \tag{4.11}$$

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

$$\begin{aligned} I_{12}(x) =& |E(x,t)|^2 \\ =& |E_1(x,t) + E_2(x,t)|^2 \\ =& I_1(x) + I_2(x) + 2E_1E_2\cos\left(\frac{2\pi d\sin\theta}{\lambda}\right) \\ =& I_1(x) + I_2(x) + 2\sqrt{I_1(x)I_2(x)}\cos\delta \end{aligned} \tag{4.12}$$

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

$$P_{12}(x) = P_1(x) + P_2(x) + 2\sqrt{P_1(x)P_2(x)}\cos\delta. \tag{4.13}$$

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

$$\Psi(x,t) = \Psi_1(x,t) + \Psi_2(x,t) \tag{4.14}$$

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

$$|\Psi(x,t)|^2\delta x \propto \text{probability of observing an electron in } x, x+\delta x \tag{4.15}$$

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

$$P = |\Psi_1 + \Psi_2|^2 \tag{4.16}$$

  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

$$P = P_1 + P_2 \tag{4.17}$$

  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.

Chapter 5

Wave Mechanics

The version of quantum mechanics based on studying the properties of the wave function is known as wave mechanics, and is the version that first found favour amongst early researchers in the quantum theory, in part because it involved setting up and solving a partial differential equation for the wave function, the famous Schrödinger equation, which was of a well-known and much studied form. As well as working in familiar mathematical territory, physicists and chemists were also working with something concrete in what was in other ways a very abstract theory – the wave function was apparently a wave in space which could be visualized, at least to a certain extent. But it must be borne in mind that the wave function is not an objectively real entity, the wave function does not represent waves occurring in some material substance. Furthermore, it turns out that the wave function is no more than one aspect of a more general theory, quantum theory, that is convenient for certain classes of problems, and entirely inappropriate (or indeed inapplicable) to others[1]. Nevertheless, wave mechanics offers a fairly direct route to some of the more important features of quantum mechanics, and for that reason, some attention is given here to some aspects of wave mechanics prior to moving on, in later chapters, to considering the more general theory.

5.1 The Probability Interpretation of the Wave Function

The probability interpretation of the wave function, due to Max Born, was introduced in the preceding Chapter. It is restated here for convenience:

If a particle is described by a wave function $\Psi(x, t)$, then

$$|\Psi(x,t)|^2 \delta x = \text{Probability of observing the particle in the small region } (x, x+\delta x) \text{ at time } t$$

with the consequent terminology that the wave function is usually referred to as a 'probability amplitude'.

What this interpretation means in a practical sense can be arrived at in a number of ways, the conventional manner making use of the notion of an 'ensemble of identically prepared systems'. By this we mean that take a vast number of identical copies of the same system, in our case a very simple system, consisting on just one particle, and put them all through exactly the same experimental procedures, so that presumably they all end up in exactly the same physical state. An example would be the two slit experiment in which every electron is prepared with the same momentum and energy. The collection of identical copies of the same system all prepared in the

[1] The wave function appears to be a wave in real space for a single particle, but that is only because it depends on x and t, in the same way as, say, the amplitude of a wave on a string can be written as a function of x and t. But, for a system of more than one particle, the wave function becomes a function of two space variables x_1 and x_2 say, as well as t: $\Psi(x_1, x_2, t)$. It then makes no sense to talk about the value of the wave function at some position in space. In fact, the wave function is a wave that 'exists' in an abstract space known as phase space.

same fashion is known as an ensemble. We then in effect assume that, at time t after the start of the preparation procedure, the state of each particle will be given by the same wave function $\Psi(x,t)$, though it is common practice (and a point of contention) in quantum mechanics to say that the wave function describes the whole ensemble, not each one of its members. We will however usually refer to the wave function as if it is associated with a single system, in part because this reflects the development of the point-of-view that the wave function represents the information that we have about a given particle (or system, in general).

So suppose we have this ensemble of particles all prepared in exactly the same fashion, and we measure the position of the particle. If the particle were described by classical physics, we would have to say that for each copy of the experiment we would get exactly the same result, and if we did not get the same answer every time, then we would presume that this is because we did not carry out our experiment as well as we might have: some error has crept in because of some flaw in the experimental apparatus, or in the measurement process. For instance, in the two slit experiment, the bullets struck the observation screen at random because the machine gun fired off the bullets in an erratic fashion. But the situation is not the same quantum mechanically – invariably we will find that we get different results for the measurement of position in spite of the fact that each run of the experiment is supposedly identical to every other run. Moreover the results vary in an entirely random fashion from one measurement to the next, as evidenced in the two slit experiment with electrons. This randomness cannot be removed by refining the experiment — it is built into the very nature of things. But the randomness is not without some kind of order — the manner in which the measured values of position are scattered is determined by the wave function $\Psi(x,t)$, i.e. the scatter of values is quantified by the probability distribution given by $P(x,t) = |\Psi(x,t)|^2$.

It is important to be clear about what this probability function gives us. It does not give the chances of a particle being observed at a precise position x. Recall, from the Born interpretation, that it gives the probability of a particle being observed to be positioned in a range x to $x + \delta x$, this probability being given by the product $P(x,t)\delta x$. So to properly understand what $P(x,t)$ is telling us, we have to first suppose that the range of x values are divided into regions of width δx. Now when we measure the position of the particle in each run of the experiment, what we take note of is the interval δx in which each particle is observed. If we do the experiment N times, we can count up the number of particles for which the value of x lies in the range $(x, x + \delta x)$. Call this number $\delta N(x)$. The fraction of particles that are observed to lie in this range will then be

$$\frac{\delta N(x)}{N}. \tag{5.1}$$

Our intuitive understanding of probability then tells us that this ratio gives us, approximately, the probability of observing the particle in this region $(x, x+\delta x)$. So if N is made a very large number, we would then expect that

$$\frac{\delta N(x)}{N} \approx P(x,t)\delta x \qquad \text{or} \qquad \frac{\delta N(x)}{N\delta x} \approx P(x,t) = |\Psi(x,t)|^2 \tag{5.2}$$

where the approximate equality will become more exact as the number of particles becomes larger. So what we would expect to find, if quantum mechanics is at all true, that if we were to do such a series of experiments for some single particle physical system, and if we knew by some means or other what the wave function is of the particle used in the experiment, then we ought to find that the distribution of results for the measurements of the position of the particle can be determined by a simple calculation based on the wave function.

So far, we have not developed any means by which we can actually calculate a wave function, so in the example below, we will make use of a wave function that can be found by solving the famous Schrödinger equation, i.e. we will just take the wave function as given.

Ex 5.1 An electron placed in the region $x > 0$ adjacent to the surface of liquid helium is attracted to the surface by its oppositely charged 'image' inside the surface. However, the electron cannot penetrate the surface (it is an infinitely high potential barrier). The wave function for the electron, in its lowest energy state, can be shown to be given by

$$\begin{aligned} \Psi(x,t) &= 2a_0^{-1/2}(x/a_0)\, e^{-x/a_0} e^{-i\omega t} && x > 0 \\ &= 0 && x < 0. \end{aligned}$$

where a_0 is a constant determined by the mass and charge of the particle and by the dielectric properties of the helium, and is ≈ 76 pm. The energy of the electron is given by $E = \hbar\omega$.

An experiment is conducted with the aim of measuring the distance of the electron from the surface. Suppose that the position x can be measured to an accuracy of $\pm a_0/4$. We can then divide the x axis into intervals of length $\delta x = a_0/2$, and record the number of times that the electron is found in the ranges $(0, a_0/2)$, $(a_0/2, a_0)$, $(a_0, 3a_0/2), \ldots, (4.5a_0, 4.74a_0)$. The experiment is repeated 300 times, yielding the results in the adjacent table. This raw data can be plotted as a histogram:

$\frac{x}{a_0}$	no. of detections $\delta N(x)$ in interval $(x, x+\delta x)$
0	28
0.5	69
1.0	76
1.5	58
2.0	32
2.5	17
3.0	11
3.5	6
4.0	2
4.5	1

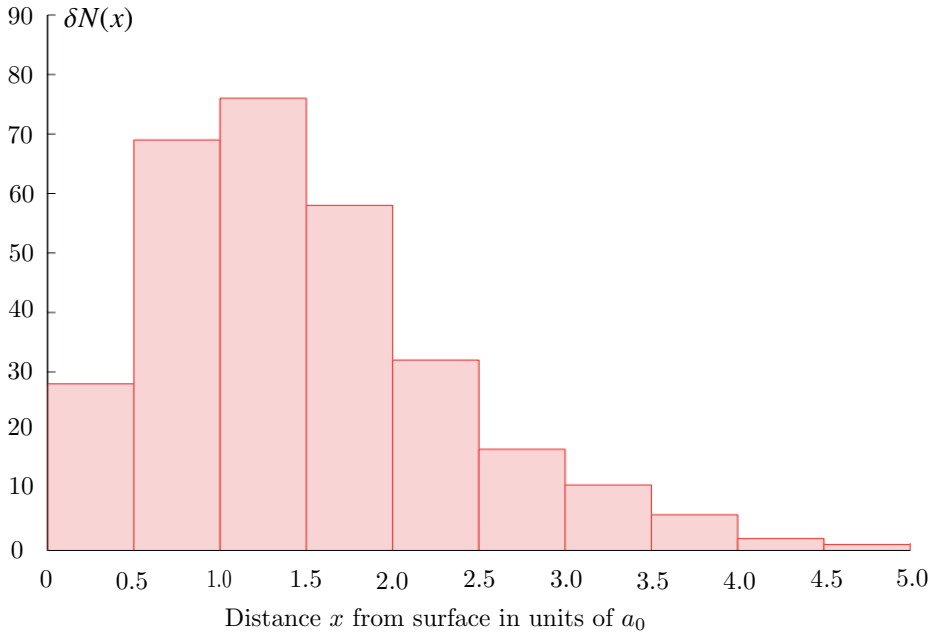


Figure 5.1: Histogram of raw data plotted vertically is the number $\delta N(x)$ of electrons measured to be in the range $(x, x+\delta x)$.

which tells us that there is a preference for the electron to be found at a distance of between a and $1.5a$ from the surface.

To relate this to the probability distribution provided by the quantum theory, we can set $\delta x = 0.5a$, and construct a histogram of the values of $\delta N(x)/N\delta x$ which, as we can see

from Eq. (5.2), should approximate to $P(x,t) = |\Psi(x,t)|^2$ where

$$P(x,t) = \left|2a_0^{-1/2}(x/a_0)\, e^{-x/a_0} e^{-i\omega t}\right|^2 = 4(x^2/a_0^3)\, e^{-2x/a_0} \tag{5.3}$$

This is illustrated in the following figure

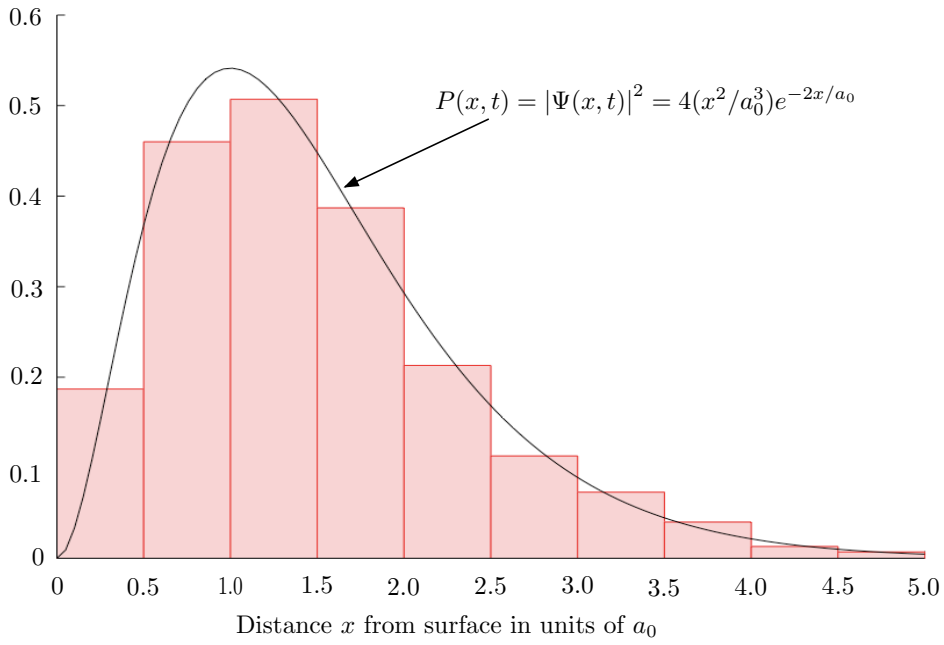


Figure 5.2: Histogram of $\delta N(x)/(N\delta x)$ and of the probability distribution $P(x,t)$ to which it approximates.

If many more measurements were made, and the interval δx was made smaller and smaller, the expectation is that the tops of the histogram would form a smooth curve that would better approximate to the form of $P(x,t)$ predicted by quantum mechanics.

5.1.1 Normalization

The probability interpretation of the wave function has an immediate consequence of profound significance. The probability interpretation given above tells us the probability of finding the particle in a small interval δx. We can calculate the probability of finding the particle in a finite range by dividing this range into segments of size δx and simply adding together the contributions from each such segment. In the limit that δx is made infinitesimally small, this amounts to evaluating an integral. To out it more precisely, the probability of finding the particle in a finite range $a < x < b$, will be given by the integral

$$\int_a^b |\Psi(x,t)|^2 \, dx \tag{5.4}$$

From this it immediately follows that the probability of finding the particle somewhere in the range $-\infty < x < \infty$ must be unity. After all, the particle is guaranteed to be found somewhere. Mathematically, this can be stated as

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2 \, dx = 1. \tag{5.5}$$

A wave function that satisfies this condition is said to be 'normalized to unity'.

When deriving a wave function, e.g. by solving the Schrödinger equation, usually the result obtained will be obtained up to an unknown, constant factor. This factor is then usually determined by imposing the normalization condition. We can illustrate this in the following example.

Ex 5.2 The wave function for the electron floating above the liquid helium surface, freshly calculated by solving the Schrödinger equation, would be given by

$$\begin{aligned}\Psi(x,t) &= C(x/a_0)\, e^{-x/a_0} e^{-i\omega t} && x > 0\\ &= 0 && x < 0.\end{aligned}$$

where C is a constant whose value is not known. We can determine its value by requiring that this wave function be normalized to unity, i.e.

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\, dx = 1.$$

We first note that the wave function vanishes for $x < 0$ so the normalization integral extends only over $0 < x < \infty$, and hence we have

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\, dx = |C|^2 \int_0^{+\infty} (x/a_0)^2 e^{-2x/a_0}\, dx = 1$$

The integral is straightforward, and we get

$$|C|^2 a_0/4 = 1$$

from which follows

$$C = 2a_0^{-1/2} e^{i\phi}$$

where $\exp(i\phi)$ is an unknown 'phase factor' which we can set equal to unity (see below) so that

$$C = 2a_0^{-1/2}$$

and the normalized wave function becomes

$$\begin{aligned}\Psi(x,t) &= 2a_0^{-1/2}(x/a_0)\, e^{-x/a_0} e^{-i\omega t} && x > 0\\ &= 0 && x < 0.\end{aligned}$$

To be noted in this above calculation is the appearance of a phase factor $\exp(i\phi)$. This comes about because the normalization condition only took us as far as giving us the value of $|C|^2$, so *any* choice of ϕ can be made without affecting the value of $|C|^2$ since $|\exp(i\phi)|^2 = 1$ for any (real) value of ϕ. Furthermore, there does not appear any way, from the calculation above, what value to give to ϕ. But it turns out that we can choose any value we like for ϕ without having any effect on the results of any calculation since, at some stage before arriving at a physical prediction based on a wave function, it is necessary to calculate a probability. At this point, we find we will have to evaluate $|\exp(i\phi)|^2 = 1$, and so ϕ drops out of the picture. For this reason, we might as well choose ϕ to have any convenient value right from the beginning. Here $\phi = 0$ is a good choice, and that is what is chosen above.

This is not to say that phase factors are unimportant in quantum mechanics. There is an intimate connection between phase factors and, believe it or not, the forces of nature. Simply allowing ϕ

to be a *function of x and t*, but at the same time demanding that such a choice have no effect on the results calculated from the wave function, leads to the requirement that the particle be acted on by forces which can be identified as electromagnetic forces. The same idea can be pushed even further, so that it is possible to 'extract' from quantum mechanics the strong and weak nuclear forces, and maybe even gravity. In a sense, these forces are 'built into' quantum mechanics, and the theory that describes these forces is known as 'gauge theory'.

An immediate consequence of the normalization condition is that the wave function must vanish as $x \to \pm\infty$ otherwise the integral will have no hope of being finite. This condition on the wave function is found to lead to one of the most important results of quantum mechanics, namely that the energy of the particle (and other observable quantities as well) is *quantized*, that is to say, it can only have certain discrete values in circumstances in which, classically, the energy can have any value.

We can note at this stage that the wave function that we have been mostly dealing with, the wave function of a free particle of given energy and momentum

$$\Psi(x,t) = A\sin(kx-\omega t), \quad A\cos(kx-\omega t), \quad Ae^{i(kx-\omega t)}, \quad \ldots, \tag{5.6}$$

does not satisfy the normalization condition Eq. (5.5) – the integral of $|\Psi(x,t)|^2$ is infinite. Thus it already appears that there is an inconsistency in what we have been doing. However, there is a place for such wave functions in the greater scheme of things, though this is an issue that cannot be considered here. It is sufficient to interpret this wave function as saying that because it has the same amplitude everywhere in space, the particle is equally likely to be found anywhere.

5.2 Expectation Values and Uncertainties

Since $|\Psi(x,t)|^2$ is a normalized probability density for the particle to be found in some region in space, it can be used to calculate various statistical properties of the position of the particle. In defining these quantities, we make use again of the notion of an 'ensemble of identically prepared systems' to build up a record of the number of particles $\delta N(x)$ for which the value of x lies in the range $(x, x+\delta x)$.

The fraction of particles that are observed to lie in this range will then be

$$\frac{\delta N(x)}{N} \tag{5.7}$$

We can then calculate the mean or average value of all these results, call it $\overline{x}(t)$, in the usual way:

$$\overline{x}(t) = \sum_{\text{All } \delta x} x\,\frac{\delta N(x)}{N}. \tag{5.8}$$

This mean will be an approximation to the mean value that would be found if the experiment were repeated an infinite number of times and in the limit in which $\delta x \to 0$. This latter mean value will be written as $\langle x\rangle$, and is given by the integral:

$$\langle x(t)\rangle = \int_{-\infty}^{+\infty} x\,P(x,t)\,dx = \int_{-\infty}^{+\infty} x\,|\Psi(x,t)|^2 dx \tag{5.9}$$

This average value is usually referred to as the expectation value of x.

Ex 5.3 Using the data from the previous example, we can calculate the average value of the distance of the electron from the surface of the liquid helium using the formula of Eq. (5.8). Thus we have

$$\begin{aligned}\langle x\rangle \approx \overline{x} =& 0\times\frac{28}{300} + 0.5a_0\times\frac{69}{300} + a_0\times\frac{76}{300} + 1.5a_0\times\frac{58}{300} + 2a_0\times\frac{32}{300}\\ &+ 2.5a_0\times\frac{17}{300} + 3a_0\times\frac{11}{300} + 3.5a_0\times\frac{6}{300} + 4a_0\times\frac{2}{300} + 4.5a_0\times\frac{1}{300}\\ &= 1.235a_0.\end{aligned}$$

This can be compared with the result that follows for the expectation value calculated from the wave function for the particle:

$$\langle x\rangle = \int_{-\infty}^{+\infty} x\,|\Psi(x,t)|^2\,dx = \frac{4}{a_0^3}\int_0^{\infty} x^3\,e^{-2x/a_0}\,dx = \frac{4.6a_0^4}{2^4a_0^4} = 1.5a_0.$$

Similarly, expectation values of functions of x can be derived. For $f(x)$, we have

$$\langle f(x)\rangle = \int_{-\infty}^{+\infty} f(x)\,|\Psi(x,t)|^2 dx. \tag{5.10}$$

In particular, we have

$$\langle x^2\rangle = \int_{-\infty}^{+\infty} x^2\,|\Psi(x,t)|^2 dx. \tag{5.11}$$

We can use this to define the uncertainty in the position of the particle. The uncertainty is a measure of how widely the results of the measurement of the position of the electron are spread around the mean value. As is the case in the analysis of statistical data, this is done in terms of the usual statistical quantity, the standard deviation, written Δx. This quantity is determined in terms of the average value of the *squares* of the distance of each value x from the average value of the data – if we did not square these differences, the final average would be zero, a useless result. For a sample obtained by N measurements of the position of the particle, it is obtained by the following formula:

$$(\Delta x)^2 \approx \sum_{\text{All }\delta x} (x-\overline{x})^2\frac{\delta N(x)}{N} \tag{5.12}$$

where $\overline{x}$ is the average value obtained from the data. The uncertainty is then found by taking the square root of this average.

In the limit of an infinite number of measurements, the uncertainty can be written

$$(\Delta x)^2 = \langle (x-\langle x\rangle)^2\rangle = \langle x^2\rangle - \langle x\rangle^2. \tag{5.13}$$

It is the uncertainty defined in this way that appears in the standard form for the Heisenberg uncertainty relation.

Ex 5.4 Once again, using the data given above for an electron close to the surface of liquid helium, we can calculate the uncertainty in the position of the electron from this surface. Using $\overline{x} = 1.235a_0$ we have

$$\begin{aligned}(\Delta x)^2 \approx\ & (0-1.235)^2a_0^2\times\frac{28}{300} + (0.5-1.235)^2a_0^2\times\frac{69}{300}\\ &+ (1-1.235)^2a_0^2\times\frac{76}{300} + (1.5-1.235)^2a_0^2\times\frac{58}{300}\\ &+ (2-1.235)^2a_0^2\times\frac{32}{300} + (2.5-1.235)^2a_0^2\times\frac{17}{300}\\ &+ (3-1.235)^2a_0^2\times\frac{11}{300} + (3.5-1.235)^2a_0^2\times\frac{6}{300}\\ &+ (4-1.235)^2a_0^2\times\frac{2}{300} + (4.5-1.235)^2a_0^2\times\frac{1}{300}\\ &= 0.751a_0^2\end{aligned}$$

so that $\Delta x \approx 0.866 a_0$. This can be compared to the uncertainty as calculated from the wave function itself. This will be given by

$$(\Delta x)^2 = \langle x^2 \rangle - 2.25 a_0^2$$

where we have used the previously calculated value for $\langle x \rangle = 1.5 a_0$. What is required is

$$\langle x^2 \rangle = \int_{-\infty}^{+\infty} x^2 \, |\Psi(x,t)|^2 \, dx = \frac{4}{a_0^3} \int_0^{\infty} x^4 e^{-2x/a_0} \, dx = \frac{4}{a_0^3} \frac{24 a_0^5}{2^5} = 3 a_0^2$$

and hence

$$\Delta x = \sqrt{3a_0^2 - 2.25 a_0^2} = 0.866 a_0.$$

So far, we have made use of the wave function to make statements concerning the position of the particle. But if we are to believe that the wave function is the repository of all information about the properties of a particle, we ought to be able to say something about the properties of, for instance, the velocity of the particle. To this end, we introduce the average velocity into the picture in the following way.

We have shown above that the expectation value of the position of a particle described by a wave function $\Psi(x,t)$ is given by Eq. (5.9). It is therefore reasonable to presume that the average value of the velocity is just the derivative of this expression, i.e.

$$\begin{aligned}
\langle v(t) \rangle &= \frac{d\langle x(t) \rangle}{dt} = \frac{d}{dt} \int_{-\infty}^{+\infty} x \, |\Psi(x,t)|^2 dx \\
&= \int_{-\infty}^{+\infty} x \, \frac{\partial}{\partial t} |\Psi(x,t)|^2 dx \\
&= \int_{-\infty}^{+\infty} x \left[\frac{\partial \Psi^*(x,t)}{\partial t} \Psi(x,t) + \Psi^*(x,t) \frac{\partial \Psi(x,t)}{\partial t} \right] dx. \qquad (5.14)
\end{aligned}$$

More usually, it is the average value of the momentum that is of deeper significance than that of velocity. So we rewrite this last result as the average of the momentum. We also note the the term in [...] can be written as the real part of a complex quantity, so that we end up with

$$\langle p \rangle = 2m \mathrm{Re} \left[\int_{-\infty}^{+\infty} x \, \Psi^*(x,t) \frac{\partial \Psi(x,t)}{\partial t} dx \right]. \qquad (5.15)$$

Later we will see that there is a far more generally useful (and fundamental) expression for the expectation value of momentum, which also allows us to define the uncertainty Δp in momentum.

The question now arises as to how the wave function can be obtained for a particle, or indeed, for a system of particles. The most general way is by solving the Schrödinger equation, but before we consider this general approach, we will consider a particularly simple example about which much can be said, even with the limited understanding that we have at this stage. The model is that of a particle in an infinitely deep potential well.

5.3 Particle in an Infinite Potential Well

Suppose we have a single particle of mass m confined to within a region $0 < x < L$ with potential energy $V = 0$ bounded by infinitely high potential barriers, i.e. $V = \infty$ for $x < 0$ and $x > L$. This simple model is sufficient to describe (in one dimension), for instance, the properties of the

conduction electrons in a metal (in the so-called free electron model), or the properties of gas particles in an ideal gas where the particles do not interact with each other. We want to learn as much about the properties of the particle using what we have learned about the wave function above.

The first point to note is that, because of the infinitely high barriers, the particle cannot be found in the regions $x > L$ and $x < 0$. Consequently, the wave function has to be zero in these regions. If we make the not unreasonable assumption that the wave function has to be continuous, then we must conclude that

$$\Psi(0,t) = \Psi(L,t) = 0. \tag{5.16}$$

These conditions on $\Psi(x,t)$ are known as boundary conditions. Between the barriers, the energy of the particle is purely kinetic. Suppose the energy of the particle is E, so that

$$E = \frac{p^2}{2m}. \tag{5.17}$$

Using the de Broglie relation $E = \hbar k$ we then have that

$$k = \pm\frac{\sqrt{2mE}}{\hbar} \tag{5.18}$$

while, from $E = \hbar\omega$ we have

$$\omega = E/\hbar. \tag{5.19}$$

In the region $0 < x < L$ the particle is free, so the wave function must be of the form Eq. (5.6), or perhaps a combination of such wave functions, in the manner that gave us the wave packets in Section 3.2. In deciding on the possible form for the wave function, we are restricted by two requirements. First, the boundary conditions Eq. (5.16) must be satisfied and secondly, we note that the wave function must be normalized to unity, Eq. (5.5). The first of these conditions immediately implies that the wave function cannot be simply $A\sin(kx-\omega t)$, $A\cos(kx-\omega t)$, or $Ae^{i(kx-\omega t)}$ or so on, as none of these will be zero at $x = 0$ and $x = L$ for all time. The next step is therefore to try a combination of these wave functions. In doing so we note two things: first, from Eq. (5.18) we see there are two possible values for k, and further we note that any sin or cos function can be written as a sum of complex exponentials:

$$\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2} \qquad \sin\theta = \frac{e^{i\theta}-e^{-i\theta}}{2i}$$

which suggests that we can try combining the lot together and see if the two conditions above pick out the combination that works. Thus, we will try

$$\Psi(x,t) = Ae^{i(kx-\omega t)} + Be^{-i(kx-\omega t)} + Ce^{i(kx+\omega t)} + De^{-i(kx+\omega t)} \tag{5.20}$$

where A, B, C, and D are coefficients that we wish to determine from the boundary conditions and from the requirement that the wave function be normalized to unity for all time.

First, consider the boundary condition at $x = 0$. Here, we must have

$$\begin{aligned}\Psi(0,t) &= Ae^{-i\omega t} + Be^{i\omega t} + Ce^{i\omega t} + De^{-i\omega t}\\ &= (A+D)e^{-i\omega t} + (B+C)e^{i\omega t}\\ &= 0.\end{aligned} \tag{5.21}$$

This must hold true for all time, which can only be the case if $A + D = 0$ and $B + C = 0$. Thus we conclude that we must have

$$\begin{aligned}\Psi(x,t) &= Ae^{i(kx-\omega t)} + Be^{-i(kx-\omega t)} - Be^{i(kx+\omega t)} - Ae^{-i(kx+\omega t)}\\ &= A(e^{ikx}-e^{-ikx})e^{-i\omega t} - B(e^{ikx}-e^{-ikx})e^{i\omega t}\\ &= 2i\sin(kx)(Ae^{-i\omega t} - Be^{i\omega t}).\end{aligned} \tag{5.22}$$

Now check for normalization:

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\, dx = 4\left|Ae^{-i\omega t} - Be^{i\omega t}\right|^2 \int_0^L \sin^2(kx)\, dx \tag{5.23}$$

where we note that the limits on the integral are $(0, L)$ since the wave function is zero outside that range.

This integral must be equal to unity for all time. But, since

$$\begin{aligned}\left|Ae^{-i\omega t} - Be^{i\omega t}\right|^2 &= (Ae^{-i\omega t} - Be^{-i\omega t})(A^* e^{i\omega t} - B^* e^{-i\omega t}) \\ &= AA^* + BB^* - AB^* e^{-2i\omega t} - A^* B e^{2i\omega t}\end{aligned} \tag{5.24}$$

what we have instead is a time dependent result, unless we have either $A = 0$ or $B = 0$. It turns out that either choice can be made – we will make the conventional choice and put $B = 0$ to give

$$\Psi(x,t) = 2iA \sin(kx) e^{-i\omega t}. \tag{5.25}$$

We can now check on the other boundary condition, i.e. that $\Psi(L,t) = 0$, which leads to:

$$\sin(kL) = 0 \tag{5.26}$$

and hence

$$kL = n\pi \quad n \text{ an integer} \tag{5.27}$$

which implies that k can have only a restricted set of values given by

$$k_n = \frac{n\pi}{L}. \tag{5.28}$$

An immediate consequence of this is that the energy of the particle is limited to the values

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\pi^2 n^2 \hbar^2}{2mL^2} = \hbar\omega_n \tag{5.29}$$

i.e. the energy is 'quantized'.

Using these values of k in the normalization condition leads to

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\, dx = 4|A|^2 \int_0^L \sin^2(k_n x) = 2|A|^2 L \tag{5.30}$$

so that by making the choice

$$A = \sqrt{\frac{1}{2L}} e^{i\phi} \tag{5.31}$$

where ϕ is an unknown phase factor, we ensure that the wave function is indeed normalized to unity. Nothing we have seen above can give us a value for ϕ, but whatever choice is made, it always found to cancel out in any calculation of a physically observable result, so its value can be set to suit our convenience. Here, we will choose $\phi = -\pi/2$ and hence

$$A = -i\sqrt{\frac{1}{2L}}. \tag{5.32}$$

The wave function therefore becomes

$$\begin{aligned}\Psi_n(x,t) &= \sqrt{\frac{2}{L}} \sin(n\pi x/L) e^{-i\omega_n t} \quad 0 < x < L \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad x < 0, \quad x > L.\end{aligned} \tag{5.33}$$

with associated energies

$$E_n = \frac{\pi^2 n^2 \hbar^2}{2mL^2} \quad n = 1, 2, 3, \ldots . \tag{5.34}$$

where the wave function and the energies has been labelled by the quantity n, known as a quantum number. It can have the values $n = 1, 2, 3, \ldots$, i.e. $n = 0$ is excluded, for then the wave function vanishes everywhere, and also excluded are the negative integers since they yield the same set of wave functions, and the same energies.

We see that the particle can only have the energies E_n, and in particular, the lowest energy, E_1 is greater than zero, as is required by the uncertainty principle. Thus the energy of the particle is quantized, in contrast to the classical situation in which the particle can have any energy ≥ 0.

5.3.1 Some Properties of Infinite Well Wave Functions

The wave functions derived above define the probability distributions for finding the particle of a given energy in some region in space. But the wave functions also possess other important properties, some of them of a purely mathematical nature that prove to be extremely important in further development of the theory, but also providing other information about the physical properties of the particle.

Energy Eigenvalues and Eigenfunctions

The above wave functions can be written in the form

$$\Psi_n(x, t) = \psi_n(x) e^{-iE_n t/\hbar} \tag{5.35}$$

where we note the time dependence factors out of the overall wave function as a complex exponential of the form $\exp[-iE_n t/\hbar]$. As will be seen later, the time dependence of the wave function for any system in a state of given energy is always of this form. The energy of the particle is limited to the values specified in Eq. (5.34). This phenomenon of energy quantization is to be found in all systems in which a particle is confined by an attractive potential such as the Coulomb potential binding an electron to a proton in the hydrogen atom, or the attractive potential of a simple harmonic oscillator. In all cases, the boundary condition that the wave function vanish at infinity guarantees that only a discrete set of wave functions are possible, and each is associated with a certain energy – hence the energy levels of the hydrogen atom, for instance.

The remaining factor $\psi_n(x)$ contains all the spatial dependence of the wave function. We also note a 'pairing' of the wave function $\psi_n(x)$ with the allowed energy E_n. The wave function $\psi_n(x)$ is known as an energy eigenfunction and the associated energy is known as the energy eigenvalue. This terminology has its origins in the more general formulation of quantum mechanics in terms of state vectors and operators that we will be considering in later Chapters.

Some illustrative plots of the wave functions are presented in Fig. 5.3

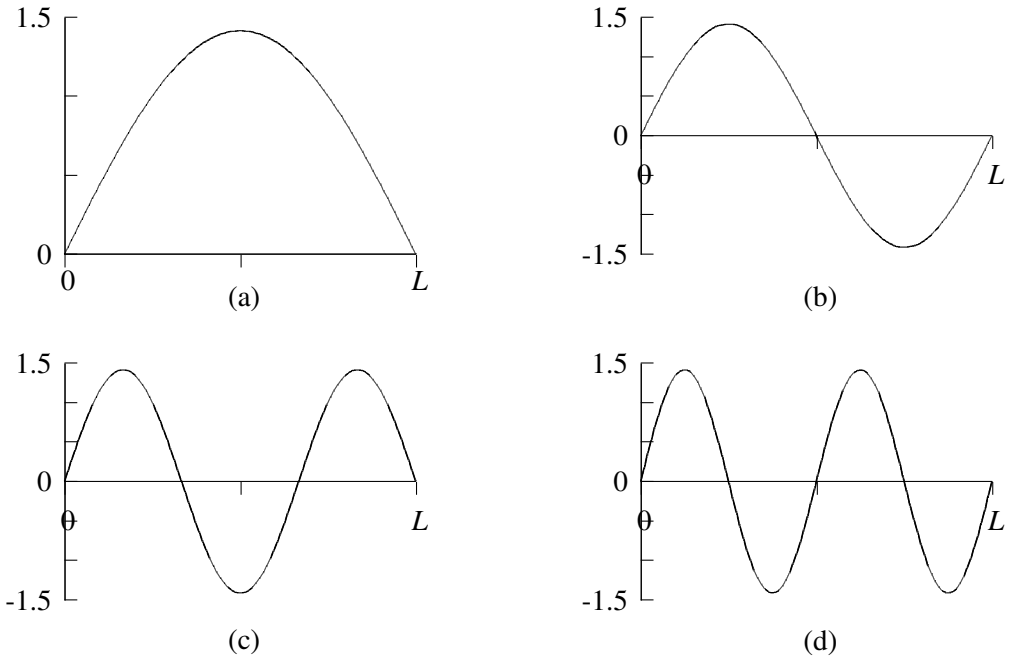


Figure 5.3: Plots of infinite potential well wave function $\psi_n(x)$ for (a) $n = 1$, (b) $n = 2$, (c) $n = 3$, and (d) $n = 4$.

Probability Distributions for Position

The probability distributions corresponding to the wave functions obtained above are

$$\begin{aligned} P_n(x) = |\Psi_n(x,t)|^2 &= \frac{2}{L}\sin^2(n\pi x/L) \quad 0 < x < L \\ &= 0 \qquad\qquad\qquad\qquad\quad x < 0, \quad x > L \end{aligned} \tag{5.36}$$

which are all independent of time, i.e. these are analogous to the stationary states of the hydrogen atom introduced by Bohr – states whose properties do not change in time. The nomenclature stationary state is retained in modern quantum mechanics for such states. We can plot $P_n(x)$ as a function of x for various values of n to see what we can learn about the properties of the particle in the well (see Fig. (5.4)). We note that P_n is not uniform across the well. In fact, there are regions where it is very unlikely to observe the particle, whereas elsewhere the chances are maximized. If n becomes very large (see Fig. (5.4)(d)), the probability oscillates very rapidly, averaging out to be $1/L$, so that the particle is equally likely to be found anywhere in the well. This is what would be found classically if the particle were simply bouncing back and forth between the walls of the well, and observations were made at random times, i.e. the chances of finding the particle in a region of size δx will be $\delta x/L$.

The expectation value of the position of the particle can be calculated directly from the above expressions for the probability distributions, using the general result Eq. (5.10). The integral is straightforward, and gives

$$\langle x\rangle = \frac{2}{L}\int_0^L x\sin^2(n\pi x/L) = \tfrac{1}{2}L \tag{5.37}$$

i.e. the expectation value is in the middle of the well. Since the probability distributions $P_n(x)$ is symmetric about $x = L/2$, this result is as expected. Note that this does not necessarily correspond to where the probability is a maximum. In fact, for, say $n = 2$, the particle is most likely to be found in the vicinity of $x = L/4$ and $x = 3L/4$.

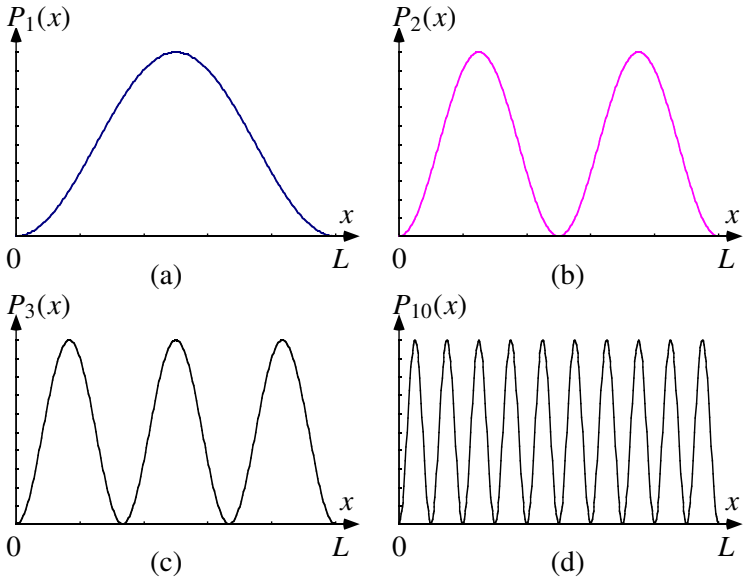


Figure 5.4: Plots of the probability distributions $P_n(x)$ for a particle in an infinite potential well of width L for (a) $n = 1$, (b) $n = 2$, (c) $n = 3$ and (d) $n = 10$. The rapid oscillations on (d) imply that the probability is averages out to be constant across the width of the well.

From the wave functions $\Psi_n(x,t)$, using the definition Eq. (5.10) to calculate $\langle x^2\rangle$ and $\langle x\rangle$, it is also possible to calculate the uncertainty in the position of the particle. We have already seen that $\langle x\rangle = L/2$, while the other expectation value $\langle x^2\rangle$ is given by

$$\langle x^2\rangle = \frac{2}{L}\int_0^L x^2 \sin^2(n\pi x/L)\,dx = L^2\frac{2n^2\pi^2-3}{6n^2\pi^2}. \tag{5.38}$$

Consequently, the uncertainty in position is

$$(\Delta x)^2 = L^2\frac{n^2\pi^2-3}{n^2\pi^2} - \frac{L^2}{4} = L^2\frac{n^2\pi^2-6}{12n^2\pi^2}. \tag{5.39}$$

Orthonormality

An important feature of the wave functions derived above follows from considering the following integral:

$$\int_{-\infty}^{+\infty}\psi_m^*(x)\psi_n(x)\,dx = \frac{2}{L}\int_0^L \sin(m\pi x/L)\sin(n\pi x/L)\,dx = \delta_{mn} \tag{5.40}$$

where δ_{mn} is known as the Kronecker delta, and has the property that

$$\begin{aligned}\delta_{mn} &= 1 \quad m = n\\ &= 0 \quad m \neq n\end{aligned} \tag{5.41}$$

Thus, if $m = n$, the integral is unity, as it should be as this is just the normalization condition that was imposed earlier on the wave functions. However, if $m \neq n$, then the integral vanishes. The two wave functions are then said to be orthogonal, a condition somewhat analogous to two vectors being orthogonal. The functions ψ_n for all $n = 1, 2, 3, \ldots$ are then said to be *orthonormal*. The property of orthonormality of eigenfunctions is found to be a general property of the states of quantum systems that will be further explored in later Chapters.

Linear Superpositions

We found earlier that if we combine together wave functions of different wavelength and frequency, corresponding to different particle momenta and energies, we produce something that had an acceptable physical meaning – a wave packet. What we will do now is combine together – but in this case only two – different wave functions and see what meaning can be given to the result obtained. Thus, we will consider the following linear combination, or linear superposition, of two wave functions:

$$\begin{aligned}\Psi(x,t) &= \frac{1}{\sqrt{2}}[\Psi_1(x,t) + \Psi_2(x,t)] \\ &= \frac{1}{\sqrt{L}}[\sin(\pi x/L)e^{-i\omega_1 t} + \sin(2\pi x/L)e^{-i\omega_2 t}] \quad 0 < x < L \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x < 0 \text{ and } x > L. \end{aligned} \tag{5.42}$$

The factor $1/\sqrt{2}$ guarantees that the wave function is normalized to unity, as can be seen by calculating the normalization integral Eq. (5.5) for the wave function defined in Eq. (5.42).

How are we to interpret this wave function? Superficially, it is seen to be made up of two wave functions associated with the particle having energies E_1 and E_2. These wave functions contribute equally to the total wave function $\Psi(x,t)$ in the sense that they both have the same amplitude, so it is tempting to believe that if we were to measure the energy of the particle rather than its position, we would get either result E_1 and E_2 with equal probability of $\frac{1}{2}$. This interpretation in fact turns out to be the case as we will see later. But the fact that the particle does not have a definite energy has important consequences as can be seen by considering the probability distribution for the position of the particle.

This probability distribution is

$$\begin{aligned}P(x,t) &= |\Psi(x,t)|^2 \\ &= \frac{1}{L}[\sin^2(\pi x/L) + \sin^2(2\pi x/L) + 2\sin(\pi x/L)\sin(2\pi x/L)\cos(\Delta\omega t)]\end{aligned} \tag{5.43}$$

where $\Delta\omega = (E_2 - E_1)/\hbar$. This is obviously a time dependent probability distribution, in contrast to what was found for the eigenfunctions $\Psi_n(x,t)$. In other words, if the wave function is made up of contributions of different energies, the particle is not in a stationary state.

In Fig. (5.5), this probability distribution is plotted at three times. At $t = 0$, the probability distribution is

$$P(x,0) = \frac{1}{L}(\sin(\pi x/L) + \sin(2\pi x/L))^2 \tag{5.44}$$

which results in the distribution being peaked on the left hand side of the well. At the time $t = \pi/2\Delta\omega$, the time dependent term vanishes and the distribution is

$$P(x,\pi/2\Delta\omega) = \frac{1}{L}(\sin^2(\pi x/L) + \sin^2(2\pi x/L)). \tag{5.45}$$

Finally, at time $t = \pi/\Delta\omega$, the distribution is

$$P(x,\pi/\Delta\omega) = \frac{1}{L}(\sin(\pi x/L) - \sin(2\pi x/L))^2 \tag{5.46}$$

which gives a peak on the right hand side of the well. Thus, the maximum probability swings from the left to the right hand side of the well (and back again), *but without the maximum moving through the centre of the well*. This is counterintuitive: the maximum would be expected to also

move back and forth between the walls, mirroring the expected classical behaviour of the particle bouncing back and forth between the walls.

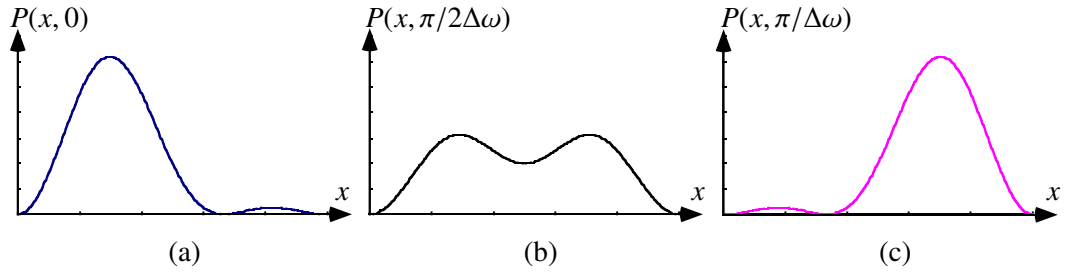


Figure 5.5: Plots of the time dependent probability distributions $P(x,t) = (\Psi_1(x,t) + \Psi_2(x,t))/\sqrt{2}$ for a particle in an infinite potential well. The times are (a) $t = 0$, (b) $t = \pi/2\Delta\omega$ and (c) $t = \pi/\Delta\omega$ where $\hbar\Delta\omega = E_2 - E_1$.

As a final illustration, we can find the total probability of finding the particle on the left hand half of the well, i.e. in the region $0 < x < L/2$:

$$P_L = \int_0^{L/2} P(x,t)\,dx = \tfrac{1}{2} + \pi^{-1}\cos(\Delta\omega t) \tag{5.47}$$

while the corresponding result for the right hand side is

$$P_R = \tfrac{1}{2} - \pi^{-1}\cos(\Delta\omega t) \tag{5.48}$$

which perhaps better illustrates the 'see-sawing' of the probability from one side to the other with a frequency of $2\pi\Delta\omega$. What can be learned from this example is that if the wave function is made up of two contributions of different energy, then the properties of the system do not stay constant in time, i.e. the system is no longer in a stationary state. Once again, the example of a particle in a well illustrates a generic feature of quantum systems, namely that if they do not have a definite energy, then the properties of the system change in time.

Particle Momentum

A final point to be considered here is that of determining what we can say about the momentum of the particle in the well. We cannot do this in an entirely correct fashion at this stage, but for the purposes of further illustrating that the wave function contains more than just information on the position of the particle, we will use slightly less rigorous arguments to arrive at essentially correct conclusions.

If we return to the eigenfunctions $\Psi_n(x,t)$, we see that they can be written in the form

$$\Psi_n(x,t) = \sqrt{\frac{2}{L}}\frac{e^{i(k_n x - \omega_n t)} - e^{-i(k_n x + \omega_n t)}}{2i} \tag{5.49}$$

i.e. a linear combination of two counterpropagating waves, one associated with the particle having momentum $p_n = \hbar k_n$, the other with the particle having momentum $p_n = -\hbar k_n$. These two contributions enter the above expression with equal weight, by which is meant that the magnitude of the coefficients of each of the exponentials is the same. It therefore seems reasonable to suspect that the particle has an equal chance of being observed to have momenta $p_n = \pm\hbar k_n$ when the wave function of the particle is $\Psi_n(x,t)$. This conjecture is consistent with what we would suppose is going on classically – the particle would be bouncing back and forth between the walls, simply

reversing the direction of its momentum on each bounce. It should be pointed out that this is the argument alluded to above which is not entirely correct. The momentum actually has a probability distribution which is peaked at the values we have obtained here. The point can still be taken that information on other than the position of the particle is to be found in the wave function.

Accepting this, we can then say that the particle has a chance of $\frac{1}{2}$ of being observed to have momentum $\hbar k_n$ and a chance of $\frac{1}{2}$ of being observed to have momentum $-\hbar k_n$, at least if the wave function is $\Psi_n(x,t)$. On this basis, we can calculate the expectation value of the momentum, that is

$$\langle p \rangle = \tfrac{1}{2}\hbar k_n + \tfrac{1}{2}(-\hbar k_n) = 0 \tag{5.50}$$

and the expectation value of the momentum squared:

$$\langle p^2 \rangle = \tfrac{1}{2}(\hbar k_n)^2 + \tfrac{1}{2}(-\hbar k_n)^2 = \hbar^2 k_n^2. \tag{5.51}$$

Both these results are, in fact, exact. The uncertainty in the momentum, Δp, follows from

$$(\Delta p)^2 = \langle p^2 \rangle - \langle p \rangle^2 = \hbar^2 k_n^2 = \hbar^2 (n\pi/L)^2. \tag{5.52}$$

We can combine this with the result Eq. (5.39) for the uncertainty in the position to give

$$(\Delta x)^2 (\Delta p)^2 = L^2 \frac{n^2\pi^2 - 6}{12 n^2 \pi^2} \hbar^2 (n\pi/L)^2 = \frac{\hbar^2}{4}\left(\frac{n^2\pi^2 - 6}{3}\right). \tag{5.53}$$

Since the term $(n^2\pi^2 - 6)/3)$ is always bigger than unity (at the smallest, when $n = 1$, it is 1.29), we have

$$(\Delta x)^2 (\Delta p)^2 \geq \frac{\hbar^2}{4} \tag{5.54}$$

or, in other words

$$\Delta x \Delta p \geq \tfrac{1}{2}\hbar \tag{5.55}$$

in agreement with the Heisenberg uncertainty principle, Eq. (3.14).

These results were obtained on the basis of making a few reasonable assumptions about how we extract properties of the momentum from the wave function. But this is not a method that can be easily generalised for arbitrary wave functions. We *could* also approach these problems by making use of the result obtained in Eq. (5.15):

$$\langle p \rangle = 2m \,\mathrm{Re}\left[\int_{-\infty}^{+\infty} x \,\Psi^*(x,t) \frac{\partial \Psi(x,t)}{\partial t} dx\right]. \tag{5.56}$$

This is a result that would presumably be applicable for any wave function, so it that sense it is more general than what we have been trying to do here. In fact, it is not difficult to show that on substituting $\Psi(x,t) = \Psi_n(x,t)$ from the expression Eq. (5.33) for the infinite potential wave functions $\Psi_n(x,t)$, that $\langle p \rangle = 0$ as before: the quantity $[\ldots]$ turns out to be purely imaginary, i.e. its real part is zero, hence the result follows.

Unfortunately, we cannot go much further than this on the basis of Eq. (5.56) since, for instance, it is not clear how $\langle p^2 \rangle$ could be calculated. To proceed further, other methods of extracting information on the particle from the wave function need to be developed. An important step in achieving this necessitates moving beyond the simple example of a particle in an infinite potential well, but the sort of arguments that were used to derive the infinite potential well wave functions and energy levels are really only suited to this one problem. What we need is a method of determining the wave function for an arbitrary potential, and what we end up with is a 'wave equation' for the wave function: the celebrated Schrödinger wave equation.

5.4 The Schrödinger Wave Equation

So far, we have made a lot of progress concerning the properties of, and interpretation of the wave function, but as yet we have had very little to say about how the wave function may be derived in a general situation, that is to say, we do not have on hand a 'wave equation' for the wave function. There is no true derivation of this equation, but its form can be motivated by physical and mathematical arguments at a wide variety of levels of sophistication. Here, we will offer a simple derivation based on what we have learned so far about the wave function.

The Schrödinger equation has two 'forms', one in which time explicitly appears, and so describes how the wave function of a particle will evolve in time. In general, the wave function behaves like a wave, and so the equation is often referred to as the time dependent Schrödinger wave equation. The other is the equation in which the time dependence has been 'removed' and hence is known as the time independent Schrödinger equation and is found to describe, amongst other things, what the allowed energies are of the particle. These are not two separate, independent equations – the time independent equation can be derived readily from the time dependent equation (except if the potential is time dependent, a development we will not be discussing here). In the following we will describe how the first, time dependent equation can be 'derived', and in then how the second follows from the first.

5.4.1 The Time Dependent Schrödinger Wave Equation

In the discussion of the particle in an infinite potential well, it was observed that the wave function of a particle of fixed energy E could most naturally be written as a linear combination of wave functions of the form

$$\Psi(x,t) = Ae^{i(kx-\omega t)} \tag{5.57}$$

representing a wave travelling in the positive x direction, and a corresponding wave travelling in the opposite direction, so giving rise to a standing wave, this being necessary in order to satisfy the boundary conditions. This corresponds intuitively to our classical notion of a particle bouncing back and forth between the walls of the potential well, which suggests that we adopt the wave function above as being the appropriate wave function for a *free* particle of momentum $p = \hbar k$ and energy $E = \hbar\omega$. With this in mind, we can then note that

$$\frac{\partial^2\Psi}{\partial x^2} = -k^2\Psi \tag{5.58}$$

which can be written, using $E = p^2/2m = \hbar^2k^2/2m$:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} = \frac{p^2}{2m}\Psi. \tag{5.59}$$

Similarly

$$\frac{\partial\Psi}{\partial t} = -i\omega\Psi \tag{5.60}$$

which can be written, using $E = \hbar\omega$:

$$i\hbar\frac{\partial\Psi}{\partial t} = \hbar\omega\Psi = E\Psi. \tag{5.61}$$

We now generalize this to the situation in which there is both a kinetic energy and a potential energy present, then $E = p^2/2m + V(x)$ so that

$$E\Psi = \frac{p^2}{2m}\Psi + V(x)\Psi \tag{5.62}$$

where Ψ is now the wave function of a particle moving in the presence of a potential $V(x)$. But if we assume that the results Eq. (5.59) and Eq. (5.61) still apply in this case then we have

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V(x)\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{5.63}$$

which is the famous time dependent Schrödinger wave equation. It is setting up and solving this equation, then analyzing the physical contents of its solutions that form the basis of that branch of quantum mechanics known as wave mechanics.

Even though this equation does not look like the familiar wave equation that describes, for instance, waves on a stretched string, it is nevertheless referred to as a 'wave equation' as it can have solutions that represent waves propagating through space. We have seen an example of this: the harmonic wave function for a free particle of energy E and momentum p, i.e.

$$\Psi(x,t) = Ae^{i(px-Et)/\hbar} \tag{5.64}$$

is a solution of this equation with, as appropriate for a free particle, $V(x) = 0$. But this equation can have distinctly non-wave like solutions whose form depends, amongst other things, on the nature of the potential $V(x)$ experienced by the particle.

In general, the solutions to the time dependent Schrödinger equation will describe the *dynamical* behaviour of the particle, in some sense similar to the way that Newton's equation $F = ma$ describes the dynamics of a particle in classical physics. However, there is an important difference. By solving Newton's equation we can determine the position of a particle as a function of time, whereas by solving Schrödinger's equation, what we get is a wave function $\Psi(x,t)$ which tells us (after we square the wave function) how the *probability* of finding the particle in some region in space varies as a function of time.

It is possible to proceed from here and look at ways and means of solving the full, time dependent Schrödinger equation in all its glory, and look for the physical meaning of the solutions that are found. This is a task that takes us into mathematical territory that is left to much later, Chapter ∞, to address, but a step towards this, and one that is also exceedingly important in its own right is that of determining the wave function for a particle of a given energy E. To do so requires 'extracting' the time dependence from the time dependent Schrödinger equation.

5.4.2 The Time Independent Schrödinger Equation

We have seen what the wave function looks like for a free particle of energy E – one or the other of the harmonic wave functions – and we have seen what it looks like for the particle in an infinitely deep potential well – see Section 5.3 – though we did not obtain that result by solving the Schrödinger equation. But in both cases, the time dependence entered into the wave function via a complex exponential factor $\exp[-iEt/\hbar]$. This suggests that to 'extract' this time dependence we guess a solution to the Schrödinger wave equation of the form

$$\Psi(x,t) = \psi(x)e^{-iEt/\hbar} \tag{5.65}$$

i.e. where the space and the time dependence of the complete wave function are contained in separate factors[2]. The idea now is to see if this guess enables us to derive an equation for $\psi(x)$, the spatial part of the wave function.

If we substitute this trial solution into the Schrödinger wave equation, and make use of the meaning of partial derivatives, we get:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2}e^{-iEt/\hbar} + V(x)\psi(x)e^{-iEt/\hbar} = i\hbar.-iE/\hbar e^{-iEt/\hbar}\psi(x) = E\psi(x)e^{-iEt/\hbar}. \tag{5.66}$$

[2]A solution of this form can be shown to arise by the method of 'the separation of variables', a well known mathematical technique used to solve equations of the form of the Schrödinger equation.

We now see that the factor $\exp[-iEt/\hbar]$ cancels from both sides of the equation, giving us

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \tag{5.67}$$

If we rearrange the terms, we end up with

$$\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + (E - V(x))\psi(x) = 0 \tag{5.68}$$

which is the time independent Schrödinger equation.

We note here that the quantity E, which we have identified as the energy of the particle, is a free parameter in this equation. In other words, at no stage has any restriction been placed on the possible values for E. Thus, if we want to determine the wave function for a particle with some specific value of E that is moving in the presence of a potential $V(x)$, all we have to do is to insert this value of E into the equation with the appropriate $V(x)$, and solve for the corresponding wave function. In doing so, we find, perhaps not surprisingly, that for different choices of E we get different solutions for $\psi(x)$. We can emphasize this fact by writing $\psi_E(x)$ as the solution associated with a particular value of E. But it turns out that it is not all quite as simple as this. To be physically acceptable, the wave function $\psi_E(x)$ must satisfy two conditions, one of which we have seen before, namely that the wave function must be normalizable (see Eq. (5.5)), and a second, that the wave function and its derivative must be continuous. Together, these two requirements, the first founded in the probability interpretation of the wave function, the second in more esoteric mathematical necessities which we will not go into here and usually only encountered in somewhat artificial problems, lead to a rather remarkable property of physical systems described by this equation that has enormous physical significance: the quantization of energy.

5.4.3 Boundary Conditions and the Quantization of Energy

At first thought it might seem to be perfectly acceptable to insert any value of E into the time independent Schrödinger equation and solve it for $\psi_E(x)$. But in doing so we must remain aware of one further requirement of a wave function which comes from its probability interpretation: to be physically acceptable a wave function must satisfy the normalization condition, Eq. (5.5)

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\, dx = 1$$

for all time t. For the particular trial solution introduced above, Eq. (5.65):

$$\Psi(x,t) = \psi_E(x)e^{-iEt/\hbar} \tag{5.69}$$

the requirement that the normalization condition must hold gives, on substituting for $\Psi(x,t)$, the result[3]

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\, dx = \int_{-\infty}^{+\infty} |\psi_E(x)|^2\, dx = 1. \tag{5.70}$$

Since this integral must be finite, (unity in fact), we *must* have $\psi_E(x) \to 0$ as $x \to \pm\infty$ in order for the integral to have any hope of converging to a finite value. The importance of this with regard to solving the time dependent Schrödinger equation is that we must check whether or not a solution $\psi_E(x)$ obtained for some chosen value of E satisfies the normalization condition. If it does, then

[3]Note that the time dependence has cancelled out because

$$|\Psi(x,t)|^2 = |\psi_E(x)e^{-iEt/\hbar}|^2 = |\psi_E(x)|^2|e^{-iEt/\hbar}|^2 = |\psi_E(x)|^2$$

since, for any complex number of the form $\exp(i\phi)$, we have $|\exp(i\phi)|^2 = 1$.

this is a physically acceptable solution, if it does not, then that solution *and the corresponding value of the energy* are not physically acceptable. The particular case of considerable physical significance is if the potential $V(x)$ is attractive, such as would be found with an electron caught by the attractive Coulomb force of an atomic nucleus, or a particle bound by a simple harmonic potential (a mass on a spring), or, as we have seen in Section 5.3, a particle trapped in an infinite potential well. In all such cases, we find that except for certain discrete values of the energy, the wave function $\psi_E(x)$ does not vanish, or even worse, diverges, as $x \to \pm\infty$. In other words, it is only for these discrete values of the energy E that we get physically acceptable wave functions $\psi_E(x)$, or to put it more bluntly, the particle can never be observed to have any energy other than these particular values, for which reason these energies are often referred to as the 'allowed' energies of the particle. This pairing off of allowed energy and normalizable wave function is referred to mathematically as $\psi_E(x)$ being an eigenfunction of the Schrödinger equation, and E the associated energy eigenvalue, a terminology that acquires more meaning when quantum mechanics is looked at from a more advanced standpoint.

So we have the amazing result that the probability interpretation of the wave function forces us to conclude that the allowed energies of a particle moving in a potential $V(x)$ are restricted to certain discrete values, these values determined by the nature of the potential. This is the phenomenon known as the quantization of energy, a result of quantum mechanics which has enormous significance for determining the structure of atoms, or, to go even further, the properties of matter overall. We have already seen an example of this quantization of energy in our earlier discussion of a particle in an infintely deep potential well, though we did not derive the results by solving the Schrödinger equation itself. We will consider how this is done shortly.

The requirement that $\psi(x) \to 0$ as $x \to \pm\infty$ is an example of a *boundary condition*. Energy quantization is, mathematically speaking, the result of a combined effort: that $\psi(x)$ be a solution to the time independent Schrödinger equation, and that the solution satisfy these boundary conditions. But both the boundary condition and the Schrödinger equation are derived from, and hence rooted in, the nature of the physical world: we have here an example of the unexpected relevance of purely mathematical ideas in formulating a physical theory.

5.4.4 Continuity Conditions

When solving the Schrödinger equation, two conditions are required to be placed on the solution: both $\psi(x)$ and its derivative $\psi'(x)$ must be continuous. There are good physical reasons for these requirements. Firstly, if the wave function were discontinuous at some point for some reason, then its derivative at this point would be undefined. For reasons that will be explained later, this leads to trouble in trying to determine the momentum of the particle since, as it turns out, the momentum of the wave function is closely related to the spatial derivative of the wave function. And if the derivative should be discontinuous at some point, then in turn, its derivative, i.e. the second derivative of the wave function, would be undefined, which leads to trouble for determining the kinetic energy of the particle.

Typically it is not necessary to explicitly worry too much about these continuity conditions except in the case in which the potential $V(x)$ itself has a discontinuity, whether it be an infinite discontinuity, as in the infinite potential well example, or a finite discontinuity as we will see later in the case of the finite potential well. Although discontinuous potentials do not occur in nature (this would imply an infinite force), they do represent useful simplified models, but in dealing with them, we have to make certain that the wave function is smoothly connected across the discontinuity in the potential, in other words, such that the wave function and its derivative are continuous across the point at which the potential is discontinuous. The one exception is when the discontinuity in the potential is infinite, as in the case of the infinite potential well. In that case, only the wave function is required to be continuous. We shall see how this extra condition is implemented when we look at the finite potential well later.

5.4.5 Bound States and Scattering States

But what about wave functions such as the harmonic wave function $\Psi(x,t) = A\exp[i(kx - \omega t)]$? These wave functions represent a particle having a definite energy $E = \hbar\omega$ and so would seem to be legitimate and necessary wave functions within the quantum theory. But the problem here, as has been pointed out before in Section 5.1.1, is that $\Psi(x,t)$ does *not* vanish as $x \to \pm\infty$, so the normalization condition, Eq. (5.70) cannot be satisfied. So what is going on here? The answer lies in the fact that there are two kinds of wave functions, those that apply for particles trapped by an attractive potential into what is known as a bound state, and those that apply for particles that are free to travel to infinity (and beyond), otherwise known as scattering states. A particle trapped in an infinitely deep potential well is an example of the former: the particle is confined to move within a restricted region of space. An electron trapped by the attractive potential due to a positively charged atomic nucleus is also an example – the electron rarely moves a distance more than ~10 nm from the nucleus. A nucleon trapped within a nucleus by attractive nuclear forces is yet another. In all these cases, the probability of finding the particle at infinity is zero. In other words, the wave function for the particle satisfies the boundary condition that it vanish at infinity. So we see that it is when a particle is trapped, or confined to a limited region of space by an attractive potential $V(x)$ (or $V(\mathbf{r})$ in three dimensions), we obtain wave functions that satisfy the above boundary condition, and hand in hand with this, we find that their energies are quantized. But if it should be the case that the particle is free to move as far as it likes in space, in other words, if it is not bound by any attractive potential, (or even repelled by a repulsive potential) then we find that the wave function need not vanish at infinity, and nor is its energy quantized. The problem of how to reconcile this with the normalization condition, and the probability interpretation of the wave function, is a delicate mathematical issue which we cannot hope to address here, but it can be done. Suffice to say that provided the wave function does not diverge at infinity (in other words it remains finite, though not zero) we can give a physical meaning of such states as being an idealized mathematical limiting case which, while it does not satisfy the normalization condition, can still be dealt with in, provided some care is taken with the physical interpretation, in much the same way as the bound state wave functions.

In order to illustrate how the time independent Schrödinger equation can be solved in practice, and some of the characteristics of its solutions, we will here briefly reconsider the infinitely deep potential well problem, already solved by making use of general properties of the wave function, in Section 5.3. We will then move on to looking at other simple applications.

5.5 Solving the Time Independent Schrödinger Equation

5.5.1 The Infinite Potential Well Revisited

In the earlier treatment of this example, we assumed we had a single particle of mass m confined to within a region $0 < x < L$ with potential energy $V = 0$ bounded by infinitely high potential barriers, i.e. $V = \infty$ for $x < 0$ and $x > L$. Here, we will make a slight change which, superficially, seems to be unnecessary, but in the long run makes it possible to highlight properties of the infinite well wave functions that are somewhat obscure in the original treatment. Here we will choose the potential well to be specified by

$$V(x) = 0 \qquad |x| < \tfrac{1}{2}L \tag{5.71}$$

$$= \infty \qquad |x| \geq -\tfrac{1}{2}L \tag{5.72}$$

i.e. the well is symmetrically positioned with respect to the origin $x = 0$. The advantage of the earlier treatment was that imposing the boundary conditions on the wave function lead to particularly simple algebraic results. Here the algebra is slightly more complicated, but there is much to gain by making the change.

In the regions for which the potential is infinite, the wave function will be zero, for exactly the same reasons that it was set to zero in Section 5.3, that is, there is zero probability of the particle being found in these regions. Thus, we must impose the boundary conditions

$$\psi(-\tfrac{1}{2}L) = \psi(\tfrac{1}{2}L) = 0. \tag{5.73}$$

Meanwhile, in the region $|x| < \frac{1}{2}L$, the potential vanishes, so the time independent Schrödinger equation becomes:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x). \tag{5.74}$$

To solve this, we define a quantity k by

$$k = \sqrt{\frac{2mE}{\hbar^2}} \tag{5.75}$$

so that Eq. (5.74) can be written

$$\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \tag{5.76}$$

whose general solution is

$$\psi(x) = A\sin(kx) + B\cos(kx). \tag{5.77}$$

It is now that we impose the boundary conditions, Eq. (5.73), to give, first at $x = -\frac{1}{2}L$:

$$\psi(-\tfrac{1}{2}L) = -A\sin(\tfrac{1}{2}kL) + B\cos(\tfrac{1}{2}kL) = 0 \tag{5.78}$$

while applying the boundary condition at $x = L$ gives

$$\psi(\tfrac{1}{2}L) = A\sin(\tfrac{1}{2}kL) + B\cos(\tfrac{1}{2}kL) = 0. \tag{5.79}$$

Combining these last two equations by adding and subtracting then gives the two conditions

$$2B\cos(\tfrac{1}{2}kL) = 0 \quad \text{or} \quad 2A\sin(\tfrac{1}{2}kL) = 0. \tag{5.80}$$

The quantities $\cos(\frac{1}{2}kL)$ and $\sin(\frac{1}{2}kL)$ cannot both be zero at the same time. By the same token, they cannot both be non-zero, because that would mean both A and B would be zero, and the wave function would then be zero, so we must have EITHER $\cos(\frac{1}{2}kL)$ OR $\sin(\frac{1}{2}kL)$ equal to zero. Suppose we choose

$$\cos(\tfrac{1}{2}kL) = 0 \tag{5.81}$$

then we must have $2A\sin(\frac{1}{2}kL) = 0$, and since $\sin(\frac{1}{2}kL)$ cannot be zero as well, we must have $A = 0$. Thus we have the possible solution

$$\psi(x) = B\cos(kx) \qquad \text{with} \qquad \cos(\tfrac{1}{2}kL) = 0. \tag{5.82}$$

The allowed values of k are then

$$k_n = \frac{n\pi}{L}, \quad n = 1, 3, 5, \ldots \tag{5.83}$$

where we have excluded the negative values of n as they will merely reproduce the same set of solutions (except with opposite sign[4]) as the positive values.

The other alternative is to have

$$\sin(\tfrac{1}{2}kL) = 0 \tag{5.84}$$

[4]The sign has no effect on probabilities as we always square the wave function.

which tells us that we must have $B = 0$, so the corresponding solution is:

$$\psi(x) = A\sin(kx) \qquad \text{with} \qquad \sin(\tfrac{1}{2}kL) = 0. \tag{5.85}$$

The allowed values of k are then

$$k_n = \frac{n\pi}{L}, \quad n = 2, 4, 6, \ldots \tag{5.86}$$

We exclude the $n = 0$ possibility as that would give us $\psi(x) = 0$, and, as above, we have also excluded the negative possible values for n. Thus we have, combining the two results for the allowed values of k_n

$$k_n = n\pi/L, \quad n = 1, 2, \ldots \tag{5.87}$$

as before in Section 5.3, this leading to, on using Eq. (5.75),

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{n^2\pi^2\hbar^2}{2mL^2}, \quad n = 1, 2, \ldots \tag{5.88}$$

also as before. Thus we see that the boundary conditions, Eq. (5.73), have the effect of restricting the values of the energy of the particle to those given by Eq. (5.88).

We have to apply the the normalization condition to determine A and B. Up to an inessential phase factor, we find that

$$\psi_n(x) = \begin{cases} \left.\begin{array}{ll} \sqrt{\dfrac{2}{L}}\sin(n\pi x/L) & n = 2, 4, 6\ldots \\ \sqrt{\dfrac{2}{L}}\cos(n\pi x/L) & n = 1, 3, 5, \ldots \end{array}\right\} & |x| < \tfrac{1}{2}L \\ 0 & |x| > \tfrac{1}{2}L. \end{cases} \tag{5.89}$$

Parity

This result illustrates another important feature of the solution to the Schrödinger equation that arises when the potential possesses symmetry. The symmetry here is that

$$V(-x) = V(x) \tag{5.90}$$

i.e. the potential is symmetric about the origin, $x = 0$. An immediate consequence of this is to be seen if we return to the Schrödinger equation, and substitute $V(x) = V(-x)$ to give

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(-x)\psi(x) = E\psi(x). \tag{5.91}$$

If we now make the change of variable $x \to -x$ we get

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(-x)}{dx^2} + V(x)\psi(-x) = E\psi(-x) \tag{5.92}$$

which appears to be the same equation, but now with a solution $\psi(-x)$. However, this cannot be a different solution to $\psi(x)$ except that, at worst

$$\psi(-x) = a\psi(x) \tag{5.93}$$

i.e. the two solutions merely differ by a factor a – it is easy enough to check that if $\psi(x)$ is a solution to the Schrödinger equation, then so is any multiple of it.

But is we now use Eq. (5.93) twice over, i.e. we substitute $x \to -x$ into Eq. (5.93), we get

$$\psi(x) = a\psi(-x) = a^2\psi(x) \tag{5.94}$$

so that $a^2 = 1$ or

$$a = \pm 1 \tag{5.95}$$

and hence we conclude that if $\psi(x)$ is a solution to the Schrödinger equation with a symmetric potential, then either

$$\psi(x) = \psi(-x) \quad \text{or} \quad \psi(-x) = -\psi(x). \tag{5.96}$$

In the first case, the solution is said to be symmetric, or to have even parity, while in the second it is said to be antisymmetric, or to have odd parity.

A quick look at the solutions we have just obtained for the infinite potential well potential illustrates this fact. The solutions $\psi_n(x)$ for n even are sine functions, which are antisymmetric functions, since $\sin(x) = -\sin(x)$, while the solutions for n odd are cosine functions which are symmetric about the origin, since $\cos(x) = \cos(-x)$.

5.5.2 The Finite Potential Well

The infinite potential well is a valuable model since, with the minimum amount of fuss, it shows immediately the way that energy quantization as potentials do not occur in nature. However, for electrons trapped in a block of metal, or gas molecules contained in a bottle, this model serves to describe very accurately the quantum character of such systems. In such cases the potential experienced by an electron as it approaches the edges of a block of metal, or as experienced by a gas molecule as it approaches the walls of its container are effectively infinite

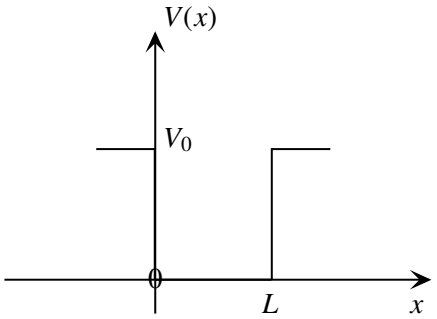


Figure 5.6: Finite potential well.

as far as these particles are concerned, at least if the particles have sufficiently low kinetic energy compared to the height of these potential barriers.

But, of course, any potential well is of finite depth, and if a particle in such a well has an energy comparable to the height of the potential barriers that define the well, there is the prospect of the particle escaping from the well. This is true both classically and quantum mechanically, though, as you might expect, the behaviour in the quantum mechanical case is not necessarily consistent with our classical physics based expectations. Thus we now proceed to look at the quantum properties of a particle in a finite potential well.

In this case, the potential will be of the form

$$V(x) = 0 \qquad 0 < x < L \tag{5.97}$$

$$= V \qquad x \geq L \quad x \leq 0 \tag{5.98}$$

i.e. we have 'lowered' the infinite barriers to a finite value V. We now want to solve the time independent Schrödinger equation for this potential.

To do this, we recognize that the problem can be split up into three parts: $x \leq 0$ where the potential is V, $0 < x < L$ where the potential is zero and $x \geq 0$ where the potential is once again V.

Therefore, to find the wave function for a particle of energy E, we have to solve three equations, one for each of the regions:

$$\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + (E - V)\psi(x) = 0 \quad x \leq 0 \tag{5.99}$$

$$\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + E\psi(x) = 0 \quad 0 < x < L \tag{5.100}$$

$$\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + (E - V)\psi(x) = 0 \quad x \geq L. \tag{5.101}$$

The solutions to these equations take different forms depending on whether $E < V$ or $E > V$. We shall consider the two cases separately.

E < V

First define

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad \text{and} \quad \alpha = \sqrt{\frac{2m(V - E)}{\hbar^2}}. \tag{5.102}$$

Note that, as $V > E$, α will be a real number, as it is square root of a positive number. We can now write these equations as

$$\frac{d^2\psi(x)}{dx^2} - \alpha^2\psi(x) = 0 \quad x \leq 0 \tag{5.103}$$

$$\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \quad 0 < x < L \tag{5.104}$$

$$\frac{d^2\psi(x)}{dx^2} - \alpha^2\psi(x) = 0 \quad x \geq L. \tag{5.105}$$

Now consider the first of these equations, which will have as its solution

$$\psi(x) = Ae^{-\alpha x} + Be^{+\alpha x} \tag{5.106}$$

where A and B are unknown constants. It is at this point that we can make use of our boundary condition, namely that $\psi(x) \to 0$ as $x \to \pm\infty$. In particular, since the solution we are currently looking at applies for $x < 0$, we should look at what this solution does for $x \to -\infty$. What it does is diverge, because of the term $A\exp(-\alpha x)$. So, in order to guarantee that our solution have the correct boundary condition for $x \to -\infty$, we must have $A = 0$. Thus, we conclude that

$$\psi(x) = Be^{\alpha x} \quad x \leq 0. \tag{5.107}$$

We can apply the same kind of argument when solving Eq. (5.105) for $x \geq L$. In that case, the solution is

$$\psi(x) = Ce^{-\alpha x} + De^{\alpha x} \tag{5.108}$$

but now we want to make certain that this solution goes to zero as $x \to \infty$. To guarantee this, we must have $D = 0$, so we conclude that

$$\psi(x) = Ce^{-\alpha x} \quad x \geq L. \tag{5.109}$$

Finally, at least for this part of the argument, we look at the region $0 < x < L$. The solution of Eq. (5.104) for this region will be

$$\psi(x) = P\cos(kx) + Q\sin(kx) \quad 0 < x < L \tag{5.110}$$

but now we have no diverging exponentials, so we have to use other means to determine the unknown coefficients P and Q.

At this point we note that we still have four unknown constants B, P, Q, and C. To determine these we note that the three contributions to $\psi(x)$ do not necessarily join together smoothly at $x = 0$ and $x = L$. This awkward state of affairs has its origins in the fact that the potential is discontinuous at $x = 0$ and $x = L$ which meant that we had to solve three separate equations for the three different regions. But these three separate solutions cannot be independent of one another, i.e. there must be a relationship between the unknown constants, so there must be other conditions that enable us to specify these constants. The extra conditions that we impose, as discussed in Section 5.4.3, are that the wave function has to be a continuous function, i.e. the three solutions:

$$\psi(x) = Be^{\alpha x} \qquad x \leq 0 \tag{5.111}$$

$$= P\cos(kx) + Q\sin(kx) \quad 0 < x < L \tag{5.112}$$

$$= Ce^{-\alpha x} \qquad x \geq L. \tag{5.113}$$

should all 'join up' smoothly at $x = 0$ and $x = L$. This means that the first two solutions and their slopes (i.e. their first derivatives) must be the same at $x = 0$, while the second and third solutions and their derivatives must be the same at $x = L$. Applying this requirement at $x = 0$ gives:

$$B = P \tag{5.114}$$

$$\alpha B = kQ \tag{5.115}$$

and then at $x = L$:

$$P\cos(kL) + Q\sin(kL) = Ce^{-\alpha L} \tag{5.116}$$

$$-kP\sin(kL) + kQ\cos(kL) = -\alpha Ce^{-\alpha L}. \tag{5.117}$$

If we eliminate B and C from these two sets of equations we get, in matrix form:

$$\begin{pmatrix} \alpha & -k \\ \alpha\cos(kL) - k\sin(kL) & \alpha\sin(kL) + k\cos(kL) \end{pmatrix} \begin{pmatrix} P \\ Q \end{pmatrix} = 0 \tag{5.118}$$

and in order that we get a non-trivial solution to this pair of homogeneous equations, the determinant of the coefficients must vanish:

$$\begin{vmatrix} \alpha & -k \\ \alpha\cos(kL) - k\sin(kL) & \alpha\sin(kL) + k\cos(kL) \end{vmatrix} = 0 \tag{5.119}$$

which becomes, after expanding the determinant and rearranging terms:

$$\tan(kL) = \frac{2\alpha k}{k^2 - \alpha^2}. \tag{5.120}$$

Solving this equation for k will give the allowed values of k for the particle in this finite potential well, and hence, using Eq. (5.102) in the form

$$E = \frac{\hbar^2 k^2}{2m} \tag{5.121}$$

we can determine the allowed values of energy for this particle. What we find is that these allowed energies are finite in number, in contrast to the infinite potential well, but to show this we must solve this equation. This is made difficult to do analytically by the fact that this is a transcendental equation – it has no solutions in terms of familiar functions. However, it is possible to get an idea of what its solutions look like either numerically, or graphically. The latter has some advantages as it allows us to see how the mathematics conspires to produce the quantized energy levels. We can first of all simplify the mathematics a little by writing Eq. (5.120) in the form

$$\tan(kL) = \frac{2(\alpha/k)}{1 - (\alpha/k)^2} \tag{5.122}$$

which, by comparison with the two trigonometric formulae

$$\tan 2\theta = \frac{2\tan\theta}{1-\tan^2\theta}$$
$$\tan 2\theta = \frac{2\cot(-\theta)}{1-\cot^2(-\theta)}$$

we see that Eq. (5.120) is equivalent to the two conditions

$$k\tan(\tfrac{1}{2}kL) = \alpha \tag{5.123}$$
$$k\cot(-\tfrac{1}{2}kL) = -k\cot(\tfrac{1}{2}kL) = \alpha. \tag{5.124}$$

The aim here is to plot the left and right hand sides of these two expressions as a function of k (well, actually as a function of $\frac{1}{2}kL$), but before we can do that we need to take account of the fact that the quantity α is given in terms of E by $\alpha = \sqrt{2m(V-E)/\hbar^2}$, and hence, since $E = \hbar^2k^2/2m$, we have

$$\alpha = \sqrt{k_0^2 - k^2}$$

where

$$k_0 = \sqrt{\frac{2mV}{\hbar^2}}. \tag{5.125}$$

As we will be plotting as a function of $\frac{1}{2}kL$, it is useful to define the new dimensionless variables

$$u = \tfrac{1}{2}kL \quad \text{and} \quad u_0 = \tfrac{1}{2}k_0L$$

to rewrite the above expressions as

$$u\tan u = \sqrt{u_0^2 - u^2} \quad \text{and} \quad -u\cot u = \sqrt{u_0^2 - u^2}. \tag{5.126}$$

We can now plot $u\tan u$, $-u\cot u$ and $(u_0^2 - u^2)^{1/2}$ as functions of u for various values for u_0, i.e. different values of V yield different values of u_0 via

$$V = \frac{2\hbar^2u_0^2}{mL^2}. \tag{5.127}$$

This is illustrated in Fig. (5.7) where five such plots are given for different values of u_0. The points of intersection of the quarter-circles $(u_0^2 - u^2)^{1/2}$ with the $u\tan u$ and $u\cot u$ curves give the u_n values for an allowed energy level of the particle in this potential. The important feature of these curves is that the number of points of intersection is finite, i.e. there are only a finite number of values of k that solve Eq. (5.120). Correspondingly, there will only be a finite number of allowed values of E for the particle for each value of V, and these energies will be given by

$$E_n = \frac{2\hbar^2u_n^2}{mL^2}. \tag{5.128}$$

There will always be at least one allowed value.

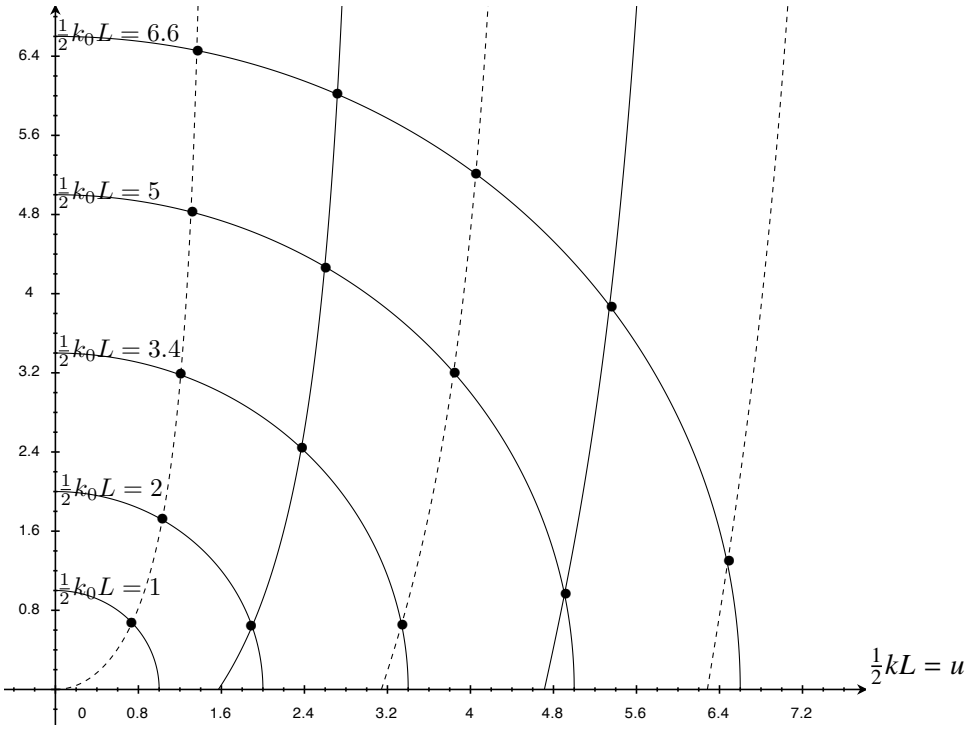


Figure 5.7: Graph to determine bound states of a finite well potential. The $u \tan u$ functions are dashed lines, which alternate with the $u \cot u$ functions which are solid lines. The points of intersection are solutions to Eq. (5.120). The quarter-circles are for increasing values of V, starting with V lowest such that $\frac{1}{2}k_0L = 1$, for which there is only one bound state, slightly higher at $\frac{1}{2}k_0L = 2$, for which there are two bound states, slightly higher again for $\frac{1}{2}k_0L = 3.4$ where there are three bound states, then $\frac{1}{2}k_0L = 5$ for which there is four bound states, and so on.

To determine the corresponding wave functions is a straightforward but complicated task. The first step is to show, by using Eq. (5.120) and the equations for B, C, P and Q that

$$C = e^{\alpha L}B \tag{5.129}$$

from which readily follows the solution

$$\psi(x) = Be^{\alpha x} \qquad x \le 0 \tag{5.130}$$

$$= B(\cos kx + \frac{\alpha}{k}\sin kx) \quad 0 < x < L \tag{5.131}$$

$$= Be^{-\alpha(x-L)} \qquad x \ge L. \tag{5.132}$$

The constant B is determined by the requirement that $\psi(x)$ be normalized, i.e. that

$$\int_{-\infty}^{+\infty} |\psi(x)|^2\, dx = 1. \tag{5.133}$$

which becomes:

$$|B|^2\left[\int_{-\infty}^{0} e^{-2\alpha x}\, dx + \int_0^L (\cos kx + \frac{\alpha}{k}\sin kx)^2\, dx + \int_L^{+\infty} e^{-2\alpha(x-L)}\, dx\right] = 1. \tag{5.134}$$

After a somewhat tedious calculation that makes liberal use of Eq. (5.120), the result found is that

$$B = \frac{k}{k_0}\sqrt{\frac{\alpha}{\frac{1}{2}\alpha L + 1}}. \tag{5.135}$$

The task of determining the wave functions is then that of determining the allowed values of k from the graphical solution, or numerically, and then substituting those values into the above expressions for the wave function. The wave functions found are similar in appearance to the infinite well wave functions, with the big difference that they are non-zero outside the well. This is true even if the particle has the lowest allowed energy, i.e. there is a non-zero probability of finding the particle outside the well. This probability can be readily calculated, being just

$$P_{outside} = |B|^2\left[\int_{-\infty}^{0} e^{-2\alpha x}\,dx + \int_{L}^{+\infty} e^{-2\alpha(x-L)}\,dx\right] = \alpha^{-1}|B|^2 \tag{5.136}$$

5.5.3 Scattering from a Potential Barrier

The above examples are of *bound states*, i.e. wherein the particles are confined to a limited region of space by some kind of attractive or confining potential. However, not all potentials are attractive (e.g. two like charges repel), and in any case, even when there is an attractive potential acting (two opposite charges attracting), it is possible that the particle can be 'free' in the sense that it is not confined to a limited region of space. A simple example of this, classically, is that of a comet orbiting around the sun. It is possible for the comet to follow an orbit in which it initially moves towards the sun, then around the sun, and then heads off into deep space, never to return. This is an example of an unbound orbit, in contrast to the orbits of comets that return repeatedly, though sometimes very infrequently, such as Halley's comet. Of course, the orbiting planets are also in bound states.

A comet behaving in the way just described – coming in from infinity and then ultimately heading off to infinity after bending around the sun – is an example of what is known as a scattering process. In this case, the potential is attractive, so we have the possibility of both scattering occurring, as well as the comet being confined to a closed orbit – a bound state. If the potential was repulsive, then only scattering would occur.

The same distinction applies in quantum mechanics. It is possible for the particle to confined to a limited region in space, in which case the wave function must satisfy the boundary condition that

$$\psi(x) \to 0 \qquad \text{as} \qquad x \to \pm\infty.$$

As we have seen, this boundary condition is enough to yield the quantization of energy. However, in the quantum analogue of scattering, it turns out that energy is not quantized. This in part can be linked to the fact that the wave function that describes the scattering of a particle of a given energy does not decrease as $x \to \pm\infty$, so that the very thing that leads to the quantization of energy for a bound particle does not apply here.

This raises the question of what to do about the quantization condition, i.e. that

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2 dx = 1.$$

If the wave function does not go to zero as $x \to \pm\infty$, then it is not possible for the wave function to satisfy this normalization condition – the integral will always diverge. So how are we to maintain the claim that the wave function must have a probability interpretation if one of the principal requirements, the normalization condition, does not hold true? Strictly speaking, a wave function that cannot be normalized to unity is not physically permitted (because it is inconsistent with the probability interpretation of the wave function). Nevertheless, it is possible to retain, and work with, such wave functions, provided a little care is taken. The answer lies in interpreting the wave

function so that $|\Psi(x,t)|^2 \propto$ particle flux[5], though we will not be developing this idea to any extent here.

To illustrate the sort of behaviour that we find with particle scattering, we will consider a simple, but important case, which is that of a particle scattered by a potential barrier. This is sufficient to show the sort of things that can happen that agree with our classical intuition, but it also enables us to see that there occurs new kinds of phenomena that have no explanation within classical physics.

Thus, we will investigate the scattering problem of a particle of energy E interacting with a potential $V(x)$ given by:

$$\begin{aligned} V(x) &= 0 \quad x < 0 \\ V(x) &= V_0 \quad x > 0. \end{aligned} \tag{5.137}$$

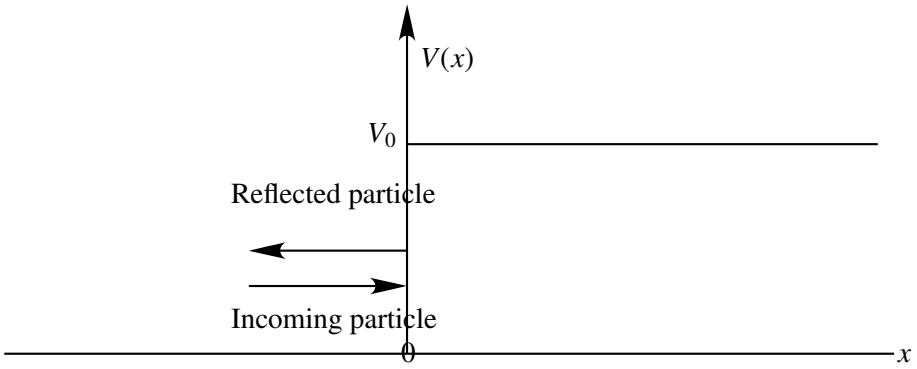


Figure 5.8: Potential barrier with particle of energy $E < V_0$ incident from the left. Classically, the particle will be reflected from the barrier.

In Fig. (5.8) is illustrated what we would expect to happen if a classical particle of energy $E < V_0$ were incident on the barrier: it would simply bounce back as it has insufficient energy to cross over to $x > 0$. Quantum mechanically we find that the situation is not so simple.

Given that the potential is as given above, the Schrödinger equation comes in two parts:

$$\begin{aligned} -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} &= E\psi \quad x < 0 \\ -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V_0\psi &= E\psi \quad x > 0 \end{aligned} \tag{5.138}$$

where E is, once again, the total energy of the particle.

We can rewrite these equations in the following way:

$$\begin{aligned} \frac{d^2\psi}{dx^2} + \frac{2mE}{\hbar^2}\psi &= 0 \quad x < 0 \\ \frac{d^2\psi}{dx^2} - \frac{2m}{\hbar^2}\Big(V_0 - E\Big)\psi &= 0 \quad x > 0 \end{aligned} \tag{5.139}$$

If we put

$$k = \frac{\sqrt{2mE}}{\hbar} \tag{5.140}$$

[5]In more advanced treatments, it is found that the usual probability interpretation does, in fact, continue to apply, though the particle is described not by a wave function corresponding to a definite energy, but rather by a wave packet, though then the particle does not have a definite energy.

then the first equation becomes

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0 \quad x < 0$$

which has the general solution

$$\psi = Ae^{ikx} + Be^{-ikx} \tag{5.141}$$

where A and B are unknown constants. We can get an idea of what this solution means if we reintroduce the time dependence (with $\omega = E/\hbar$):

$$\begin{aligned} \Psi(x,t) =& \psi(x)e^{-iEt/\hbar} &=& \psi(x)e^{-i\omega t} \\ =& Ae^{i(kx-\omega t)} &+& Be^{-i(kx+\omega t)} \\ =& \text{wave travelling to right} &+& \text{wave travelling to left} \end{aligned} \tag{5.142}$$

i.e. this solution represents a wave associated with the particle heading towards the barrier and a reflected wave associated with the particle heading away from the barrier. Later we will see that these two waves have the same amplitude, implying that the particle is perfectly reflected at the barrier.

In the region $x > 0$, we write

$$\alpha = \sqrt{2m(V_0 - E)}/\hbar > 0 \tag{5.143}$$

so that the Schrödinger equation becomes

$$\frac{d^2\psi}{dx^2} - \alpha^2\psi = 0 \quad x > 0 \tag{5.144}$$

which has the solution

$$\psi = Ce^{-\alpha x} + De^{\alpha x} \tag{5.145}$$

where C and D are also unknown constants.

The problem here is that the $\exp(\alpha x)$ solution grows exponentially with x, and we do not want wave functions that become infinite: it would essentially mean that the particle is forever to be found at $x = \infty$, which does not make physical sense. So we must put $D = 0$. Thus, if we put together our two solutions for $x < 0$ and $x > 0$, we have

$$\begin{aligned} \psi =& Ae^{ikx} + Be^{-ikx} \quad x < 0 \\ =& Ce^{-\alpha x} \quad x > 0. \end{aligned} \tag{5.146}$$

If we reintroduce the time dependent factor, we get

$$\Psi(x,t) = \psi(x)e^{-i\omega t} = Ce^{-\alpha x}e^{-i\omega t} \tag{5.147}$$

which is *not* a travelling wave at all. It is a stationary wave that simply diminishes in amplitude for increasing x.

We still need to determine the constants A, B, and C. To do this we note that for arbitrary choice of these coefficients, the wave function will be discontinuous at $x = 0$. For reasons connected with the requirement that probability interpretation of the wave function continue to make physical sense, we will require that the wave function *and* its first derivative both be continuous[6] at $x = 0$.

[6]To properly justify these conditions requires introducing the notion of 'probability flux', that is the rate of flow of probability carried by the wave function. The flux must be such that the point $x = 0$, where the potential is discontinuous, does not act as a 'source' or 'sink' of probability. What this means, as is shown later, is that we end up with $|A| = |B|$, i.e. the amplitude of the wave arriving at the barrier is the same as the amplitude of the wave reflected at the barrier. If they were different, it would mean that there is more probability flowing into the barrier than is flowing out (or vice versa) which does not make physical sense.

These conditions yield the two equations

$$\begin{aligned} C =& A + B \\ -\alpha C =& ik(A - B) \end{aligned} \tag{5.148}$$

which can be solved to give

$$\begin{aligned} B =& \frac{ik + a}{ik - a} A \\ C =& \frac{2ik}{ik - \alpha} A \end{aligned} \tag{5.149}$$

and hence

$$\begin{aligned} \psi(x) =& Ae^{ikx} + \frac{ik + a}{ik - a} Ae^{-ikx} \quad x < 0 \\ =& \frac{2ik}{ik - \alpha} Ae^{-\alpha x} \quad x < 0. \end{aligned} \tag{5.150}$$

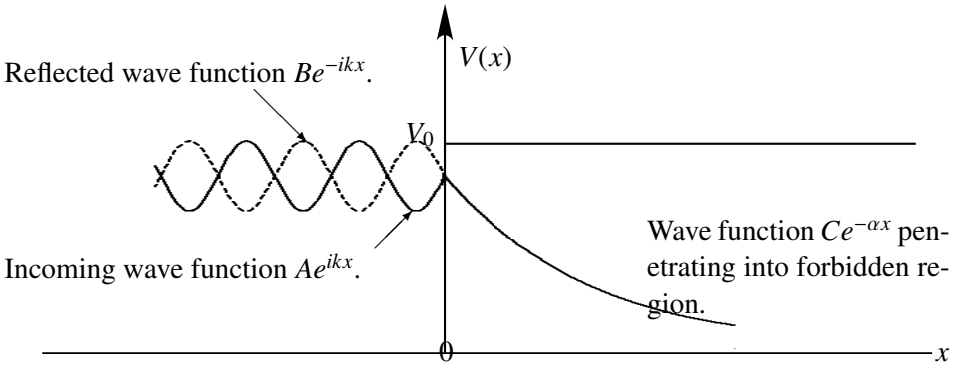


Figure 5.9: Potential barrier with wave function of particle of energy $E < V_0$ incident from the left (solid curve) and reflected wave function (dotted curve) of particle bouncing off barrier. In the clasically 'forbidden' region of $x > 0$ there is a decaying wave function. Note that the complex wave functions have been represented by their real parts.

Having obtained the mathematical solution, what we need to do is provide a physical interpretation of the result.

First we note that we cannot impose the normalization condition as the wave function does not decrease to zero as $x \to -\infty$. But, in keeping with comments made above, we can still learn something from this solution about the behaviour of the particle.

Secondly, we note that the incident and reflected waves have the same 'intensity'

$$\begin{aligned} \text{Incident intensity} =& |A|^2 \\ \text{Reflected intensity} =& |A|^2 \left| \frac{ik + \alpha}{ik - \alpha} \right|^2 = |A|^2 \end{aligned} \tag{5.151}$$

and hence they have the same amplitude. This suggests that the incident de Broglie wave is totally reflected, i.e. that the particle merely travels towards the barrier where it 'bounces off', as would be expected classically. However, if we look at the wave function for $x > 0$ we find that

$$\begin{aligned} |\psi(x)|^2 \propto& \left| \frac{2ik}{ik - \alpha} \right|^2 e^{-2\alpha x} \\ =& \frac{4k^2}{\alpha^2 + k^2} e^{-2\alpha x} \end{aligned} \tag{5.152}$$

which is an exponentially decreasing probability.

This last result tells us that there is a non-zero probability of finding the particle in the region $x > 0$ where, classically, the particle has no chance of ever reaching. The distance that the particle can penetrate into this 'forbidden' region is given roughly by $1/2\alpha$ which, for a subatomic particle can be a few nanometers, while for a macroscopic particle, this distance is immeasurably small.

The way to interpret this result is along the following lines. If we imagine that a particle is fired at the barrier, and we are waiting a short distance on the far side of the barrier in the forbidden region with a 'catcher's mitt' poised to grab the particle then we find that *either* the particle hits the barrier and bounces off with the same energy as it arrived with, but with the opposite momentum – it never lands in the mitt, *or* it lands in the mitt and we catch it – it does not bounce off in the opposite direction. The chances of the latter occurring are generally very tiny, but it occurs often enough in microscopic systems that it is a phenomenon that is exploited, particularly in solid state devices. Typically this is done, not with a single barrier, but with a barrier of finite width, in which case the particle can penetrate through the barrier and reappear on the far side, in a process known as quantum tunnelling.

5.6 Expectation Value of Momentum

We can make use of Schrödinger's equation to obtain an alternative expression for the expectation value of momentum given earlier in Eq. (5.15). This expression is

$$\langle p\rangle = m\langle v(t)\rangle = m\int_{-\infty}^{+\infty} x\left[\frac{\partial\Psi^*(x,t)}{\partial t}\Psi(x,t) + \Psi^*(x,t)\frac{\partial\Psi(x,t)}{\partial t}\right]dx. \tag{5.153}$$

We note that the appearance of time derivatives in this expression. If we multiply both sides by $i\hbar$ and make use of Schrödinger's equation, we can substitute for these time derivatives to give

$$\begin{aligned} i\hbar\langle p\rangle =& m\int_{-\infty}^{+\infty} x\left[\left\{\frac{\hbar^2}{2m}\frac{\partial^2\Psi^*(x,t)}{\partial x^2} - V(x)\Psi^*(x,t)\right\}\Psi(x,t)\right. & (5.154)\\ &\left. + \Psi^*(x,t)\left\{-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + V(x)\Psi(x,t)\right\}\right]dx. & (5.155)\end{aligned}$$

The terms involving the potential cancel. The common factor $\hbar^2/2m$ can be moved outside the integral, while both sides of the equation can be divided through by $i\hbar$, yielding a slightly less complicated expression for $\langle p\rangle$:

$$\langle p\rangle = -\tfrac{1}{2}i\hbar\int_{-\infty}^{+\infty} x\left[\frac{\partial^2\Psi^*(x,t)}{\partial x^2}\Psi(x,t) - \Psi^*(x,t)\frac{\partial^2\Psi(x,t)}{\partial x^2}\right]dx. \tag{5.156}$$

Integrating both terms in the integrand by parts then gives

$$\begin{aligned}\langle p\rangle =& \tfrac{1}{2}i\hbar\int_{-\infty}^{+\infty}\left[\frac{\partial\Psi^*(x,t)}{\partial x}\frac{\partial x\Psi(x,t)}{\partial x} - \frac{\partial x\Psi^*(x,t)}{\partial x}\frac{\partial\Psi(x,t)}{\partial x}\right]dx\\ &+ \tfrac{1}{2}i\hbar\left[\frac{\partial\Psi^*(x,t)}{\partial x}\Psi(x,t) - \Psi^*(x,t)\frac{\partial\Psi(x,t)}{\partial x}\right]_{-\infty}^{+\infty} \end{aligned}\tag{5.157}$$

As the wave function vanishes for $x \to \pm\infty$, the final term here will vanish. Carrying out the derivatives in the integrand then gives

$$\langle p\rangle = \tfrac{1}{2}i\hbar\int_{-\infty}^{+\infty}\left[\frac{\partial\Psi^*(x,t)}{\partial x}\Psi(x,t) - \Psi^*(x,t)\frac{\partial\Psi(x,t)}{\partial x}\right]dx \tag{5.158}$$

Integrating the first term only by parts once again then gives

$$\langle p\rangle = -i\hbar \int_{-\infty}^{+\infty} \Psi^*(x,t)\frac{\partial \Psi(x,t)}{\partial x}dx + \tfrac{1}{2}i\hbar\Psi^*(x,t)\Psi(x,t)\Big|_{-\infty}^{+\infty}. \tag{5.159}$$

Once again, the last term here will vanish as the wave function itself vanishes for $x \to \pm\infty$ and we are left with

$$\langle p\rangle = -i\hbar \int_{-\infty}^{+\infty} \Psi^*(x,t)\frac{\partial \Psi(x,t)}{\partial x}dx. \tag{5.160}$$

This is a particularly significant result as it shows that the expectation value of momentum can be determined directly from the wave function – i.e. information on the momentum of the particle is contained within the wave function, along with information on the position of the particle. This calculation suggests making the identification

$$p \to -i\hbar\frac{\partial}{\partial x} \tag{5.161}$$

which further suggests that we can make the replacement

$$p^n \to \Big(-i\hbar\frac{\partial}{\partial x}\Big)^n \tag{5.162}$$

so that, for instance

$$\langle p^2\rangle = -\hbar^2 \int_{-\infty}^{+\infty} \Psi^*(x,t)\frac{\partial^2 \Psi(x,t}{\partial x^2}dx \tag{5.163}$$

and hence the expectation value of the kinetic energy of the particle is

$$\langle K\rangle = \frac{\langle p^2\rangle}{2m} = -\frac{\hbar^2}{2m}\int_{-\infty}^{+\infty} \Psi^*(x,t)\frac{\partial^2 \Psi(x,t}{\partial x^2}dx. \tag{5.164}$$

We can check this idea by turning to the classical formula for the total energy of a particle

$$\frac{p^2}{2m} + V(x) = E. \tag{5.165}$$

If we multiply this equation by $\Psi(x,t) = \psi(x)\exp(-iEt/\hbar)$ and make the replacement given in Eq. (5.163) we end up with

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \tag{5.166}$$

which is just the time independent Schrödinger equation. So there is some truth in the ad hoc procedure outlined above.

This association of the physical quantity p with the derivative i.e. is an example of a physical observable, in this case momentum, being represented by a differential operator. This correspondence between physical observables and operators is to be found throughout quantum mechanics. In the simplest case of position, the operator corresponding to position x is itself just x, so there is no subtleties in this case, but as we have just seen this simple state of affairs changes substantially for other observables. Thus, for instance, the observable quantity K, the kinetic energy, is represented by the differential operator

$$K \to \hat{K} = -\hbar^2\frac{\partial^2}{\partial x^2}. \tag{5.167}$$

while the operator associated with the position of the particle is $\hat{x}$ with

$$x \to \hat{x} = x. \tag{5.168}$$

In this last case, the identification is trivial.

5.7 Is the wave function all that is needed?

So far, we have been concerned almost entirely with the quantum mechanics associated with the motion of particles through space, which may give rise to the mistaken notion that all of quantum mechanics revolves around the idea of the wave function. However, this is far from the case. There is no objective reality that can be attached to the wave function – it is just one of many ways that the quantum nature of matter can be expressed. It is useful for many kinds of problems, but entirely inappropriate for others, and a singular focus on the wave function can give the impression that the wave function *is* the heart of quantum mechanics. But the quantum mechanical nature of matter makes itself known in ways that have little to do with waves and wave functions. Correspondingly, the mathematical theory of quantum mechanics is much more than just the mathematics of waves.

But what are the other aspects of quantum mechanics that need more than wave mechanics? For one thing, it can be shown there is no such thing as a wave function for a photon. There are many physical processes in which the physically observable quantity is not where a particle is, or how fast it is moving, but rather how many particles there are of a certain kind – an important consideration as particles can be created or destroyed in a random way. The whole subject of the quantum theory of fields is expressed in the language of particle numbers. But one of the most important examples of a circumstance in which wave mechanics falls short is in describing the intrinsic spin of elementary particles. In the following Chapter, attention will be focussed on the example of particle spin and the Stern-Gerlach experiment as being a way of illustrating some of the general principles of quantum mechanics.

Chapter 6

Particle Spin and the Stern-Gerlach Experiment

The spin of an elementary particle would appear, on the surface, to be little different from the spin of a macroscopic object – the image of a microscopic sphere spinning around some axis comes to mind. However, there is far more going on here than what this simple picture might suggest. But first, a brief look at what the classical properties are of angular momentum is needed.

6.1 Classical Spin Angular Momentum

A particle moving through space possesses angular momentum, a vector, defined by

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \tag{6.1}$$

where $\mathbf{r}$ and $\mathbf{p}$ are the position vector and momentum respectively of the particle. This is sometimes referred to as *orbital* angular momentum since, in particular, it is an important consideration in describing the properties of a particle orbiting around some centre of attraction such as, in the classical picture of an atom, electrons orbiting around an atomic nucleus. Classically there is no restriction on the magnitude or direction of orbital angular momentum.

From a classical perspective, as an electron carries a charge, its orbital motion will result in a tiny current loop which will produce a dipolar magnetic field. The strength of this dipole field is measured by the magnetic moment $\boldsymbol{\mu}$ which is related to the orbital angular momentum by

$$\boldsymbol{\mu}_{\boldsymbol{L}} = \frac{q}{2m}\mathbf{L}. \tag{6.2}$$

Thus, the expectation on the basis of this classical picture is that atoms can behave as tiny little magnets.

The classical idea of spin follows directly from the above considerations. Spin is the angular momentum we associate with a rotating object such as a spinning golf ball, or the spinning Earth. The angular momentum of such a body can be calculated by integrating over the contributions to the angular momentum due to the motion of each of the infinitesimal masses making up the body. The well known result is that the total angular momentum or spin $\mathbf{S}$ is given by

$$\mathbf{S} = I\boldsymbol{\omega} \tag{6.3}$$

where I is the moment of inertia of the body, and $\boldsymbol{\omega}$ is its angular velocity. Spin is a vector which points along the axis of rotation in a direction determined by the right hand rule: curl the fingers of the right hand in the direction of rotation and the thumb points in the direction of $\mathbf{S}$. The moment of inertia is determined by the distribution of mass in the rotating body relative to the axis of rotation. If the object were a solid uniform sphere of mass m and radius a, and rotation were about a diameter of the sphere, then the moment of inertia can be shown to be

$$I = \tfrac{2}{5}Ma^2. \tag{6.4}$$

If the sphere possesses an electric charge, then the circulation of the charge around the axis of rotation will constitute a current and hence will give rise to a magnetic field. This field is a dipole field whose strength is measured by the dipole moment which can be shown, for a uniformly charged sphere of total charge q, to be given by

$$\boldsymbol{\mu_S} = \frac{q}{2m}\mathbf{S}, \tag{6.5}$$

exactly the same as in the orbital case.

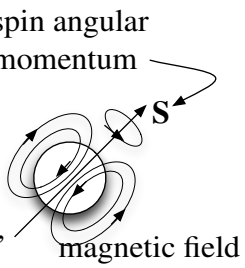


Figure 6.1: Magnetic field produced by a spinning charged sphere.

The point to be made here is that the spinning object is extended in space, i.e. the spinning sphere example has a non-zero radius. If we try to extend the idea to a *point* particle by taking the limit of $a \to 0$ we immediately see that the spin angular momentum must vanish unless ω is allowed to be infinitely large. If we exclude this last possibility, then classically a point particle can only have a spin angular momentum of zero and so it cannot have a magnetic moment. Thus, from the point-of-view of classical physics, elementary particles such as an electron, which are known to possess spin angular momentum, cannot be viewed as point objects – they must be considered as tiny spinning spheres. But as far as it has been possible to determine by high energy scattering experiments, elementary particles such as the electron behave very much as point particles. Whatever radius they might have, it is certainly very tiny: experiment suggests it is $< 10^{-17}$ m. Yet they are found to possess spin angular momentum of a magnitude equal (for the electron) to $\sqrt{3}/\hbar/2$ which requires the surface of the particle to be moving at a speed greater than that of light. This conflict with special relativity makes this classical picture of an elementary particle as a tiny, rapidly rotating sphere obviously untenable. The resolution of this problem can be found within quantum mechanics, though this requires considering the relativistic version of quantum mechanics: the spin of a point particle is identified as a relativistic effect. We shall be making use of what quantum mechanics tells us about particle spin, though we will not be looking at its relativistic underpinnings. On thing we do learn, however, is that spin is not describable in terms of the wave function idea that we have been working with up till now.

6.2 Quantum Spin Angular Momentum

Wave mechanics and the wave function describe the properties of a particle moving through space, giving, as we have seen, information on its position, momentum, energy. In addition it also provides, via the quantum mechanical version of $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ a quantum description of the *orbital* angular momentum of a particle, such as that associated with an electron moving in an orbit around an atomic nucleus. The general results found are that the magnitude of the angular momentum is limited to the values

$$L = \sqrt{l(l+1)}\hbar, \qquad l = 0, 1, 2, 3, \ldots, \tag{6.6}$$

which can be looked on as an 'improved' version of the result used by Bohr, the one subsequently 'justified' by the de Broglie hypothesis, that is $L = n\hbar$, Eq. (2.5). The quantum theory of orbital angular momentum also tells us that any one vector component of $\mathbf{L}$, L_z say, is restricted to the values

$$L_z = m_l\hbar, \qquad m_l = -l, -l+1, -l+2, \ldots l-2, l-1, l. \tag{6.7}$$

This restriction on the possible values of L_z mean that the angular momentum vector can have only certain orientations in space – a result known as 'space quantization'.

All this is built around the quantum mechanical version of $\mathbf{L} = \mathbf{r}\times\mathbf{p}$, and so implicitly is concerned with the angular momentum of a particle moving through space. But a more general perspective yields some surprises. If special relativity and quantum mechanics are combined, it is found that

even if a particle, a point object, has zero momentum, so that the orbital angular momentum is zero, its total angular momentum is, in general, *not zero*. The only interpretation that can be offered is that this angular momentum is due to the intrinsic spin of the particle. The possible values for the magnitude S of the spin angular momentum turn out to be

$$S = \sqrt{s(s+1)}\hbar, \qquad s = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2\ldots, \tag{6.8}$$

and any one vector component of $\mathbf{S}$, S_z say, is restricted to the values

$$S_z = m_l\hbar, \qquad m_s = -s, -s+1, -s+2, \ldots s-2, s-1, s \tag{6.9}$$

i.e. similar to orbital angular momentum, but with the significant difference of the appearance of half integer values for the spin quantum number s in addition to the integer values. This theoretical result is confirmed by experiment. In nature there exist elementary particles for which $s = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}\ldots$ such as the electron, proton, neutron, quark (all of which have spin $s = \frac{1}{2}$), and more exotic particles of higher half-integer spin, while there exist many particles with integer spin, the photon, for which $s = 1$, being the most well-known example, though because it is a zero rest mass particle, it turns out that S_z can only have the values ± 1. Of particular interest here is the case of $s = \frac{1}{2}$ for which there are two possible values for S_z, that is $S_z = \pm\frac{1}{2}\hbar$.

Particle spin is what is left after the contribution to the angular momentum due to motion through space has been removed. It is angular momentum associated with the internal degrees of freedom of a point particle, whatever they may be, and cannot be described mathematically in terms of a wave function. It also has no classical analogue: we have already seen that a point particle cannot have spin angular momentum. Thus, particle spin is a truly quantum property that cannot be described in the language of wave functions – a more general mathematical language is required. It was in fact the discovery of particle spin, in particular the spin of the electron, that lead to the development of a more general version of quantum mechanics than that implied by wave mechanics.

There is one classical property of angular momentum that does carry over to quantum mechanics. If the particle is charged, and if it possesses either orbital or spin angular momentum, then there arises a dipole magnetic field. In the case of the electron, the dipole moment is found to be given by

$$\boldsymbol{\mu_S} = -\frac{e}{2m_e} g\mathbf{S} \tag{6.10}$$

where m_e and $-e$ are the mass and charge of the electron, $\mathbf{S}$ is the spin angular momentum of the electron, and g is the so-called gyromagnetic ratio, which classically is exactly equal to one, but is known (both from measurement and as derived from relativistic quantum mechanics) to be approximately equal to two for an electron. It is the fact that electrons possess a magnetic moment that has made it possible to perform experiments involving the spin of electrons, in a way that reveals the intrinsically quantum properties of spin.

6.3 The Stern-Gerlach Experiment

This experiment, first performed in 1922, has long been considered as the quintessential experiment that illustrates the fact that the electron possesses intrinsic angular momentum, i.e. spin. It is actually the case that the original experiment had nothing to do with the discovery that the electron possessed spin: the first proposal concerning the spin of the electron, made in 1925 by Uhlenbach and Goudsmit, was based on the analysis of atomic spectra. What the experiment was intended to test was 'space-quantization' associated with the orbital angular momentum of atomic electrons. The prediction, already made by the 'old' quantum theory that developed out of Bohr's work, was that the spatial components of angular momentum could only take discrete values, so

that the direction of the angular momentum vector was restricted to only a limited number of possibilities, and this could be tested by making use of the fact that an orbiting electron will give rise to a magnetic moment proportional to the orbital angular momentum of the electron. So, by measuring the magnetic moment of an atom, it should be possible to determine whether or not space quantization existed. In fact, the results of the experiment were in agreement with the then existing (incorrect) quantum theory – the existence of electron spin was not at that time suspected. Later, it was realized that the interpretation of the results of the experiment were incorrect, and that what was seen in the experiment was direct evidence that electrons possess spin. It is in this way that the Stern-Gerlach experiment has subsequently been used, i.e. to illustrate the fact that electrons have spin. But it is also valuable in another way. The simplicity of the results of the experiment (only two possible outcomes), and the fact that the experiment produces results that are directly evidence of the laws of quantum mechanics in action makes it an ideal means by which the essential features of quantum mechanics can be seen and, perhaps, 'understood'.

The original experimental arrangement took the form of a collimated beam of silver atoms heading in, say, the y direction, and passing through a non-uniform magnetic field directed (mostly) in the z-direction. Assuming the silver atoms possess a non-zero magnetic moment $\boldsymbol{\mu}$, the magnetic field will have two effects. First, the magnetic field will exert a torque on the magnetic dipole, so that the magnetic moment vector will *precess* about the direction of the magnetic field. This will not affect the z component of $\boldsymbol{\mu}$, but the x and y components of $\boldsymbol{\mu}$ will change with time. Secondly, and more importantly here, the non-uniformity of the field means that the atoms experience a sideways force given by

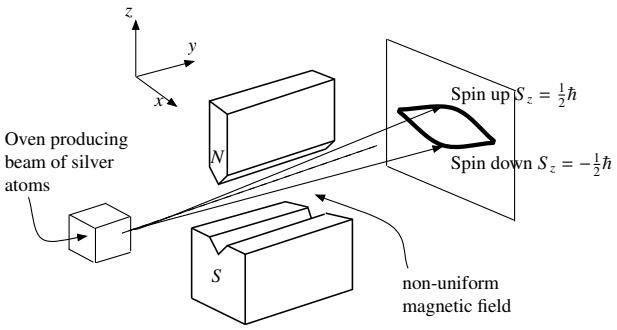


Figure 6.2: The Stern-Gerlach apparatus.

$$F_z = -\frac{\partial U}{\partial z} \tag{6.11}$$

where $U = -\boldsymbol{\mu} \cdot \mathbf{B} = -\mu_z B$ is the potential energy of the silver atom in the magnetic field. Thus

$$F_z = \mu_z \frac{\partial B}{\partial z}. \tag{6.12}$$

Different orientations of the magnetic moment vector $\boldsymbol{\mu}$ will lead to different values of μ_z, which in turn will mean that there will be forces acting on the atoms which will differ depending on the value of μ_z.

The expectation based on classical physics is that due to random thermal effects in the oven, the magnetic dipole moment vectors of the atoms will be randomly oriented in space, so there should be a continuous spread in the z component of the magnetic moments of the silver atoms as they emerge from the oven, ranging from $-|\mu_z|$ to $|\mu_z|$. A line should then appear on the observation screen along the z direction. Instead, what was found was that the silver atoms arrived on the screen at only two points that corresponded to magnetic moments of

$$\mu_z = \pm\mu_B; \qquad \mu_B = \frac{e\hbar}{2m_e} \tag{6.13}$$

where μ_B is known as the Bohr magneton.

Space quantization was clearly confirmed by this experiment, but the full significance of their results was not realized until some time later, after the proposal by Uhlenbach and Goudsmit that the

electron possessed intrinsic spin, and a magnetic moment. The full explanation based on what is now known about the structure of the silver atom is as follows. There are 47 electrons surrounding the silver atom nucleus, of which 46 form a closed inner core of total angular momentum zero – there is no orbital angular momentum, and the electrons with opposite spins pair off, so the total angular momentum is zero, and hence there is no magnetic moment due to the core. The one remaining electron also has zero orbital angular momentum, so the sole source of any magnetic moment is that due to the intrinsic spin of the electron as given by Eq. (6.10).

Thus, the experiment represents a direct measurement of one component of the spin of the electron, this component being determined by the direction of the magnetic field, here taken to be in the z direction.

There are two possible values for S_z, corresponding to the two spots on the observation screen, as required by the fact that $s = \frac{1}{2}$ for electrons, i.e. they are spin-$\frac{1}{2}$ particles. The allowed values for the z component of spin are

$$S_z = \pm\tfrac{1}{2}\hbar \tag{6.14}$$

which, with the gyromagnetic value of two, yields the two values given in Eq. (6.13) for μ_z.

Of course there is nothing special about the direction z, i.e. there is nothing to distinguish the z direction from any other direction in space. What this means is that *any* component of the spin of an electron will have only two values, i.e.

$$S_x = \pm\tfrac{1}{2}\hbar, \qquad S_y = \pm\tfrac{1}{2}\hbar \tag{6.15}$$

and indeed, if $\hat{\mathbf{n}}$ is a unit vector specifying some arbitrary direction in space, then

$$\mathbf{S}\cdot\hat{\mathbf{n}} = \pm\tfrac{1}{2}\hbar. \tag{6.16}$$

Thus, by orienting the magnetic field in a Stern-Gerlach device in some direction $\hat{\mathbf{n}}$ perpendicular to the direction of motion of the atoms in the beam, the atoms will emerge in two possible beams, corresponding to $\mathbf{S}\cdot\hat{\mathbf{n}} = \pm\frac{1}{2}\hbar$. The positive sign is usually referred to as spin up in the $\hat{\mathbf{n}}$ direction, the negative sign as spin down in the $\hat{\mathbf{n}}$ direction. In the examples considered so far, the separation has always been in the z direction, i.e. $\hat{\mathbf{n}} = \hat{\mathbf{k}}$, but it is equally well possible to orient the magnetic field to lie in the x direction, i.e. $\hat{\mathbf{n}} = \hat{\mathbf{i}}$, so that the atomic beam is split into two beams with $S_x = \pm\frac{1}{2}\hbar$. In order to represent these possibilities in a diagram of the Stern-Gerlach device, a label will be included on the diagram to indicate the direction in which the magnetic field is oriented, and hence the component of the spin that is being measured. This is illustrated in the diagram Fig. 6.3.

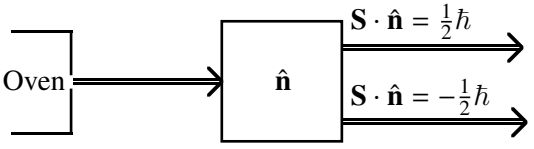


Figure 6.3: Stern-Gerlach device set to separate an atomic beam according to the $\hat{\mathbf{n}}$ component of spin. Separation according to the x component would be represented by the same diagram, except with an X within the rectangle, and similarly for other directions.

We will now use the above stripped-down picture of a Stern-Gerlach device to examine some purely quantum features of particle spin. Although the fact that particle spin is a purely quantum phenomenon, it is not the fact that particle spin exists and is of quantum origin that is of interest here. It is the properties that spin possesses when subject to various measurements that is of importance here – features that all quantum mechanical systems exhibit such as probabilistic outcomes of measurements, interference of probability amplitudes and so on are found to arise, but in circumstances in which there are only a handful of parameters needed to describe what is happening.

6.4 Quantum Properties of Spin

We shall now make use of the Stern-Gerlach apparatus to analyse the quantum properties of spin half in a way analogous to the two slit experiment. In this regard, we will consider repeated spin measurements, quantum randomness for spin and quantum interference for spin.

6.4.1 Spin Preparation and Measurement

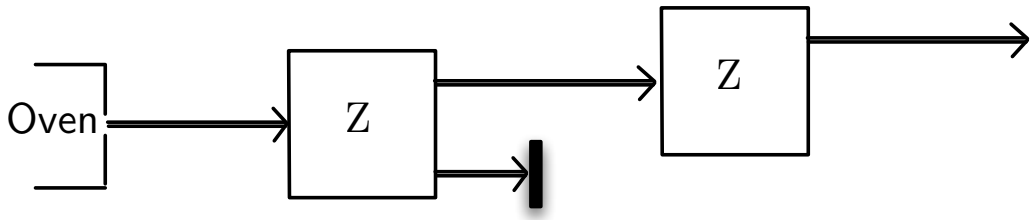
sorted out. The first point to recognize is that the Stern-Gerlach apparatus is both a spin preparation device and a spin measurement device. Thus, if an atom should emerge in a beam corresponding to a spin component $S_z = \frac{1}{2}\hbar$, then we can claim that the Stern-Gerlach apparatus has *prepared* the atom to have this specific value for S_z. More than that, we can also claim that the apparatus is a spin measuring device, i.e. if we wish to determine what the z component of the atomic spin happens to be for a given atom, we would pass that atom through a Stern-Gerlach apparatus, and the beam in which it emerges will tell us what the value is of this component. This relationship between preparation and measurement is of course not purely classical, but it acquires a heightened level of significance in quantum mechanics.

6.4.2 Repeated spin measurements

The Stern-Gerlach device presents a possible way of both preparing and measuring the various components of atomic spin. Thus, if a silver atom emerges in the $S_z = \frac{1}{2}\hbar$ beam, then the statement can be made that an atom has been prepared such that the z component of the spin of the valence electron is $S_z = \frac{1}{2}\hbar$.

If we pass one atom through the apparatus, then, given that the effect of the oven is to completely randomize the direction of the spin of each atom in the oven, it would come as no surprise that the atom is equally likely to emerge in either the spin up or the spin down beam. This outcome has nothing much to do with quantum mechanics, it is what we would more or less expect on the basis of classical physics, apart from the fact, of course, that quantum mechanically, any spin has only two values.

Now suppose that the atom exits in the $S_z = \frac{1}{2}\hbar$ beam. We now know the S_z component of the total spin of that atom, i.e. we have *prepared* an atom whose spin component has the value $S_z = \frac{1}{2}\hbar$. If we then immediately pass this atom into a second Stern-Gerlach apparatus with is magnetic field in the z direction, we find that the atom re-emerges in the $S_z = \frac{1}{2}\hbar$ beam, i.e. this second measurement merely confirms the first result.

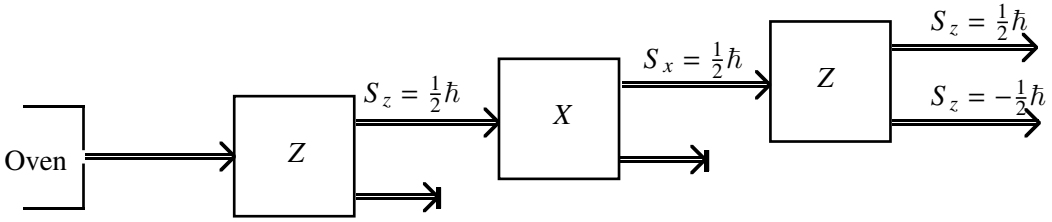


6.4.3 Quantum randomness

One of the features of quantum mechanics is that it is not possible, even in principle, to have complete knowledge of all the physical variables that characterize the state of a system. Thus, for instance, exact knowledge of the position of a particle means that there is total uncertainty in the knowledge of its momentum, and vice versa. The same is true for particle spin, except that here it us the various components of spin that cannot be known simultaneously with complete accuracy.

That this is the case has is built into quantum mechanics in a fundamental way, but the manner in which it expresses itself varies depending on the circumstances under which an attempt is made to measure more than one component of spin. It is found, in the example to be discussed below, that the uncertainty principle plays a fundamental role in that it appears to provide a mechanism, or at least an explanation, as to why only one component can be known exactly at any one time. But the real 'why' is that it is a fundamental law of nature.

Consider a series of spin measurements using a sequence of Stern-Gerlach devices, as illustrated in following diagram:



In this experiment, atoms are separated in the first device according to their z component of spin. Those for which $S_z = \frac{1}{2}\hbar$ are then passed through a second device in which atoms are separated according to their x component of spin. Those for which $S_x = \frac{1}{2}\hbar$ are passed through a third device which separates the atoms according to their z component, once again. The naive expectation is that, since these atoms have already been preselected to have $S_z = \frac{1}{2}\hbar$, then they will all emerge from the final device in the $S_z = \frac{1}{2}\hbar$ beam. It turns out that this is not what is observed. The atoms emerge randomly in either beam, but with equal probability. The interpretation that immediately comes to mind is that the intervening measurement of the x component of spin has in some way scrambled the z component of spin, but according to classical physics, it should be possible either to arrange the experiment such that any such scrambling be made negligibly small, or else be able to correct for the apparent scrambling in some fashion. It turns out that the quantum effects prevent this from happening – this scrambling, and the consequent introduction of randomness into the outcome of the experiment cannot be avoided, except at the cast of not being able to measure the x component of spin at all! Thus we see again an example of intrinsic randomness in the behaviour of macroscopic systems.

In the following section, an argument is presented which shows how it is that quantum effects prevent the simultaneous exact measurement of both the x and the z components of spin i.e. that it is uncontrollable quantum effects that give rise to the scrambling of the z component of spin during the measurement of the x component.

Incompatible Measurements of Spin Components

The obvious question to ask is whether or not the experiment can be refined in some way to avoid this scrambling. From the perspective of classical physics, the answer is definitely yes, at least in principle. The problem is that the atoms, as they pass through the second Stern-Gerlach device, will experience precession about the x axis which will have the effect of changing the z component of the spin. But by suitable fiddling with the beam, the magnetic field strengths and so on it should be possible in principle, at least from the point of view of classical physics, to minimize this effect, or at least determine exactly how much precession occurs, and take account of it. But in practice, it turns out that all these attempts fail. If the experiment is refined in such a manner that the precession is made negligible, (e.g. by using faster atoms, or a weaker magnetic field), the result is that the two emerging beams overlap so much that it is impossible to tell which beam an atom belongs to, i.e. we retain exact information on the z component of spin, but learn nothing about the

x component! In general, it appears that it is not possible to measure both S_z and S_x (or, indeed any pair of components of the particle spin), precisely. This kind of behaviour is reminiscent of what is found to happen when we attempt to measure both the position and the momentum of a particle. According to the uncertainty principle, the more precisely we determine the position of a particle, the less we know about the momentum of the particle. The difference here is that the quantities being measured are discrete – they have only two possible values, whereas the position and momentum of a free particle (even in quantum mechanics) can assume a continuous range of values.

A Detailed Analysis By use of a hybrid mixture of classical and quantum mechanical arguments, it is possible to come to some 'understanding' of why this occurs. Consider the atoms that have left the first Stern-Gerlach device with $S_z = \frac{1}{2}\hbar$ and enter the next device which has a magnetic field $\mathbf{B} = B\mathbf{i}$ oriented in the x direction. This magnetic field is non-uniform, in other words B is a function of position – it could be written $B(x, y, z)$. The experimental arrangement is such that the non-uniformity is most marked in the x direction – it is this non-uniformity that is responsible for the forces acting to deflect the atoms as they move through the device. In fact, the interaction of the magnetic moment $\boldsymbol{\mu}$ of the atoms with this non-uniform magnetic field has two consequences. First, as just mentioned, the atoms feel a force in the x direction given by

$$F_x = -\mu_x \frac{\partial B}{\partial x} \tag{6.17}$$

where μ_x is the x component of the magnetic moment of the atoms. If we accept in what is otherwise a classical argument, that the values of the spin of an electron are restricted to their quantized values, then

$$\mu_x = \pm \frac{e\hbar}{2m} \tag{6.18}$$

corresponding to the two possible values of $S_x = \mp\frac{1}{2}\hbar$, and leading to the formation of two separate beams corresponding to the two values of S_x.

Second, the magnetic moment of the atoms will *precess* about the direction of the magnetic field with an angular frequency given by

$$\omega = \frac{\mu_x B}{\hbar}. \tag{6.19}$$

As a consequence of this precession, the y and z components of $\boldsymbol{\mu}$ and hence of $\mathbf{S}$ will change with time, while the x component will remain unchanged.

This precession is one ingredient in the explanation of the 'scrambling' of the z component of the spin. The second ingredient is based on the fact that the atomic beam that leaves the oven, and passes through the various Stern-Gerlach devices will have a non-zero cross-section, or, in other words, atoms in the beam will, in general, pass through the magnetic field along trajectories with different values of x and hence each atom will experience different magnetic field strengths, and consequently will have different precession rates. The nett result of this is that after the atoms leave the magnetic field, the various atoms will have had their magnetic moments rotated through a range of different angles, so that there will be, in consequence, a spread in the possible values of S_z. Translated into the quantum picture, this means that S_z can, with equal probability, be observed to be $\pm\frac{1}{2}\hbar$, and hence the result that is seen in the experiment.

If we are to believe that this argument has some truth in it then it seems that the 'scrambling' of the z component of the atomic magnetic moment can be minimized simply by making sure that all the atoms pass along the same trajectory through the magnetic fields. If this were possible, and classical physics claims that it is, then the effect of precession on the z component of spin would be the same for every atom, and so could be accounted for. The net result is that, in effect, the measurement of the x component of spin will not interfere with the results of the preceding

measurement of the z component. However, quantum mechanics, in the form of the uncertainty principle, prevents this from happening, as the following argument shows.

The fact that the atomic beam has a finite width means that there is uncertainty in the cross-sectional position of the atoms in the beam. In the x direction, the uncertainty in position is Δx, which implies, by the uncertainty principle, that there is an uncertainty Δp_x in the x component of the momentum of the atom given by

$$\Delta p_x \approx \frac{\hbar}{\Delta x}. \tag{6.20}$$

This translates into an uncertainty in the x velocity of the atom given by

$$v_x \approx \frac{\hbar}{m\Delta x}. \tag{6.21}$$

As a consequence, during the time of flight t of the atoms through the device, the uncertainty in the width of the beam will grow by an amount δx given by

$$\delta x = \Delta v_x t \approx \frac{\hbar}{m\Delta x} t. \tag{6.22}$$

So, the width of the beams is growing linearly in time. Meanwhile the two beams are separating at a rate determined by the force F_x given in Eq. (6.17). Assuming that this force is constant, then the separation between the beams will be, after a time t

$$2 \times \tfrac{1}{2} \frac{F_x}{m} t^2 = m^{-1} \mu_x \frac{\partial B}{\partial x} t^2 \tag{6.23}$$

where the factor of 2 comes from the fact that the two beams are pulling away from each other at the same rate. The crucial part of the argument is then this: the separation of the two beams must be greater than the widths of the beams otherwise the two beams will overlap, and it will be impossible to distinguish which beam a particle belongs to, in other words it will be impossible to know what the x component of the spin of the atom is. Thus, in order to be able to determine the x component of spin, we must have

$$\delta x << m^{-1} \mu_x \frac{\partial B}{\partial x} t^2 \tag{6.24}$$

which becomes, after substituting for δx

$$\hbar^{-1} \mu_x \Delta x \frac{\partial B}{\partial x} t >> 1. \tag{6.25}$$

The quantity $\Delta x \partial B/\partial x$ is the variation in the strength of the magnetic field across the width of the beam as experienced by the atoms as they pass through the device. This means that the atoms will precess at rates that cover a range of values $\Delta\omega$ given by, from Eq. (6.19)

$$\Delta\omega = \frac{\mu_x}{\hbar} \Delta x \frac{\partial B}{\partial x}. \tag{6.26}$$

Substituted into the inequality Eq. (6.25), this gives

$$\Delta\omega t >> 1. \tag{6.27}$$

In other words, the spread in the angle $\Delta\omega t$ through which the magnetic moments precess is so large that the z component of the spin, roughly speaking, is equally likely to have any value, in other words, it is completely randomized.

This argument shows that it is not possible to measure both the z and the x components of spin, or indeed any pair of components of spin. If we have determined a value for S_z say, and we want to

then measure S_x, then, in order to make the latter measurement possible, we have to separate the beams enough that they are distinguishable. But this unavoidably results in total randomization of the value of S_z. If we arrange the experimental conditions to be such that S_z is not changed by the measurement of S_x, we find that the two beams exiting from the Stern-Gerlach device overlap to such an extent that it is not possible to say which beam an atom belongs to, i.e. we have not, in fact, measured S_x. The preceding argument is not wholly satisfactory as it is a mixture of classical and quantum concepts, and should be viewed purely as aid to understanding what is taking place. The central, quantum mechanical fact, is that the intervening measurement of the x component randomizes the previously exactly known value of S_z. It might be argued that the fault lies with the Stern-Gerlach device, and that by using some other method of measuring the components of spin, we can get around the sort of problems encountered here. Even if S_x and S_z were measured by means that have nothing whatsoever to do with the Stern-Gerlach experiment, the same result would be obtained: an intervening measurement of the x component will randomize the previously exactly known value of S_z. A different argument based on the uncertainty relation could undoubtedly be formulated in each case to 'explain' the result, as discussed in Chapter 4, but the fact that the same kind of behaviour is always observed irrespective of the circumstances is telling us that there is a basic physical principle in action here, in effect a law of nature – one of the laws of quantum mechanics – that guarantees that under no circumstances is it possible to have exact knowledge of more than one component of the spin of a particle.

6.4.4 Probabilities for Spin

A crucial feature of the above result was that the intervening measurement of the x component of spin had the effect of randomizing the outcome of the remeasurement of the z component. By symmetry it is expected that if the z component of spin has been measured as $S_z = \frac{1}{2}\hbar$ say, then in the following measurement of S_x, there is an equal chance of the atoms emerging in either of the $S_x = \pm\frac{1}{2}\hbar$ beams. However, for later purposes, it is useful to have on hand an expression for the probabilities in the case in which the magnetic fields in the Stern-Gerlach devices are set in some arbitrary direction in the XZ plane (the atoms are travelling in the direction of the positive Y axis). It is possible to use arguments based on symmetry and geometry to arrive at the required results, but here, the result will be simply presented as something that can be measured.

To begin with, we will look at the following Stern-Gerlach experiment, illustrated in Fig. (6.4).

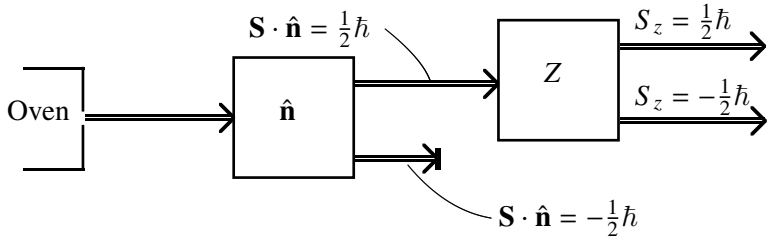


Figure 6.4: Atoms with random spin orientation filtered through a Stern-Gerlach device with magnetic field in $\hat{\mathbf{n}}$ direction, and the $S_i = \mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar$ beam passed through a second device with magnetic field in z direction.

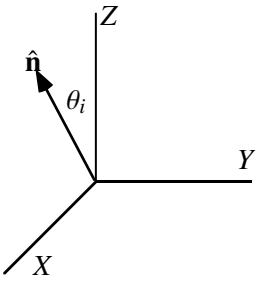


Figure 6.5: The unit vector $\hat{\mathbf{n}}$ specifies the direction of the magnetic field in a Stern-Gerlach device. This vector lies in the XZ plane, and the atomic beam travels in the direction of the positive Y axis.

In this experiment, the atoms, after they leave the oven, pass through a Stern-Gerlach device in which the magnetic field is oriented in the direction specified by the unit vector $\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ lies in the XZ plane, at an angle of θ_i to the Z axis, see Fig. (6.5). Atoms will leave this device in one of two beams, corresponding to the component of spin $\mathbf{S}$ in the direction of $\hat{\mathbf{n}}$ having one or the other of the two values $S_i = \mathbf{S}\cdot\hat{\mathbf{n}} = \pm\frac{1}{2}\hbar$. For the purposes of the experiment, atoms exiting from the lower beam, for which $S_i = -\frac{1}{2}\hbar$ are blocked, while those exiting in the upper beam, for which $S = \frac{1}{2}\hbar$ pass through a second Stern-Gerlach device, this time with its magnetic field oriented to separate the atoms according to their z component of spin. In general, the atoms will exit from this second device, once again, in one or the other of two beams, the upper one in the diagram being the beam for which $S_z = \frac{1}{2}\hbar$, the lower one being the one for which $S_z = -\frac{1}{2}\hbar$.

Let us suppose that the experiment is repeated many times over for each setting of the angle θ_i in order to obtain, experimentally, the fraction, or in other words, the probability, of atoms emerging from the final Stern-Gerlach device in either of the two beams. The experimental result obtained is that

$$\begin{aligned}\text{Probability of atoms emerging in the } S_z = \tfrac{1}{2}\hbar \text{ beam} &= \cos^2(\theta_i/2)\\ \text{Probability of atoms emerging in the } S_z = -\tfrac{1}{2}\hbar \text{ beam} &= \sin^2(\theta_i/2)\end{aligned} \tag{6.28}$$

At this point it is useful to introduce a new notation for this probability. First we note that the atoms, as they exit from the first Stern-Gerlach device, are such that $S_i = \frac{1}{2}\hbar$. Next we note that this is the maximum amount of information that we can have about the spin of these atoms – any attempt to measure another component will scramble this component is an uncontrollable way. So, to the best that we can manage, we can characterize the physical state of the atoms by $S_i = \frac{1}{2}\hbar$. When they exit from the second Stern-Gerlach device, they are either in a state for which $S_z = \frac{1}{2}\hbar$, or for which $S_z = -\frac{1}{2}\hbar$. We will now adopt the notation

$$P(A|B) = \text{Probability of observing a system in a state for which information } A \text{ is known given that it was in a state for which information } B \text{ is known.}$$

We can now write

$$\begin{aligned}P(S_z = \tfrac{1}{2}\hbar | S = \tfrac{1}{2}\hbar) &= \cos^2(\theta_i/2)\\ P(S_z = -\tfrac{1}{2}\hbar | S = \tfrac{1}{2}\hbar) &= \sin^2(\theta_i/2)\end{aligned} \tag{6.29}$$

We can check to see if this makes physical sense by looking at some special cases. Thus, if $\hat{\mathbf{n}} = \hat{\mathbf{k}}$, i.e. the first Stern-Gerlach device has the magnetic field oriented in the z direction, then $S_i = S_z$ and $\theta_i = 0$ so that the device is equivalent to the set-up given in Fig. (6.6)

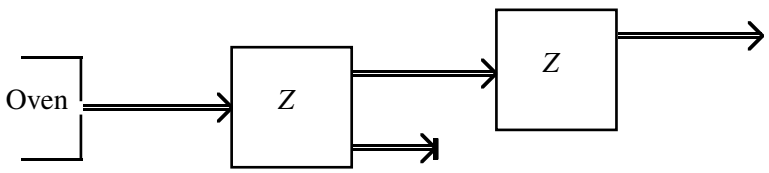


Figure 6.6: Same as Fig. (6.4) but with $\hat{\mathbf{n}}$ in z direction.

and the probabilities become, from Eq. (6.28) with $\theta_i = 0$

$$\begin{aligned} P(S_z = \tfrac{1}{2}\hbar | S_z = \tfrac{1}{2}\hbar) &= 1 \\ P(S_z = -\tfrac{1}{2}\hbar | S_z = \tfrac{1}{2}\hbar) &= 0 \end{aligned} \tag{6.30}$$

which is as it should be – if an atom has been measured to have $S_z = \frac{1}{2}\hbar$, then a subsequent measurement of S_z should simply confirm this result.

Next, if we look at the case of $\hat{\mathbf{n}} = \hat{\mathbf{i}}$, so that the magnetic field is oriented in the x direction in the first Stern-Gerlach device, then we have $S_i = S_x$ and $\theta_i = \pi/2$. The set-up is then as illustrated in Fig. (6.7)

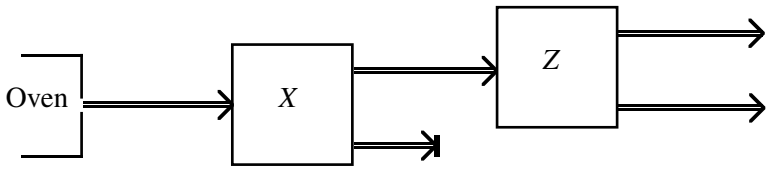


Figure 6.7: Same as Fig. (6.4) but with $\hat{\mathbf{n}}$ in x direction.

and the probabilities are, from Eq. (6.28) with $\theta_i = \pi/2$

$$\begin{aligned} P(S_z = \tfrac{1}{2}\hbar | S_x = \tfrac{1}{2}\hbar) &= \tfrac{1}{2} \\ P(S_z = -\tfrac{1}{2}\hbar | S_x = \tfrac{1}{2}\hbar) &= \tfrac{1}{2} \end{aligned} \tag{6.31}$$

which is also consistent with what we have seen before – if the atom has been measured to have $S_x = \frac{1}{2}\hbar$, then there is an equal chance that it will be measured to have $S_z = \pm\frac{1}{2}\hbar$. Finally, we will consider the case in which $\hat{\mathbf{n}} = -\hat{\mathbf{k}}$, i.e. $\theta_i = \pi$. In this case, $S_i = -\mathbf{S}\cdot\hat{\mathbf{k}} = -S_z$ and the set-up is as in Fig. (6.8).

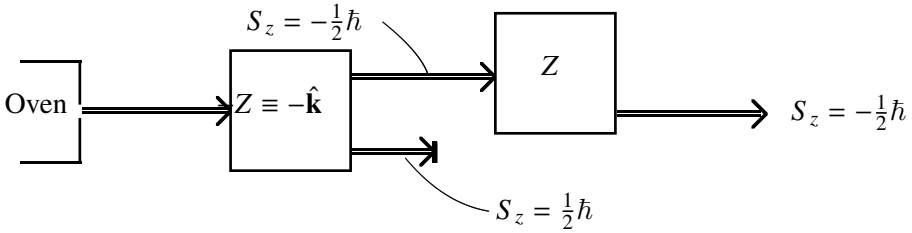


Figure 6.8: Atoms with random spin orientation filtered through a Stern-Gerlach device with magnetic field in $\hat{\mathbf{n}} = -\hat{\mathbf{k}}$ direction. The atoms in the upper beam exiting from this Stern-Gerlach device are those for which $S_i = \mathbf{S}\cdot\hat{\mathbf{n}} = -S_z = \frac{1}{2}\hbar$.

As the field is in the negative z direction, the upper beam leaving the first Stern-Gerlach device in Fig. (6.8) will have $S_i = -S_z = \frac{1}{2}\hbar$, i.e. $S_z = -\frac{1}{2}\hbar$. Consequently, when this beam enters the next Stern-Gerlach device with the field oriented in the z direction, all the atoms will emerge in the $S_z = -\frac{1}{2}\hbar$ beam. This is in agreement with the probabilities that follow from Eq. (6.28) with $\theta_i = \pi$, i.e.

$$\begin{aligned} P(S_z = \tfrac{1}{2}\hbar | S_z = -\tfrac{1}{2}\hbar) &= \cos^2(\tfrac{1}{2}\pi) = 0 \\ P(S_z = -\tfrac{1}{2}\hbar | S_x = \tfrac{1}{2}\hbar) &= \sin^2(\tfrac{1}{2}\pi) = 1 \end{aligned} \tag{6.32}$$

6.5 Quantum Interference for Spin

In the last Chapter, what is identified as the essential 'mystery' of quantum mechanics was illustrated in the two slit experiment using particles. In this experiment, there are two ways that a particle can pass from the particle source to the observation screen i.e. via one slit or the other, but provided the slit through which the particle passes is *not* observed, the particles do not strike the screen in a way that is consistent with our intuitive notion of the way a particle should behave: the particles strike the observation screen at random, but with a preference to accumulate in certain regions, and not at all in other regions, so as to form a pattern identical to the interference pattern that would be associated with waves passing through the slits. In contrast, if the slit through which each particle passes is observed in some fashion, the interference pattern is replaced by the expected result for particles. It is the lack of any explanation for this kind of behaviour in terms of everyday intuition and/or classical physics that is seen as the fundamental mystery of quantum mechanics.

It was inferred from this experiment that associated with the particle was some kind of wave, a probability amplitude wave or wave function which, for a point x on the observation screen, could be written as the sum of two contributions originating from each slit – $\Psi(x,t) = \Psi_1(x,t) + \Psi_2(x,t)$ – and whose intensity $|\Psi(x,t)|^2$ gave the probability density of observing a particle at a particular position x on the observation screen. All these results referred to the measurement of the position of the particle, a continuously variable quantity. The aim here is to show that interference is a signature of quantum mechanics even when, as in the case of particle spin, the property of the particle being observed is not its position, but rather its spin, which can only have discrete values. Moreover, it is intended to show that interference arises when there is more than one 'path' that a particle can follow between its source and its final observation. This demonstration provides further evidence that there is an underlying commonality between different examples of quantum behaviour, evidence of some fundamental law or laws that apply to all physical systems, though superficially realized in different ways for different systems. In this experiment, atoms emerges from the oven, and are then passed through a Stern-Gerlach device whose magnetic field is oriented so as to separate the atoms into two beams according to their x component of spin. The atoms emerge in two separate beams corresponding to the atomic spin component $S_x = \mathbf{S}\cdot\hat{\mathbf{i}} = \pm\frac{1}{2}\hbar$. The atoms in one of the beams ($S_x = \frac{1}{2}\hbar$) is then selected and passed through a Stern-Gerlach device where the magnetic field further separates this beam according to its z component of spin. The atoms emerge in one or the other of two beams corresponding to $S_z = \mathbf{S}\cdot\hat{\mathbf{k}} = \pm\frac{1}{2}\hbar$. The two beams are then recombined into a single beam. This is done using a third Stern-Gerlach device in which the magnetic field is equal and opposite to the preceding device. This does not scramble the spins of the atoms – the sole purpose is to recombine the beams and could equally well have been done by some other technique. Finally, this beam is passed through a further Stern-Gerlach device with its magnetic field oriented in the x direction so that atoms will emerge from this device with either $S_x = \pm\frac{1}{2}\hbar$.

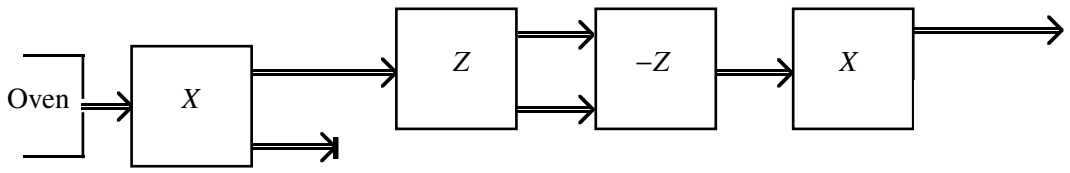


Figure 6.9: Atomic beam for which $S_x = \frac{1}{2}\hbar$ split into $S_z = \pm\frac{1}{2}\hbar$ beams and then recombined before passing through a final Stern-Gerlach device with magnetic field in x direction.

It is important to see the analogy between this setup and the two slit interference experiment. The

oven plus the first Stern-Gerlach device is the equivalent of the source of identically prepared particles in the two slit experiment. Here the atoms are all identically prepared to have $S_x = \frac{1}{2}\hbar$. The next two Stern-Gerlach devices are analogous to the two slits in that the atoms can, in principle, follow two different paths corresponding to $S_z = \pm\frac{1}{2}\hbar$ before they are recombined to emerge in one beam. The analogue is, of course, with a particle passing through one or the other of two slits before the position where it strikes the observation screen is observed. We can tell which path an atom follows (i.e. via the $S_z = \frac{1}{2}\hbar$ or the $S_z = -\frac{1}{2}\hbar$ beam) by monitoring which beam an atom emerges from after it passes through the first z oriented Stern-Gerlach device in much the same way that we can monitor which slit a particle passes through in the two slit experiment. Watching to see in which beam an atom finally emerges after passing through the last Stern-Gerlach device is then analogous to seeing where on the observation screen a particle lands after passing through the two slit device.

The results found are as follows. If the intervening state of the atoms is *not* observed, the results obtained are the same as if the beam splitter-recombiner were not there, i.e. the results are the same as in Fig. (6.6), and Eq. (6.30), though here for the x component. However, if the z component of the spin *is* observed, then it is effectively an atom with a known z component of spin that enters the last Stern-Gerlach device, as for Fig. (6.7), and hence the probability of the atom having either value of S_x becomes $\frac{1}{2}$, as in Eq. (6.31).

This behaviour is reminiscent of what was observed in the two slit experiment – if we do not observe through which slit the particles pass, then we observe an interference pattern. If we do observe through which slit the particles pass, then there is no interference pattern. So, is there a sense in which the results found above for the Stern-Gerlach experiment can be interpreted as the presence of interference in the first case, and no interference in the second? We can present a persuasive, but non-rigorous argument that this is the case. A much sharper argument is presented later in Section 7.3.

Suppose, for the present that probability amplitudes can indeed be associated with the atoms passing through either of the two $S_z = \pm\frac{1}{2}\hbar$ beams before their x component of spin is observed. So let $\Psi_\pm(S_x)$ be the amplitudes for the spin to be measured to be S_x, given that they passed through either the $S_z = \frac{1}{2}\hbar$ or the $S_z = -\frac{1}{2}\hbar$ beam. This is analogous to the probability amplitudes $\Psi_n(x)$ of observing the particle at position x given that they passed through slit n. From the results presented above if we do not observe through which intervening beam the atoms passed, we should add the probability amplitudes and then take the square:

$$\begin{aligned} &\text{Probability of atom emerging in } S_x = \tfrac{1}{2}\hbar \text{ beam} &&= |\Psi_+(\tfrac{1}{2}\hbar) + \Psi_-(\tfrac{1}{2}\hbar)|^2 = 1 \\ &\text{Probability of atom emerging in } S_x = -\tfrac{1}{2}\hbar \text{ beam} &&= |\Psi_+(-\tfrac{1}{2}\hbar) + \Psi_-(-\tfrac{1}{2}\hbar)|^2 = 0 \end{aligned} \tag{6.33}$$

While, if we *do* observe through which beam they pass, we should add the probabilities:

$$\begin{aligned} &\text{Probability of atom emerging in } S_x = \tfrac{1}{2}\hbar \text{ beam} &&= |\Psi_+(\tfrac{1}{2}\hbar)|^2 + |\Psi_-(\tfrac{1}{2}\hbar)|^2 = \tfrac{1}{2} \\ &\text{Probability of atom emerging in } S_x = -\tfrac{1}{2}\hbar \text{ beam} &&= |\Psi_+(-\tfrac{1}{2}\hbar)|^2 + |\Psi_-(-\tfrac{1}{2}\hbar)|^2 = \tfrac{1}{2}. \end{aligned} \tag{6.34}$$

By symmetry we should also have that

$$|\Psi_\pm(\tfrac{1}{2}\hbar)|^2 = |\Psi_\pm(-\tfrac{1}{2}\hbar)|^2 \tag{6.35}$$

i.e. whether the atom comes through via the $S_z = \frac{1}{2}\hbar$ or the $S_z = -\frac{1}{2}\hbar$ beams, they should still have an equal chance of emerging in either of the $S_x = \pm\frac{1}{2}\hbar$ beams. A quick calculation shows

that these equations are satisfied by

$$\Psi_{\pm}(\tfrac{1}{2}\hbar) = \tfrac{1}{2} \qquad \Psi_{\pm}(-\tfrac{1}{2}\hbar) = \pm\tfrac{1}{2}. \tag{6.36}$$

In other words, the possibility exists of interpreting the observed results as being the consequence of interference taking place. Thus, we have

$$\begin{array}{l}\text{Probability of atom emerging in} \\ S_x = \pm\frac{1}{2}\hbar \text{ beam}\end{array} = |\tfrac{1}{2} \pm \tfrac{1}{2}|^2 = \tfrac{1}{4} + \tfrac{1}{4} \pm \tfrac{1}{2} \tag{6.37}$$

where the term $\pm\frac{1}{2}$ is the 'interference' term. We have constructive interference when this term is positive, giving unit probability of finding the atom exiting in the $S_x = \frac{1}{2}\hbar$ beam, and destructive interference when this term is negative, giving zero probability of the atom emerging in the $S_x = -\frac{1}{2}\hbar$ beam. If the intervening beam through which the atoms pass is observed, the results are just a half for either the $S_x = \frac{1}{2}\hbar$ or the $S_x = -\frac{1}{2}\hbar$ beam, which is just the result that is obtained if the interference term in Eq. (6.37) is removed. Thus, there indeed appear to be two contributions to the total probability amplitude of observing the atom to have the x component of spin equal to $\pm\frac{1}{2}\hbar$, these being associated with the probability amplitudes of the atoms passing along one or the other of the two S_z beams.

There is a complete analogue here with the two slit experiment. In that experiment, the aim was to provide two paths along which the particles could pass: from the source through either slit 1 or 2, and then to the final measurement of the x position on the screen. Here, we want to provide two possible 'paths' for the *spin* of the atoms: from initial spin $S = \frac{1}{2}\hbar$, through either of $S_z = \frac{1}{2}\hbar$ or $S_z = -\frac{1}{2}\hbar$, until finally a measurement is made of S_x. The spin of the atoms therefore follows paths in what might be called 'spin space', rather than in real space. Experimentally these paths in spin space are produced by providing different paths in real space for the atoms to follow, depending on their spin, but this is a feature of the experiment only, and largely irrelevant to the argument being developed here.

The first Stern-Gerlach device plays the same role here as the source of particles in the two-slit experiment, and provides a source of atoms for which $S_x = \frac{1}{2}\hbar$. The Stern-Gerlach device that separates the beams in the z direction is then the equivalent of the two slits as it provides two different 'paths' that the atomic spin can follow prior to the final measurement. By then recombining the two beams, we lose all information concerning the path that the atoms follow. Thus, when the final measurement of the x component of spin is performed, we have no way of knowing whether an atom exited from the second Stern-Gerlach device with $S_z = \frac{1}{2}\hbar$ or $S_z = -\frac{1}{2}\hbar$, unless we explicitly observe which beam an atom belongs to immediately as it exits the device. This is analogous to not knowing which slit a particle passes through before its x position is measured on the observation screen in the usual two slit experiment.

We therefore find, once again, that if we have information on which 'path' the system of interest follows as it makes its way from some initial state to some final measurement, we get a different result from what we get if we do not have this information. In the case of the two slit experiment, lack of 'which path' information leads to wave-like interference effects which are absent if we do know which slit the particle passes through. In the case of particle spin the result obtained when the intermediate spin S_z is not observed can also be interpreted as being due to interference effects which are absent if the spin of the atoms is observed. For the present it is sufficient to note that the outcome of the experiment does depend on whether or not the intermediate observation of S_z is made. It therefore appears that there is much in common between the two slit experiment and the spin experiment, in spite of the manifestly different physical character of the experiments. Put in another way, there appears to be some fundamental laws in action here, the laws of quantum mechanics, that are expressed in slightly different ways in different physical systems: interference and randomness observed in the measurement of particle position in the two slit experiment, and

similar behaviour in the measurement of particle spin. The laws of quantum mechanics, and the mathematical language in terms of which these laws are stated, is the subject of the following Chapter.

Chapter 7

Probability Amplitudes

EVIDENCE was presented in the preceding Chapter that under certain circumstances physical systems of quite disparate nature, but usually on the atomic scale, have the common properties of randomness and interference effects, that have no classical explanation. These general properties are the indicators, or the signatures, of the existence of basic physical laws that are not part of our normal everyday view of the way the world behaves, which immediately poses the problem of determining, and clearly stating, what these fundamental laws are. But there is in fact a second problem that has to be dealt with here. In order to state these basic laws, the question arises as to whether or not the concepts and mathematical language of classical physics is adequate for the task. We have seen that the mathematics of waves works fine for particles moving through space: from the wave function, which can be derived by solving the Schrödinger equation, information on the position and momentum, angular momentum, energy, and much else besides can be obtained. But we have also seen that the wave function cannot be used to describe spin. Apart from anything else, spin is a discrete variable, with a finite number of values, unlike the position of a particle which can vary in a continuous way. So what is needed is a mathematical language in terms of which these laws can be expressed, but which can assume outwardly different forms when applied to different physical systems. The two problems are not independent: the process of determining what the physical laws are guides us in the development of the mathematical language needed to express these laws. The outcome of all this, in another example of what has been termed 'the unreasonable effectiveness of mathematics in physics' is a mathematical language that was already well known to mathematicians in the early days of the quantum theory: the mathematics of linear vector spaces. It was a British theoretical physicist, Paul Dirac, who played a major role in formulating quantum mechanics in this way. He was inspired to do so, in part, because of the above-mentioned need for a theory that could cope with both particle spin and particle motion through space. What this work does, amongst other things, is bring into sharp focus the problem of defining what is meant by the state of a quantum system, and how a quantum state is to be represented mathematically. A further consequence of his work is a succinct notation which encapsulates both the physics and the mathematics, the Dirac bra-ket notation which is used, amongst other things, to represent the state of a quantum system. The first issue to be dealt with here is then that of coming to terms with what we mean by the state of a system, following which we look again at the two-slit experiment, which sets the scene for the development of the Dirac/Feynman version of quantum mechanics.

7.1 The State of a System

The notion of the state of a system is a central one in both classical and quantum physics, though it is often possible to live with only an intuitive idea of what it means. However, it proves to be important here to have the concept of the state of a system clearly defined. Ideally, specifying the state of a system would amount to gathering together all the information that it is possible

to know about the system at any instant in time. The information should be enough so that, in principle, it would be possible to reconstruct the system at some other time and place and have the reconstructed system behave in exactly the same way as the original. From the point of view of physics, this information would take the form of numerical data that specifies, for instance for a single particle, its position and momentum at any instant in time. How this information is gained is, of course, another matter, though in classical physics the accepted point of view is that the information is there to be found; it is up to us to be clever enough to get at it. Likewise, preparing a system in a particular state, that is, preparing it such that the various quantities characterising its state are given desired values, such as preparing a particle in a state with a certain position and momentum, is assumed to be possible, though not necessarily easy. Of course, the simple matter of measuring the value of, say, the position of a particle is tantamount to preparing it to have that particular value, provided the particle is not destroyed in the process of having its position measured!

Of course, many systems when considered in all their detail are forbiddingly complex, but fortunately it is not always necessary to specify *everything* about a system. For instance, if we are interested in describing the orbit of the Earth around the sun, it would be sufficient to specify the state only in so far as the position and momentum of the centre of mass of the Earth. The fact that the Earth rotates, or that grass is green is not information that is required to deal with the orbital dynamics of the Earth. Knowing what is relevant and what is not is important in setting up good models of complex systems, but given that this has been decided upon for a given system, a good definition of the state of the system consistent with our intuitive notions is then to say that the state of a system is defined by specifying the maximum amount of data that can, in principle, be known simultaneously without mutual interference or contradiction about the system.

According to classical physics, it is possible in principle, if not always in practice, to determine *exactly* all the quantities needed to specify the state of a physical system. Thus, for a system consisting of a single particle, the state of the system is specified by giving the position and the momentum of the particle at the instant of interest. For a multi-particle system, the state could be specified by giving the positions and momenta of the individual particles, or else, if there are constraints of some kind, e.g. if the particles are organized into a single rigid body then it is probably sufficient to give the position and momentum of the centre of mass, the orientation of the body, and a few other quantities such as the linear and angular momentum. In practice, of course, there will always be uncertainty in our knowledge of these quantities, but this can be put down to inadequacies in our experimental technique, measuring apparatus, or the sheer size and complexity of the system under consideration.

When it comes to actually representing the state of a classical system, there is no way of doing this that has any significance beyond simply being a list of the values of all the physical parameters that can be determined for the system, though different ways of presenting this information offer different advantages in different circumstances. An example of one way of representing the information is, for a single particle, to plot the position and momentum of the particle as a point (x, p) in what is known as *phase space*. As a function of time, this point will trace out a path in phase space, this path then representing the evolution of the state of the system as a function of time.

The situation is somewhat different when quantum effects become important. It is not possible, even in principle, to specify the position *and* the momentum of a particle with total precision. The uncertainty principle tells us that we either have to compromise on the accuracy with which we specify each of the quantities, or else have to deal with the consequences of knowing, say, the position exactly, implying that the momentum is totally unknown, or vice versa. So what then are we to do with the notion of the state of a quantum system? The situation is one of accepting the information that we have about the system, and taking that as specifying the state. Doing so is consistent with the definition of state presented above, and is a reflection of what

knowledge will really do possess about the system. This is information we have without mutual interference – i.e. we specify the position of a particle, but not its momentum, or vice versa. Thus we cannot represent the state of the system in the same way that we can for a classical system, as a point in phase space, for instance. Similarly we can only specify one component of the spin of a particle, say the x component of the spin of a particle, in which case we cannot specify its y or z component. The information should also be without contradiction, i.e. we do not claim a particle to be at one position *and* at some other. Of course, there remains the question of how we obtain this information about a quantum system, though, as in the classical case, measuring the value of some observable property of a quantum system is also tantamount to preparing it in a state in which it definitely has the measured value, provided, once again, that the system is not destroyed by the measuring process.

Here we will introduce a notation for the state of a quantum system which, for the present, is nothing more than a fancy way of writing down all the information that we can know about a quantum system, but which turns out to be very useful in describing the mathematical properties of quantum systems. The notation looks like this:

$$\left| \begin{array}{l} \text{All the data concerning the system that can be known with-} \\ \text{out mutual interference or contradiction.} \end{array} \right\rangle \tag{7.1}$$

The symbol $|\ \rangle$ is known as a *ket*. Contained within the ket is a summary of the data specifying the state of a system, and hence a ket is also referred to as the state of the system. Thus, for instance, if we know the position of a particle is $x = 3$ cm with respect to some origin, then the state would be $|x = 3\text{ cm}\rangle$. If we know that a particle has a z component of spin equal to $\frac{1}{2}\hbar$, then the state would be $|S_z = \frac{1}{2}\hbar\rangle$. There would then be no such state as $|S_x = \frac{1}{2}\hbar, S_z = \frac{1}{2}\hbar\rangle$ since it is not possible to known both S_x and S_z simultaneously – the measurement of one component interferes with the value of any other previously measured component. We also cannot have a state like $|S_x = \frac{1}{2}\hbar, S_x = -\frac{1}{2}\hbar\rangle$, or $|x = 3cm, x = 7cm\rangle$, either of these being a contradiction.

At this stage there seems to be no point to enclosing the description of the state of a system within the symbol $|\ \rangle$, but, as will be seen later, quantum systems have the schizophrenic property of behaving as if they are *simultaneously* in a number of different states, and the way that this property of quantum systems is represented mathematically is by treating the states of a quantum system as if they are vectors in a way that we will be discussing later. Hence the above symbol is also known as a *ket vector* or *state vector*.

At times, when we want to be less specific, we would write $|x\rangle$ as being the state in which the x position of a particle is known, or $|p_x\rangle$ to be the state in which the x component of momentum is known, though the value in each case is unspecified. Finally, if we want to talk about a state without trying to spell out just what it is that is known precisely, i.e. an arbitrary state, we will write $|\psi\rangle$ or $|\phi\rangle$ or some such symbol. As a companion to this notation we will introduce a second way of writing down what this state is:

$$\left\langle \begin{array}{l} \text{All the data concerning the system that can be known with-} \\ \text{out mutual interference or contradiction.} \end{array} \right|. \tag{7.2}$$

This symbol is known as a *bra* or *bra vector* and is equally well a way of representing the state of a quantum system. The distinction between a bra and a ket lies in the mathematics of quantum mechanics, which we will be dealing with later.

To see how this notation is motivated and employed, and how it acquires a mathematical meaning, we return to the two slit experiment in Section 7.2, but before doing so, we will have a brief look at some 'philosophical' issues connected with the notion of the state of a quantum system.

7.1.1 Limits to knowledge

There is a subtle issue tucked away inside this superficially innocent argument. It would appear that we are merely doing for the state of a quantum system exactly the same as we do for a classical system, with the proviso that we are a little bit restricted in what we can do in the quantum case. But this is an important distinction. In the case of a classical system, the state is the totality of all that it is in principle possible to know about the system — it is is a perfect summary of what the system is 'really' doing, the data we have is a perfect reflection of the objectively real state of affairs that pertains to the system. However, in the case of a quantum system, there is a limit to what we can know — any missing information is not knowable, even in principle. So, in a very real sense, we do not know and cannot know what a quantum system is actually doing, if indeed it is 'actually doing' anything in any precise way. But could we not take the attitude that the system really is doing something well-defined, and we simply do not know what it is? That is to say, just because we have precise knowledge of a particle's position, could not that also mean that the particle really has a specific position, and for some perverse reason known only to Mother Nature, we are not allowed to also partake in this information? The uncomfortable fact is that there are very good reasons to believe that any unknown information concerning a quantum system really does mean that the information does not exist in the first place, i.e. if we know the momentum of a particle, then according to quantum mechanics it not that the particle really does have a position that we are not 'allowed' to know, but that it does not have a position in the first place.

7.2 The Two Slit Experiment Revisited

We now want to recast the two slit experiment in a fashion that leads to a new way of formulating quantum mechanics. The argument to be presented here is not meant to be rigorous, but more to suggest a way of thinking about quantum mechanics that can be made more precise, and very general in that it can be applied to any physical system. The experimental set up will be as before. Particles, which have all passed through exactly the same preparation procedure and hence are presumably all in the same state, are produced at a source S. These particles are then incident on a screen in which there are two narrow slits, 1 and 2, and beyond this screen is a further screen which the particles will ultimately strike, the point at which they hit this screen being registered in some way.

Since the particles are all prepared in the same way, they will presumably all be associated with the same wave function. The detailed form of the wave function will depend on many things, not the least of which is that the particles are all produced at the source S. So, to remind us of this, we will introduce the notation for the wave function

$$\Psi_S(x) = \text{probability amplitude of finding the particle at } x \text{ given that it originated at the source } S. \tag{7.3}$$

We will assume this wave function is unity at the source, i.e.

$$\Psi_S(S) = 1. \tag{7.4}$$

It is not important that we do this. We could choose otherwise, only to find that it cancels out at the end. Furthermore, we will not be concerning ourselves with the possible time dependence of the wave function here – in effect we are assuming some kind of steady state. This wave function will propagate through space and at the positions of the two slits will have values $\Psi_S(n)$, n =1, 2. These waves will then pass through these slits on their way towards the observation screen. The task then is to determine the amplitude of the waves at a point x on the observation screen due to waves that originated at each of the slits. To do that we will first of all suppose that, if the amplitude of the wave incident on slit 1 is unity, then the resulting wave amplitude at a position

x on the screen is $\Psi_1(x)$. But since the amplitude of the wave at slit 1 is $\Psi_S(1)$ then it appears reasonable to suppose that we scale up the amplitude of the wave from slit 1 incident on the observation screen at x by the same factor, i.e. $\Psi_S(1)$, so that the amplitude of the wave incident on the screen at x will be $\Psi_S(1)\Psi_1(x)$. Of course, this is an assumption, but it is what is observed for, say, light waves passing through a slit – if the amplitude of a light wave incident on a slit is doubled, then the amplitude of the wave that emerges from the slit is also doubled. Light waves are *linear*. Probability amplitudes are hence also assumed to be linear.

In a similar way, the amplitude at x due to waves from slit 2 will be $\Psi_S(2)\Psi_2(x)$. Consequently, the total amplitude of the wave at x will be

$$\Psi_S(x) = \Psi_S(1)\Psi_1(x) + \Psi_S(2)\Psi_2(x) \tag{7.5}$$

where $\Psi_S(x)$ is the amplitude of waves at x that originated from the source S. This then is the probability amplitude of observing a particle at x given that it originated from the source S, i.e. by the Born interpretation Eq. (4.15), $|\Psi_S(x)|^2 dx$ is the probability density of observing a particle in the region x to $x + dx$. This result is the same as the one presented in Eq. (4.14), but here we have expressed this total probability amplitude as a sum of two contributions of the form

$$\begin{aligned}
\Psi_S(n)\Psi_n(x) = &\text{ Probability amplitude of observing the particle at slit } n \\
&\text{ given that it originated from the source } S. \\
\times &\text{ Probability amplitude of observing the particle at } x \text{ given} \\
&\text{ that it originated from slit } n. \\
= &\text{ Probability amplitude for the particle to pass from the} \\
&\text{ source } S \text{ to point } x \text{ through slit } n.
\end{aligned} \tag{7.6}$$

7.2.1 Sum of Amplitudes in Bra(c)ket Notation

It is at this point that we make use of the new notation we introduced in the preceding Section. Consider, for example, $\Psi_S(n)$, which we stated above as being the probability amplitude of observing the particle at slit n given that it originated from the source S. We could equivalently say that $\Psi_S(n)$ is the probability amplitude of observing the particle at the position of slit n, given that it was originally at the position of the source S, or even more precisely, that it is the probability amplitude of observing the particle to be in the state in which it is at the position of slit n, given that it was in a state in which it was at the position of the source S. If we use our notation from the previous Section to write

$$\begin{aligned}
|S\rangle &\equiv \text{state of the particle when at the position of the source } S \\
|n\rangle &\equiv \text{state of the particle when at the position of slit } n \\
|x\rangle &\equiv \text{state of the particle when at the position } x
\end{aligned} \tag{7.7}$$

we can then write, for instance,

$$\begin{aligned}
\Psi_S(n) = &\text{ Probability amplitude of observing the particle in state } |n\rangle \\
&\text{ given that it was in state } |S\rangle.
\end{aligned} \tag{7.8}$$

This we will, finally, write as

$$\Psi_S(n) = \langle n||S\rangle = \langle n|S\rangle \tag{7.9}$$

i.e. we have written the final state as a bra, and where we have replaced the double vertical bar by a single bar. We can make similar replacements for the other probability amplitudes:

$$\Psi_S(x) \to \langle x|S\rangle; \quad \Psi_n(x) \to \langle x|n\rangle \tag{7.10}$$

i.e., for instance, $\langle x|S\rangle$ is the probability amplitude of finding the particle at x, i.e. in the state $|x\rangle$, given that it was initially at the source S, i.e. in the state $|S\rangle$. Recalling that a symbol such as $|S\rangle$ is also known as a ket, we can trace the origin of the names bra and ket to the fact that $\langle x|S\rangle$ can be thought of as a quantity enclosed in a pair of angled brackets, or 'bra(c)kets'. In terms of this new notation, the above result Eq. (7.5) becomes

$$\langle x|S\rangle = \langle x|1\rangle\langle 1|S\rangle + \langle x|2\rangle\langle 2|S\rangle. \tag{7.11}$$

Being able to write the probability amplitude in this way is a particularly important result as it leads directly to a new way of looking at the idea of the state of a physical system that lies at the heart of quantum mechanics.

7.2.2 Superposition of States for Two Slit Experiment

We can note that the expression Eq. (7.11) will hold true for all values of the variable x, i.e. it does not hold true for just one value of x. Because of this, we can do something rather radical and that is 'cancel' the $\langle x|$ to leave

$$|S\rangle = |1\rangle\langle 1|S\rangle + |2\rangle\langle 2|S\rangle \tag{7.12}$$

with the understanding that what we have created by doing so is a template into which we insert $\langle x|$ when we so desire, and thereby regain the expression Eq. (7.11). As a result of this step, we have apparently given a new meaning to the ket $|S\rangle$ as being more than just a way of summarizing all the information on the state of the system. The expression just obtained seems to suggest that there is a deeper mathematical and perhaps physical meaning that can be assigned to $|S\rangle$. For instance, we are free to manipulate Eq. (7.12) as we see fit. For instance we could solve for $|2\rangle$ to give

$$|2\rangle = \frac{|S\rangle - |1\rangle\langle 1|S\rangle}{\langle 2|S\rangle} \tag{7.13}$$

(recall that the $\langle \ldots | \ldots \rangle$ are all just complex numbers) and so on, and then put $\langle x|$ back in to give

$$\langle x|2\rangle = \frac{\langle x|S\rangle - \langle x|1\rangle\langle 1|S\rangle}{\langle 2|S\rangle}. \tag{7.14}$$

However, this development in the notation has more to offer than just this. Having created this new kind of expression, it is worthwhile to see whether or not it can be given any useful meaning. The new interpretation is a very potent one, and constitutes the central feature of quantum mechanics, the idea that a system, in some sense, can be simultaneously in a number of different physical states, otherwise known as a 'superposition of states'.

To see how this interpretation can be arrived at, we first of all note that since $\langle n|S\rangle$ is the probability amplitude of finding the particle at slit n, given that it was initially at the source S, and similarly for $\langle x|S\rangle$, $\langle x|n\rangle$, $\langle n|S\rangle$, then

$$\begin{aligned}
|\langle x|S\rangle|^2 &= \text{probability of finding the particle in state } |x\rangle \text{ given that it was in state } |S\rangle \\
|\langle x|n\rangle|^2 &= \text{probability of finding the particle in state } |x\rangle \text{ given that it was in state } |n\rangle \\
|\langle n|S\rangle|^2 &= \text{probability of finding the particle in state } |n\rangle \text{ given that it was in state } |S\rangle,
\end{aligned} \tag{7.15}$$

so that the coefficients $\langle 1|S\rangle$ and $\langle 2|S\rangle$ in Eq. (7.12) in some sense determine 'how much' of the state $|1\rangle$ is to be found in the initial state $|S\rangle$ and 'how much' of state $|2\rangle$ is to be found in $|S\rangle$. Put another way, the expression Eq. (7.12) indicates in a symbolic way the fact that when the particle is prepared in its initial state $|S\rangle$, there is built into this state the *potential* of the particle

being found in the other states $|n\rangle$, with the chances of the particle being found in these other states being given by the coefficients $\langle n|S\rangle$. It is this sort of result that lies at the heart of quantum mechanics – the idea that a system in a certain state can behave as if it is in one or another of a number of other states, and to do so in a way that is probabilistic in nature. It is also out of this expression that the basic mathematical formalism of quantum mechanics is built, the idea being that the kets $|n\rangle$ can be considered to be, in some sense, vectors, much like the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ encountered in mechanics, and the coefficients $\langle n|S\rangle$ are the components of the total vector $|S\rangle$ along these two directions. It is this analogue that we will pursue in some detail in later Chapters. But for the present, we will show that the result Eq. (7.11) which we have 'derived' in the case of the two slit experiment can also be shown (and much more rigorously) to be an intrinsic part of the Stern-Gerlach experiment. In fact Eq. (7.11) is of a general form that is taken, ultimately, as a central postulate of quantum mechanics, one that is suggested by circumstantial evidence, though not proven. In effect, it is a law of nature.

7.3 The Stern-Gerlach Experiment Revisited

The aim here is to show that a result equivalent to the sum of amplitudes for different paths result obtained for the two slit experiment can also be shown to arise in the Stern-Gerlach experiment. However, unlike the previous result for the two slit arrangement, the result to be obtained here can be shown to be more rigorously based and more useful in the long run as providing the insight necessary to generalize the result to any physical system.

7.3.1 Probability amplitudes for particle spin

Further development then follows what was done in Section 7.2 in that we use the sum of probability amplitudes result to arrive at the notion of the spin state of the system being expressible as a 'superposition' of other possible states.

Without affecting the generality of the argument, we will specialize to the case illustrated in Fig. (7.1).

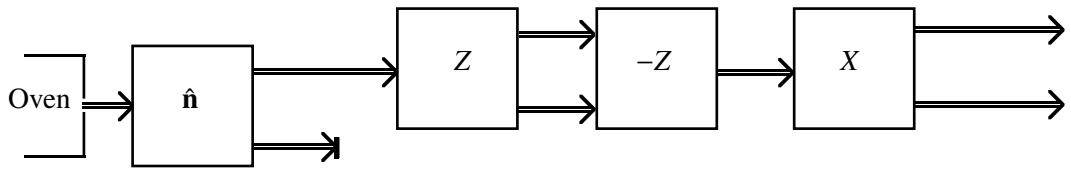


Figure 7.1: Atomic beam for which $S_i = \frac{1}{2}\hbar$ split into $S_z = \pm\frac{1}{2}\hbar$ beams and then recombined before passing through final Stern-Gerlach device with magnetic field in x direction.

This is the same setup as presented in Fig. (6.9) except that here the atom can emerge from the first Stern-Gerlach device in one or the other of two separate beams corresponding to the atomic spin component $S_i = \mathbf{S}\cdot\hat{\mathbf{n}} = \pm\frac{1}{2}\hbar$ where $\hat{\mathbf{n}}$ is some arbitrary orientation of the magnetic field (in the XZ plane) in the device. The atoms in one of the beams ($S_i = \frac{1}{2}\hbar$) is then selected and passed through a Stern-Gerlach device where the magnetic field further separates this beam according to its z component of spin. The beams are then recombined before entering the final Stern-Gerlach apparatus, this arrangement being made so as to erase the information on which beam the each atom emerged from the preceding apparatus.

We are now going to make use of the analogy of this set-up with the two slit interference experiment, and the assumption that the observed measurement outcomes at the final Stern-Gerlach apparatus are of quantum origin, to construct assign a probability interpretation to the observed

behaviour of particle spin. Thus, we will assume that the oven plus the first Stern-Gerlach device is the equivalent of the source of identically prepared particles in the two slit experiment, except that here the atoms are all identically prepared to have $S_i = \mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar$. The next Stern-Gerlach device is analogous to the two slits in that the atoms can, in principle, emerge along two different paths corresponding to $S_z = \pm\frac{1}{2}\hbar$. We can tell which path an atom follows (i.e. via the $S_z = \frac{1}{2}\hbar$ or the $S_z = -\frac{1}{2}\hbar$ beam) by monitoring which beam an atom emerges from after it passes through the first z oriented Stern-Gerlach device in much the same way that we can monitor which slit a particle passes through in the two slit experiment.

We will then assign a probability amplitude for an atom to pass along either of the $S_z = \pm\frac{1}{2}\hbar$ beams, written:

$$\begin{aligned}\langle S_z = \pm\tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle = &\quad \text{Probability amplitude of observing the atom to have } S_z = \\ &\quad \pm\tfrac{1}{2}\hbar \text{ given that originally it had an } \hat{\mathbf{n}} \text{ component of spin} \\ &\quad S_i = \tfrac{1}{2}\hbar. \\ = &\langle \pm | S\rangle\end{aligned} \tag{7.16}$$

where we have simplified the notation a little: $\langle S_z = \pm\frac{1}{2}\hbar | S_i = \frac{1}{2}\hbar\rangle \rightarrow \langle \pm | S\rangle$.

The atomic beams are then recombined and finally, after passing through the last Stern-Gerlach device, emerge with $S_x = \pm\frac{1}{2}\hbar$. We then write

$$\begin{aligned}\langle S_x = \pm\tfrac{1}{2}\hbar | S_z = \tfrac{1}{2}\hbar\rangle = &\quad \text{Probability amplitude of observing the atom to have } S_x = \\ &\quad \pm\tfrac{1}{2}\hbar \text{ given that it had a } z \text{ component of spin } S_z = \tfrac{1}{2}\hbar. \\ = &\langle S' | +\rangle\end{aligned} \tag{7.17}$$

with a similar definition of $\langle S_x = \pm\frac{1}{2}\hbar | S_z = -\frac{1}{2}\hbar\rangle \rightarrow \langle S'|-\rangle$ where S' can have the values of $\pm\frac{1}{2}\hbar$. We can then construct the probability amplitude of measuring the x component of spin to have the value S' given that it initially the $\hat{\mathbf{n}}$ component of spin $\mathbf{S}\cdot\hat{\mathbf{n}} = S = \frac{1}{2}\hbar$ either by analogy with what applied in the two slit experiment, or by use of the more general argument of Sections 7.3.3 and 8.3.2. Either way, this probability amplitude is given by

$$\langle S' | S\rangle = \langle S' | +\rangle\langle +| S\rangle + \langle S'|-\rangle\langle -| S\rangle. \tag{7.18}$$

We now want to obtain form this result the idea of the spin state of the atoms emerging from the first Stern-Gerlach apparatus as being 'made up of' the states $|\pm\rangle$.

7.3.2 Superposition of States for Spin Half

Proceeding as in the two slit result, we can argue that this is a result that holds for all final states $|S'\rangle$, so that we might as well 'cancel' the common factor '$\langle S'|$' in Eq. (7.18) to give a new expression for the state $|S\rangle$, that is

$$|S\rangle = |+\rangle\langle +| S\rangle + |-\rangle\langle -| S\rangle. \tag{7.19}$$

with the understanding that we can reintroduce '$\langle S'|$' for any chosen final state, yielding an expression for the probability amplitudes as needed. What Eq. (7.19) effectively represents is a 'template' into which we insert the appropriate information in order to recover the required probability amplitudes.

We have once again shown how the state of a physical system can, in a sense, be expressed in terms of other possible states of the system, with a weighting that determines the probability of observing the system in each of these other states. To see how this interpretation can be arrived at,

we first of all note that since $\langle \pm | S \rangle$ is the probability amplitude of the z component of spin having the values $\pm\frac{1}{2}\hbar$ given that it initially the $\hat{\mathbf{n}}$ component of spin was $S = \frac{1}{2}\hbar$, and similarly for $\langle S' | S \rangle$, $\langle S' | \pm \rangle$, $\langle \pm | S \rangle$, then

$$\begin{aligned} |\langle S' | S \rangle|^2 &= \text{probability of the atomic spin being in state } |S'\rangle \text{ given that} \\ &\quad \text{it was in state } |S\rangle \\ |\langle S' | \pm \rangle|^2 &= \text{probability of the atomic spin being in the state } |S'\rangle \text{ given} \\ &\quad \text{that it was in state } |\pm\rangle \\ |\langle \pm | S \rangle|^2 &= \text{probability of the atomic spin being in the state } |\pm\rangle \text{ given} \\ &\quad \text{that it was in state } |S\rangle, \end{aligned} \tag{7.20}$$

so that the coefficients $\langle + | S \rangle$ and $\langle - | S \rangle$ in Eq. (7.19) in some sense determine 'how much' of the state $|+\rangle$ is to be found in the initial state $|S\rangle$ and 'how much' of state $|-\rangle$ is to be found in $|S\rangle$. Put another way, the expression Eq. (7.19) indicates in a symbolic way the fact that when the atomic spin is prepared in its initial state $|S\rangle$, there is built into this state the *potential* of the spin to be found in the other states $|\pm\rangle$, with the chances of the particle being found in these other states being given by the coefficients $\langle \pm | S \rangle$. Once again, we see the possibility of a system in a certain state behaving as if it is in one or another of a number of other states, and to do so in a way that is probabilistic in nature.

7.3.3 A derivation of sum-over-probability-amplitudes for spin half

The task here is to show that the sum of probability amplitudes expression Eq. (7.18) can be extracted in the case of spin half from the assumed experimentally observed spin probability. This is a purely mathematical exercise which, nevertheless, has a clear physical interpretation in terms of a particular kind of Stern-Gerlach experimental arrangement that is analogous of the two slit experiment. As a bonus, it turns out that it is also possible to derive explicit expressions for spin half probability amplitudes.

Here will consider a fairly general kind of experimental arrangement illustrated in Fig. (7.2) in which an atomic beam is passed through two consecutive Stern-Gerlach devices. It should be noted here that in contrast to, for instance, Fig. (6.9), a slightly different choice of axes has been made. The choice made is such that the angle of orientation of the magnetic fields is the ϕ angular coordinate of a set of spherical polar coordinates. Thus, it will be assumed that each Stern-Gerlach device will have their magnetic fields lying in the XY plane, so that the beam of atoms is heading in the z direction. Apart from that, the magnetic fields of the two devices can be oriented in an arbitrary direction, $\hat{\mathbf{n}}$ for the first, and $\hat{\mathbf{m}}$ for the second.

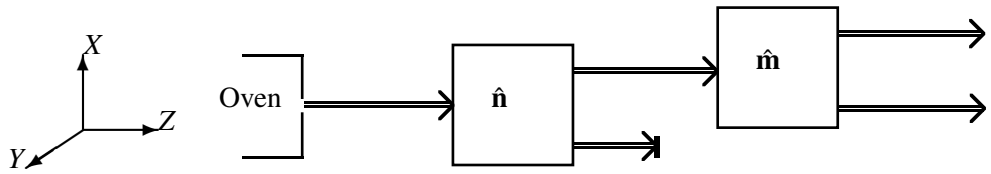


Figure 7.2: Similar to Fig. (6.4) but with magnetic field of second Stern-Gerlach device allowed to be oriented in an arbitrary direction $\hat{\mathbf{m}}$ direction.

Here, the first magnetic field makes an angle of ϕ_i with the x direction and the second an angle of ϕ_f with the x direction, see Fig. (7.3).

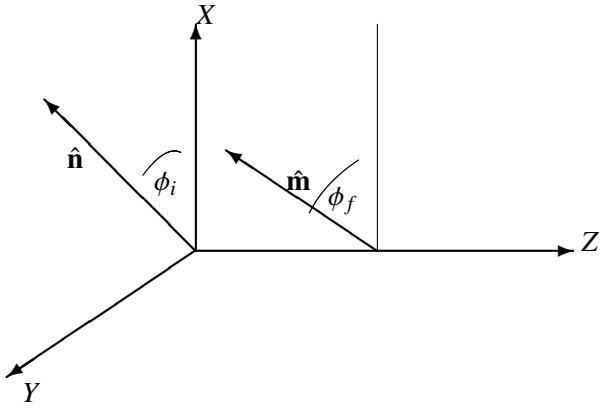


Figure 7.3: Orientation of the magnetic fields in the two Stern-Gerlach devices in Fig. (7.2)

The angle between the directions of the two magnetic fields is $\phi = \phi_i - \phi_f$. Consequently, the probability of an atom emerging in the beam for which $S_f = \mathbf{S}\cdot\hat{\mathbf{m}} = \frac{1}{2}\hbar$, given that it entered the last Stern-Gerlach device with $S_i = \mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar$ will be, from Eq. (6.29) (using $\cos\phi = \cos(-\phi)$)

$$P(S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar) = \cos^2(\tfrac{1}{2}\phi) = \cos^2[(\phi_f - \phi_i)/2]. \tag{7.21}$$

The question we want to address here is the following: is it possible to show that this probability (which we are assuming has been determined by experiment) is in any way to be understood as being of purely quantum mechanical origin?

One way of doing this is to show that this probability is consistent with its interpretation as the modulus squared of a possibly complex probability amplitude, and moreover, that this complex probability amplitude can be written as the sum over probability amplitudes associated with different ways that the spin of the atom can arrive at its final observed value, in a way analogous to the two slit experiment. Thus, we begin this investigation by writing

$$P(S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar) = |\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle|^2 \tag{7.22}$$

where $\langle S_f = \frac{1}{2}\hbar|S_i = \frac{1}{2}\hbar\rangle$ is, following the notation used in the previous Section, the probability amplitude of observing the $\hat{\mathbf{m}}$ component of the spin of an atom to be $\frac{1}{2}\hbar$ given that the $\hat{\mathbf{n}}$ component is known to be $\frac{1}{2}\hbar$.

We now want to show that $\langle S_f = \frac{1}{2}\hbar|S_i = \frac{1}{2}\hbar\rangle$ can indeed be written as a sum of contributions associated with probability amplitudes of the spins 'passing through' some other states, as in Eq. (7.11) for the two slit case, prior to the final measurement of S_f. These 'other states' would be those in which the atom has a spin component $S_I = \mathbf{S}\cdot\hat{\mathbf{l}} = \pm\frac{1}{2}\hbar$ in some direction $\hat{\mathbf{l}}$. So, by analogy with the two slit result, what we are aiming to do here is to show that

$$\begin{aligned}\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle =& \langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle \\ &+ \langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle\langle S_I = -\tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle\end{aligned} \tag{7.23}$$

which is to be interpreted as meaning that the probability amplitude $\langle S_f = \frac{1}{2}\hbar|S_i = \frac{1}{2}\hbar\rangle$ of an atom being found in state $|S_f = \frac{1}{2}\hbar\rangle$ given that it was initially in the state $|S_i = \frac{1}{2}\hbar\rangle$ is the sum of the probability amplitudes of the atomic spin 'passing through', without observation, the two intermediate states $|S_I = \pm\frac{1}{2}\hbar\rangle$.

This is, of course, a purely mathematical exercise, but rather remarkably it nevertheless leads to a result that can be given meaning in terms of the experimental set-up shown in Fig. (7.4)). This is a set-up which involves using a Stern-Gerlach device with magnetic field in the direction $\hat{\mathbf{l}}$ aligned at some angle ϕ_I. Atoms exiting such a device will emerge with spin components $S_I = \mathbf{S}\cdot\hat{\mathbf{l}}$ that, as

usual, can have the two values $\pm\frac{1}{2}\hbar$. In a sense, these two beams are much like the slits in the two slit experiment. A second device in the direction $-\hat{\mathbf{l}}$ merely serves to bring the two beams back together again prior to entering the final Stern-Gerlach device.

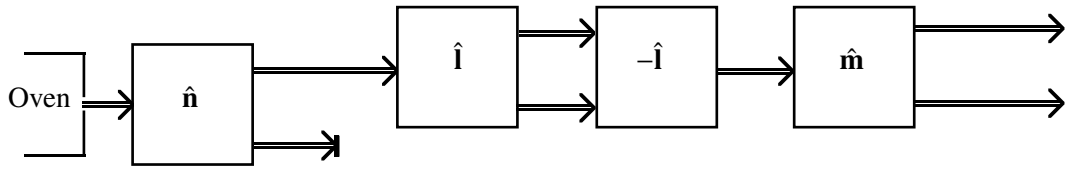


Figure 7.4: Atoms exiting from the first Stern-Gerlach apparatus with $S_i = \frac{1}{2}\hbar$ and entering the second can emerge in either of two ($S_l = \pm\frac{1}{2}\hbar$) beams. The beams are then recombined, so erasing information on which beam any individual atom might emerge in, before passing through a final Stern-Gerlach device which measures the component of the atomic spin in the $\hat{\mathbf{m}}$ direction.

What this setup does is to present the experimenter with the opportunity of measuring any component of atomic spin (in the XY plane only, in this case) of an atom, prior to a final measurement of S_f. As we shall see, if no observation is made as to which beam each atom emerges from the second Stern-Gerlach device, then the final observed results for S_f are the same as if the intervening Stern-Gerlach apparatus was not present at all, while if an intervening observation *is* made, then the observed result is consistent with the absence of interference effects.

That the result Eq. (7.23) is implied by the probability Eq. (7.21) requires us to write down the possible mathematical form for the probability amplitude $\langle S_f = \frac{1}{2}\hbar | S_i = \frac{1}{2}\hbar\rangle$. Starting with Eq. (7.21) and Eq. (7.22)

$$P(S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar) = \cos^2[(\phi_f - \phi_i)/2] = |\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle|^2 \tag{7.24}$$

it follows that

$$\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle = \cos[(\phi_f - \phi_i)/2] e^{i\varphi(\phi_f - \phi_i)} \tag{7.25}$$

where $\exp(i\varphi(\phi_f - \phi_i))$ is a phase factor which is unknown except that φ is a function of the angular separation $\phi_f - \phi_i$ only — it is only the difference in the angular settings of the Stern-Gerlach devices that can be of physical significance. From this we have

$$\cos[(\phi_f - \phi_i)/2] = \langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle e^{-i\varphi(\phi_f - \phi_i)} \tag{7.26}$$

a result that we will make use of shortly. But before that, we need to write this cos factor in a way that will lead to an interpretation of the spin system passing through an intermediate Stern-Gerlach apparatus with its magnetic field oriented at an angle ϕ_I. We can do this by expanding out the cos term by writing

$$\cos[\tfrac{1}{2}(\phi_f - \phi_i)] = \cos[\tfrac{1}{2}(\phi_f - \phi_I) + \tfrac{1}{2}(\phi_I - \phi_i)] \tag{7.27}$$

and using standard trigonometric formulae to give

$$\begin{aligned}\cos[\tfrac{1}{2}(\phi_f - \phi_I) + \tfrac{1}{2}(\phi_I - \phi_i)] \\ = \cos[\tfrac{1}{2}(\phi_f - \phi_I)]\cos[\tfrac{1}{2}(\phi_I - \phi_i)] - \sin[\tfrac{1}{2}(\phi_f - \phi_I)]\sin[\tfrac{1}{2}(\phi_I - \phi_i)].\end{aligned} \tag{7.28}$$

The second, sin dependent term becomes, on further use of simple trigonometry:

$$\sin[\tfrac{1}{2}(\phi_f - \phi_I)]\sin[\tfrac{1}{2}(\phi_I - \phi_i)] = -\cos[\tfrac{1}{2}(\phi_f - (\phi_I + \pi))]\cos[\tfrac{1}{2}((\phi_I + \pi) - \phi_i)] \tag{7.29}$$

so that overall we get

$$\begin{aligned}\cos[(\phi_f - \phi_i)/2] \\ = \cos[\tfrac{1}{2}(\phi_f - \phi_I)]\cos[\tfrac{1}{2}(\phi_I - \phi_i)] + \cos[\tfrac{1}{2}(\phi_f - (\phi_I + \pi))]\cos[\tfrac{1}{2}((\phi_I + \pi) - \phi_i)].\end{aligned} \tag{7.30}$$

If we now make use of Eq. (7.26) (with corresponding expressions for $\langle S_I = \frac{1}{2}\hbar|S_i = \pm\frac{1}{2}\hbar\rangle$) and so on, we can write this as

$$\begin{aligned}
&\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle e^{-i\varphi(\phi_f-\phi_i)} \\
&= \langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle e^{-i\varphi(\phi_f-\phi_I)}\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle e^{-i\varphi(\phi_I-\phi_i)} \\
&+ \langle S_f = \tfrac{1}{2}\hbar|S_i = -\tfrac{1}{2}\hbar\rangle e^{-i\varphi(\phi_f-(\phi_I+\pi))}\langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle e^{-i\varphi((\phi_I+\pi)-\phi_i)}.
\end{aligned} \tag{7.31}$$

To get the required result Eq. (7.23), the various phase factors must cancel, which requires

$$e^{-i\varphi(\phi_f-\phi_i)} = e^{-i\varphi(\phi_f-\phi_I)}e^{-i\varphi(\phi_I-\phi_i)} = e^{-i\varphi(\phi_f-(\phi_I+\pi))}e^{-i\varphi((\phi_I+\pi)-\phi_i)}. \tag{7.32}$$

If we, for instance, put $\alpha = \phi_f - \phi_I$ and $\beta = \phi_i - \phi_I$, and put $F(x) = \exp(-i\varphi(x))$, then the first of these equations can be written

$$F(\alpha - \beta) = F(\alpha)F(\beta) \tag{7.33}$$

and hence $F(x)$ must be an exponential function whose exponent is linear in x:

$$F(x) = e^{-iax} = e^{-i\varphi(x)} \tag{7.34}$$

where a is a constant. Hence

$$\varphi(x) = ax + 2n\pi \tag{7.35}$$

where n is an integer. The term $2n\pi$ is the same in all the phase factors, so it will always cancel out, so we might as well set n to zero. Thus we end up with

$$\cos[(\phi_f - \phi_i)/2] = \langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle e^{-ia(\phi_f-\phi_i)} \tag{7.36}$$

or

$$\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle = \cos[(\phi_f - \phi_i)/2]e^{ia(\phi_f-\phi_i)} \tag{7.37}$$

and similarly for the other cos factors.

We can pin down the phase factor a little further by noting that in $\langle S_f = -\frac{1}{2}\hbar|S_i = \frac{1}{2}\hbar\rangle$ we could replace $\phi_f \to \phi_f + 2\pi$ without changing the physical meaning of the result. Thus we must have

$$\cos[(\phi_f - \phi_i)/2]e^{ia(\phi_f-\phi_i)} = \cos[(\phi_f - \phi_i + 2\pi)/2]e^{ia(\phi_f-\phi_i+2\pi)} \tag{7.38}$$

and hence that

$$e^{2\pi ai} = -1. \tag{7.39}$$

which can be satisfied if we put $a = n + \frac{1}{2}$ where n is an integer that can be chosen freely and is conventionally set equal to -1 so that

$$\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle = e^{-i(\phi_f-\phi_i)/2}\cos[\tfrac{1}{2}(\phi_f - \phi_i)] \tag{7.40}$$

So overall we have simply shown that Eq. (7.30) can be written

$$\begin{aligned}
e^{-i(\phi_f-\phi_i)/2}\cos[\tfrac{1}{2}(\phi_f - \phi_i)] &= e^{-i(\phi_f-\phi_I)/2}\cos[\tfrac{1}{2}(\phi_f - \phi_I)]e^{-i(\phi_I-\phi_i)/2}\cos[\tfrac{1}{2}(\phi_I - \phi_i)] \\
&+ e^{-i(\phi_f-\phi_I-\pi)/2}\cos[\tfrac{1}{2}(\phi_f - (\phi_I + \pi))]e^{-i(\phi_I+\pi-\phi_i)/2}\cos[\tfrac{1}{2}((\phi_I + \pi) - \phi_i)]
\end{aligned} \tag{7.41}$$

which becomes, using Eq. (7.40)

$$\begin{aligned}
\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle =& \langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle \\
&+ \langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle\langle S_I = -\tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle
\end{aligned} \tag{7.42}$$

as required. We also have from this the probability of measuring an atom to be in a state for which $S_f = \frac{1}{2}\hbar$:

$$\begin{aligned}
&P(S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar) \\
&= \left|\langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle + \langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle\langle S_I = -\tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle\right|^2 \\
&= \left|\langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle\right|^2 + \left|\langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle\langle S_I = -\tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle\right|^2 \\
&\quad + 2\mathrm{Re}\left[\langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle(\langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle\langle S_I = -\tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle)^*\right]
\end{aligned} \tag{7.43}$$

where the third term is an *interference* term. Thus we have managed to show that the probability distribution associated with passing spin half systems through a Stern-Gerlach device to prepare the spin of the atoms to have a value $S_i = \frac{1}{2}\hbar$, and then through a second to measure the spin components S_f can be directly interpreted as being the consequence of the interference of probability amplitudes in a way analogous to the two slit experiment.

An important feature of the result Eqs. (7.23) and (7.42) is that it is expressed in terms of the probability amplitudes of the system passing through one or the other of the intermediate states for which $S_I = \mathbf{S}\cdot\hat{\mathbf{I}} = \pm\frac{1}{2}\hbar$, i.e. there appears here probability amplitudes that would be associated with the use of an intermediate Stern-Gerlach device to separate the atoms according to their S_I component of spin. But this result was in a sense 'built-in' to the original mathematical expression for the probability Eq. (7.21), i.e. no mention was made of any actual intermediate Stern-Gerlach device when deriving the result Eq. (7.42). In other words, the observation probability, Eq. (7.43) can be written in a way that looks like there are interference contributions associated with the possibility of passing the atoms through an intervening Stern-Gerlach device, *even though no such Stern-Gerlach apparatus need ever appear in the actual experiment*. In fact, if we *do* insert the appropriate device with magnetic field in the $\hat{\mathbf{I}}$ direction to separate and then recombine the two beams, we find that *if there is no information available that specifies through which of the* $S_I = \pm\frac{1}{2}\hbar$ *beams the atoms pass*, then the probabilities for observing $S_f = \pm\frac{1}{2}\hbar$ for the recombined beam are exactly the same state as if the extra device were not present, i.e. the probability is given by $\cos^2(\frac{1}{2}(\phi_f - \phi_i))$, Eq. (7.21), which is independent of the intermediate states that appear in Eq. (7.42). Put another way, having the intermediate device present, but not observing what it does, is the same as not having it there at all. On the other hand if we *do* have information as to which beam the atoms come through, then the observation probability is

$$\begin{aligned}
P(S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar) = |\langle S_f = \tfrac{1}{2}\hbar|S_I = \tfrac{1}{2}\hbar\rangle\langle S_I = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle|^2 \\
+ |\langle S_f = \tfrac{1}{2}\hbar|S_I = -\tfrac{1}{2}\hbar\rangle\langle S_I = -\tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle|^2
\end{aligned} \tag{7.44}$$

where we have lost the interference terms of Eq. (7.43). This result *does* depend on the intermediate state that happens to be observed. These considerations show that built into the properties of the state of the spin system is the *potential* for the system to be observed in *any* other state $|S_I = \frac{1}{2}\hbar\rangle$, for which the relevant probability amplitude $\langle S_I = \frac{1}{2}\hbar|S_i = \frac{1}{2}\hbar\rangle$ is non-zero. It is this property of the states of quantum systems that leads to the formulation of the state of a quantum system as a vector belonging to a complex vector space: the language of vectors makes it possible to express one vector (i.e. state) in terms of other vectors (i.e. states).

7.4 The General Case of Many Intermediate States

The above analysis has only been presented for two cases, the two slit experiment and spin half, but both these examples can be readily generalized. Thus, for instance, we could consider an interference experiment in which there are multiple slits in the first barrier. The final result for the

probability amplitude of observing an electron striking the screen at position x after having set out from a source S will then be given by

$$\langle x|S\rangle = \sum_{n=1}^{N} \langle x|n\rangle\langle n|S\rangle \tag{7.45}$$

where we have supposed that there are N slits in the first barrier, so that the electron can pass through N intermediate states. Likewise, if we repeat the Stern-Gerlach experiment with spin 1 atoms, we would find that the atoms would emerge in one or the other of three beams corresponding to $S_z = -\hbar, 0, \hbar$. More generally, if the atoms have spin s (where $s = 0$ or $\frac{1}{2}$ or 1 or $\frac{3}{2}, \ldots$) then they will emerge in one of a total of $2s + 1$ different possible beams corresponding to $S_z = -s\hbar$, or $(-s + 1)\hbar$, or $\ldots$, or $(s - 1)\hbar$, or $s\hbar$. We can then expect that we could write

$$\langle S'|S\rangle = \sum_{n=1}^{2s+1} \langle S'|n - s - 1\rangle\langle n - s - 1|S\rangle \tag{7.46}$$

i.e. the atom can pass through $2s + 1$ intermediate states, where we have written $|n\rangle$ for the state in which the atom has a z component of spin $S_z = n\hbar$. In either case, the 'cancellation' trick could then be applied to give, for instance in the latter case,

$$|S\rangle = \sum_{n=-s}^{s} |n\rangle\langle n|S\rangle \tag{7.47}$$

so we recover a generalization of the result found for spin half, Eq. (7.19) that we ultimately will interpret as a relationship between vectors.

We now want to move beyond these two particular cases. In effect, we want to extract from the above results concepts that can be applied to *any* physical system in the hope that what is obtained is a fundamental quantum principle that would apply to all physical systems, i.e. it would constitute a new law of nature.

It is clearly the case in the above examples that it is not the probability of occurrence of random events that is fundamental, rather it is probability amplitudes. The probabilities are derived from probability amplitudes – they are obtained by squaring an appropriately determined probability amplitude. Furthermore, these probability amplitudes satisfy a particular law concerning the role of intermediate states: the probability amplitude $\langle\phi|\psi\rangle$ for finding a system in a state $|\phi\rangle$ given that it was in the state $|\psi\rangle$ can be written in a way analogous to the particular examples considered here, that is, as a sum over the probability amplitudes associated with all the different 'pathways' by which a system can make its way from the initial state $|\psi\rangle$ to the final state $|\phi\rangle$ via a collection of intermediate states, i.e. we can always write

$$\langle\phi|\psi\rangle = \sum_{n} \langle\phi|n\rangle\langle n|\psi\rangle \tag{7.48}$$

It is this relationship that constitutes the fundamental law of quantum mechanics. Its natural consequence, obtained by taking the sum over intermediate states concept to its logical extreme, is the path integral formulation of quantum mechanics due to Feynman, the 'version' of quantum mechanics that can be applied to essentially any physical system whatsoever, including the universe itself, though it can yield mathematical formulae to which it can be very difficult to give a meaning. But it also yields the familiar forms of quantum mechanics such as wave mechanics and the Schrödinger equation.

Our aims here are much more modest, so we turn to the important consequence of Eq. (7.48) that we have already made much use of, that is the result obtained by the cancellation trick:

$$|\phi\rangle = \sum_{n} |n\rangle\langle n|\psi\rangle. \tag{7.49}$$

This is an expression that can be interpreted as describing the physical fact that any system in a given state can behave as if it is in some sense made up of other distinct states, here, the 'intermediate states' $|n\rangle$ with the probability amplitudes providing a weighting to be given to the different possibilities in the sense that, $|\langle n|\psi\rangle|^2$ is the probability that the system could be observed to be in the state $|n\rangle$. The challenge is to identify the 'intermediate states' for any given system. Moreover, this relationship can be given an interpretation in which the state of a quantum system can be considered as an abstract vector in some kind of vector space, i.e. we can set up a mathematical description of the quantum properties of a physical system in terms of the mathematical formalism of complex vector spaces.

7.5 Probabilities vs probability amplitudes

The fundamental formula of quantum mechanics, Eq. (7.48) has a very close resemblance to a well known formula of classical probability theory. We can analyse the 'sum-over-paths' picture according to these rules. So suppose, as we have above we have a system initially prepared in some initial state $|\psi\rangle$, and we are interested in calculating the probability of finding the system in some final state $|\phi\rangle$. (If we had never heard of quantum mechanics, we would have to consider the Dirac notation manner of writing down the state of a system as nothing much more than a rather fancy way of writing out some data). Further suppose that the system can access this final state by different pathways via some intermediate states $|n\rangle$. Then, according to classical probability theory, the probability of finding the system in state $|\phi\rangle$ given that it was originally in the state $|\psi\rangle$ is given by

$$P_{cl}(\phi|\psi) = \sum_n P(\phi|n)P(n|\psi)$$

i.e. the system starts in the state $|\psi\rangle$, and has the probability $P(n|\psi)$ that it will end up in the state $|n\rangle$, and once it 'arrives there' we then have the probability $P(\phi|n)$ that it will then end up in the state $|\phi\rangle$ given that it was in the state $|n\rangle$. Multiplying these two probabilities will then give the probability of the system starting from $|\psi\rangle$ and passing through $|n\rangle$ on its way to the final state $|\phi\rangle$. We then sum over all the possible intermediate states $|n\rangle$ to give the total probability of arriving in the state $|\phi\rangle$. However, what is found experimentally is that *if the system is never observed in any of the intermediate states* $|n\rangle$, this probability is not given by this classical result – the measurements show evidence of interference effects, which can only be understood if this probability is given as the square of a 'probability amplitude'.

Thus, we find we must write $P(\phi|\psi) = |\langle\phi|\psi\rangle|^2$, where $\langle\phi|\psi\rangle$ is a complex number, generally referred to as the probability amplitude of finding the system in state $|\phi\rangle$ given that it was in state $|\psi\rangle$, and this probability amplitude is then given by

$$\langle\phi|\psi\rangle = \sum_n \langle\phi|n\rangle\langle n|\psi\rangle \tag{7.50}$$

where the sum is over the probability amplitudes of the system passing through all the possible intermediate states $|n\rangle$.

If we square this result we find that

$$\begin{aligned} P(\phi|\psi) = |\langle\phi|\psi\rangle|^2 &= \sum_n |\langle\phi|n\rangle|^2|\langle n|\psi\rangle|^2 + \text{cross terms} \\ &= \sum_n P(\phi|n)P(n|\psi) + \text{cross terms}. \end{aligned} \tag{7.51}$$

We note that this expression consists, in part, of the classical result Eq. (7.5) but there is, in addition, cross terms. It is these terms that are responsible for the interference effects such as

those observed in the two slit experiment, or in the Stern-Gerlach experiment. If we observe which intermediate state the system 'passes through' it is always found that these interference terms are washed out, reducing the result to the classical result Eqn. (7.5). This we have seen in the case of two slit interference wherein if the slit through which the particle passes is observed, then the interference pattern on the observation screen is washed out.

We can see how the general comments made above look in the case of a spin half system for which we have

$$\langle S'|S\rangle = \langle S'|+\rangle\langle +|S\rangle + \langle S'|-\rangle\langle -|S\rangle. \tag{7.52}$$

To then calculate the probability of measuring the atomic spin to have the value $S' = \frac{1}{2}\hbar$, given that it was initially known to have the value $S = \frac{1}{2}\hbar$ requires us to calculate the square of the probability amplitude $\langle S'|S\rangle$:

$$\begin{aligned} P(S'|S) &= \left|\langle S'|S\rangle\right|^2 \\ &= \left|\langle S'|+\rangle\langle +|S\rangle\right|^2 + \left|\langle S'|-\rangle\langle -|S\rangle\right|^2 + 2\mathrm{Re}\left[\langle S'|+\rangle\langle +|S\rangle\langle S|-\rangle\langle -|S'\rangle\right] \end{aligned} \tag{7.53}$$

where the last term is the interference term. But if we observe, or measure, the z component of the atomic spin *before* we measure its final spin S', then we have seen above that the interference term is wiped out, so that we get

$$P_{\mathrm{cl}}(S'|S) = \left|\langle S'|+\rangle\langle +|S\rangle\right|^2 + \left|\langle S'|-\rangle\langle -|S\rangle\right|^2 = \left|\langle S'|+\rangle\right|^2 |\langle +|S\rangle|^2 + \left|\langle S'|-\rangle\right|^2 |\langle -|S\rangle|^2 \tag{7.54}$$

$$= P(S'|+)P(+|S) + P(S'|-)P(-|S). \tag{7.55}$$

It is this result that has an immediate interpretation according to the rules of classical probability theory:

Probability of measuring the spin to be $S' = \frac{1}{2}\hbar$ given that it was initially $S' = \frac{1}{2}\hbar$	=	Probability of measuring the spin to be $S' = \frac{1}{2}\hbar$ given it was observed to be $S_z = \frac{1}{2}\hbar$	× Probability of measuring the spin to be $S_z = \frac{1}{2}\hbar$ given it was initially $S' = \frac{1}{2}\hbar$
	+	Probability of measuring the spin to be $S' = \frac{1}{2}\hbar$ given it was observed to be $S_z = -\frac{1}{2}\hbar$	× Probability of measuring the spin to be $S_z = -\frac{1}{2}\hbar$ given it was initially $S = \frac{1}{2}\hbar$

i.e. this is the way we would have calculated the probability if we assumed the familiar rules of classical physics applied. Instead, quantum physics tells us that we must use what is almost the same rule, except that the rule is applied to probability amplitudes, not probabilities, i.e. Eq. (7.5) is replaced by

Probability amplitude of measuring the spin to be $S' = \frac{1}{2}\hbar$ given that it was initially $S' = \frac{1}{2}\hbar$	=	Probability amplitude of measuring the spin to be $S' = \frac{1}{2}\hbar$ given it was observed to be $S_z = \frac{1}{2}\hbar$	× Probability amplitude of measuring the spin to be $S_z = \frac{1}{2}\hbar$ given it was initially $S' = \frac{1}{2}\hbar$
	+	Probability amplitude of measuring the spin to be $S' = \frac{1}{2}\hbar$ given it was observed to be $S_z = -\frac{1}{2}\hbar$	× Probability amplitude of measuring the spin to be $S_z = -\frac{1}{2}\hbar$ given it was initially $S = \frac{1}{2}\hbar$

which is just Eq. (7.52) written out in words. Thus, in a sense, what quantum mechanics provides us with is a different set of rules for calculating the probability of outcome of random events. But it is more than that. The classical rule refers to outcomes that actually occur, the quantum rule refers to possibilities, or potentials for possible outcomes which are never realized in actuality. But these possibilities nevertheless exert a 'ghostly influence', in that they give rise to interference effects which distinguish the fully quantum result from the classically expected result.

Chapter 8

Vector Spaces in Quantum Mechanics

We have seen in the previous Chapter that there is a sense in which the state of a quantum system can be thought of as being made up of other possible states. The aim here is to use the example of the Stern-Gerlach experiment to develop this idea further, and to show that the states of a quantum system can be represented by vectors in a complex vector space. To begin, we review some of the basic ideas of vectors, using the example of vectors in real two dimensional space.

8.1 Vectors in Two Dimensional Space

The primitive prototype of what we mean by a vector is provided by the position vector in ordinary space. Below is a summary of the important properties of such vectors in physical space based on their interpretation as mathematical objects that have both magnitude and direction. As well as position, velocity and force and so on are also examples of such vectors. The intention is not to give a complete discussion, but to highlight a number of important properties of such vectors that have analogues in the case of quantum states, including the property that two vectors can be combined to produce another vector, and that 'how much' of one vector is contained in another can be measured via the inner product of two vectors.

8.1.1 Linear Combinations of Vectors – Vector Addition

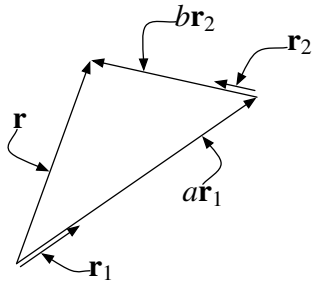


Figure 8.1: An example of vector addition. Two basis vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ combined to give $\mathbf{r}$.

Consider two non-collinear vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, as illustrated in Fig. (8.1). Multiples of these pair of vectors can be added together in accordance with the triangle addition rule illustrated in Fig. (8.1) to form another vector. Conversely, any other vector $\mathbf{r}$ can be expressed in terms of $\mathbf{r}_1$ and $\mathbf{r}_2$ using appropriate values for the (real number) components a and b of $\mathbf{r}$, i.e.

$$\mathbf{r} = a\mathbf{r}_1 + b\mathbf{r}_2. \tag{8.1}$$

The right hand side of this equation is known as a *linear combination* of the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. The particular points to take away from this is that combining vectors produces other vectors, analogous to our observation above that combining states produces other states, and that the components a and b are a measure of how much of $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively go towards making up the vector $\mathbf{r}$.

The two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ introduced above are arbitrary except insofar as they are not collinear. What we mean by collinear is that there is no way that multiples of these vectors can be combined to produce the zero vector, or, in other words,

$$a\mathbf{r}_1 + b\mathbf{r}_2 = 0 \quad \Longrightarrow \quad a, b = 0. \tag{8.2}$$

The usual terminology is to say that these two vectors are *linearly independent*.

Further, as any vector in the plane can be written as a linear combination of $\mathbf{r}_1$ and $\mathbf{r}_2$, they act as the basic building blocks for any vector in the plane, and hence are known as *basis vectors*. There is effectively an infinite number of choices for the basis vectors, and in fact it is possible to choose three or more vectors to be basis vectors. But the minimum number is two, if we wish to be able to describe any vector in the plane as a linear combination of basis vectors. The collection of all the vectors that can be constructed by taking linear combinations of these basis vectors using any real numbers a and b as components is known as a real vector space, and since two basis vectors are needed, the vector space is said to be of dimension two.

This vector space possess more structure than that implied by simply forming various linear combinations. The various vectors that can be drawn in a plane, as in Fig. (8.1), i.e. for which the coefficients a and b take any real value, can be clearly seen to have different lengths and relative orientations. These properties of vectors are encompassed in the definition of the inner, scalar or dot product of pairs of vectors.

8.1.2 Inner or Scalar Products

The inner or scalar product of the two position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ is defined by

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = r_1 r_2 \cos\theta \tag{8.3}$$

where r_1 and r_2 are the lengths of $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively, and θ is the angle between them.

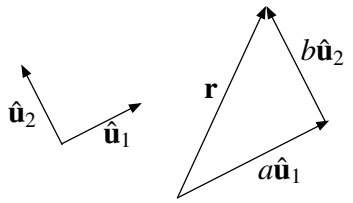


Figure 8.2: An example of a pair of unit vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ combining to give $\mathbf{r}$.

Using the idea of an inner or scalar product $\mathbf{r}_1 \cdot \mathbf{r}_2$ it is possible to introduce a particularly useful pair of basis vectors. To this end, consider two vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ that satisfy

$$\hat{\mathbf{u}}_1 \cdot \hat{\mathbf{u}}_1 = \hat{\mathbf{u}}_2 \cdot \hat{\mathbf{u}}_2 = 1 \tag{8.4}$$

i.e. they have unit length, hence they are known as unit vectors, and

$$\hat{\mathbf{u}}_1 \cdot \hat{\mathbf{u}}_2 = 0 \tag{8.5}$$

i.e. they are orthogonal. This pair of vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ is said to be orthonormal. As before, they form a basis in that multiples of each vector can be added together in the fashion illustrated in Fig. (8.2) to form another vector. Conversely, any vector $\mathbf{r}$ can be expressed using appropriate values for the components of $\mathbf{r}$, i.e.

$$\mathbf{r} = a\hat{\mathbf{u}}_1 + b\hat{\mathbf{u}}_2. \tag{8.6}$$

The components a and b, which represent 'how much' of the vector $\mathbf{r}$ is made up of the vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ are given by

$$a = \hat{\mathbf{u}}_1 \cdot \mathbf{r} \quad \text{and} \quad b = \hat{\mathbf{u}}_2 \cdot \mathbf{r} \tag{8.7}$$

A well known example of a vector expressed as a linear combination of a pair of unit vectors is given by the position vector $\mathbf{r}$ written with respect to the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$:

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}}. \tag{8.8}$$

The position vector is the 'original' or prototypical vector in the sense that the properties of position vectors can be generalized, essentially as a creative exercise in pure mathematics, so as to arrive at the notion of an abstract vector which has nothing to do with position in space, but which nevertheless has mathematical properties firmly founded on the intuitive properties assigned to position vectors. This generalization is central to arriving at the mathematical language of quantum mechanics.

8.2 Generalization to higher dimensions and complex vectors

The above properties of position vectors are built around basic concepts of two, or perhaps, three dimensional physical space. But the mathematics itself, considered as an abstract entity in its own right, can be extended in a number of ways.

Arbitrary dimensions Space has three dimensions: e.g. for the x, y, and z directions in space we require three basis vectors, the familiar unit vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$, which we will here refer to as $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$ respectively. These basis vectors are sufficient to construct any arbitrary vector by taking appropriate linear combinations. But we can imagine a 'space' with four, five or more dimensions, including infinity. This would require defining more orthonormal unit vectors, $\hat{\mathbf{u}}_4, \hat{\mathbf{u}}_5, \ldots, \hat{\mathbf{u}}_N$ where

$$\begin{aligned} \hat{\mathbf{u}}_n \cdot \hat{\mathbf{u}}_m &= 1 \quad n = m \\ &= 0 \quad n \neq m \end{aligned} \tag{8.9}$$

so that a typical vector would look like

$$\mathbf{v} = a_1\hat{\mathbf{u}}_1 + a_2\hat{\mathbf{u}}_2 + \ldots + a_N\hat{\mathbf{u}}_N \tag{8.10}$$

and by the orthonormality property of the basis vectors, we can specify 'how much' of each of these basis vectors contributes towards the final vector $\mathbf{v}$, i.e.

$$a_n = \hat{\mathbf{u}}_n \cdot \mathbf{v}. \tag{8.11}$$

Clearly we are no longer dealing with ordinary three dimensional space; in fact we are dealing with a space of dimension N.

Complex components The next generalization we can make is to relax the requirement that the coefficients or 'components' a_n be real numbers; we can take them to be complex in general. Thus, for instance, we are going to admit vectors like

$$\mathbf{v} = (1 + 2i)\hat{\mathbf{i}} + (3 - 4i)\hat{\mathbf{j}} \tag{8.12}$$

or more generally, we would allow the coefficients a_n appearing in Eq. (8.10) to be complex.

Inner product Finally, the idea of an inner product has to be looked at again. If we had two complex vectors, such as $\mathbf{v}$ in Eq. (8.12), and some other complex vector $\mathbf{u}$, then it might seem obvious to take the inner product to be simply $\mathbf{u} \cdot \mathbf{v}$. However, there are good reasons to not follow this simple path, the most obvious concerning the issue of defining the 'length' of such a vector. In the case of a real position vector, the length r of vector $\mathbf{r}$ can be written in terms of the scalar product of $\mathbf{r}$ with itself as $r = \sqrt{\mathbf{r} \cdot \mathbf{r}}$. But if we were to adopt this for the complex vector $\mathbf{v}$ given above then we find that

$$\mathbf{v} \cdot \mathbf{v} = (1 + 2i)^2 + (3 - 4i)^2 = -10 - 20i$$

which is a complex number. So, the 'length' $v = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ would also be a complex number, which is not a satisfactory state of affairs if we needed to have the concept of length for complex vectors: there is no way to be able to say that one vector was 'longer' than another. In contrast, however

$$\mathbf{v}^* \cdot \mathbf{v} = |1 + 2i|^2 + |3 - 4i|^2 = 30$$

which is real, and positive, so we can define a length of a complex vector as $v = \sqrt{\mathbf{v}^* \cdot \mathbf{v}}$. This then suggests that we adopt the following definition:

$$\text{The inner product of the two complex vectors } \mathbf{u} \text{ and } \mathbf{v} = \mathbf{u}^* \cdot \mathbf{v}. \tag{8.13}$$

One obvious consequence of this is that the order of the factors matters, i.e.

$$\mathbf{u}^* \cdot \mathbf{v} \neq \mathbf{v}^* \cdot \mathbf{u}. \tag{8.14}$$

In fact,

$$\mathbf{u}^* \cdot \mathbf{v} = (\mathbf{v}^* \cdot \mathbf{u})^*. \tag{8.15}$$

Orthonormal vectors now satisfy

$$\begin{aligned} \hat{\mathbf{u}}_n^* \cdot \hat{\mathbf{u}}_m &= 1 \quad n = m \\ &= 0 \quad n \neq m \end{aligned} \tag{8.16}$$

Complex basis vectors are not exotic mathematical entities. A simple example is

$$\begin{aligned} \hat{\mathbf{u}}_1 &= \frac{3\mathbf{i} + 4i\mathbf{j}}{5} \\ \hat{\mathbf{u}}_2 &= \frac{4\mathbf{i} - 3i\mathbf{j}}{5} \end{aligned} \tag{8.17}$$

which can be easily shown to be orthonormal. For instance

$$\hat{\mathbf{u}}_1^* \cdot \hat{\mathbf{u}}_1 = \frac{3\mathbf{i} - 4i\mathbf{j}}{5} \cdot \frac{3\mathbf{i} + 4i\mathbf{j}}{5} = \frac{9 + 16}{25} = 1$$

and so on.

Thus by fairly straightforward generalizations of the basic concept of a position vector it is possible to construct a more abstract notion of a vector which lives in a vector space of arbitrary dimensions, and which can have complex components. In other words, the study of the properties of ordinary vectors and their generalizations as outlined above, leads to a set of rules that these vectors obey. If it turns out that in some context which *a priori* has nothing to do with vectors in ordinary space, there arises mathematical quantities that obey these same set of rules, then these mathematical quantities will be looked upon as vectors. If we call these quantities $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ ..., then what we must have is

- A rule that enables us to 'add' multiples of these quantities together so as to generate another such object

  $$a\mathbf{u} + b\mathbf{v} = \mathbf{w}$$

  where a and b are complex numbers. For ordinary vectors, the rule was simply the triangle rule for adding position vectors. For more abstract vectors, the rule could be much more complicated.

- A rule that tells us how to define, and hence calculate, the inner product $\mathbf{u}^* \cdot \mathbf{v}$. For ordinary vectors, we made use of the pre-existing notions of the length of a vector and the angle between vectors to give us our definition. In the general case, we provide a definition, and the notion of the length and angle between vectors emerges from the definition!! Often, in pure mathematical contexts, the notation $(\mathbf{u}, \mathbf{v})$ is used to indicate an inner product rather than the dot product notation. It is this bracket notation that segues into the 'bra-ket' notation due to Dirac. Thus we require that the inner product satisfy the requirements:

  $$\begin{aligned} (\mathbf{u}, \mathbf{v}) &= \text{a complex number} \\ (\mathbf{u}, \mathbf{v})^* &= (\mathbf{v}, \mathbf{u}) \\ (\mathbf{v}, \mathbf{v}) &\geq 0 \\ (\mathbf{w}, a\mathbf{u} + b\mathbf{v}) &= a(\mathbf{w}, \mathbf{u}) + b(\mathbf{w}, \mathbf{v}) \end{aligned}$$

What we will see in the case of quantum states is that the mathematical properties of these states *as determined by their physical meaning* provide us with both these set of rules: i.e. we find that we can 'add' states to generate other states, and we can 'multiply' these states in a way that obeys the rules of an inner product. Thus we conclude that quantum states can be treated as abstract vectors. All the mathematical machinery of vectors then follows: vector spaces, unit vectors, linear independence, dimension ....

In the following Section, these parallels between what we have seen here and the behaviour of quantum states is made explicit.

8.3 Spin Half Quantum States as Vectors

We now need to examine the way in which the quantum states for a spin half system can be seen to fit in with the idea of being considered as vectors. To see how this comes about, we will show that there is a perfect analogy between Eqs. (7.18) and (7.19) and corresponding relationships for ordinary (complex) vectors. Returning to the Stern-Gerlach example discussed in the preceding Chapter we obtained there an expression Eq. (7.19)

$$|S\rangle = |+\rangle\langle +|S\rangle + |-\rangle\langle -|S\rangle.$$

for the state of a spin half atom expressed in terms of the states $|\pm\rangle$, which are states for which the atom has a z component of spin equal to $\pm\frac{1}{2}\hbar$, and $\langle\pm|S\rangle$ are probability amplitudes (complex numbers) whose magnitude in some sense tells us 'how much' of the states $|\pm\rangle$ are to be found in the state $|S\rangle$. This result was obtained by 'cancelling' the common factor '$\langle S'|$' from the result

$$\langle S'|S\rangle = \langle S'|+\rangle\langle +|S\rangle + \langle S'|-\rangle\langle -|S\rangle.$$

What we should bear in mind here is that we can recover this relationship between probability amplitudes by reintroducing '$\langle S'|$' into Eq. (7.19) for any chosen final state, yielding an expression for the probability amplitudes as needed. Thus, as has been said before, Eq. (7.19) effectively represents a 'template' into which we insert the appropriate information in order to recover the required probability amplitudes. We can also note that there is nothing sacred about choosing to cancel the the common factor $|S\rangle$ – we could equally as well cancel the factor $|S\rangle$, yielding

$$\langle S'| = \langle S'|+\rangle\langle +| + \langle S'|-\rangle\langle -|. \tag{8.18}$$

Having carried out this cancellation procedure, what has reappeared is the state of a quantum system i.e. $|S\rangle$ which was introduced earlier in a different context, specifically as being nothing more than a way of writing down all that we knew about the state of a quantum system. There, the notation had no mathematical significance, but in the manner in which it appears here, it seems to have acquired a mathematical meaning of some kind. The aim is to see what this meaning might be, and in doing so, we will show that the expression for $|S\rangle$ has many of the properties that we associate with expressing a vector as a sum of its components.

We begin by considering the probability amplitudes $\langle S'|S\rangle$ themselves. These are complex numbers in general for arbitrary spin directions, (but they were real in the particular Stern-Gerlach example used above), such that their modulus squared $|\langle S'|S\rangle|^2$ is the probability $P(S'|S)$ of observing the spin to be in the state $|S'\rangle$ given that it was in the state $|S\rangle$. In particular, $\langle S|S\rangle$ is the probability amplitude of observing the spin to be in the state $|S\rangle$ given that it was in the state $|S\rangle$. This will have to be unity, i.e. $P(S|S) = |\langle S|S\rangle|^2 = 1$. Thus we can conclude that

$$\langle S|S\rangle = e^{i\eta} \tag{8.19}$$

where η is an arbitrary phase. It turns out that this phase always cancels out in any calculation of observable quantities, so it is conventionally set to zero, and hence

$$\langle S|S\rangle = 1. \tag{8.20}$$

The state $|S\rangle$ is said to be normalized to unity. As a particular case, this last result implies that

$$\langle +|+\rangle = 1. \tag{8.21}$$

We can now consider the probability amplitude $\langle +|S\rangle$ obtained by replacing S' by + in the above expression for $\langle S'|S\rangle$:

$$\langle +|S\rangle = \langle +|+\rangle\langle +|S\rangle + \langle +|-\rangle\langle -|S\rangle. \tag{8.22}$$

We have seen that we can put $\langle +|+\rangle = 1$, so we have

$$\langle +|-\rangle\langle -|S\rangle = 0 \tag{8.23}$$

which has to be true no matter what the state $|S\rangle$ happens to be, i.e. no matter what value the probability amplitude $\langle -|S\rangle$ is. Thus we conclude that

$$\langle +|-\rangle = 0. \tag{8.24}$$

What this is telling us is that if an atom has been measured as being in the state $|-\rangle$, for instance, it has emerged from a Stern-Gerlach apparatus in the $S_z = -\frac{1}{2}\hbar$ beam, and if it is then passed through a second such apparatus, then there will be zero probability of it emerging in the $S_z = \frac{1}{2}\hbar$ beam (it has a 100% chance of emerging in the $S_z = -\frac{1}{2}\hbar$ beam). The two states $|+\rangle$ and $|-\rangle$ represent *mutually exclusive* possibilities.

Similarly we can show that

$$\langle -|-\rangle = 1 \quad \text{and} \quad \langle -|+\rangle = 0. \tag{8.25}$$

Thus we can set up a comparison:

$$\begin{aligned} \langle +|+\rangle = 1 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_1^* \cdot \hat{\mathbf{u}}_1 = 1 \\ \langle +|-\rangle = 0 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_2^* \cdot \hat{\mathbf{u}}_1 = 0 \\ \langle -|-\rangle = 1 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_2^* \cdot \hat{\mathbf{u}}_2 = 1 \\ \langle -|+\rangle = 0 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_2^* \cdot \hat{\mathbf{u}}_1 = 0 \end{aligned} \tag{8.26}$$

where we have chosen to make the comparison between the probability amplitudes and the inner product of complex unit vectors as we are dealing with probability amplitudes that are, in general, complex numbers. This comparison implies the following correspondences:

$$\begin{aligned} |+\rangle &\longleftrightarrow \hat{\mathbf{u}}_1 \quad |-\rangle \quad \longleftrightarrow \hat{\mathbf{u}}_2 \\ \langle +| &\longleftrightarrow \hat{\mathbf{u}}_1^* \quad \langle -| \quad \longleftrightarrow \hat{\mathbf{u}}_2^*. \end{aligned} \tag{8.27}$$

We know that $\langle +|S\rangle$ and $\langle -|S\rangle$ are both just complex numbers, so call them a and b respectively. If we now write

$$|S\rangle = a|+\rangle + b|-\rangle \tag{8.28}$$

we establish a perfect correspondence with the expression

$$\mathbf{v} = a\,\hat{\mathbf{u}}_1 + b\,\hat{\mathbf{u}}_2. \tag{8.29}$$

On the basis of this result, we are then tempted to interpret the ket $|S\rangle$ as a vector expressed as a linear combination of two orthonormal basis vectors $|\pm\rangle$. We can push the analogy further if we once again use the fact that $\langle S|S\rangle = 1$, so that

$$\langle S|S\rangle = 1 = \langle S|-\rangle\langle -|S\rangle + \langle S|+\rangle\langle +|S\rangle \tag{8.30}$$

On the other hand, the total probability of observing the system in either of the states $|\pm\rangle$ must add up to unity, which means that

$$P(+|S) + P(-|S) = |\langle +|S\rangle|^2 + |\langle -|S\rangle|^2 = 1. \tag{8.31}$$

By comparing the last two equations, and noting that

$$|\langle \pm|S\rangle|^2 = \langle \pm|S\rangle\langle \pm|S\rangle^* \tag{8.32}$$

we conclude that

$$\langle \pm|S\rangle = \langle S|\pm\rangle^*. \tag{8.33}$$

If we now consider

$$\langle S'| = \langle S'|+\rangle\langle +| + \langle S'|-\rangle\langle -|$$

and use the result, Eq. (8.33), $\langle \pm|S'\rangle = \langle S'|\pm\rangle^*$, we can write this as

$$\langle S'| = \langle +|S'\rangle^*\langle +| + \langle -|S'\rangle^*\langle -| \tag{8.34}$$

or, expressed in terms of $a' = \langle +|S'\rangle$ and $b' = \langle -|S'\rangle$, we have

$$\langle S'| = a'^*\langle +| + b'^*\langle -| \tag{8.35}$$

which has a perfect correspondence with an ordinary vector $\hat{\mathbf{v}}'$ in the form

$$\mathbf{v}'^* = a'^*\hat{\mathbf{u}}_1^* + b'^*\hat{\mathbf{u}}_2^*. \tag{8.36}$$

So the bra $\langle S'|$ is itself a vector, a bra vector, which can be thought of as being just the complex conjugate of the corresponding ket vector $|S'\rangle$. But while it is occasionally useful to think this way, it is not strictly true mathematically, and this way of viewing a bra vector will not be employed here. Instead, as will be shown shortly, an interpretation of ket vectors as column vectors leads to the interpretation of bra vectors as row vectors. A more mathematical standpoint also leads to interpretation of bra vectors as 'linear functionals', that is, a bra is a mathematical operator that acts on a ket vector to produce a complex number.

Finally, to complete the correspondence, we note that the probability amplitude $\langle S'|S\rangle$ can be written

$$\langle S'|S\rangle = a'^*a + b'^*b \tag{8.37}$$

which can be compared with the inner product $\mathbf{v}'^*\cdot\mathbf{v}$

$$\mathbf{v}'^*\cdot\mathbf{v} = a'^*a + b'^*b \tag{8.38}$$

which tells us that the probability amplitude can be considered as being simply the inner product of the two vectors $|S'\rangle$ and $|S\rangle$, i.e.

$$\langle S'|S\rangle = \langle S'|\cdot|S\rangle. \tag{8.39}$$

In other words, we have a perfect analogy between the two dimensional complex vector space formed by linear combinations of the unit vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ discussed in Section 8.1 and a complex vector space consisting of all the linear combinations of the states $|\pm\rangle$. The ket vectors $|\pm\rangle$ are referred to as *basis states*, analogous to $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ being referred to as basis vectors.

Different spin states can be constructed by forming linear combinations $|S\rangle = a|+\rangle + b|-\rangle$ of these basis states, with a and b being allowed to range over all the complex numbers, though we have only been looking at linear combinations with real coefficients. By limiting a and b to be real numbers, we are constructing states whose measured component of spin all lie in the same plane, which, with the system of axes we have been using here, is the XZ plane, corresponding to polar angle $\phi = 0$. If the coefficients a and b are complex, the state $|S\rangle$ represents a state in which the measured component $S = \mathbf{S}\cdot\mathbf{m}$ is along a direction $\hat{\mathbf{m}}$ that points out of this plane, as discussed in Section 8.3.2. So, any linear combination in which a and b are any complex numbers also defines a possible spin state of the spin half system. Thus all the possible linear combinations of $|\pm\rangle$, i.e.

combinations of the form $a|+\rangle + b|-\rangle$ where a and b are complex numbers form a *complex vector space* known as the *state space* of the system.

The quantum state vectors can also be 'multiplied' together – the inner product of the two vectors $|S'\rangle$ and $|S\rangle$ is just the probability amplitude $\langle S'|S\rangle$. In particular, the basis states are normalized to unity, i.e. they are unit vectors, and they are orthogonal to each other, i.e. they form a pair of *orthonormal basis states*.

The terminology often adopted is to say that the state vector $|S\rangle = a|+\rangle + b|-\rangle$ is a 'linear superposition' of the two states $|\pm\rangle$. The probability amplitudes $a = \langle +|S\rangle$ and $b = \langle -|S\rangle$ represent 'how much' of the states $|\pm\rangle$ are contained within the state $|S\rangle$ to the extent that $|\langle \pm|S\rangle|^2$ is the probability of the z component of spin being found to have the values $\pm\frac{1}{2}\hbar$.

8.3.1 The Normalization Condition

One difference between ordinary vectors and quantum state vectors is the importance of the 'normalization condition', i.e. the requirement that $\langle S|S\rangle = 1$, which must hold true given the interpretation of the inner product as a probability amplitude. But how can this be reconciled with the statement above that *any* linear combination of the basis states is a possible state of the system? How can a state vector such as $|\widetilde{S}\rangle = |+\rangle + |-\rangle$ which has the property

$$\langle \widetilde{S}|\widetilde{S}\rangle = 2 \tag{8.40}$$

be a physically acceptable state vector as it seems to be saying that the probability of finding the system in the state $|\widetilde{S}\rangle$ given that it is in the state $|\widetilde{S}\rangle$ is 4, which does not make sense. But, if we define a new vector $|S\rangle$ by

$$|S\rangle = \frac{|\widetilde{S}\rangle}{\sqrt{\langle \widetilde{S}|\widetilde{S}\rangle}} = \frac{1}{\sqrt{2}}|\widetilde{S}\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle + |-\rangle\right) \tag{8.41}$$

then automatically $|S\rangle$ will have the required normalization property – it is said to be normalized to unity. So, rather than abandoning giving a physical interpretation of state vectors which are not normalized to unity, the approach adopted is that we can multiply any state vector by any factor and say that it still represents the same *physical* state, i.e. $|S\rangle$ and $|\widetilde{S}\rangle = a|S\rangle$, where a is any complex number, represent the same physical state. However, it is only the normalized state $|S\rangle$ that should be used in any calculations in order to be certain that probability is properly accounted for.

8.3.2 The General Spin Half State

We can generalize the result obtained in the preceding Chapter for the state of a spin half system. That result applied only to the case of spin directions oriented in the XY plane. But it ought not matter how the vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ are oriented in space as the probability amplitude in Eq. (7.40) depends only on the angle between these two vectors. For arbitrary orientations in space, we can specify these vectors by their spherical polar coordinates as illustrated in Fig. (8.3).

If we suppose that $\hat{\mathbf{n}}$ has the polar coordinates (θ_i, ϕ_i) then we have:

$$\hat{\mathbf{n}} = \sin\theta_i \cos\phi_i \,\hat{\mathbf{i}} + \sin\theta_i \sin\phi_i \,\hat{\mathbf{j}} + \cos\theta_i \,\hat{\mathbf{k}} \quad (8.42)$$

and similarly for $\hat{\mathbf{m}}$, with polar coordinates (θ_f, ϕ_f). The angle γ between these two vectors can be determined from the inner product $\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}$:

$$\begin{aligned}\hat{\mathbf{n}} \cdot \hat{\mathbf{m}} &= \cos[\gamma(\theta_f, \phi_f, \theta_i, \phi_i)] \\ &= \cos\theta_i \cos\theta_f + \cos(\phi_f - \phi_i) \sin\theta_i \sin\theta_f.\end{aligned} \quad (8.43)$$

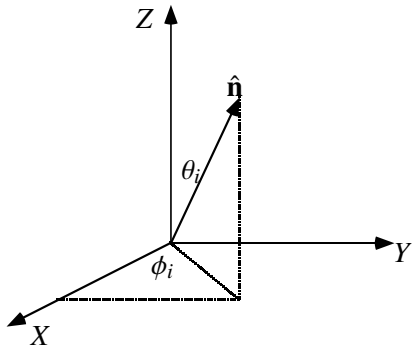


Figure 8.3: Polar angles for defining direction of unit vector $\hat{\mathbf{n}}$

We can now write for the probability amplitude $\langle S_f = \frac{1}{2}\hbar | S_i = \frac{1}{2}\hbar\rangle$ the expression

$$\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle = \cos(\tfrac{1}{2}\gamma) e^{i\xi(\theta_f, \phi_f, \theta_i, \phi_i)} \quad (8.44)$$

where $\exp[i\xi(\theta_f, \phi_f, \theta_i, \phi_i)]$ is a phase factor to be determined. It is necessary to introduce this phase factor as there is no requirement that the phases of the probability amplitudes for arbitrary orientations in space will be necessarily the same as those for the vectors in the XY plane.

We can determine this phase factor by requiring that the probability amplitudes defined in Eq. (8.44) satisfy the sum over amplitudes formula, Eq. (7.42). If we specialize this expression to the particular case of $\hat{\mathbf{l}} = \hat{\mathbf{k}}$, i.e. in the z direction, then we have

$$\begin{aligned}\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle =& \langle S_f = \tfrac{1}{2}\hbar | S_z = \tfrac{1}{2}\hbar\rangle\langle S_z = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle \\ &+ \langle S_f = \tfrac{1}{2}\hbar | S_z = -\tfrac{1}{2}\hbar\rangle\langle S_z = -\tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle \\ =& \cos[\tfrac{1}{2}\gamma(\theta_f, \phi_f, 0, 0)] \cos[\tfrac{1}{2}\gamma(0, 0, \theta_i, \phi_i)] e^{i(\xi(\theta_f, \phi_f, 0, 0) + \xi(0, 0, \theta_i, \phi_i))} \\ &+ \cos[\tfrac{1}{2}\gamma(\theta_f, \phi_f, \pi, 0)] \cos[\tfrac{1}{2}\gamma(\pi, 0, \theta_i, \phi_i)] e^{i(\xi(\theta_f, \phi_f, \pi, 0) + \xi(\pi, 0, \theta_i, \phi_i))}\end{aligned} \quad (8.45)$$

The value of γ for the various values of $(\theta_f, \phi_f, \theta_i, \phi_i)$ can be evaluated by use of Eq. (8.43) to give

$$\begin{aligned}&\gamma(\theta_f, \phi_f, 0, 0) = \theta_f \quad \gamma(0, 0, \theta_i, \phi_i) = \theta_i \\ &\gamma(\theta_f, \phi_f, \pi, 0) = \pi + \theta_f \quad \gamma(\pi, 0, \theta_i, \phi_i) = \pi + \theta_i\end{aligned} \quad (8.46)$$

so that

$$\begin{aligned}\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle =& \cos(\tfrac{1}{2}\theta_f) \cos(\tfrac{1}{2}\theta_i) e^{i(\xi(\theta_f, \phi_f, 0, 0) + \xi(0, 0, \theta_i, \phi_i))} \\ &+ \sin(\tfrac{1}{2}\theta_f) \sin(\tfrac{1}{2}\theta_i) e^{i(\xi(\theta_f, \phi_f, \pi, 0) + \xi(\pi, 0, \theta_i, \phi_i))} \\ =& e^{i\alpha} \left[\cos(\tfrac{1}{2}\theta_f) \cos(\tfrac{1}{2}\theta_i) + \sin(\tfrac{1}{2}\theta_f) \sin(\tfrac{1}{2}\theta_i) e^{i\beta}\right].\end{aligned} \quad (8.47)$$

where $\alpha = \xi(\theta_f, \phi_f, 0, 0) + \xi(0, 0, \theta_i, \phi_i)$ and $\beta = \xi(\theta_f, \phi_f, \pi, 0) + \xi(\pi, 0, \theta_i, \phi_i) - \alpha$. We can absorb $\exp(i\alpha)$ into the 'yet-to-be-determined' phase factor $\exp(i\xi)$, which amounts to putting $\alpha = 0$, so the final step is to determine β. This we do by noting that

$$\begin{aligned}\left|\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle\right|^2 =& \cos^2(\tfrac{1}{2}\gamma) \\ =& \tfrac{1}{2}(1 + \cos\gamma) \\ =& \tfrac{1}{2}(1 + \cos\theta_i \cos\theta_f + \cos(\phi_f - \phi_i) \sin\theta_i \sin\theta_f).\end{aligned} \quad (8.48)$$

But this is also given, from Eq. (8.47), by

$$\begin{aligned}\left|\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle\right|^2 =& \left|\cos(\tfrac{1}{2}\theta_f) \cos(\tfrac{1}{2}\theta_i) + \sin(\tfrac{1}{2}\theta_f) \sin(\tfrac{1}{2}\theta_i) e^{i\beta}\right|^2 \\ =& \cos^2(\tfrac{1}{2}\theta_f) \cos^2(\tfrac{1}{2}\theta_i) + \sin^2(\tfrac{1}{2}\theta_f) \sin^2(\tfrac{1}{2}\theta_i) + \tfrac{1}{2} \sin\theta_i \sin\theta_f \cos\beta.\end{aligned} \quad (8.49)$$

Using standard trigonometric relations, this last expression can be rearranged to read

$$\left|\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle\right|^2 = \tfrac{1}{2}(1 + \cos\theta_i\cos\theta_f + \cos\beta\sin\theta_i\sin\theta_f) \tag{8.50}$$

which can be compared with Eq. (8.48) to give us $\beta = \pm(\phi_f - \phi_i)$. Choosing the negative sign (in order to have a result consistent with the earlier result Eq. (7.40) for $\theta_i = \theta_f = \pi/2$) gives the required result

$$\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle = \cos(\tfrac{1}{2}\theta_f)\cos(\tfrac{1}{2}\theta_i) + \sin(\tfrac{1}{2}\theta_f)\sin(\tfrac{1}{2}\theta_i)e^{-i(\phi_f-\phi_i)}. \tag{8.51}$$

which can be readily shown to reduce to the earlier result Eq. (7.40) if we put $\theta_i = \theta_f = \pi/2$, this corresponding to the unit vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ lying in the XY plane.

We can make use of our general formula for the probability amplitudes, Eq. (8.51) to make explicit the probability amplitudes appearing in expression Eq. (7.19). Since here the states $|\pm\rangle$ correspond to the 'intermediate' states $|S_I = \pm\frac{1}{2}\hbar\rangle$ for $\hat{\mathbf{l}} = \hat{\mathbf{k}}$, i.e. in the z direction, so these are the states $|S_z = \pm\frac{1}{2}\hbar\rangle$. The state $|S_z = \frac{1}{2}\hbar\rangle = |+\rangle$, is associated with polar angles $\theta = 0, \phi = 0$ and $\theta = \pi, \phi = 0$ for the state $|S_z = -\frac{1}{2}\hbar\rangle = |-\rangle$. Finally, $|S\rangle \equiv |S_i = \frac{1}{2}\hbar\rangle$ with $\hat{\mathbf{n}}$ in the direction corresponding to the polar angles θ, ϕ. Overall, we have

$$\begin{aligned}
\langle +|S\rangle &= \cos(\tfrac{1}{2}\,0)\cos(\tfrac{1}{2}\theta) + e^{-i(0-\phi)}\sin(\tfrac{1}{2}\,0)\sin(\tfrac{1}{2}\theta)\\
&= \cos(\tfrac{1}{2}\theta)\\
\langle -|S\rangle &= \cos(\tfrac{1}{2}\pi)\cos(\tfrac{1}{2}\theta) + e^{-i(0-\phi)}\sin(\tfrac{1}{2}\pi)\sin(\tfrac{1}{2}\theta)\\
&= e^{i\phi}\sin(\tfrac{1}{2}\theta)
\end{aligned} \tag{8.52}$$

and hence

$$|S\rangle = \cos(\tfrac{1}{2}\theta)|+\rangle + e^{i\phi}\sin(\tfrac{1}{2}\theta)|-\rangle. \tag{8.53}$$

The associated probabilities are then

$$|\langle +|S\rangle|^2 = \cos^2(\tfrac{1}{2}\theta) \qquad |\langle -|S\rangle|^2 = \sin^2(\tfrac{1}{2}\theta). \tag{8.54}$$

Eq. (8.53) is a remarkably important formula used throughout quantum mechanics, finding application not only to explicitly spin half systems, but in fact to *any* system that possesses two basis states, otherwise known as a *qubit*.

8.3.3 Is every linear combination a state of the system?

One of the important properties of vectors is that two or more of them can be combined as a 'linear combination' to produce a third. If we are to consider quantum states as vectors, then this very basic property must also be possessed by quantum states, i.e. we ought to be able to show that by taking a linear combination of two basis states, the result is a vector that represents a possible state of the system. To see how this comes about, at least in the case of spin half, we make use of the general expression for the state $|S\rangle$ Eq. (8.53):

$$|S\rangle = \cos(\tfrac{1}{2}\theta)|+\rangle + e^{i\phi}\sin(\tfrac{1}{2}\theta)|-\rangle. \tag{8.55}$$

The atomic spin has been prepared in a state $|S\rangle$ where $S = \mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar$ by passing the atom through a Stern-Gerlach apparatus with its magnetic field in the $\hat{\mathbf{n}}$ direction, and the atom emerges in the $S = \frac{1}{2}\hbar$ beam. What this equation is then saying is that the combination $\cos(\frac{1}{2}\theta)|+\rangle + \exp(i\phi)\sin(\frac{1}{2}\theta)|-\rangle$, and $|S\rangle$, both represent the same thing – the atomic spin is in a state for which $S = \frac{1}{2}\hbar$. In other words, if we were presented with some arbitrary linear combination of the states $|\pm\rangle$, we ought to be able to work out exactly what state $|S\rangle$ this linear combination represents. Thus, for instance. consider the combination:

$$\frac{1}{\sqrt{2}}|+\rangle + i\frac{1}{\sqrt{2}}|-\rangle \tag{8.56}$$

we immediately see that $\cos(\frac{1}{2}\theta) = 1/\sqrt{2}$, $\exp(i\phi) = i$, and $\sin(\frac{1}{2}\theta) = 1/\sqrt{2}$, and hence $\theta = \phi = \pi/2$. Thus the magnetic field is pointing in the y direction, and hence the spin state of the atom is the state $|S\rangle = |S_y = \frac{1}{2}\hbar\rangle$.

What if we were presented with the combination $2|+\rangle + i2|-\rangle$? Here, we cannot find any angle θ, so it appears that this combination is not a possible state of the atomic spin. But we can write this as

$$2\sqrt{2}\left[\frac{1}{\sqrt{2}}|+\rangle + i\frac{1}{\sqrt{2}}|-\rangle\right] \tag{8.57}$$

which we can now understand as representing $2\sqrt{2}|S_y = \frac{1}{2}\hbar\rangle$. As we have argued above, we consider $2\sqrt{2}|S_y = \frac{1}{2}\hbar\rangle$ as describing the same physical state as $|S_y = \frac{1}{2}\hbar\rangle$. Thus, we can say that $2|+\rangle + i2|-\rangle$ is also a state of the system, namely $2\sqrt{2}|S_x = \frac{1}{2}\hbar\rangle$, which represents the same *physical* information about the state of the system as $|S_x = \frac{1}{2}\hbar\rangle$.

In the same way any combination $C_+|+\rangle + C_-|-\rangle$ where $C_\pm$ are complex numbers in general will always represent some state of the system, in general given by

$$\sqrt{|C_+|^2 + |C_-|^2}\,|S = \tfrac{1}{2}\hbar\rangle \tag{8.58}$$

where

$$S = \mathbf{S}\cdot\hat{\mathbf{m}} \tag{8.59}$$

and where $\hat{\mathbf{m}}$ is a unit vector in the direction defined by the polar angles

$$\theta = 2\tan^{-1}\left|C_-/C_+\right| \qquad \phi = \arg[C_-/C_+]. \tag{8.60}$$

Conversely, given any state of the system, we can work out how to write it in the form $C_+|+\rangle + C_-|-\rangle$. Further, we can repeat the whole of the above discussion for any other choice of the intermediate states $|S_I = \pm\frac{1}{2}\hbar\rangle$.

Ex 8.1 A spin half particle is prepared in the spin state

$$|S\rangle = 2|+\rangle + (1 + \sqrt{3}i)|-\rangle.$$

where $|S\rangle \equiv |\mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar\rangle$. What is the component $\hat{\mathbf{n}}$ of spin of the particle which is known to be $\frac{1}{2}\hbar$?

We first need to check that this state is normalized to unity. The bra vector corresponding to this state is

$$\langle S| = 2\langle +| + (1 - \sqrt{3}i)\langle -|$$

so that the inner product $\langle S|S\rangle$ is given by

$$\begin{aligned}\langle S|S\rangle &= \Big(2\langle +| + (1 - \sqrt{3}i)\langle -|\Big)\Big(2|+\rangle + (1 + \sqrt{3}i)|-\rangle\Big)\\ &= 4\langle +|+\rangle + 2(1 + \sqrt{3}i)\langle +|-\rangle + 2(1 - \sqrt{3}i)\langle -|+\rangle\\ &\quad + (1 - \sqrt{3}i)(1 + \sqrt{3}i)\langle -|-\rangle.\end{aligned}$$

Using the fact that $\{|+\rangle, |-\rangle\}$ form an orthonormal basis for the state space for the spin of the particle, we have $\langle +|+\rangle = \langle -|-\rangle = 1$ and $\langle -|+\rangle = \langle +|-\rangle = 0$ so that

$$\langle S|S\rangle = 8.$$

This state is not normalized to unity, so in order to calculate probabilities correctly, we must renormalize $|S\rangle$. We do this by defining

$$|\widetilde{S}\rangle = \frac{|S\rangle}{\sqrt{8}} = \frac{1}{2\sqrt{2}}\Big(2|+\rangle + (1+\sqrt{3}i)|-\rangle\Big)$$

which now satisfies $\langle\widetilde{S}|\widetilde{S}\rangle = 1$. But note that the new state vector $|\widetilde{S}\rangle$ still represents the *same physical state* as the original vector $|S\rangle$.

We can now compare the coefficients of $|\widetilde{S}\rangle$ with the general expression for the state of a spin half particle, that is

$$|S\rangle = \cos(\tfrac{1}{2}\theta)|+\rangle + e^{i\phi}\sin(\tfrac{1}{2}\theta)|-\rangle$$

and immediately read off

$$\cos(\tfrac{1}{2}\theta) = \frac{1}{\sqrt{2}} \qquad \sin(\tfrac{1}{2}\theta)e^{i\phi} = \frac{1+\sqrt{3}i}{2\sqrt{2}}$$

from which we extract $\theta = \pi/2$ and $\phi = \pi/3$, the polar angle components of $\hat{\mathbf{n}}$.

It is this last fact that a state $|S\rangle$ can be written as the linear combination or linear superposition $C_+|+\rangle + C_-|-\rangle$ of two other states, analogous to Eq. (8.1) for the arbitrary vector $\mathbf{v}$, and conversely that any linear superposition of states is itself another state is the essential property that these states need to possess in order for them to be interpreted as vectors belonging to some vector vector space, specifically here, since the coefficients $C_\pm$ are in general complex, a complex vector space. This more general case is what is usually encountered in quantum mechanics, and below we will assume that the probability amplitudes are, in general, complex.

8.4 Constructing State Spaces

The approach adopted here starts with the sort of result that is obtained when the 'cancellation trick' is employed. In the spin half case, what is obtained is the expression

$$|S\rangle = |+\rangle\langle+|S\rangle + |-\rangle\langle-|S\rangle$$

which is interpreted as a vector equation, with the arbitrary state $|S\rangle$ expressed as a linear combination of the pair of orthonormal basis states $|\pm\rangle$. This brings us to the general idea that lies behind setting up a quantum mechanical description of a physical system, and that is to identify a set of orthonormal *basis* states for the system. These basis states have to satisfy a number of properties that we can extract from our study of the two slit and spin half cases.

In the case of two slit interference, the states $|1\rangle$ and $|2\rangle$ are states associated with the two possible positions of the particle: at the position of one or the other of the slits cut into a screen. We have also seen that the electron could be observed to be at one slit *or* the other – the two possibilities are mutually exclusive. Finally, these two possibilities cover *all* the possibilities, at least within the constraints of the two slit model. So we take these states as our basis states, and construct an arbitrary state of our system by taking linear combinations of these two states.

In the case of particle spin, the basis states are, for instance, $|\pm\rangle$, or written out more fully, $|S_z = \pm\frac{1}{2}\hbar\rangle$, corresponding to the two possible values of the z component of the spin of the particle. These two possibilities are mutually exclusive – the atom emerges in either the $S_z = \frac{1}{2}\hbar$ *or* the $S_z = -\frac{1}{2}\hbar$ beam, and complete in the sense that all possibilities are covered – the atom never emerges in any other beam.

In each case, note was taken of the possible values of some measurable property of the system: the position of the particle at any one of the slits, or S_z, the z component of spin. The collection of possibilities was exhaustive — all possibilities were included — and mutually exclusive — the system could be observed to be in one state or the other. In fact, we label these states by the possible values that could be measured, e.g. spin states $|S_z = \pm\frac{1}{2}\hbar\rangle$. In other words, we identify a measurable property of the system, otherwise known as an *observable*, determine all the possible values that this observable can have, either by experiment or by theoretical argument, and for each possible value of this observable, we associate a state labelled by the observable's value. This collection of states so determined we take as our basis states for the system. From these two examples, we can propose a number of properties that our basis states would have to have in the general case:

1. Each basis state represents a mutually exclusive possibility, that is, if the system is observed to be in one of the basis states, it is definitely the case that it will not be observed in any of the others
2. The basis states are associated with the possible values of some measurable property of the system.
3. The basis states must be complete in that they cover all possible such values.

Note that the basis states are not unique. This is most readily seen in the case of spin half. We could have taken as our basis states *any* pair of states $|S = \pm\frac{1}{2}\hbar\rangle$ where $S = \mathbf{S}\cdot\hat{\mathbf{n}}$ with $\hat{\mathbf{n}}$ an arbitrary unit vector.

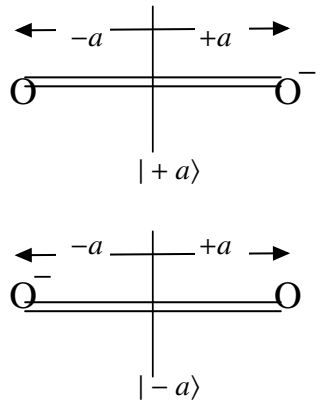


Figure 8.4: O_2^- ion with the two possible positions of the electron, corresponding to the two states $|\pm a\rangle$.

So let us see how these ideas can be applied to a completely different system. This system we will take to be an O_2^- as illustrated in the adjacent figure. Is there any natural choice for what observable we might look upon as determining the basis states of this system? One possibility that suggests itself is the position of the extra electron. For a real O_2^- ion the electron could be found in myriad different possible places, but for our purposes we will assume that there are only two possibilities: at the position of either of the oxygen atoms, that is at $x = \pm a$. The two possibilities correspond to two possible states for the ion, $|\pm a\rangle$. These states we recognize as having all the properties that we have listed above for basis states, i.e. the electron is observed to be in *either* state $|+a\rangle$ *or* state $|-a\rangle$, and that there is nowhere else for the electron to be found (within the limits of our model). By analogy with the spin half example, we then say that

$$\langle a|a\rangle = \langle -a|-a\rangle = 1, \quad \langle -a|a\rangle = \langle a|-a\rangle = 0 \tag{8.61}$$

i.e. the states $|\pm a\rangle$ form a pair of orthonormal basis states for the state space of the ion. This state space has dimension 2. An arbitrary state of the ion will be given by

$$|\psi\rangle = |+a\rangle\langle +a|\psi\rangle + |-a\rangle\langle -a|\psi\rangle \tag{8.62}$$

As another example we can consider a (fictitious) two level atom, an atom which, if its energy is measured, is found to have only two values, E_1 or E_2, with $E_1 > E_2$. Such an atom is, of course, an idealization, but one that has proved to extremely valuable one in understanding the details of the interaction of quasimonochromatic light fields, such as that produced by a laser, with a real atom.

The nature of the interaction is such that the atom behaves as if it has only two energy levels, so that the simplification being considered here is taken as the basis of often used theoretical models.

Given that the energy can only have two values, and moreover that the energy is measured to be *either* E_1 *or* E_2 in a way totally analogous to measuring a component of the spin of a spin half system, we can then assign to the atom two possible states, call them $|E_1\rangle$ and $|E_2\rangle$, or $|e\rangle$ and $|g\rangle$, where $e \equiv$ excited state and $g \equiv$ ground state. We then have

$$\begin{aligned} \langle e|e\rangle &= \langle g|g\rangle = 1 \\ \langle e|g\rangle &= \langle g|e\rangle = 0 \end{aligned}. \tag{8.63}$$

These states then act as the orthonormal basis states of our two level atom, so that any state of the two level atom can be written as a linear combination

$$|\psi\rangle = a|e\rangle + b|g\rangle. \tag{8.64}$$

8.4.1 A General Formulation

We will now see how these ideas can be applied to more general kinds of physical systems. To begin, we have to set up these basis states for the given system. So suppose we perform an *exhaustive* series of measurements of some observable property of the system – call it Q. For example, we could determine through which slits it is possible for an electron to pass in the two slit experiment above in which case $Q \equiv$ 'the positions of the two slits', or the possible values of the Z component of the spin of a particle, in which case $Q \equiv S_z$, or the possible values of the position of an electron on an O_2^- ion, as discussed above, in which case $Q \equiv$ 'the position of the electron on the O_2^- ion'. The generic name given to Q is that it is an 'observable'. We will give many more examples of observables later in this Chapter, and look at the concept again in a later Chapter.

Whatever the system, what we mean by exhaustive is that we determine *all* the possible values that the observed quantity Q might have. For instance, we determine that, in the two slit interference experiment, the electron can pass through, (that is, be observed to be at) the position of one or the other of the two slits, or that a spin half particle would be observed to have either of the values $S_z = \pm\frac{1}{2}\hbar$, and no other value. Or for the O_2^- ion, the electron can be found *either* on the atom at position $x = -a$ *or* on the atom at position $x = +a$. In general, for an arbitrary observable Q, let us represent these observed values by $q_1, q_2, q_3, \ldots, q_N$, i.e. N in total, all of which will be real numbers. Of course, there might be other observable properties of the system that we might be able to measure, but for the present we will suppose that we only need concern ourselves with just one. In keeping with the way that we developed the idea of state earlier, we then let $|q_1\rangle$ represent the state for which Q definitely has the value q_1, and similarly for all the others possibilities.

We now turn to Eq. (7.48), the result obtained earlier when considering generalizations of the two slit experiment to the multiple slit case, or the generalization of the Stern-Gerlach experiment to the arbitrary spin case. There, a sum was made over probability amplitudes for the different 'pathways' from a given initial state $|\psi\rangle$ to some final state $|\phi\rangle$ via a set of intermediate slit or spin states. Here, we generalize that result for an arbitrary set of intermediate states $\{|q_1\rangle, |q_2\rangle, \ldots\}$ as defined above, and make the following claim, the fundamental rule of quantum mechanics, that if the system is initially prepared in the state $|\psi\rangle$, then the probability amplitude of finding it in the state $|\phi\rangle$ is given by

$$\langle\phi|\psi\rangle = \sum_{n=1}^{N} \langle\phi|q_n\rangle\langle q_n|\psi\rangle \tag{8.65}$$

which tells us that the total probability amplitude of finding the system in the final state $|\phi\rangle$ is just the sum of the probability amplitudes of the system 'passing through' any of the states $\{|q_n\rangle; n =$

$1, 2, \ldots, N\}$. The expression Eq. (8.65) is often referred to as a *closure relation* though confusingly, it is also sometimes referred to as a completeness relation, a term which we apply to another expression below.

We further claim, from the basic meaning of probability amplitude, that for any state $|\psi\rangle$ of the system, we must have

$$\langle\psi|\psi\rangle = 1 \tag{8.66}$$

known as the normalization condition. This follows from the requirement that $|\langle\psi|\psi\rangle|^2 = 1$, i.e. that if the system is in the state $|\psi\rangle$, then it is definitely (i.e. with probability one) in the state $|\psi\rangle$.

We then recognize that the probability amplitudes $\langle q_m|q_n\rangle$ must have the following properties:

- $\langle q_m|q_n\rangle = 0$ if $m \neq n$. This amounts to stating that if the system is in the state $|q_n\rangle$, i.e. wherein the observable Q is known to have the value q_n, then there is zero possibility of finding it in the state $|q_m\rangle$. Thus the states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ are *mutually exclusive*.
- $\langle q_n|q_n\rangle = 1$. This asserts that if the system is in the state for which the quantity Q has the value q_n, then it is certain to be found in the state in which it has the value q_n.
- The states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ are also exhaustive in that they cover *all* the possible values that could be observed of the observable Q. These states are said to be *complete* — simply because they cover all possibilities.

These three properties are analogous to the properties that we associate with the inner products of members of an orthonormal set of basis vectors for a complex inner product space, which suggests we interpret the states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ as an orthonormal set of basis vectors, or basis states, for our system. We then interpret that fundamental law, Eq. (8.65) as an expression for the inner product of two state vectors, $|\psi\rangle$ and $|\phi\rangle$ in terms of the components of these vectors with respect to the basis states $|q_n\rangle$. Expressions for these state vectors in terms of the basis states can then be obtained by what we have referred to earlier as the 'cancellation trick' to give

$$\begin{aligned} |\psi\rangle &= \sum_{n=1}^{N} |q_n\rangle\langle q_n|\psi\rangle \\ \langle\phi| &= \sum_{n=1}^{N} \langle\phi|q_n\rangle\langle q_n| \end{aligned} \quad . \tag{8.67}$$

We have argued above that the states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ are, in a sense, complete. We can use this to argue that *any* state of the system can be expressed in the form Eq. (8.67). To see how this follows, suppose $|\psi\rangle$ is some arbitrary state of the system. Then, for at least one of the states $|q_n\rangle$, we must have $\langle q_n|\psi\rangle \neq 0$, i.e. we must have a non-zero probability of observing the system in *one* of the states $|q_n\rangle$. If this were not the case, then for the system in a state such as $|\psi\rangle$ it would be saying that if we measure Q, we don't get an answer! This does not make physical sense. So physical consistency means that it must be the case that any state of the system can be written as in Eq. (8.67) and for that reason, these expressions are referred to as *completeness relations*.

We can also make the inverse claim that *any* such linear combination represents a possible state of the system. The justification for this is not as clear cut as there are physical systems for which there are limitations on allowed linear combinations (so-called super-selection rules), but it appears to be a rule that holds true unless there are good physical reasons, on a case-by-case basis, why it should not. We will assume it to be true here.

If we choose $|\phi\rangle = |\psi\rangle$ in Eq. (8.65) and we use the normalization condition we find that

$$\langle\psi|\psi\rangle = \sum_{n=1}^{N} \langle\psi|q_n\rangle\langle q_n|\psi\rangle = 1. \tag{8.68}$$

But we must also have, since the probability of finding the system in any of the states $|q_n\rangle$ must add up to unity that

$$\sum_{n} |\langle q_n|\psi\rangle|^2 = 1 \tag{8.69}$$

This can also be understood as being a consequence of our interpretation of the states $\{|q_n\rangle; n = 1, 2, \ldots N\}$ as a complete set of mutually exclusive possibilities, complete in the sense that the total probability of ending up in any of the mutually exclusive possible final states $|q_n\rangle$ adds up to unity — there is nowhere else for the system to be found. By subtracting the last two expressions we arrive at

$$\sum_{n=1}^{N} \left(\langle\psi|q_n\rangle - \langle q_n|\psi\rangle^*\right) \langle q_n|\psi\rangle = 0. \tag{8.70}$$

A sufficient condition for this result to hold true is

$$\langle\psi|q_n\rangle = \langle q_n|\psi\rangle^* \tag{8.71}$$

and hence that, in general,

$$\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*. \tag{8.72}$$

Thus, by a simple extension of the arguments presented in Section 8.3 in the case of spin half quantum states it can be seen that the general results above are also completely analogous to the properties of vectors in a complex vector space. This mathematical formalism will be discussed more fully in the next Chapter, but for the present we can summarize the essential ideas based on what we have already put forward earlier. The important points then are as follows:

1. The collection of all the possible state vectors of a quantum system forms a complex vector space known as the *state space* of the system.
2. The probability amplitudes are identified as the inner product of these state vectors.
3. The intermediate states $\{|q_n\rangle; n = 1, 2, \ldots\}$ form a complete orthonormal set of basis states of this state space, i.e. any state vector $|\psi\rangle$ can be written as a linear combination of these basis states.
4. The number of basis states is known as the dimension of the state space.

8.4.2 Further Examples of State Spaces

The ideas developed above can now be applied to constructing a state space for a physical system. The basic idea is as discussed in Section 7.4 which enables us to define a set of basis states for the state space of the system. By establishing a set of basis states, in a sense, we 'bring the state space into existence', and once this is done, we are free to use all the mathematical machinery available for analysing the properties of the state space so constructed. The question can be asked as to whether or not the ideas presented in Section 7.4, admittedly extracted from only a handful of examples, can be applied with success to any other system. This is a question that can only be answered by applying the rules formulated there and considering the consequences. In Section 8.5 we will discuss where these ideas, if naively applied, fail to work. Otherwise, these ideas, when fully formed, constitute the basis of quantum physics.

In accordance with the ideas developed in Section 7.4, constructing a state space for a physical system can be carried out by recognizing the intermediate states through which a system can pass as it makes its way from some initial state to some observed final state, as was done in the case of the two slit, or spin half systems. Thus, in the two slit example, the two possible intermediate states are those for which the particle is to be found at the position of either of the two slits. In the spin half example, the two intermediate states are those in which the spin is observed to have either of the two values $S_z = \pm\frac{1}{2}\hbar$; these are the states we have been calling $|\pm\rangle$. These intermediate states are states of the system that can be identified through an argument based on the idea that some physical property of the system can be exhaustively measured to yield a set of values that we then use to label a complete set of basis states for the state space of the system.

Negatively Charged Ions Here the system is a molecule which has acquired an extra electron, which can be assumed to found only on any one of the atoms making up the molecule. This is, of course, an approximation. The electron could be found anywhere around the atoms, or in the space between the atoms, in a way that depends on the nature of the chemical bond between the atoms. Here we are making use of a coarse notion of position, i.e. we are assuming that the electron can be observed to reside on one atom or the other, and we do not really care about exactly where on each atom the electron might be found. The idea is best illustrated by the simple example of the O_2^- ion in which the electron can be found on one or the other of the oxygen atoms (see Fig. (8.4)) as discussed on p117.

This kind of model can be generalized to situations involving different geometries, such as atoms arranged in a ring e.g. an ozone ion O_3^-. In this case, the state space will be spanned by three basis states corresponding to the three possible positions at which the electron can be observed. This model (and its generalizations to an arbitrary number of atoms arranged in a ring) is valuable as it gives rise to results that serve as an approximate treatment of angular momentum in quantum mechanics.

Spin Flipping In this case, we have a spin half particle (for instance) in a constant magnetic field, so the two possible states are the familiar spin up or spin down states. If, in addition, we add a *rotating* magnetic field at right angles to the constant field, there arises a time dependent probability of the spin flipping from one orientation to the other. As the spin up and spin down states are of different energies, this represents a change in energy of the particle, a change that can be detected, and is the basis of the electron spin and nuclear magnetic resonance imaging much used in medical work. Obviously this is a state space of dimension two.

Ammonia molecule Here the system is the ammonia molecule NH_3 in which the nitrogen atom is at the apex of a triangular pyramid with the three hydrogen atoms forming an equilateral triangle as the base. The nitrogen atom can be positioned either above or below the plane of the hydrogen atoms, these two possibilities we take as two possible states of the ammonia molecule. (The N atom can move between these two positions by 'quantum tunnelling' through the potential barrier lying in the plane of the hydrogen atoms.) Once again, this is a state space of dimension 2.

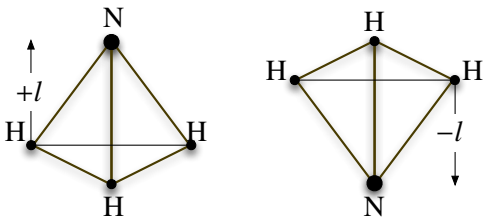


Figure 8.5: Ammonia molecule in two states distinguished by the position of the nitrogen atom, either above or below the plane of the hydrogen atoms, corresponding to the states $|+l\rangle$ and $|-l\rangle$ respectively.

Benzene Molecule An example of quite a different character is that of the benzene molecule, illustrated in Fig. 8.6. The two states of the molecule are distinguished by the positioning of the double bonds between pairs of carbon atoms. The molecule, at least with regard to the arrangements of double bonds can be found in two different states which, for want of a better name, we will call $|\alpha\rangle$ and $|\beta\rangle$. The state space is therefore of dimension 2, and an arbitrary state of the molecule would be given by

$$|\psi\rangle = a|\alpha\rangle + b|\beta\rangle. \tag{8.73}$$

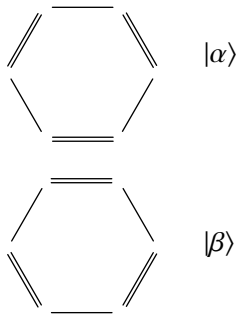


Figure 8.6: Two arrangements of the double bonds in a benzene molecule corresponding to two states $|\alpha\rangle$ and $|\beta\rangle$.

In all the cases considered above, the states were labelled by one piece of data only. It is possible, under certain circumstances, to generalise this to situations where two or more labels are needed to specify the basis states.

8.4.3 States with multiple labels

We have been solely concerned with states with a single label, Recall the general definition of a ket:

$$\Big|\ \text{All the data concerning the system that can be known without mutual interference or contradiction.}\ \Big\rangle$$

As we have seen, for simple systems, the state can be labelled with a single piece of information, e.g. $|S_z = \frac{1}{2}\hbar\rangle$ for a spin half particle for which $S_z = \frac{1}{2}\hbar$, $|x\rangle$ for a particle positioned at the point x, $|n\rangle$ for a single mode cavity with n photons. But, for more complex systems, it is reasonable to expect that more information is needed to fully specify the state of the system — this is certainly the case for classical systems. But, perhaps not unexpectedly, this is not a straightforward procedure for quantum systems. The important difference is embodied in the condition 'without mutual interference': some observables interfere with others in that the measurement of one 'scrambles' the predetermined value of the other, as was seen in the case of the components of spin, or as seen when trying to measure the position and momentum of a particle. If there is no such interference between two observables, then they are said to be *compatible*, and two (or more) labels can be legitimately used to specify the state e.g. $|S_x = \frac{1}{2}\hbar, x\rangle$ for a spin half particle with x component of spin $S_x = \frac{1}{2}\hbar$ AND at the position x. Thus, in this example, we are assuming that the spin of a particle and its position are compatible.

As another example, we can consider a system made up of more than one particle e.g. a system of two spin half particles. The state could then be written as

$$\big|\text{data for first particle,data for second particle}\big\rangle.$$

Its possible states would then be:

$$|S_z = \tfrac{1}{2}\hbar, S_z = \tfrac{1}{2}\hbar\rangle,\ \ |S_z = \tfrac{1}{2}\hbar, S_z = -\tfrac{1}{2}\hbar\rangle,\ \ |S_z = -\tfrac{1}{2}\hbar, S_z = \tfrac{1}{2}\hbar\rangle,\ \ |S_z = -\tfrac{1}{2}\hbar, S_z = -\tfrac{1}{2}\hbar\rangle. \tag{8.74}$$

We can write this more simply as $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ by making the identifications $S_z = \frac{1}{2}\hbar \to 0$ and $S_z = -\frac{1}{2}\hbar \to 1$. The states are orthonormal:

$$\begin{aligned} \langle 00|00\rangle &= 1\langle 01|01\rangle = 1 \ \text{ etc} \\ \langle 00|01\rangle &= 0\langle 01|10\rangle = 0 \ \text{ etc} \end{aligned} \tag{8.75}$$

An arbitrary state is $|\psi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$. These states form a set of orthonormal basis states for a state space of dimension 4. The idea can be extended to many particles, or other more complex systems e.g. whole atoms, or molecules or solid state systems and so on.

The notion of 'compatible observables' and determining whether or not two (or more) observables are compatible, is studied later in a more rigorous way once the idea of an observable being represented by Hermitean operators has been formulated.

Qubits Note that a single spin half has the basis states $|0\rangle$ and $|1\rangle$ in our new notation. These two states can be looked on as corresponding to the two binary numbers 0 and 1. A linear combination

$$|\psi\rangle = c_0|0\rangle + c_1|1\rangle.$$

can be formed which represents the possibility of a memory registering a bit of information simultaneously as both a 0 and a 1. This is in contrast with a classical bit, which can be registered as *either* a 0 *or* a 1. The spin half system is an example of a *qubit*. But spin half is not in any way special: *any* two state system: a spin half particle, a two level atom, the two configurations of a benzene molecule, a quantum dot either occupied or not occupied by an electron .... Quantum computation then involves manipulating the whole state $|\psi\rangle$, which, in effect, amounts to performing two calculations at once, differing by the initial setting of the memory bit.

This idea introduced can be readily extended. Thus, if we have two spin half particles, we have the possible states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. The data labelling the state $|00\rangle$ represents the number zero, $|01\rangle$ the number one, $|10\rangle$ the number two, and $|11\rangle$ the number three. We now have two qubits, and a state space of dimension four, and we can set up linear combinations such as

$$|\psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle \tag{8.76}$$

and we can then perform calculations making use, simultaneously, of four different possible values for whatever quantity the states are intended to represent. With three atoms, or four and so on, the state space becomes much larger: of dimension 2^N in fact where N is the number of qubits, and the basis states represent the numbers ranging from 0 to $2^N - 1$ in binary notation.

The viability of this scheme relies on the linear combination not being destroyed by decoherence. Decoherence is the consequence of noise destroying a linear combination of quantum states, turning a quantum state into a mixture of alternate classical possibilities.

8.5 States of Macroscopic Systems — the role of decoherence

In the examples given above, it was assumed that an exhaustive list of results that could be obtained in the measurement of some observable of a quantum system could be used to set up the basis states for the state space of the system. The value of doing this is, of course, to be determine by the success or otherwise of these ideas. That quantum mechanics is such an overwhelmingly successful theory indicates that there is something correct in this procedure, but the question that arises is this: why does it not appear to work for macroscopic systems, i.e. for systems which we know can be fully adequately explained by standard classical physics? The answer appears to lie in the fact that in all the examples discussed above, whether or not the Hilbert space is of finite or infinite dimension, i.e. whether or not we are talking about spin up or spin down of a spin half particle, or the position of a particle in space, the implicit assumption is that the system we are considering is totally isolated from all other systems, in particular from any influence of the surrounding environment. After all, when we talked about a system, such as an O_2^- ion, we are ignoring all the other physical influences that could act on this system, i.e. we do not need to mention, in our specification of the state of the system, anything other than properties that directly pertain to the system of interest. The assumption is made, as it is in classical physics, that such

influences are sufficiently weak that they can be ignored to a good approximation. In effect, we are supposing that the systems under consideration are isolated systems, that is, systems that are isolated from the effect of any external perturbations.

Classically, at the macroscopic level, we can usually continue to ignore weak perturbing influences when specifying the state of a system. In fact, when defining a 'system' we typically include in what we refer to as the 'system', all the participants in the physical process being described that interact strongly with each other. Anything else that weakly affects these constituents is ignored. For instance, when describing the orbital dynamics of the Earth as it revolves around the Sun, we might need to take into account the gravitational pull of the Moon – the system is the Earth, the Sun and the Moon. But we do not really need to take into account the effect of the background microwave radiation left over after the Big Bang. Or, when describing the collision between two billiard balls, it is probably necessary to include the effect of rolling friction, but it not really necessary to take into account the frictional drag due to air resistance. Of course, sometimes it is necessary to include external influences even when weak: to describe a system coming to thermal equilibrium with its surroundings it is necessary to extend the system by including the environment in the dynamical model. In any of these examples, the same classical physics methods and philosophy applies.

There is a subtle difference when it comes to trying to apply the quantum ideas developed so far to macroscopic systems. The same, weak perturbations that can be put to one side in a classical description of a macroscopic system turn out to have a far-reaching effect if included in a quantum description of the same system. If we were to attempt to describe a macroscopic system according to the laws of quantum mechanics, we would find that any linear superposition of different possible states of the system evolves on a fantastically short time scale to a classical mixture of the different possibilities. For instance, if we were to attempt to describe the state of a set of car keys in terms of two possibilities: in your pocket $|p\rangle$ or in your brief case $|b\rangle$, then a state of the form

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|p\rangle + |b\rangle) \tag{8.77}$$

could be used to represent a possible 'quantum state' of the keys. But this quantum state would be exceedingly short lived (on a time scale $\sim 10^{-40}$ sec), and would evolve into the two alternative possibilities: a 50% chance of the keys being in the state $|p\rangle$, i.e. a 50% chance of finding your keys in your pocket, and a 50% chance of being in the state $|b\rangle$, i.e. a 50% chance of finding them in your brief case. But this is no longer a superposition of these two states. Instead, the keys are *either* in the state $|p\rangle$ *or* the state $|b\rangle$. What this effectively means is that randomness is still there, i.e. repeating an experiment under identical conditions can give randomly varying results. But the state of the keys is no longer represented by a state vector, so there are no longer any quantum interference effects present. The randomness can then be looked upon as being totally classical in nature, i.e. as being due to our ignorance of information that is in principle there, but impossibly difficult to access. In effect, the quantum system behaves like a noisy classical system.

The process that washes out the purely quantum effects is known as decoherence. Since it is effectively impossible to isolate any macroscopic system from the influence of its surrounding environment[1], all macroscopic systems are subject to decoherence. This process is believed to play a crucial role in why, at the macroscopic level, physical systems, which are *all* intrinsically quantum mechanical, behave in accordance with the classical laws of physics. It is also one of the main corrupting influences that prevent a quantum computer from functioning as it should. Quantum computers rely for their functioning on the 'qubits' remaining in linear superpositions of states, but the ever-present decohering effects of the environment will tend to destroy these delicate quantum states before a computation is completed, or else at the very least introduce errors as the

[1] 'No man is an Iland, intire of itselfe' – J. Donne, *Devotions upon Emergent Occasions* Meditation XVII (1693)

computation proceeds. Controlling decoherence is therefore one of the major challenges in the development of viable quantum computers.

So, the bottom line is that it is only for protected isolated systems that quantum effects are most readily observed, and it is for microscopic systems that this state of affairs is to be found. But that is not to say that quantum effects are not present at the macroscopic level. Peculiar quantum effects associated with the superposition of states are not to be seen, but the properties of matter in general, and indeed the properties of the forces of nature, are all intrinsically quantum in origin.

Chapter 9

General Mathematical Description of a Quantum System

It was shown in preceding Chapter that the mathematical description of this sum of probability amplitudes admits an interpretation of the state of the system as being a vector in a complex vector space, the state space of the system. It is this mathematical picture that is summarized here in the general case introduced there. This idea that the state of a quantum system is to be considered a vector belonging to a complex vector space, which we have developed here in the case of a spin half system, and which has its roots in the sum over paths point of view, is the basis of all of modern quantum mechanics and is used to describe any quantum mechanical system. Below is a summary of the main points as they are used for a general quantum system whose state spaces are of arbitrary dimension (including state spaces of infinite dimension). The emphasis here is on the mathematical features of the theory.

9.1 State Space

We have indicated a number of times that in quantum mechanics, the state of a physical system is represented by a vector belonging to a complex vector space known as the state space of the system. Here we will give a list of the defining conditions of a state space, though we will not be concerning ourselves too much with the formalities. The following definitions and concepts set up the state space of a quantum system.

1. Every physical state of a quantum system is specified by a symbol knonw as a *ket* written $|\ldots\rangle$ where $\ldots$ is a label specifying the physical information known about the state. An arbitrary state is written $|\psi\rangle$, or $|\phi\rangle$ and so on.

2. The set of all state vectors describing a given physical system forms a complex inner product space $\mathcal{H}$ (actually a Hilbert space, see Sec. 9.2) also known as the state space or ket space for the system. A ket is also referred to as a state vector, ket vector, or sometimes just state.

   Thus every linear combination (or superposition) of two or more state vectors $|\phi_1\rangle$, $|\phi_2\rangle$, $|\phi_3\rangle$, $\ldots$, is also a state of the quantum system i.e. the state $|\psi\rangle$ given by

   $$|\psi\rangle = c_1|\phi_1\rangle + c_2|\phi_2\rangle + c_3|\phi_3\rangle + \ldots$$

   is a state of the system for all complex numbers $c_1, c_2, c_3, \ldots$.

This last point amounts to saying that every physical state of a system is represented by a vector in the state space of the system, and every vector in the state space represents a possible physical state of the system. To guarantee this, the following condition is also imposed:

3. If a physical state of the system is represented by a vector $|\psi\rangle$, then the same physical state is represented by the vector $c|\psi\rangle$ where c is any non-zero complex number.

Next, we need the concept of *completeness*:

4. A set of vectors $|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle, \ldots$ is said to be complete if every state of the quantum system can be represented as a linear combination of the $|\varphi_i\rangle$'s, i.e. for any state $|\psi\rangle$ we can write

$$|\psi\rangle = \sum_i c_1 |\varphi_i\rangle.$$

   The set of vectors $|\varphi_i\rangle$ are said to *span* the space.

For example, returning to the spin half system, the two states $|\pm\rangle$ are all that is needed to describe any state of the system, i.e. there are no spin states that cannot be described in terms of these basis states. Thus, these states are said to be complete.

Finally, we need the concept of a set of basis states, and of the dimension of the state space.

5. A set of vectors $\{|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle, \ldots\}$ is said to form a basis for the state space if the set of vectors is complete, and if they are linearly independent. The vectors are also termed the *base* states for the vector space.

   Linear independence means that if $\sum_i c_i|\varphi_i\rangle = 0$ then $c_i = 0$ for all i.

The states $|\pm\rangle$ for a spin half system can be shown to be linearly independent, and thus form a basis for the state space of the system.

6. The minimum number of vectors needed to form a complete set of basis states is known as the dimension of the state space. [In many, if not most cases of interest in quantum mechanics, the dimension of the state space is infinite.]

It should be noted that there is an infinite number of possible sets of basis states for any state space. The arguments presented in the preceding Chapter by which we arrive at a set of basis states serves as a physically motivated starting point to construct the state space for the system. But once we have defined the state space in this way, there is no reason why we cannot, at least mathematically, construct other sets of basis states. These basis states that we start with are particularly useful as they have an immediate physical meaning; this might not be the case for an arbitrary basis set. But there are other means by which other physically meaningful basis states can be determined: often the choice of basis states is suggested by the physics (such as the set of eigenstates of an observable, see Chapter 11).

9.2 Probability Amplitudes and the Inner Product of State Vectors

We obtained a number of properties of probability amplitudes when looking at the case of a spin half system. Some of the results obtained there, and a few more that were not, are summarized in the following.

If $|\phi\rangle$ and $|\psi\rangle$ are any two state vectors belonging to the state space $\mathcal{H}$, then

1. $\langle\phi|\psi\rangle$, a complex number, is the probability amplitude of observing the system to be in the state $|\phi\rangle$ given that it is in the state $|\psi\rangle$.

2. The probability of observing the system to be in the state $|\phi\rangle$ given that it is in the state $|\psi\rangle$ is $|\langle\phi|\psi\rangle|^2$.

The probability amplitude $\langle\phi|\psi\rangle$, can then be shown to have the properties

3. $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$.
4. $\langle\phi|\{c_1|\psi_1\rangle + c_2|\psi_2\rangle\} = c_1\langle\phi|\psi_1\rangle + c_2\langle\phi|\psi_2\rangle$ where c_1 and c_2 are complex numbers.
5. $\langle\psi|\psi\rangle \geq 0$. If $\langle\psi|\psi\rangle$=0 then $|\psi\rangle = 0$, the zero vector.

This last statement is related to the physically reasonable requirement that the probability of a system being found in a state $|\psi\rangle$ given that it is in the state $|\psi\rangle$ has to be unity, i.e. $|\langle\psi|\psi\rangle|^2 = 1$ which means that $\langle\psi|\psi\rangle = \exp(i\eta)$. We now *choose* $\eta = 0$ so that $\langle\psi|\psi\rangle = 1$. But recall that any multiple of a state vector still represents the same physical state of the system, i.e. $|\widetilde{\psi}\rangle = a|\psi\rangle$ still represents the same physical state as $|\psi\rangle$. However, in this case, $\langle\widetilde{\psi}|\widetilde{\psi}\rangle = |a|^2$ which is not necessarily unity, but is is certainly bigger than zero.

6. The quantity $\sqrt{\langle\psi|\psi\rangle}$ is known as the *length* or *norm* of $|\psi\rangle$.
7. A state $|\psi\rangle$ is normalized, or normalized to unity, if $\langle\psi|\psi\rangle = 1$.

Normalized states are states which have a direct probability interpretation. It is mathematically convenient to permit the use of states whose norms are not equal to unity, but it is necessary in order to make use of the probability interpretation to deal only with the normalized state which has norm of unity. Any state that cannot be normalized to unity (i.e. it is of infinite length) cannot represent a physically acceptable state.

8. Two states $|\phi\rangle$ and $|\psi\rangle$ are orthogonal if $\langle\phi|\psi\rangle = 0$.

The physical significance of two states being orthogonal should be understood: for a system in a certain state, there is zero probability of it being observed in a state with which it is orthogonal. In this sense, two orthogonal states are as distinct as it is possible for two states to be.

Finally, a set of orthonormal basis vectors $\{|\varphi_n\rangle; n = 1, 2, \dots\}$ will have the property

9. $\langle\varphi_m|\varphi_n\rangle = \delta_{mn}$ where δ_{mn} is known as the Kronecker delta, and equals unity if $m = n$ and zero if $m \neq n$.

All the above conditions satisfied by probability amplitudes were to a greater or lesser extent physically motivated, but it nevertheless turns out that these conditions are identical to the conditions that are used to define the inner product of two vectors in a complex vector space, in this case, the state space of the system, i.e. we could write, using the usual mathematical notation for an inner product, $\langle\phi|\psi\rangle = (|\phi\rangle, |\psi\rangle)$. The state space of a physical system is thus more than just a complex vector space, it is a vector space on which there is defined an inner product, and so is more correctly termed a complex 'inner product' space. Further, it is usually required in quantum mechanics that certain convergency criteria, defined in terms of the norms of sequences of vectors belonging to the state space, must be satisfied. This is not of any concern for spaces of finite dimension, but are important for spaces of infinite dimension. If these criteria are satisfied then the state space is said to be a Hilbert space. Thus rather than referring to the state space of a system, reference is made to the Hilbert space of the system.

What this means, mathematically, is that for every state $|\phi\rangle$ say, at least one of the inner products $\langle\varphi_n|\phi\rangle$ will be non-zero, or conversely, there does not exist a state $|\xi\rangle$ for which $\langle\varphi_n|\xi\rangle = 0$ for all the basis states $|\varphi_n\rangle$. Completeness clearly means that no more basis states are needed to describe any possible physical state of a system.

It is important to recognize that all the vectors belonging to a Hilbert space have finite norm, or, putting it another way, all the state vectors can be normalized to unity – this state of affairs is physically necessary if we want to be able to apply the probability interpretation in a consistent way. However, as we shall see, we will encounter states which do not have a finite norm and hence neither represent physically realizable states, nor do they belong to the state or Hilbert space of the system. Nevertheless, with proper care regarding their use and interpretation, such states turn out to be essential, and play a crucial role throughout quantum mechanics.

Recognizing that a probability amplitude is nothing but an inner product on the state space of the system, leads to a more general way of defining what is meant by a bra vector. The following discussion emphasizes the fact that a bra vector, while it shares many characteristices of a ket vector, is actually a different mathematical entity.

9.2.1 Bra Vectors

We have consistently used the notation $\langle\phi|\psi\rangle$ to represent a probability amplitude, but we have just seen that this quantity is in fact nothing more than the inner product of two state vectors, which can be written in a different notation, $(|\phi\rangle, |\psi\rangle)$, that is more commonly encountered in pure mathematics. But the inner product can be viewed in another way, which leads to a new interpretation of the expression $\langle\phi|\psi\rangle$, and the introduction of a new class of state vectors. If we consider the equation

$$\langle\phi|\psi\rangle = (|\phi\rangle, |\psi\rangle) \tag{9.1}$$

and 'cancel' the $|\psi\rangle$, we get the result

$$\langle\phi|\bullet = (|\phi\rangle,\ \bullet) \tag{9.2}$$

where the '$\bullet$' is inserted, temporarily, to remind us that in order to complete the equation, a ket vector has to be inserted. By carrying out this procedure, we have introduced a new quantity $\langle\phi|$ which is known as a bra or bra vector, essentially because $\langle\phi|\psi\rangle$ looks like quantities enclosed between a pair of 'bra(c)kets'. It is a vector because, as can be readily shown, the collection of all possible bras form a vector space. For instance, by the properties of the inner product, if

$$|\psi\rangle = a_1|\varphi_1\rangle + a_2|\varphi_2\rangle \tag{9.3}$$

then

$$(|\psi\rangle,\ \bullet) = \langle\psi|\bullet = (a_1|\varphi_1\rangle + a_2|\varphi_2\rangle,\ \bullet) \tag{9.4}$$

$$= a_1^*(|\varphi_1\rangle,\ \bullet) + a_2^*(|\varphi_2\rangle,\ \bullet) = a_1^*\langle\varphi_1|\bullet + a_2^*\langle\varphi_2|\bullet \tag{9.5}$$

i.e., dropping the '$\bullet$' symbols, we have

$$\langle\psi| = a_1^*\langle\varphi_1| + a_2^*\langle\varphi_2| \tag{9.6}$$

so that a linear combination of two bras is also a bra, from which follows (after a bit more work checking that the other requirements of a vector space are also satisfied) the result that the set of all bras is a vector space. Incidentally, this last calculation above shows, once again, that if $|\psi\rangle = a_1|\varphi_1\rangle + a_2|\varphi_2\rangle$ then the corresponding bra is $\langle\psi| = a_1^*\langle\varphi_1| + a_2^*\langle\varphi_2|$. So, in a sense, the bra vectors are the 'complex conjugates' of the ket vectors.

The vector space of all bra vectors is obviously closely linked to the vector space of all the kets $\mathcal{H}$, and is in fact usually referred to as the dual space, and represented by $\mathcal{H}^*$. To each ket vector $|\psi\rangle$ belonging to $\mathcal{H}$, there is then an associated bra vector $\langle\psi|$ belonging to the dual space $\mathcal{H}^*$. However, the reverse is not necessarily true: there are bra vectors that do not necessarily have a corresponding ket vector, and therein lies the difference between bras and kets. It turns out that the difference only matters for Hilbert spaces of infinite dimension, in which case there can arise bra vectors whose corresponding ket vector is of infinite length, i.e. has infinite norm, and hence cannot be normalized to unity. Such ket vectors can therefore never represent a possible physical state of a system. But these issues will not be of any concern here. The point to be taken away from all this is that a bra vector is not the same kind of mathematical object as a ket vector. In fact, it has all the attributes of an operator in the sense that it acts on a ket vector to produce a complex number, this complex number being given by the appropriate inner product. This is in contrast to the more usual sort of operators encountered in quantum mechanics that act on ket vectors to produce other ket vectors. In mathematical texts a bra vector is usually referred to as a 'linear functional'. Nevertheless, in spite of the mathematical distinction that can be made between bra and ket vectors, the correspondence between the two kinds of vectors is in most circumstances so complete that a bra vector equally well represents the state of a quantum system as a ket vector. Thus, we can talk of a system being in the state $\langle\psi|$.

We can summarize all this in the general case as follows: The inner product $(|\psi\rangle, |\phi\rangle)$ defines, for all states $|\psi\rangle$, the set of functions (or linear functionals) $(|\psi\rangle, \ )$. The linear functional $(|\psi\rangle, \ )$ maps any ket vector $|\phi\rangle$ into the complex number given by the inner product $(|\psi\rangle, |\phi\rangle)$.

1. The set of all linear functionals $(|\psi\rangle, \ )$ forms a complex vector space $\mathcal{H}^*$, the dual space of $\mathcal{H}$.

2. The linear functional $(|\psi\rangle, \ )$ is written $\langle\psi|$ and is known as a bra vector.

3. To each ket vector $|\psi\rangle$ there corresponds a bra vector $\langle\psi|$ such that if $|\phi_1\rangle \rightarrow \langle\phi_1|$ and $|\phi_2\rangle \rightarrow \langle\phi_2|$ then
$$c_1|\phi_1\rangle + c_2|\phi_2\rangle \rightarrow c_1^*\langle\phi_1| + c_2^*\langle\phi_2|.$$

Chapter 10

State Spaces of Infinite Dimension

So far we have limited the discussion to state spaces of finite dimensions, but it turns out that, in practice, state spaces of infinite dimension are fundamental to a quantum description of almost all physical systems. The simple fact that a particle moving in space requires for its quantum mechanical description a state space of infinite dimension shows the importance of being able to work with such state spaces. This would not be of any concern if doing so merely required transferring over the concepts already introduced in the finite case, but infinite dimensional state spaces have mathematical peculiarities and associated physical interpretations that are not found in the case of finite dimension state spaces. Some of these issues are addressed in this Chapter, while other features of infinite dimensional state spaces are discussed as the need arises in later Chapters.

10.1 Examples of state spaces of infinite dimension

All the examples given in Chapter 8 yield state spaces of finite dimension. Much the same argument can be applied to construct state spaces of infinite dimension. A couple of examples follow.

The Tight-Binding Model of a Crystalline Metal The examples given above of an electron being positioned on one of a (finite) number of atoms can be readily generalized to a situation in which there are an infinite number of such atoms. This is not a contrived model in any sense, as it is a good first approximation to modelling the properties of the conduction electrons in a crystalline solid. In the free electron model of a conducting solid, the conduction electrons are assumed to be able to move freely (and without mutual interaction) through the crystal, i.e. the effects of the background positive potentials of the positive ions left is ignored. A further development of this model is to take into account the fact that the electrons will experience some attraction to the periodically positioned positive ions, and so there will be a tendency for the electrons to be found in the neighbourhood of these ions. The resultant model – with the basis states consisting of a conduction electron being found on any one of the ion sites – is obviously similar to the one above for the molecular ion. Here however, the number of basis states is infinite (for an infinite crystal), so the state space is of infinite dimension. Representing the set of basis states by $\{|n\rangle, n = 0, \pm 1, \pm 2, \dots\}$ where na is the position of the n^{th} atom, and a is the separation between neighbouring atoms, then any state of the system can then be written as

$$|\psi\rangle = \sum_{n=-\infty}^{+\infty} c_n |n\rangle. \tag{10.1}$$

By taking into account the fact that the electrons can make their way from an ion to one of its neighbours, much of the band structure of semiconducting solids can be obtained.

Free Particle We can generalize the preceding model by supposing that the spacing between the neighbouring atoms is allowed to go to zero, so that the positions at which the electron can be found become continuous. This then acts as a model for the description of a particle free to move anywhere in one dimension, and is considered in greater detail later in Section 10.2.2. In setting up this model, we find that as well as there being an infinite number of basis states — something we have already encountered — we see that these basis states are not discrete, i.e. a particle at position x will be in the basis state $|x\rangle$, and as x can vary continuously over the range $-\infty < x < \infty$, there will be a non-denumerably infinite, that is, a continuous range of such basis states. As a consequence, the completeness relation ought to be written as an integral:

$$|\psi\rangle = \int_{-\infty}^{+\infty} |x\rangle\langle x|\psi\rangle \, dx. \tag{10.2}$$

The states $|x\rangle$ and $|x'\rangle$ will be orthonormal if $x \neq x'$, but in order to be able to retain the completeness relation in the form of an integral, it turns out that these basis states have to have an infinite norm. However, there is a sense in which we can continue to work with such states, as will be discussed in Section 10.2.2.

Particle in an Infinitely Deep Potential Well We saw in Section 5.3 that a particle of mass m in an infinitely deep potential well of width L can have the energies $E_n = n^2\pi^2\hbar^2/2mL^2$ where n is a positive integer. This suggests that the basis states of the particle in the well be the states $|n\rangle$ such that if the particle is in state $|n\rangle$, then it has energy E_n. The probability amplitude of finding the particle at position x when in state $|n\rangle$ is then $\langle x|n\rangle$ which, from Section 5.3 we can identify with the wave function ψ_n, i.e.

$$\begin{aligned} \psi_n(x) = \langle x|n\rangle &= \sqrt{\frac{2}{L}} \sin(n\pi x/L) \quad 0 < x < L \\ &= 0 \qquad\qquad\qquad\qquad x < 0, \quad x > L. \end{aligned} \tag{10.3}$$

The state space is obviously of infinite dimension.

It has been pointed out before that a state space can have any number of sets of basis states, i.e. the states $|n\rangle$ introduced here do not form the sole possible set of basis states for the state space of this system. In this particular case, it is worthwhile noting that we could have used as the base states the states labelled by the position of the particle in the well, i.e. the states $|x\rangle$.

As we have seen, there are an infinite number of such states which is to be expected as we have already seen that the state space is of infinite dimension. But the difference between this set of states and the set of states $|n\rangle$ is that in the latter case, these states are discrete, i.e. they can be labelled by the integers, while the states $|x\rangle$ are continuous, they are labelled by the continuous variable x. Thus, something new emerges from this example: for state spaces of infinite dimension, it is possible to have a denumerably infinite number of basis states (i.e. the discrete states $|n\rangle$) or non-denumerably infinite number of basis states (i.e. the states $|x\rangle$.) This feature of state spaces of infinite dimension, plus others, are discussed separately below in Section 10.2.

A System of Identical Photons Many other features of a quantum system not related to the position or energy of the system can be used as a means by which a set of basis states can be set up. An important example is one in which the system consists of a possibly variable number of identical particles. One example is a 'gas' of photons, all of the same frequency and polarization. Such a situation is routinely achieved in the laboratory using suitably constructed hollow superconducting metallic cavities designed to support just one mode (i.e. a single frequency and polarization) of the electromagnetic field. The state of the electromagnetic field can then be characterized by the number n of photons in the field which can range from zero to positive infinity, so that the states of the field (known as number states) can be written $|n\rangle$ with $n = 0, 1, 2, \ldots$. The state $|0\rangle$ is often

referred to as the vacuum state. These states will then constitute a complete, orthonormal set of basis states (called Fock or number states), i.e.

$$\langle n|m\rangle = \delta_{nm} \tag{10.4}$$

and as n can range up to infinity, the state space for the system will be infinite dimensional. An arbitrary state of the cavity field can be then be written

$$|\psi\rangle = \sum_{n=0}^{\infty} c_n|n\rangle \tag{10.5}$$

so that $|c_n|^2$ will be the probability of finding n photons in the field. In terms of these basis states, it is possible to describe the processes in which particles are created or destroyed. For instance if there is a single atom in an excited energy state in the cavity, and the cavity is in the vacuum state $|0\rangle$, then the state of the combined atom field system can be written $|e, 0\rangle$, where the e indicates that the atom is in an excited state. The atom can later lose this energy by emitting it as a photon, so that at some later time the state of the system will be $a|e, 0\rangle + b|g, 1\rangle$, where now there is the possibility, with probability $|b|^2$, of the atom being found in its ground state, and a photon having been created.

10.2 Some Mathematical Issues

Some examples of physical systems with state spaces of infinite dimension were provided in the previous Section. In these examples, we were able to proceed, at least as far as constructing the state space was concerned, largely as was done in the case of finite dimensional state spaces. However, further investigation shows that there are features of the mathematics, and the corresponding physical interpretation in the infinite dimensional case that do not arise for systems with finite dimensional state spaces. Firstly, it is possible to construct state vectors that cannot represent a state of the system and secondly, the possibility arises of the basis states being continuously infinite. This latter state of affairs is not at all a rare and special case — it is just the situation needed to describe the motion of a particle in space, and hence gives rise to the wave function, and wave mechanics.

10.2.1 States of Infinite Norm

To illustrate the first of the difficulties mentioned above, consider the example of a system of identical photons in the state $|\psi\rangle$ defined by Eq. (10.5). As the basis states are orthonormal we have for $\langle\psi|\psi\rangle$

$$\langle\psi|\psi\rangle = \sum_{n=0}^{\infty} |c_n|^2 \tag{10.6}$$

If the probabilities $|c_n|^2$ form a convergent infinite series, then the state $|\psi\rangle$ has a finite norm, i.e. it can be normalized to unity. However, if this series does not converge, then it is not possible to supply a probability interpretation to the state vector as it is not normalizable to unity. For instance, if $c_0 = 0$ and $c_n = 1/\sqrt{n}, n = 1, 2, \ldots$, then

$$\langle\psi|\psi\rangle = \sum_{n=1}^{\infty} \frac{1}{n} \tag{10.7}$$

which is a divergent series, i.e. this state cannot be normalized to unity. In contrast, if $c_n = 1/n, n = 1, 2, \ldots$, then

$$\langle\psi|\psi\rangle = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{10.8}$$

which means we can normalize this state to unity by defining

$$|\tilde{\psi}\rangle = \frac{\sqrt{6}}{\pi}|\psi\rangle. \tag{10.9}$$

This shows that there are some linear combination of states that do not represent possible physical states of the system. Such states do not belong to the Hilbert space $\mathcal{H}$ of the system, i.e. the Hilbert space consists only of those states for which the coefficients c_n satisfy Eq. (10.6) [1]. This is a new feature: the possibility of constructing vectors that do not represent possible physical states of the system. It turns out that some very useful basis states have this apparently undesirable property, as we will now consider.

10.2.2 Continuous Basis States

In Section 10.1 an infinite one-dimensional model of a crystal was used as an illustrative model for a state space of infinite dimension. We can now consider what happens if we suppose that the separation between the neighbouring atoms in the crystal goes to zero, so that the electron can be found anywhere over a range extending from $-\infty$ to ∞. This, in effect, is the continuous limit of the infinite crystal model just presented, and represents the possible positions that a particle free to move anywhere in one dimension, the X axis say, can have.

In this case, we could label the possible states of the particle by its X position, i.e. $|x\rangle$, where now, instead of having the discrete values of the crystal model, the position can now assume any of a continuous range of values, $-\infty < x < \infty$. It would seem that we could then proceed in the same way as we have done with the discrete states above, but it turns out that such states cannot be normalized to unity and hence do not represent (except in an idealised sense) physically allowable states of the system.

The aim here is to try to develop a description of the possible basis states for a particle that is not limited to being found only at discrete positions on the X axis. After all, in principle, we would expect that a particle in free space could be found at any position x in the range $-\infty < x < \infty$. We will get at this description by a limiting procedure which is not at all mathematically rigorous, but nevertheless yields results that turn out to be valid. Suppose we return to the completeness relation for the states $|na\rangle$ for the one dimensional crystal

$$|\psi\rangle = \sum_{n=-\infty}^{+\infty} |na\rangle\langle na|\psi\rangle. \tag{10.10}$$

If we now put $a = \Delta x$ and $na = x_n$, and write $|na\rangle = \sqrt{a}|x_n\rangle$, this becomes

$$|\psi\rangle = \sum_{n=-\infty}^{+\infty} |x_n\rangle\langle x_n|\psi\rangle\Delta x \tag{10.11}$$

where now

$$\langle x_n|x_m\rangle = \frac{\delta_{nm}}{a} \tag{10.12}$$

[1] Note however, that we can still construct a bra vector

$$\langle\psi| = \sum_{n=0}^{n=\infty} c_n^*\langle n|$$

without placing any restrictions on the convergence of the c_n's such as the one in Eq. (10.6). The corresponding ket cannot then represent a possible state of the system, but such inner products as $\langle\psi|\phi\rangle$ where $|\phi\rangle$ is a normalized ket can still be evaluated. The point being made here is that if $\mathcal{H}$ is of infinite dimension, the dual space $\mathcal{H}^*$ can also include bra vectors that do not correspond to normalized ket vectors in $\mathcal{H}$, which emphasizes the fact that $\mathcal{H}^*$ is defined as a set of linear functionals, and not simply as a 'complex conjugate' version of $\mathcal{H}$. The distinction is important in some circumstances, but we will not have to deal with such cases.

i.e. each of the states $|x_n\rangle$ is not normalized to unity, but we can nevertheless identify such a state as being that state for which the particle is at position x_n – recall if a state vector is multiplied by a constant, it still represents the same physical state of the system.

If we put to one side any concerns about the mathematical legitimacy of what follows, we can now take the limit $\Delta x \to 0$, i.e. $a \to 0$, then Eq. (10.11) can be written as an integral, i.e.

$$|\psi\rangle = \int_{-\infty}^{+\infty} |x\rangle\langle x|\psi\rangle \, dx \tag{10.13}$$

We can identify the state $|x\rangle$ with the physical state of affairs in which the particle is at the position x, and the expression Eq. (10.13) is consistent with the completeness requirement i.e. that the states $\{|x\rangle, -\infty < x < \infty\}$ form a complete set of basis states, so that any state of the one particle system can be written as a superposition of the states $|x\rangle$, though the fact that the label x is continuous has forced us to write the completeness relation as an integral. The difficulty with this procedure is that the states $|x\rangle$ are no longer normalized to unity. This we can see from Eq. (10.12) which tells us that $\langle x|x'\rangle$ will vanish if $x \neq x'$, but for $x = x'$ we see that

$$\langle x|x\rangle = \lim_{a\to 0} \frac{1}{a} = \infty \tag{10.14}$$

i.e. the state $|x\rangle$ has infinite norm! This means that there is a price to pay for trying to set up the mathematics in such a manner as to produce the completeness expression Eq. (10.13), which is that we are forced to introduce basis states which have infinite norm, and hence cannot represent a possible physical state of the particle! Nevertheless, provided care is taken, it is still possible to work with these states as if they represent the state in which the particle is at a definite position. To see this, we need to look at the orthonormality properties of these states, and in doing so we are lead to introduce a new kind of function, the Dirac delta function.

10.2.3 The Dirac Delta Function

We have just seen that the inner product $\langle x|x'\rangle$ vanishes if $x \neq x'$, but appears to be infinite if $x = x'$. In order to give some mathematical sense to this result, we return to Eq. (10.13) and look more closely at the properties that $\langle x|x'\rangle$ must have in order for the completeness relation also to make sense.

The probability amplitude $\langle x|\psi\rangle$ appearing in Eq. (10.13) are functions of the continuous variable x, and is often written $\langle x|\psi\rangle = \psi(x)$, which we identify as the wave function of the particle. If we now consider the inner product

$$\langle x'|\psi\rangle = \int_{-\infty}^{+\infty} \langle x'|x\rangle\langle x|\psi\rangle dx \tag{10.15}$$

or

$$\psi(x') = \int_{-\infty}^{+\infty} \langle x'|x\rangle\psi(x) dx \tag{10.16}$$

we now see that we have an interesting difficulty. We know that $\langle x'|x\rangle = 0$ if $x' \neq x$, so if $\langle x|x\rangle$ is assigned a finite value, the integral on the right hand side will vanish, so that $\psi(x) = 0$ for all x!! But if $\psi(x)$ is to be a non-trivial quantity, i.e. if it is not to be zero for all x, then it cannot be the case that $\langle x|x\rangle$ is finite. In other words, $\langle x'|x\rangle$ must be infinite for $x = x'$ in some sense in order to guarantee a non-zero integral. The way in which this can be done involves introducing a new 'function', the Dirac delta function, which has some rather unusual properties.

What we are after is a 'function' $\delta(x - x_0)$ with the property that

$$f(x_0) = \int_{-\infty}^{+\infty} \delta(x - x_0) f(x) dx \tag{10.17}$$

for all (reasonable) functions $f(x)$.

So what is $\delta(x - x_0)$? Perhaps the simplest way to get at what this function looks like is to examine beforehand a sequence of functions defined by

$$\begin{aligned} D(x, \epsilon) &= \epsilon^{-1} \qquad -\epsilon/2 < x < \epsilon/2 \\ &= 0 \qquad x < -\epsilon/2, x > \epsilon/2. \end{aligned} \tag{10.18}$$

What we first notice about this function is that it defines a rectangle whose area is always unity for any (non-zero) value of ϵ, i.e.

$$\int_{-\infty}^{+\infty} D(x, \epsilon) dx = 1. \tag{10.19}$$

Secondly, we note that as ϵ is made smaller, the rectangle becomes taller and narrower. Thus, if we look at an integral

$$\int_{-\infty}^{+\infty} D(x, \epsilon) f(x) dx = \epsilon^{-1} \int_{-\epsilon/2}^{\epsilon/2} f(x) dx \tag{10.20}$$

where $f(x)$ is a reasonably well behaved function (i.e. it is continuous in the neighbourhood of $x = 0$), we see that as $\epsilon \to 0$, this tends to the limit $f(0)$. We can summarize this by the equation

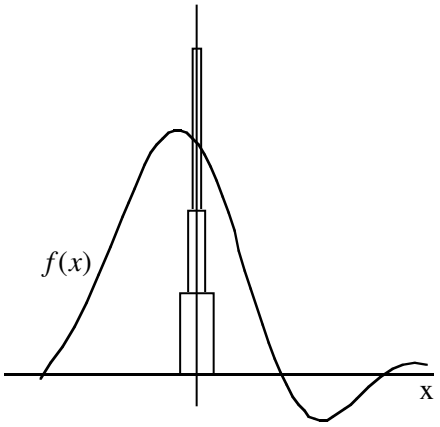


Figure 10.1: A sequence of rectangles of decreasing width but increasing height, maintaining a constant area of unity approaches an infinitely high 'spike' at $x = 0$..

$$\lim_{\epsilon \to 0} \int_{-\infty}^{+\infty} D(x, \epsilon) f(x) dx = f(0). \tag{10.21}$$

Taking the limit inside the integral sign (an illegal mathematical operation, by the way), we can write this as

$$\int_{-\infty}^{+\infty} \lim_{\epsilon \to 0} D(x, \epsilon) f(x) dx = \int_{-\infty}^{+\infty} \delta(x) f(x) dx = f(0) \tag{10.22}$$

where we have introduced the 'Dirac delta function' $\delta(x)$ defined as the limit

$$\delta(x) = \lim_{\epsilon \to 0} D(x, \epsilon), \tag{10.23}$$

a function with the unusual property that it is zero everywhere except for $x = 0$, where it is infinite.

The above defined function $D(x, \epsilon)$ is but one 'representation' of the Dirac delta function. There are in effect an infinite number of different functions that in an appropriate limit behave as the rectangular function here. Some examples are

$$\begin{aligned} \delta(x - x_0) &= \lim_{L \to \infty} \frac{1}{\pi} \frac{\sin L(x - x_0)}{x - x_0} \\ &= \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{(x - x_0)^2 + \epsilon^2} \\ &= \lim_{\lambda \to \infty} \tfrac{1}{2} \lambda e^{-\lambda |x - x_0|}. \end{aligned} \tag{10.24}$$

In all cases, the function on the right hand side becomes narrower and taller as the limit is taken, while the area under the various curves remains the same, that is, unity.

The first representation above is of particular importance. It arises by via the following integral:

$$\frac{1}{2\pi}\int_{-L}^{+L} e^{ik(x-x_0)}dk = \frac{e^{iL(x-x_0)} - e^{iL(x-x_0)}}{2\pi i(x-x_0)} = \frac{1}{\pi}\frac{\sin L(x-x_0)}{x-x_0} \tag{10.25}$$

In the limit of $L \to \infty$, this then becomes

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ik(x-x_0)}dk = \delta(x-x_0). \tag{10.26}$$

The delta function is not to be thought of as a function as it is usually defined in pure mathematics, but rather it is to be understood that a limit of the kind outlined above is implied whenever the delta function appears in an integral[2]. However, such mathematical niceties do not normally need to be a source of concern in most instances. It is usually sufficient to be aware of the basic property Eq. (10.17) and a few other rules that can be proven using the limiting process, such as

$$\begin{aligned}
\delta(x) &= \delta(-x) \\
\delta(ax) &= \frac{1}{|a|}\delta(x) \\
x\delta(x) &= 0 \\
\int_{-\infty}^{+\infty} \delta(x-x_0)\delta(x-x_1) &= \delta(x_0-x_1) \\
\int_{-\infty}^{+\infty} f(x)\delta'(x-x_0)dx &= -f'(x_0).
\end{aligned}$$

The limiting process should be employed if there is some doubt about any result obtained. For instance, it can be shown that the square of a delta function cannot be given a satisfactory meaning.

Delta Function Normalization

Returning to the result

$$\psi(x') = \int_{-\infty}^{+\infty} \langle x'|x\rangle\psi(x)dx \tag{10.27}$$

we see that the inner product $\langle x'|x\rangle$, must be interpreted as a delta function:

$$\langle x'|x\rangle = \delta(x-x'). \tag{10.28}$$

The states $|x\rangle$ are said to be delta function normalized, in contrast to the orthonormal property of discrete basis states. One result of this, as has been pointed out earlier, is that states such as $|x\rangle$ are of infinite norm and so cannot be normalized to unity. Such states cannot represent possible physical states of a system, though it is often convenient, with caution, to speak of such states as if they were physically realizable. Mathematical (and physical) paradoxes can arise if care is not taken. However, linear combinations of these states can be normalized to unity, as this following example illustrates. If we consider a state $|\psi\rangle$ given by

$$|\psi\rangle = \int_{-\infty}^{+\infty} |x\rangle\langle x|\psi\rangle dx, \tag{10.29}$$

[2]This raises the question as to whether or not it would matter what representation of the delta function is used. Provided the function $f(x)$ is bounded over $(-\infty, +\infty)$ there should be no problem, but if the function $f(x)$ is unbounded over this interval, e.g. $f(x) = \exp(x^2)$, then only the rectangular representation of the delta function will give a sensible answer.

then

$$\langle\psi|\psi\rangle = \int_{-\infty}^{+\infty} \langle\psi|x\rangle\langle x|\psi\rangle dx. \tag{10.30}$$

But $\langle x|\psi\rangle = \psi(x)$ and $\langle\psi|x\rangle = \psi(x)^*$, so that

$$\langle\psi|\psi\rangle = \int_{-\infty}^{\infty} |\psi(x)|^2 dx. \tag{10.31}$$

Provided $|\psi(x)|^2$ is a well behaved function, in particular that it vanish as $x \to \pm\infty$, this integral will converge to a finite result, so that the state $|\psi\rangle$ can indeed be normalized to unity, and if so, then we can interpret $|\psi(x)|^2 dx$ as the probability of finding the particle in the region $(x, x + dx)$, which is just the standard Born interpretation of the wave function.

10.2.4 Separable State Spaces

We have seen that state spaces of infinite dimension can be set up with either a denumerably infinite number of basis states, i.e. the basis states are discrete but infinite in number, or else a non-denumerably infinite number of basis states, i.e. the basis states are labelled by a continuous parameter. Since a state space can be spanned by more than one set of basis states, it is worthwhile investigating whether or not a space of infinite dimension can be spanned by a set of denumerable basis states, as well as a set of non-denumerable basis states. An example of where this is the case was given earlier, that of a particle in an infinitely deep potential well, see p 132. It transpires that not all vector spaces of infinite dimension have this property, i.e. that they can have both a denumerable and a non-denumerable set of basis states. Vector spaces which can have both kinds of basis states are said to be separable, and in quantum mechanics it is assumed that state spaces are separable.

Chapter 11

Operations on States

We have seen in the preceding Chapter that the appropriate mathematical language for describing the states of a physical system is that of vectors belonging to a complex vector space. But the state space by itself is insufficient to fully describe the properties of a physical system. Describing such basic physical processes as how the state of a system evolves in time, or how to represent fundamental physical quantities such as position, momentum or energy, requires making use of further developments in the mathematics of vector spaces. These developments involve introducing a new mathematical entity known as an operator whose role it is to 'operate' on state vectors and map them into other state vectors. In fact, the earliest version of modern quantum mechanics, that put forward by Heisenberg, was formulated by him in terms of operators represented by matrices, at that time a not particularly well known (even to Heisenberg) development in pure mathematics. It was Born who recognized, and pointed out to Heisenberg, that he was using matrices in his work – another example of a purely mathematical construct that has proven to be of immediate value in describing the physical world.

Operators play many different roles in quantum mechanics. They can be used to represent physical processes that result in the change of state of the system, such as the evolution of the state of a system in time, or the creation or destruction of particles such as occurs, for instance in the emission or absorption of photons – particles of light – by matter. But operators have a further role, the one recognized by Heisenberg, which is to represent the physical properties of a system that can be, in principle, experimentally measured, such as energy, momentum, position and so on, so-called observable properties of a system. It is the aim of this Chapter to introduce the mathematical concept of an operator, and to show what the physical significance is of operators in quantum mechanics.

In general, in what follows, results explicitly making use of basis states will be presented only for the case in which these basis states are discrete, though in most cases, the same results hold true if a continuous set of basis states, as would arise for state spaces of infinite dimension, were used. The modifications that are needed when this is not the case, are considered towards the end of the Chapter.

11.1 Definition and Properties of Operators

11.1.1 Definition of an Operator

The need to introduce the mathematical notion of an operator into quantum mechanics can be motivated by the recognition of a fairly obvious characteristic of a physical system: it will evolve in time i.e. its state will be time dependent. But a change in state could also come about because some action is performed on the system, such as the system being displaced or rotated in space. The state of a multiparticle system can change as the consequence of particles making up a system

being created or destroyed. Changes in state can also be brought about by processes which have a less physically direct meaning. Whatever the example, the fact that the state of a system in quantum mechanics is represented by a vector immediately suggests the possibility of describing a physical process by an exhaustive list of all the changes that the physical process induces on *every* state of the system, i.e. a table of all possible before and after states. A mathematical device can then be constructed which represents this list of before and after states. This mathematical device is, of course, known as an operator.

Thus the operator representing the above physical process is a mathematical object that acts on the state of a system and maps this state into some other state in accordance with the exhaustive list proposed above. If we represent an operator by a symbol $\hat{A}$ – note the presence of the $\hat{}$ – and suppose that the system is in a state $|\psi\rangle$, then the outcome of $\hat{A}$ acting on $|\psi\rangle$, written $\hat{A}|\psi\rangle$, (*not* $|\psi\rangle\hat{A}$, which is not a combination of symbols that has been assigned any meaning) defines a new state $|\phi\rangle$ say, so that

$$\hat{A}|\psi\rangle = |\phi\rangle. \tag{11.1}$$

As has already been mentioned, to fully characterize an operator, the effects of the operator acting on *every* possible state of the system must be specified. The extreme case, indicated above, requires, in effect, a complete tabulation of each state of the system, and the corresponding result of the operator acting on each state. More typically this formidable (impossible?) task would be unnecessary – all that would be needed is some sort of rule that enables $|\phi\rangle$ to be determined for any given $|\psi\rangle$.

Ex 11.1 Consider the operator $\hat{A}$ acting on the states of a spin half system, and suppose, for the arbitrary state $|S\rangle = a|+\rangle + b|-\rangle$, that the action of the operator $\hat{A}$ is such that $\hat{A}|S\rangle = b|+\rangle + a|-\rangle$, i.e. the action of the operator $\hat{A}$ on the state $|S\rangle$ is to exchange the coefficients $\langle\pm|S\rangle \leftrightarrow \langle\mp|S\rangle$. This rule is then enough to define the result of $\hat{A}$ acting on any state of the system, i.e. the operator is fully specified.

Ex 11.2 A slightly more complicated example is one for which the action of an operator $\hat{N}$ on the state $|S\rangle$ is given by $\hat{N}|S\rangle = a^2|+\rangle + b^2|-\rangle$. As we shall see below, the latter operator is of a kind not usually encountered in quantum mechanics, that is, it is non-linear, see Section 11.1.2 below.

While the above discussion provides motivation on physical grounds for introducing the idea of an operator as something that acts to change the state of a quantum system, it is in fact the case that many operators that arise in quantum mechanics, whilst mathematically they can act on a state vector to map it into another state vector, do not represent a physical process acting on the system. In fact, the operators that represent such physical processes as the evolution of the system in time, are but one kind of operator important in quantum mechanics known as unitary operators. Another very important kind of operator is that which represents the physically observable properties of a system, such as momentum or energy. Each such observable property, or *observable*, is represented by a particular kind of operator known as a Hermitean operator. Mathematically, such an operator can be made to act on the state of a system, thereby yielding a new state, but the interpretation of this as representing an actual physical process is much less direct. Instead, a Hermitean operator acts, in a sense, as a repository of all the possible results that can be obtained when performing a measurement of the physical observable that the operator represents. Curiously enough, it nevertheless turns out that Hermitean operators representing observables of a system, and unitary operators representing possible actions performed on a system are very closely related in a way that will be examined in Chapter 18.

In quantum mechanics, the task of fully characterizing an operator is actually made much simpler through the fact that most operators in quantum mechanics have a very important property: they are linear, or, at worst, anti-linear.

11.1.2 Linear and Antilinear Operators

There is essentially no limit to the way in which operators could be defined, but of particular importance are operators that have the following property. If $\hat{A}$ is an operator such that for any arbitrary pair of states $|\psi_1\rangle$ and $|\psi_2\rangle$ and for any complex numbers c_1 and c_2:

$$\hat{A}[c_1|\psi_1\rangle + c_2|\psi_2\rangle] = c_1\hat{A}|\psi_1\rangle + c_2\hat{A}|\psi_2\rangle, \tag{11.2}$$

then $\hat{A}$ is said to be a *linear* operator. In quantum mechanics, operators are, with one exception, linear. The exception is the time reversal operator $\hat{T}$ which has the property

$$\hat{T}[c_1|\psi_1\rangle + c_2|\psi_2\rangle] = c_1^*\hat{T}|\psi_1\rangle + c_2^*\hat{T}|\psi_2\rangle \tag{11.3}$$

and is said to be *anti-linear*. We will not have any need to concern ourselves with the time reversal operator, so any operator that we will be encountering here will be tacitly assumed to be linear.

Ex 11.3 Consider the operator $\hat{A}$ acting on the states of a spin half system, such that for the arbitrary state $|S\rangle = a|+\rangle + b|-\rangle$, $\hat{A}|S\rangle = b|+\rangle + a|-\rangle$. Show that this operator is linear.

Introduce another state $|S'\rangle = a'|+\rangle + b'|-\rangle$ and consider

$$\begin{aligned}\hat{A}[\alpha|S\rangle + \beta|S'\rangle] &= \hat{A}[(\alpha a + \beta a')|+\rangle + (\alpha b + \beta b')|-\rangle] \\ &= (\alpha b + \beta b')|+\rangle + (\alpha a + \beta a')|-\rangle \\ &= \alpha(b|+\rangle + a|-\rangle) + \beta(b'|+\rangle + a'|-\rangle) \\ &= \alpha\hat{A}|S\rangle + \beta\hat{A}|S'\rangle.\end{aligned}$$

and hence the operator $\hat{A}$ is linear.

Ex 11.4 Consider the operator $\hat{N}$ defined such that if $|S\rangle = a|+\rangle + b|-\rangle$ then $\hat{N}|S\rangle = a^2|+\rangle + b^2|-\rangle$. Show that this operator is non-linear.

If we have another state $|S'\rangle = a'|+\rangle + b'|-\rangle$, then

$$\hat{N}[|S\rangle + |S'\rangle] = \hat{N}[(a + a')|+\rangle + (b + b')|-\rangle] = (a + a')^2|+\rangle + (b + b')^2|-\rangle.$$

But

$$\hat{N}|S\rangle + \hat{N}|S'\rangle = (a^2 + a'^2)|+\rangle + (b^2 + b'^2)|-\rangle$$

which is certainly not equal to $\hat{N}[|S\rangle + |S'\rangle]$. Thus the operator $\hat{N}$ is non-linear.

The importance of linearity lies in the fact that since any state vector $|\psi\rangle$ can be written as a linear combination of a complete set of basis states, $\{|\varphi_n\rangle, n = 1, 2, \dots\}$:

$$|\psi\rangle = \sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle$$

then

$$\hat{A}|\psi\rangle = \hat{A}\sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle = \sum_n \hat{A}|\varphi_n\rangle\langle\varphi_n|\psi\rangle \tag{11.4}$$

so that provided we know what an operator $\hat{A}$ does to each basis state, we can determine what $\hat{A}$ does to any vector belonging to the state space.

Ex 11.5 Consider the spin states $|+\rangle$ and $|-\rangle$, basis states for a spin half system, and suppose an operator $\hat{A}$ has the properties

$$\hat{A}|+\rangle = \tfrac{1}{2}i\hbar|-\rangle$$
$$\hat{A}|-\rangle = -\tfrac{1}{2}i\hbar|+\rangle.$$

Then if a spin half system is in the state

$$|S\rangle = \frac{1}{\sqrt{2}}[|+\rangle + |-\rangle]$$

then

$$\begin{aligned}\hat{A}|S\rangle &= \frac{1}{\sqrt{2}}\hat{A}|+\rangle + \frac{1}{\sqrt{2}}\hat{A}|-\rangle \\ &= \tfrac{1}{2}\hbar\frac{i}{\sqrt{2}}[|-\rangle - |+\rangle] \\ &= -\tfrac{1}{2}i\hbar\frac{1}{\sqrt{2}}[|+\rangle - |-\rangle].\end{aligned}$$

So the state vector $|S\rangle = \frac{1}{\sqrt{2}}[|+\rangle + |-\rangle]$ is mapped into the state vector $-\frac{1}{2}i\hbar\frac{1}{\sqrt{2}}[|+\rangle - |-\rangle]$, which represents a different physical state, and one which, incidentally, is not normalized to unity.

Ex 11.6 Suppose an operator $\hat{B}$ is defined so that

$$\hat{B}|+\rangle = \tfrac{1}{2}\hbar|-\rangle$$
$$\hat{B}|-\rangle = \tfrac{1}{2}\hbar|+\rangle.$$

If we let $\hat{B}$ act on the state $|S\rangle = [|+\rangle + |-\rangle]/\sqrt{2}$ then we find that

$$\hat{B}|S\rangle = \tfrac{1}{2}\hbar|S\rangle \tag{11.5}$$

i.e. in this case, we regain the same state vector $|S\rangle$, though multiplied by a factor $\frac{1}{2}\hbar$. This last equation is an example of an eigenvalue equation: $|S\rangle$ is said to be an eigenvector of the operator $\hat{B}$, and $\frac{1}{2}\hbar$ is its eigenvalue. The concept of an eigenvalue and eigenvector is very important in quantum mechanics, and much more will be said about it later.

In the following Sections we look in more detail some of the more important properties of operators in quantum mechanics.

11.1.3 Properties of Operators

In this discussion, a general perspective is adopted, but the properties will be encountered again and in a more concrete fashion when we look at representations of operators by matrices.

Equality of Operators

If two operators, $\hat{A}$ and $\hat{B}$ say, are such that

$$\hat{A}|\psi\rangle = \hat{B}|\psi\rangle \tag{11.6}$$

for all state vectors $|\psi\rangle$ belonging to the state space of the system then the two operators are said to be equal, written

$$\hat{A} = \hat{B}. \tag{11.7}$$

Linearity also makes it possible to set up a direct way of proving the equality of two operators. Above it was stated that two operators, $\hat{A}$ and $\hat{B}$ say, will be equal if $\hat{A}|\psi\rangle = \hat{B}|\psi\rangle$ for all states $|\psi\rangle$. However, it is sufficient to note that if for all the basis vectors $\{|\varphi_n\rangle, n = 1, 2, \dots\}$

$$\hat{A}|\varphi_n\rangle = \hat{B}|\varphi_n\rangle \tag{11.8}$$

then we immediately have, for any arbitrary state $|\psi\rangle$ that

$$\begin{aligned}
\hat{A}|\psi\rangle &= \hat{A}\sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle \\
&= \sum_n \hat{A}|\varphi_n\rangle\langle\varphi_n|\psi\rangle \\
&= \sum_n \hat{B}|\varphi_n\rangle\langle\varphi_n|\psi\rangle \\
&= \hat{B}\sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle \\
&= \hat{B}|\psi\rangle
\end{aligned} \tag{11.9}$$

so that $\hat{A} = \hat{B}$. Thus, to prove the equality of two operators, it is sufficient to show that the action of the operators on each member of a basis set gives the same result.

The Unit Operator and the Zero Operator

Of all the operators that can be defined, there are two whose properties are particularly simple – the unit operator $\hat{1}$ and the zero operator $\hat{0}$. The unit operator is the operator such that

$$\hat{1}|\psi\rangle = |\psi\rangle \tag{11.10}$$

for all states $|\psi\rangle$, and the zero operator is such that

$$\hat{0}|\psi\rangle = 0 \tag{11.11}$$

for all kets $|\psi\rangle$.

Addition of Operators

The sum of two operators $\hat{A}$ and $\hat{B}$, written $\hat{A} + \hat{B}$ is defined in the obvious way, that is

$$(\hat{A} + \hat{B})|\psi\rangle = \hat{A}|\psi\rangle + \hat{B}|\psi\rangle \tag{11.12}$$

for all vectors $|\psi\rangle$. The sum of two operators is, of course, another operator, $\hat{S}$ say, written $\hat{S} = \hat{A} + \hat{B}$, such that

$$\hat{S}|\psi\rangle = (\hat{A} + \hat{B})|\psi\rangle = \hat{A}|\psi\rangle + \hat{B}|\psi\rangle \tag{11.13}$$

for all states $|\psi\rangle$.

Ex 11.7 Consider the two operators $\hat{A}$ and $\hat{B}$ defined by

$$\begin{aligned} \hat{A}|+\rangle &= \tfrac{1}{2}i\hbar|-\rangle \qquad \hat{B}|+\rangle = \tfrac{1}{2}\hbar|-\rangle \\ \hat{A}|-\rangle &= -\tfrac{1}{2}i\hbar|+\rangle \quad \hat{B}|-\rangle = \tfrac{1}{2}\hbar|+\rangle. \end{aligned} \tag{11.14}$$

Their sum $\hat{S}$ will then be such that

$$\begin{aligned} \hat{S}|+\rangle &= \tfrac{1}{2}(1+i)\hbar|-\rangle \\ \hat{S}|-\rangle &= \tfrac{1}{2}(1-i)\hbar|+\rangle. \end{aligned} \tag{11.15}$$

Multiplication of an Operator by a Complex Number

This too is defined in the obvious way. Thus, if $\hat{A}|\psi\rangle = |\phi\rangle$ then we can define the operator $\lambda\hat{A}$ where λ is a complex number to be such that

$$(\lambda\hat{A})|\psi\rangle = \lambda(\hat{A}|\psi\rangle) = \lambda|\phi\rangle. \tag{11.16}$$

Combining this with the previous definition of the sum of two operators, we can then make say that in general

$$(\lambda\hat{A} + \mu\hat{B})|\psi\rangle = \lambda(\hat{A}|\psi\rangle) + \mu(\hat{B}|\psi\rangle) \tag{11.17}$$

where λ and μ are both complex numbers.

Multiplication of Operators

Given that an operator $\hat{A}$ say, acting on a ket vector $|\psi\rangle$ maps it into another ket vector $|\phi\rangle$, then it is possible to allow a second operator, $\hat{B}$ say, to act on $|\phi\rangle$, producing yet another ket vector $|\xi\rangle$ say. This we can write as

$$\hat{B}\{\hat{A}|\psi\rangle\} = \hat{B}|\phi\rangle = |\xi\rangle. \tag{11.18}$$

This can be written

$$\hat{B}\{\hat{A}|\psi\rangle\} = \hat{B}\hat{A}|\psi\rangle \tag{11.19}$$

i.e. without the braces $\{\dots\}$, with the understanding that the term on the right hand side is to be interpreted as meaning that first $\hat{A}$ acts on the state to its right, and then $\hat{B}$, in the sense specified in Eq. (11.18). The combination $\hat{B}\hat{A}$ is said to be the product of the two operators $\hat{A}$ and $\hat{B}$. The product of two operators is, of course, another operator. Thus we can write $\hat{C} = \hat{B}\hat{A}$ where the operator $\hat{C}$ is such that

$$\hat{C}|\psi\rangle = \hat{B}\hat{A}|\psi\rangle \tag{11.20}$$

for all states $|\psi\rangle$.

Ex 11.8 Consider the products of the two operators defined in Eq. (11.14). First $\hat{C} = \hat{B}\hat{A}$:

$$\begin{aligned} \hat{C}|+\rangle &= \hat{B}\hat{A}|+\rangle \;\; = \hat{B}(\tfrac{1}{2}i\hbar|-\rangle) = \tfrac{1}{4}i\hbar^2|+\rangle \\ \hat{C}|-\rangle &= \hat{B}\hat{A}|-\rangle = \hat{B}(-\tfrac{1}{2}i\hbar|+\rangle) = -\tfrac{1}{4}i\hbar^2|-\rangle, \end{aligned} \tag{11.21}$$

and next $\hat{D} = \hat{A}\hat{B}$:

$$\begin{aligned} \hat{D}|+\rangle &= \hat{A}\hat{B}|+\rangle \;\; = \hat{A}(\tfrac{1}{2}\hbar|-\rangle) = -\tfrac{1}{4}i\hbar^2|+\rangle \\ \hat{D}|-\rangle &= \hat{A}\hat{B}|-\rangle = \hat{A}(-\tfrac{1}{2}\hbar|+\rangle) = \tfrac{1}{4}i\hbar^2|-\rangle. \end{aligned} \tag{11.22}$$

Commutators

Apart from illustrating how to implement the definition of the product of two operators, this last example also shows a further important result, namely that, in general, $\hat{A}\hat{B} \neq \hat{B}\hat{A}$. In other words, the order in which two operators are multiplied is important. The difference between the two, written

$$\hat{A}\hat{B} - \hat{B}\hat{A} = [\hat{A}, \hat{B}] \tag{11.23}$$

is known as the commutator of $\hat{A}$ and $\hat{B}$. If the commutator vanishes, the operators are said to commute. The commutator plays a fundamental role in the physical interpretation of quantum mechanics, being a bridge between the classical description of a physical system and its quantum description, important in describing the consequences of sequences of measurements performed on a quantum system, and, in a related way, whether or not two observable properties of a system can be known simultaneously with precision. The commutator has a number of properties that are straightforward to prove:

$$[\hat{A}, \hat{B}] = -[\hat{B}, \hat{A}] \tag{11.24a}$$

$$[\hat{A}, \alpha\hat{B} + \beta\hat{C}] = \alpha[\hat{A}, \hat{B}] + \beta[\hat{A}, \hat{C}] \tag{11.24b}$$

$$[\hat{A}\hat{B}, \hat{C}] = \hat{A}[\hat{B}, \hat{C}] + [\hat{A}, \hat{C}]\hat{B} \tag{11.24c}$$

$$[\hat{A}, [\hat{B}, \hat{C}]] + [\hat{C}, [\hat{A}, \hat{B}]] + [\hat{B}, [\hat{C}, \hat{A}]] = 0. \tag{11.24d}$$

Projection Operators

An operator $\hat{P}$ that has the property

$$\hat{P}^2 = \hat{P} \tag{11.25}$$

is said to be a *projection operator*. An important example of a projection operator is the operator $\hat{P}_n$ defined, for a given set of orthonormal basis states $\{|\varphi_n\rangle; n = 1, 2, 3 \dots\}$ by

$$\hat{P}_n|\varphi_m\rangle = \delta_{nm}|\varphi_n\rangle. \tag{11.26}$$

That this operator is a projection operator can be readily confirmed:

$$\hat{P}_n^2|\varphi_m\rangle = \hat{P}_n\{\hat{P}_n|\varphi_m\rangle\} = \delta_{nm}\hat{P}_n|\varphi_m\rangle = \delta_{nm}^2|\varphi_m\rangle. \tag{11.27}$$

But since $\delta_{nm}^2 = \delta_{nm}$, (recall that the Kronecker delta δ_{nm} is either unity for $n = m$ or zero for $n \neq m$) we immediately have that

$$\hat{P}_n^2|\varphi_m\rangle = \delta_{nm}|\varphi_m\rangle = \hat{P}_n|\varphi_m\rangle \tag{11.28}$$

from which we conclude that $\hat{P}_n^2 = \hat{P}_n$. The importance of this operator lies in the fact that if we let it act on an arbitrary vector $|\psi\rangle$, then we see that

$$\hat{P}_n|\psi\rangle = \hat{P}_n \sum_m |\varphi_m\rangle\langle\varphi_m|\psi\rangle = \sum_m \hat{P}_n|\varphi_m\rangle\langle\varphi_m|\psi\rangle = \sum_m \delta_{nm}|\varphi_m\rangle\langle\varphi_m|\psi\rangle = |\varphi_n\rangle\langle\varphi_n|\psi\rangle \tag{11.29}$$

i.e. it 'projects' out the component of $|\psi\rangle$ in the direction of the basis state $|\varphi_n\rangle$.

Ex 11.9 Consider a spin half particle in the state $|\psi\rangle = a|+\rangle + b|-\rangle$ where a and b are both real numbers, and define the operators $\hat{P}_\pm$ such that $\hat{P}_-|-\rangle = |-\rangle$, $\hat{P}_-|+\rangle = 0$, $P_+|-\rangle = 0$, and $\hat{P}_+|+\rangle = |+\rangle$. Show that $\hat{P}_\pm$ are projection operators, and evaluate $\hat{P}_\pm|\psi\rangle$.

To show that the $\hat{P}_\pm$ are projection operators, it is necessary to show that

$$\hat{P}_\pm^2 = \hat{P}_\pm.$$

It is sufficient to let $\hat{P}_-^2$ to act on the basis states $|\pm\rangle$:

$$\hat{P}_-^2|-\rangle = \hat{P}_-\{\hat{P}_-|-\rangle\} = \hat{P}_-|-\rangle = |-\rangle$$

and

$$\hat{P}_-^2|+\rangle = \hat{P}_-\{\hat{P}_-|+\rangle\} = 0 = \hat{P}_-|+\rangle.$$

Thus we have shown that $\hat{P}_-^2|\pm\rangle = \hat{P}_-|\pm\rangle$, so that, since the states $|\pm\rangle$ are a pair of basis states, we can conclude that $\hat{P}_-^2 = \hat{P}_-$, so that $\hat{P}_-$ is a projection operator. A similar argument can be followed through for $\hat{P}_+$.

By the properties of $\hat{P}_-$ it follows that

$$\hat{P}_-|\psi\rangle = \hat{P}_-\,[a|+\rangle + b|-\rangle] = b|-\rangle.$$

and similarly

$$\hat{P}_+|\psi\rangle = \hat{P}_+\,[a|+\rangle + b|-\rangle] = a|+\rangle.$$

This result is illustrated in Fig. (11.1) where the projections of $|\psi\rangle$ on to the $|+\rangle$ and $|-\rangle$ basis states are depicted.

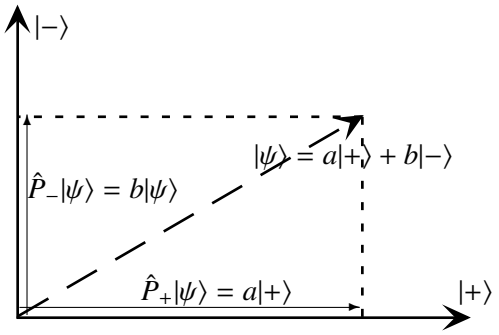


Figure 11.1: An illustration of the action of projection operators on a state $|\psi\rangle = a|+\rangle + b|-\rangle$ of a spin half system where a and b are real.

Functions of Operators

Having defined what is meant by adding and multiplying operators, we can now define the idea of a function of an operator. If we have a function $f(x)$ which we can expand as a power series in x:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots = \sum_{n=0}^{\infty} a_n x^n \tag{11.30}$$

then we define $f(\hat{A})$, a function of the operator $\hat{A}$, to be also given by the same power series, i.e.

$$f(\hat{A}) = a_0 + a_1\hat{A} + a_2\hat{A}^2 + \cdots = \sum_{n=0}^{\infty} a_n\hat{A}^n. \tag{11.31}$$

Questions such as the convergence of such a series (if it is an infinite series) will not be addressed here.

Ex 11.10 The most important example of a function of an operator that we will have to deal with here is the exponential function:

$$f(\hat{A}) = e^{\hat{A}} = 1 + \hat{A} + \frac{1}{2!}\hat{A}^2 + \ldots. \tag{11.32}$$

Many important operators encountered in quantum mechanics, in particular the time evolution operator which specifies how the state of a system evolves in time, is given as an exponential function of an operator.

It is important to note that in general, the usual rules for manipulating exponential functions do not apply for exponentiated operators. In particular, it should be noted that in general

$$e^{\hat{A}}e^{\hat{B}} \neq e^{\hat{A}+\hat{B}} \tag{11.33}$$

unless $\hat{A}$ commutes with $\hat{B}$.

The Inverse of an Operator

If, for some operator $\hat{A}$ there exists another operator $\hat{B}$ with the property that

$$\hat{A}\hat{B} = \hat{B}\hat{A} = \hat{1} \tag{11.34}$$

then $\hat{B}$ is said to be the inverse of $\hat{A}$ and is written

$$\hat{B} = \hat{A}^{-1}. \tag{11.35}$$

Ex 11.11 An important example is the inverse of the operator $\exp(\hat{A})$ defined by the power series in Eq. (11.32) above. The inverse of this operator is readily seen to be just $\exp(-\hat{A})$.

Ex 11.12 Another useful result is the inverse of the product of two operators i.e. $(\hat{A}\hat{B})^{-1}$, that is

$$(\hat{A}\hat{B})^{-1} = \hat{B}^{-1}\hat{A}^{-1}. \tag{11.36}$$

provided, of course, that both $\hat{A}$ and $\hat{B}$ have inverses. This result can be readily shown by noting that, by the definition of the inverse of an operator

$$(\hat{A}\hat{B})^{-1}(\hat{A}\hat{B}) = \hat{1}. \tag{11.37}$$

Multiplying this on the right, first by $\hat{B}^{-1}$, and then by $\hat{A}^{-1}$ then gives the required result.

11.2 Action of Operators on Bra Vectors

Given that an operator maps a ket vector into another ket, as summarized in the defining equation $\hat{A}|\psi\rangle = |\phi\rangle$, we can then take the inner product of $|\phi\rangle$ with any other state vector $|\xi\rangle$ say to yield the complex number $\langle\xi|\phi\rangle$. This we can obviously also write as

$$\langle\xi|\phi\rangle = \langle\xi|(\hat{A}|\psi\rangle). \tag{11.38}$$

This then raises the interesting question, since a bra vector is juxtaposed with an operator in Eq. (11.38), whether we could give a meaning to an operator acting on a bra vector. In other words, can we give a meaning to $\langle\xi|\hat{A}$? Presumably, the outcome of $\hat{A}$ acting on a bra vector is to produce another bra vector, i.e. we can write $\langle\xi|\hat{A} = \langle\chi|$, though as yet we have not specified how to determine what the bra vector $\langle\chi|$ might be. But since operators were originally defined above

in terms of their action on ket vectors, it makes sense to define the action of an operator on a bra in a way that makes use of what we know about the action of an operator on any ket vector. So, we define $\langle\xi|\hat{A}$ such that

$$(\langle\xi|\hat{A})|\psi\rangle = \langle\xi|(\hat{A}|\psi\rangle) \quad \text{for all ket vectors } |\psi\rangle. \tag{11.39}$$

The value of this definition, apart from the fact that it relates the action of operators on bra vectors back to the action of operators of operators on ket vectors, is that $\langle\xi|(\hat{A}|\psi\rangle)$ will always give the same result as $\langle\xi|(\hat{A}|\psi\rangle)$ i.e. it is immaterial whether we let $\hat{A}$ act on the ket vector first, and then take the inner product with $|\xi\rangle$, or to let $\hat{A}$ act on $\langle\xi|$ first, and then take the inner product with $|\psi\rangle$. Thus the brackets are not needed, and we can write:

$$(\langle\xi|\hat{A})|\psi\rangle = \langle\xi|(\hat{A}|\psi\rangle) = \langle\xi|\hat{A}|\psi\rangle. \tag{11.40}$$

This way of defining the action of an operator on a bra vector, Eq. (11.40), is rather back-handed, so it is important to see that it does in fact do the job! To see that the definition actually works, we will look at the particular case of the spin half state space again. Suppose we have an operator $\hat{A}$ defined such that

$$\begin{aligned} \hat{A}|+\rangle &= |+\rangle + i|-\rangle \\ \hat{A}|-\rangle &= i|+\rangle + |-\rangle \end{aligned} \tag{11.41}$$

and we want to determine $\langle+|\hat{A}$ using the above definition. Let $\langle\chi|$ be the bra vector we are after, i.e. $\langle+|\hat{A} = \langle\chi|$. We know that we can always write

$$\langle\chi| = \langle\chi|+\rangle\langle+| + \langle\chi|-\rangle\langle-| \tag{11.42}$$

so the problem becomes evaluating $\langle\chi|\pm\rangle$. It is at this point that we make use of the defining condition above. Thus, we write

$$\langle\chi|\pm\rangle = \{\langle+|\hat{A}\}|\pm\rangle = \langle+|\{\hat{A}|\pm\rangle\}. \tag{11.43}$$

Using Eq. (11.41) this gives

$$\langle\chi|+\rangle = \langle+|\{\hat{A}|+\rangle\} = 1 \quad \text{and} \quad \langle\chi|-\rangle = \langle+|\{\hat{A}|-\rangle\} = i \tag{11.44}$$

and hence

$$\langle\chi| = \langle+| + i\langle-|. \tag{11.45}$$

Consequently, we conclude that

$$\langle+|\hat{A} = \langle+| + i\langle-|. \tag{11.46}$$

If we note that $\hat{A}|+\rangle = |+\rangle + i|-\rangle$ we can see that $\langle+|\hat{A} \neq \langle+| - i\langle-|$. This example illustrates the result that if $\hat{A}|\psi\rangle = |\phi\rangle$ then, in general, $\langle\psi|\hat{A} \neq \langle\phi|$. This example shows that the above 'indirect' definition of the action of an operator on a bra vector in terms of the action of the operator on ket vectors does indeed give us the result of the operator acting on a bra vector.

The general method used in this example can be extended to the general case. So suppose we have a state space for some system spanned by a complete orthonormal set of basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$, and assume that we know the action of an operator $\hat{A}$ on an arbitrary basis state $|\varphi_n\rangle$:

$$\hat{A}|\varphi_n\rangle = \sum_m |\varphi_m\rangle A_{mn} \tag{11.47}$$

where the A_{mn} are complex numbers. This equation is analogous to Eq. (11.41) in the example above. Now suppose we allow $\hat{A}$ to act on an arbitrary bra vector $\langle\xi|$:

$$\langle\xi|\hat{A} = \langle\chi| \tag{11.48}$$

We can express $\langle\chi|$ in terms of the basis states introduced above:

$$\langle\chi| = \sum_n \langle\chi|\varphi_n\rangle\langle\varphi_n|. \tag{11.49}$$

Thus, the problem reduces to showing that we can indeed calculate the coefficients $\langle\chi|\varphi_n\rangle$. These coefficients are given by

$$\langle\chi|\varphi_n\rangle = \{\langle\xi|\hat{A}\}|\varphi_n\rangle = \langle\xi|\{\hat{A}|\varphi_n\rangle\} \tag{11.50}$$

where we have used the defining condition Eq. (11.39) to allow $\hat{A}$ to act on the basis state $|\varphi_n\rangle$. Using Eq. (11.47) we can write

$$\begin{aligned}\langle\chi|\varphi_n\rangle &= \langle\xi|\{\hat{A}|\varphi_n\rangle\} \\ &= \langle\xi|\Big[\sum_m |\varphi_m\rangle A_{mn}\Big] \\ &= \sum_m A_{mn}\langle\xi|\varphi_m\rangle.\end{aligned} \tag{11.51}$$

If we substitute this into the expression Eq. (11.49) we find that

$$\langle\chi| = \langle\xi|\hat{A} = \sum_n \Big[\sum_m \langle\xi|\varphi_m\rangle A_{mn}\Big]\langle\varphi_n|. \tag{11.52}$$

The quantity within the brackets is a complex number which we can always evaluate since we know the A_{mn} and can evaluate the inner product $\langle\xi|\varphi_m\rangle$. Thus, by use of the defining condition Eq. (11.39), we are able to calculate the result of an operator acting on a bra vector. Of particular interest is the case in which $\langle\xi| = \langle\varphi_k|$ for which

$$\langle\varphi_k|\hat{A} = \sum_n \Big[\sum_m \langle\varphi_k|\varphi_m\rangle A_{mn}\Big]\langle\varphi_n|. \tag{11.53}$$

Since the basis states are orthonormal, i.e. $\langle\varphi_k|\varphi_m\rangle = \delta_{km}$, then

$$\begin{aligned}\langle\varphi_k|\hat{A} &= \sum_n \Big[\sum_m \delta_{km} A_{mn}\Big]\langle\varphi_n| \\ &= \sum_n A_{kn}\langle\varphi_n|.\end{aligned} \tag{11.54}$$

It is useful to compare this result with Eq. (11.47):

$$\begin{aligned}\hat{A}|\varphi_n\rangle &= \sum_m |\varphi_m\rangle A_{mn} \\ \langle\varphi_n|\hat{A} &= \sum_m A_{nm}\langle\varphi_m|.\end{aligned} \tag{11.55}$$

Either of these expressions lead to the result

$$A_{mn} = \langle\varphi_m|\hat{A}|\varphi_n\rangle. \tag{11.56}$$

For reasons which will become clearer later, the quantities A_{mn} are known as the matrix elements of the operator $\hat{A}$ with respect to the set of basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$.

Ex 11.13 With respect to a pair of orthonormal vectors $|\varphi_1\rangle$ and $|\varphi_2\rangle$ that span the Hilbert space $\mathcal{H}$ of a certain system, the operator $\hat{A}$ is defined by its action on these base states as follows:

$$\hat{A}|\varphi_1\rangle = 3|\varphi_1\rangle - 4i|\varphi_2\rangle$$
$$\hat{A}|\varphi_2\rangle = -4i|\varphi_1\rangle - 3|\varphi_2\rangle.$$

Evaluate $\langle\varphi_1|\hat{A}$ and $\langle\varphi_2|\hat{A}$.

We proceed by considering the product $\{\langle\varphi_1|\hat{A}\}|\psi\rangle$ where $|\psi\rangle$ is an arbitrary state vector which we can expand with respect to the pair of basis states $\{|\varphi_1\rangle, |\varphi_2\rangle\}$ as $|\psi\rangle = a|\varphi_1\rangle + b|\varphi_2\rangle$.

Using the defining condition Eq. (11.39) we have

$$\{\langle\varphi_1|\hat{A}\}|\psi\rangle = \langle\varphi_1|\{\hat{A}|\psi\rangle\} = \langle\varphi_1|\{a\hat{A}|\varphi_1\rangle + b\hat{A}|\varphi_2\rangle\}$$

Using the properties of $\hat{A}$ as given, and using the fact that $a = \langle\varphi_1|\psi\rangle$ and $b = \langle\varphi_2|\psi\rangle$ we get

$$\{\langle\varphi_1|\hat{A}\}|\psi\rangle = 3a - 4ib = 3\langle\varphi_1|\psi\rangle - 4i\langle\varphi_2|\psi\rangle.$$

Extracting the common factor $|\psi\rangle$ yields

$$\{\langle\varphi_1|\hat{A}\}|\psi\rangle = \{3\langle\varphi_1| - 4i\langle\varphi_2|\}|\psi\rangle$$

from which we conclude, since $|\psi\rangle$ is arbitrary, that

$$\langle\varphi_1|\hat{A} = 3\langle\varphi_1| - 4i\langle\varphi_2|.$$

In a similar way, we find that

$$\langle\varphi_2|\hat{A} = -4i\langle\varphi_1| - 3\langle\varphi_2|.$$

Ex 11.14 A useful an important physical system with which to illustrate some of these ideas is that of the electromagnetic field inside a cavity designed to support a single mode of the field, see p 132. In this case, the basis states of the electromagnetic field are the so-called number states $\{|n\rangle, n = 0, 1, 2, \ldots\}$ where the state $|n\rangle$ is the state of the field in which there are n photons present. We can now introduce an operator $\hat{a}$ defined such that

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle, \qquad \hat{a}|0\rangle = 0. \tag{11.57}$$

Understandably, this operator is known as the photon annihilation operator as it transforms a state with n photons into one in which there are $n - 1$ photons. The prefactor $\sqrt{n}$ is there for later purposes i.e. it is possible to define an operator $\hat{a}'$ such that $\hat{a}'|n\rangle = |n-1\rangle$, but this operator turns out not to arise in as natural a way as $\hat{a}$, and is not as useful in practice. We can ask the question: what is $\langle n|\hat{a}$?

To address this, we must make use of the manner in which we define the action of an operator on a bra, that is, it must be such that

$$\{\langle n|\hat{a}\}|\psi\rangle = \langle n|\{\hat{a}|\psi\rangle\}$$

holds true for all states $|\psi\rangle$.

If we expand $|\psi\rangle$ in terms of the basis states $\{|n\rangle, n = 0, 1, 2, \ldots\}$ we have

$$|\psi\rangle = \sum_{m=0}^{\infty} |m\rangle\langle m|\psi\rangle$$

where we have used a summation index (m) to distinguish it from the label n used on the bra vector $\langle n|$. From this we have

$$\begin{aligned}
\{\langle n|\hat{a}\}|\psi\rangle &= \{\langle n|\hat{a}\} \sum_{m=0}^{\infty} |m\rangle\langle m|\psi\rangle \\
&= \sum_{m=0}^{\infty} \{\langle n|\hat{a}\}|m\rangle\langle m|\psi\rangle \\
&= \sum_{m=0}^{\infty} \langle n|\{\hat{a}|m\rangle\langle m|\psi\rangle\} \\
&= \sum_{m=1}^{\infty} \sqrt{m}\langle n|m-1\rangle\langle m|\psi\rangle\}
\end{aligned}$$

where the sum now begins at $m = 1$ as the $m = 0$ term will vanish. This sum can be written as

$$\begin{aligned}
\{\langle n|\hat{a}\}|\psi\rangle &= \sum_{m=0}^{\infty} \sqrt{m+1}\langle n|m\rangle\langle m+1|\psi\rangle\} \\
&= \sum_{m=0}^{\infty} \sqrt{m+1}\delta_{nm}\langle m+1|\psi\rangle \\
&= \left\{\sqrt{n+1}\langle n+1|\right\}|\psi\rangle
\end{aligned}$$

By comparing the left and right hand sides of this last equation, and recognizing that $|\psi\rangle$ is arbitrary, we conclude that

$$\langle n|\hat{a} = \sqrt{n+1}\langle n+1| \tag{11.58}$$

A further consequence of the above definition of the action of operators on bra vectors, which is actually implicit in the derivation of the result Eq. (11.52) is the fact that an operator $\hat{A}$ that is linear with respect to ket vectors, is also linear with respect to bra vectors i.e.

$$[\lambda\langle\psi_1| + \mu\langle\psi_2|]\hat{A} = \lambda\langle\psi_1|\hat{A} + \mu\langle\psi_2|\hat{A} \tag{11.59}$$

which further emphasizes the symmetry between the action of operators on bras and kets.

Much of what has been presented above is recast in terms of matrices and column and row vectors in a later Section.

11.3 The Hermitean Adjoint of an Operator

We have seen above that if $\hat{A}|\psi\rangle = |\phi\rangle$ then, in general, $\langle\psi|\hat{A} \neq \langle\phi|$. This then suggests the possibility of introducing an operator related to $\hat{A}$, which we will write $\hat{A}^\dagger$ which is such that

$$\text{if} \quad \hat{A}|\psi\rangle = |\phi\rangle \quad \text{then} \quad \langle\psi|\hat{A}^\dagger = \langle\phi|. \tag{11.60}$$

The operator $\hat{A}^\dagger$ so defined is known as the *Hermitean adjoint* of $\hat{A}$. There are issues concerning the definition of the Hermitean adjoint that require careful consideration if the state space is of infinite dimension. We will not be concerning ourselves with these matters here. Thus we see we have introduced a new operator which has been defined in terms of its actions on bra vectors. In keeping with our point of view that operators should be defined in terms of their action on ket vectors, it should be the case that this above definition should unambiguously tell us what the action of $\hat{A}^\dagger$ will be on any ket vector. In other words, the task at hand is to show that we can evaluate $\hat{A}^\dagger|\psi\rangle$ for any arbitrary ket vector $|\psi\rangle$. In order to do this, we need a useful property of the Hermitean adjoint which can be readily derived from the above definition. Thus, consider $\langle\xi|\hat{A}|\psi\rangle$, which we recognize is simply a complex number given by

$$\langle\xi|\hat{A}|\psi\rangle = \langle\xi|(\hat{A}|\psi\rangle) = \langle\xi|\phi\rangle \tag{11.61}$$

where $\hat{A}|\psi\rangle = |\phi\rangle$. Thus, if we take the complex conjugate. we have

$$\langle\xi|\hat{A}|\psi\rangle^* = \langle\xi|\phi\rangle^* = \langle\phi|\xi\rangle. \tag{11.62}$$

But, since $\hat{A}|\psi\rangle = |\phi\rangle$ then $\langle\psi|\hat{A}^\dagger = \langle\phi|$ so we have

$$\langle\xi|\hat{A}|\psi\rangle^* = \langle\phi|\xi\rangle = (\langle\psi|\hat{A}^\dagger)|\xi\rangle = \langle\psi|\hat{A}^\dagger|\xi\rangle \tag{11.63}$$

where in the last step the brackets have been dropped since it does not matter whether an operator acts on the ket or the bra vector. Thus, taking the complex conjugate of $\langle\xi|\hat{A}|\psi\rangle$ amounts to reversing the order of the factors, and replacing the operator by its Hermitean conjugate. Using this it is then possible to determine the action of $\hat{A}^\dagger$ on a ket vector. The situation here is analogous to that which was encountered in Section 11.2. But before considering the general case, we will look at an example.

Ex 11.15 Suppose an operator $\hat{B}$ is defined, for the two orthonormal states $|\varphi_1\rangle$ and $|\varphi_2\rangle$, by

$$\hat{B}|\varphi_1\rangle = 2|\varphi_2\rangle \quad \text{and} \quad \hat{B}|\varphi_2\rangle = i|\varphi_1\rangle.$$

What are $\hat{B}^\dagger|\varphi_1\rangle$ and $\hat{B}^\dagger|\varphi_2\rangle$?

First consider $\hat{B}^\dagger|\varphi_1\rangle$. We begin by looking at $\langle\chi|\hat{B}^\dagger|\varphi_1\rangle$ where $|\chi\rangle = C_1|\varphi_1\rangle + C_2|\varphi_2\rangle$ is an arbitrary state vector. We then have, by the property proven above:

$$\begin{aligned}
\langle\chi|\hat{B}^\dagger|\varphi_1\rangle^* =& \langle\varphi_1|\hat{B}|\chi\rangle \\
=& \langle\varphi_1|\hat{B}\left[C_1|\varphi_1\rangle + C_2|\varphi_2\rangle\right] \\
=& \langle\varphi_1|\left[C_1\hat{B}|\varphi_1\rangle + C_2\hat{B}|\varphi_2\rangle\right] \\
=& \langle\varphi_1|\left[2C_1|\varphi_2\rangle + iC_2|1\rangle\right] \\
=& iC_2 \\
=& i\langle\varphi_2|\chi\rangle.
\end{aligned}$$

Thus we have shown that

$$\langle\chi|\hat{B}^\dagger|\varphi_1\rangle^* = i\langle\varphi_2|\chi\rangle$$

which becomes, on taking the complex conjugate

$$\langle\chi|\hat{B}^\dagger|\varphi_1\rangle = -i\langle\varphi_2|\chi\rangle^* = -i\langle\chi|\varphi_2\rangle.$$

Since $|\chi\rangle$ is arbitrary, we must conclude that

$$\hat{B}^\dagger|\varphi_1\rangle = -i|\varphi_2\rangle.$$

More generally, suppose we are dealing with a state space spanned by a complete orthonormal set of basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$, and suppose we know that action of an operator $\hat{A}$ on each of the basis states:

$$\hat{A}|\varphi_n\rangle = \sum_m |\varphi_m\rangle A_{mn} \tag{11.64}$$

and we want to determine $\hat{A}^\dagger|\psi\rangle$ where $|\psi\rangle$ is an arbitrary ket vector. If we let $\hat{A}^\dagger|\psi\rangle = |\zeta\rangle$, then we can, as usual, make the expansion:

$$|\zeta\rangle = \sum_n |\varphi_n\rangle\langle\varphi_n|\zeta\rangle. \tag{11.65}$$

The coefficients $\langle\varphi_n|\zeta\rangle$ can then be written:

$$\begin{aligned}\langle\varphi_n|\zeta\rangle &= \langle\varphi_n|\hat{A}^\dagger|\psi\rangle \\ &= \langle\psi|\hat{A}|\varphi_n\rangle^* \\ &= \left(\langle\psi|\Big[\sum_m |\varphi_m\rangle A_{mn}\Big]\right)^* \\ &= \sum_m \langle\psi|\varphi_m\rangle^* A^*_{mn}\end{aligned} \tag{11.66}$$

so that

$$\begin{aligned}|\zeta\rangle &= \hat{A}^\dagger|\psi\rangle \\ &= \sum_n \sum_m |\varphi_n\rangle\langle\psi|\varphi_m\rangle^* A^*_{mn} \\ &= \sum_n |n\rangle\Big[\sum_m A^*_{mn}\langle\varphi_m|\psi\rangle\Big].\end{aligned} \tag{11.67}$$

The quantity within the brackets is a complex number which we can always evaluate since we know the A_{mn} and can evaluate the inner product $\langle\varphi_m|\psi\rangle$. Thus, we have shown that the action of the Hermitean adjoint on a ket vector can be readily calculated. Of particular interest is the case in which $|\psi\rangle = |\varphi_k\rangle$ so that

$$\hat{A}^\dagger|\varphi_k\rangle = \sum_n |\varphi_n\rangle\Big[\sum_m A^*_{mn}\langle\varphi_m|\varphi_k\rangle\Big]. \tag{11.68}$$

Using the orthonormality of the basis states, i.e. $\langle\varphi_m|\varphi_k\rangle = \delta_{mk}$ we have

$$\begin{aligned}\hat{A}^\dagger|\varphi_k\rangle &= \sum_n |n\rangle\Big[\sum_m A^*_{mn}\delta_{mk}\Big] \\ &= \sum_n |\varphi_n\rangle A^*_{kn}\end{aligned} \tag{11.69}$$

It is useful to compare this with Eq. (11.47):

$$\begin{aligned}\hat{A}|\varphi_n\rangle &= \sum_m |\varphi_m\rangle A_{mn} \\ \hat{A}^\dagger|\varphi_n\rangle &= \sum_m |\varphi_m\rangle A^*_{nm}\end{aligned} \tag{11.70}$$

From these two results, we see that

$$\langle\varphi_m|\hat{A}|\varphi_n\rangle^* = A^*_{mn} = \langle\varphi_n|\hat{A}^\dagger|\varphi_m\rangle. \tag{11.71}$$

Ex 11.16 We can illustrate the results obtained here using the photon annihilation operator defined in Eq. (11.57). There we showed that the operator $\hat{a}$ defined such that $\hat{a}|n\rangle = \sqrt{n}|n\rangle$ where $|n\rangle$ was the state of a collection of identical photons in which there were n photons present. The question to be addressed here is then: what is $\hat{a}^\dagger|n\rangle$?

This can be determined by considering $\langle\chi|\hat{a}^\dagger|n\rangle$ where $|\chi\rangle$ is an arbitrary state. We then have

$$\langle\chi|\hat{a}^\dagger|n\rangle^* = \langle n|\hat{a}|\chi\rangle = \{\langle n|\hat{a}\}|\chi\rangle.$$

If we now note, from Eq. (11.58), that $\langle n|\hat{a} = \sqrt{n+1}\langle n+1|$, we have

$$\langle\chi|\hat{a}^\dagger|n\rangle^* = \sqrt{n+1}\langle n+1|\chi\rangle$$

and hence

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle. \tag{11.72}$$

As this operator increases the photon number by unity, it is known as a *creation* operator.

Ex 11.17 Show that $(\hat{A}^\dagger)^\dagger = \hat{A}$.

This is shown to be true by forming the quantity $\langle\phi|(\hat{A}^\dagger)^\dagger|\psi\rangle$ where $|\phi\rangle$ and $|\psi\rangle$ are both arbitrary. We then have

$$\langle\phi|(\hat{A}^\dagger)^\dagger|\psi\rangle = \langle\psi|\hat{A}^\dagger|\phi\rangle^* = (\langle\phi|\hat{A}|\psi\rangle^*)^* = \langle\phi|\hat{A}|\psi\rangle.$$

since, for any complex number z we have $(z^*)^* = z$. The required result then follows by noting that $|\phi\rangle$ and $|\psi\rangle$ are both arbitrary.

Ex 11.18 An important and useful result is that $(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger$. To prove this, we once again form the quantity $\langle\phi|(\hat{A}\hat{B})^\dagger|\psi\rangle$ where $|\phi\rangle$ and $|\psi\rangle$ are both arbitrary. We then have

$$\langle\phi|(\hat{A}\hat{B})^\dagger|\psi\rangle^* = \langle\psi|\hat{A}\hat{B}|\phi\rangle = \langle\alpha|\beta\rangle$$

where $\langle\psi|\hat{A} = \langle\alpha|$ and $\hat{B}|\phi\rangle = |\beta\rangle$. Taking the complex conjugate of both sides then gives

$$\langle\phi|(\hat{A}\hat{B})^\dagger|\psi\rangle = \langle\alpha|\beta\rangle^* = \langle\beta|\alpha\rangle = \langle\phi|\hat{B}^\dagger\hat{A}^\dagger|\psi\rangle$$

The required result then follows as $|\phi\rangle$ and $|\psi\rangle$ are both arbitrary.

Much of this discussion on Hermitean operators is recast in terms of matrices and column and row vectors in a later Section.

11.3.1 Hermitean and Unitary Operators

These are two special kinds of operators that play very important roles in the physical interpretation of quantum mechanics.

Hermitean Operators If an operator $\hat{A}$ has the property that

$$\hat{A} = \hat{A}^\dagger \tag{11.73}$$

then the operator is said to be Hermitean. If $\hat{A}$ is Hermitean, then

$$\langle\psi|\hat{A}|\phi\rangle^* = \langle\phi|\hat{A}|\psi\rangle \tag{11.74}$$

and, in particular, for states belonging to a complete set of orthonormal basis states $\{|\varphi_n\rangle; n = 1, 2, 3, \ldots\}$ we have

$$A_{mn} = \langle\varphi_m|\hat{A}|\varphi_n\rangle = \langle\varphi_n|\hat{A}^\dagger|\varphi_m\rangle^* = \langle\varphi_n|\hat{A}|\varphi_m\rangle^* = A^*_{nm}. \tag{11.75}$$

Hermitean operators have a number of important mathematical properties that are discussed in detail in Section 11.4.2. It is because of these properties that Hermitean operators place a central role in quantum mechanics in that the observable properties of a physical system such as position, momentum, spin, energy and so on are represented by Hermitean operators. The physical significance of Hermitean operators will be described in the following Chapter.

Unitary Operators If the operator $\hat{U}$ is such that

$$\hat{U}^\dagger = \hat{U}^{-1} \tag{11.76}$$

then the operator is said to be unitary. Unitary operators have the important property that they map normalized states into normalized states. Thus, for instance, suppose the state $|\psi\rangle$ is normalized to unity, $\langle\psi|\psi\rangle = 1$. We then find that the state $|\phi\rangle = \hat{U}|\psi\rangle$ is also normalized to unity:

$$\langle\phi|\phi\rangle = \langle\psi|\hat{U}^\dagger\hat{U}|\psi\rangle = \langle\psi|\hat{1}|\psi\rangle = \langle\psi|\psi\rangle = 1. \tag{11.77}$$

It is because of this last property that unitary operators play a central role in quantum mechanics in that such operators represent performing actions on a system, such as displacing the system in space or time. The time evolution operator with which the evolution in time of the state of a quantum system can be determined is an important example of a unitary operator. If some action is performed on a quantum system, then the probability interpretation of quantum mechanics makes it physically reasonable to expect that the state of the system after the action is performed should be normalized if the state was initially normalized. If this were not the case, it would mean that this physical action in some way results in the system losing or gaining probability.

Ex 11.19 Consider a negatively charged ozone molecule O_3^-. The oxygen atoms, labelled A, B, and C are arranged in an equilateral triangle. The electron can be found on any one of these atoms, the corresponding state vectors being $|A\rangle$, $|B\rangle$, and $|C\rangle$. An operator $\hat{E}$ can be defined with the properties that

$$\hat{E}|A\rangle = |B\rangle,\ \hat{E}|B\rangle = |C\rangle,\ \hat{E}|C\rangle = |A\rangle$$

i.e. it represents the physical process in which the electron 'jumps' from one atom to its neighbour, as in $|A\rangle \to |B\rangle$, $|B\rangle \to |C\rangle$ and $|C\rangle \to |A\rangle$. Show that the operator $\hat{E}$ is unitary.

To show that $\hat{E}$ is unitary requires showing that $\hat{E}^\dagger\hat{E} = \hat{1}$ which amounts to showing that $\langle\psi|\hat{E}^\dagger\hat{E}|\phi\rangle = \langle\psi|\phi\rangle$ for all states $|\psi\rangle$ and $|\phi\rangle$. As the states $\{|A\rangle, |B\rangle, |C\rangle\}$ form a complete orthonormal set of basis states, we can write

$$|\psi\rangle = a|A\rangle + b|B\rangle + c|C\rangle$$

so that

$$\hat{E}|\psi\rangle = a|B\rangle + b|C\rangle + c|A\rangle.$$

Likewise, if we write

$$|\phi\rangle = \alpha|A\rangle + \beta|B\rangle + \gamma|C\rangle$$

then

$$\hat{E}|\phi\rangle = \alpha|B\rangle + \beta|C\rangle + \gamma|A\rangle$$

and hence

$$\langle\phi|\hat{E}^\dagger = \alpha^*\langle B| + \beta^*\langle C| + \gamma^*\langle A|$$

which gives

$$\langle\psi|\hat{E}^\dagger\hat{E}|\phi\rangle = \alpha^* a + \beta^* b + \gamma^* c = \langle\psi|\phi\rangle.$$

Thus $\hat{E}$ is unitary.

An Analogue with Complex Numbers

It can be a useful mnemonic to note the following analogues between Hermitean and unitary operators and real and unimodular complex numbers respectively. Thus we find that Hermitean operators are the operator analogues of real numbers:

$$\hat{A} = \hat{A}^\dagger \quad \leftrightarrow \quad z = z^* \tag{11.78}$$

while unitary operators are the analogue of complex numbers of unit modulus, i.e. of the form $\exp(i\theta)$ where θ is real:

$$\hat{U}^\dagger = \hat{U}^{-1} \quad \leftrightarrow \quad (e^{i\theta})^* = e^{-i\theta} = (e^{i\theta})^{-1}. \tag{11.79}$$

The analogue goes further that. It turns out that a unitary operator $\hat{U}$ can be written in the form

$$\hat{U} = e^{-i\hat{A}}$$

where $\hat{A}$ is Hermitean. Results of this form will be seen to arise in the cases of unitary operators representing time translation, space displacement, and rotation.

11.4 Eigenvalues and Eigenvectors

It can happen that, for some operator $\hat{A}$, there exists a state vector $|\phi\rangle$ that has the property

$$\hat{A}|\phi\rangle = a_\phi|\phi\rangle \tag{11.80}$$

where a_ϕ is, in general, a complex number. We have seen an example of such a situation in Eq. (11.5). If a situation such as that presented in Eq. (11.80) occurs, then the state $|\phi\rangle$ is said to be an eigenstate or eigenket of the operator $\hat{A}$ with a_ϕ the associated eigenvalue. Often the notation

$$\hat{A}|a\rangle = a|a\rangle \tag{11.81}$$

is used in which the eigenvector is labelled by its associated eigenvalue. This notation will be used almost exclusively here. Determining the eigenvalues and eigenvectors of a given operator $\hat{A}$, occasionally referred to as solving the eigenvalue problem for the operator, amounts to finding solutions to the eigenvalue equation Eq. (11.80). If the vector space is of finite dimension, then this can be done by matrix methods, while if the state space is of infinite dimension, then solving the eigenvalue problem can require solving a differential equation. Examples of both possibilities will be looked at later. An operator $\hat{A}$ may have

1. no eigenstates (for state spaces of infinite dimension);
2. real or complex eigenvalues;

3. a discrete collection of eigenvalues $a_1, a_2, \ldots$ and associated eigenvectors $|a_1\rangle, |a_2\rangle, \ldots$;
4. a continuous range of eigenvalues and associated eigenvectors;
5. a combination of both discrete and continuous eigenvalues.

The collection of all the eigenvalues of an operator is called the *eigenvalue spectrum* of the operator. Note also that more than one eigenvector can have the same eigenvalue. Such an eigenvalue is said to be *degenerate*.

Ex 11.20 An interesting example of an operator with complex eigenvalues is the annihilation operator $\hat{a}$ introduced in Eq. (11.57). This operator maps the state of a system of identical photons in which there is exactly n photons present, $|n\rangle$, into the state $|n-1\rangle$: $\hat{a}|n\rangle = \sqrt{n}|n-1\rangle$.

The eigenstates of this operator can be found by looking for the solutions to the eigenvalue equation $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$ where α and $|\alpha\rangle$ are the eigenvalue and associated eigenstate to be determined. Expanding $|\alpha\rangle$ in terms of the number state basis $\{|n\rangle; n = 0, 1, 2, 3, \ldots\}$ gives

$$|\alpha\rangle = \sum_{n=0}^{\infty} c_n |n\rangle$$

where, by using the orthonormality condition $\langle n|m\rangle = \delta_{nm}$, we have $c_n = \langle n|\alpha\rangle$. Thus it follows that

$$\hat{a}|\alpha\rangle = \sum_{n=1}^{\infty} c_n \sqrt{n}|n-1\rangle = \sum_{n=0}^{\infty} c_{n+1}\sqrt{n+1}|n\rangle = \sum_{n=0}^{\infty} \alpha c_n |n\rangle$$

where the last expression is just $\alpha|\alpha\rangle$.

Equating coefficients of $|n\rangle$ gives

$$c_{n+1} = \frac{\alpha c_n}{\sqrt{n+1}}$$

a recurrence relation, from which we can build up each coefficient from the one before, i.e. assuming $c_0 \neq 0$ we have

$$\begin{aligned}
c_1 &= \frac{\alpha}{\sqrt{1}} c_0 \\
c_2 &= \frac{\alpha}{\sqrt{2}} c_1 = \frac{\alpha^2}{\sqrt{2.1}} c_0 \\
c_3 &= \frac{\alpha}{\sqrt{3}} c_2 = \frac{\alpha^3}{\sqrt{3!}} c_0 \\
&\vdots \\
c_n &= \frac{\alpha^n}{\sqrt{n!}} c_0
\end{aligned}$$

and hence

$$|\alpha\rangle = c_0 \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle.$$

Requiring this state to be normalized to unity gives

$$\langle\alpha|\alpha\rangle = \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\langle\alpha|n\rangle c_0.$$

But $c_n = \langle n|\alpha\rangle$, so $\langle\alpha|n\rangle = c_n^*$ and we have

$$\langle\alpha|n\rangle = \frac{\alpha^{*n}}{\sqrt{n!}}c_0^*$$

and hence

$$\langle\alpha|\alpha\rangle = 1 = |c_0|^2 \sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!} = |c_0|^2 e^{|\alpha|^2}.$$

Thus

$$c_0 = e^{-|\alpha|^2/2}$$

where we have set an arbitrary phase factor to unity. Thus we end up with

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}|n\rangle. \tag{11.82}$$

This then is the required eigenstate of $\hat{a}$. It is known as a *coherent state*, and plays a very important role, amongst other things, as the 'most classical' state possible for a harmonic oscillator, which includes the electromagnetic field.

It should be noted in this derivation that no restriction was needed on the value of α. In other words all the states $|\alpha\rangle$ for any value of the complex number α will be an eigenstate of $\hat{a}$. It can also be shown that the states $|\alpha\rangle$ are not orthogonal for different values of α, i.e. $\langle\alpha'|\alpha\rangle \neq 0$ if $\alpha' \neq \alpha$, a fact to be contrasted with what is seen later when the eigenstates of Hermitean operators are considered.

Attempting a similar calculation to the above to try to determine what are the eigenstates of the creation operator $\hat{a}^\dagger$ for which $\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n\rangle$ (see p 154) quickly shows that this operator has, in fact, no eigenstates.

Ex 11.21 If $|a\rangle$ is an eigenstate of $\hat{A}$ with eigenvalue a, then a function $f(\hat{A})$ of $\hat{A}$ (where the function can be expanded as a power series) will also have $|a\rangle$ as an eigenstate with eigenvalue $f(a)$. This can be easily shown by first noting that

$$\hat{A}^n|a\rangle = \hat{A}^{(n-1)}\hat{A}|a\rangle = \hat{A}^{(n-1)}a|a\rangle = a\hat{A}^{(n-1)}|a\rangle.$$

Repeating this a further $n-1$ times then yields

$$\hat{A}^n|a\rangle = a^n|a\rangle.$$

If we then apply this to

$$f(\hat{A})|a\rangle$$

where $f(\hat{A})$ has the power series expansion

$$f(\hat{A}) = c_0 + c_1\hat{A} + c_2\hat{A}^2 + \ldots$$

then

$$\begin{aligned} f(\hat{A})|a\rangle &= (c_0 + c_1\hat{A} + c_2\hat{A}^2 + \ldots)|a\rangle \\ &= c_0|a\rangle + c_1 a|a\rangle + c_2 a^2|a\rangle + \ldots \\ &= (c_0 + c_1 a + c_2 a^2 + \ldots)|a\rangle \\ &= f(a)|a\rangle. \end{aligned} \tag{11.83}$$

This turns out to be a very valuable result as we will often encounter functions of operators when we deal, in particular, with the time evolution operator. The time evolution operator is expressed as an exponential function of another operator (the Hamiltonian) whose eigenvalues and eigenvectors are central to the basic formalism of quantum mechanics.

11.4.1 Eigenkets and Eigenbras

The notion of an eigenstate has been introduced above with respect to ket vectors, in which case the eigenstates could be referred to as *eigenkets*. The same idea of course can be applied to bra vectors, i.e. if for some operator $\hat{A}$ there exists a bra vector $\langle\phi|$ such that

$$\langle\phi|\hat{A} = a_\phi\langle\phi| \tag{11.84}$$

then $\langle\phi|$ is said to be an eigenbra of $\hat{A}$ and a_ϕ the associated eigenvalue. The eigenvalues in this case can have the same array of possible properties as listed above for eigenkets.

An important and useful result is the following. If $\hat{A}$ has an eigenket $|\phi\rangle$ with associated eigenvalue a_ϕ, then, for any arbitrary state $|\psi\rangle$:

$$\langle\psi|\hat{A}|\phi\rangle = a_\phi\langle\psi|\phi\rangle \tag{11.85}$$

so that, on taking the complex conjugate we get

$$\langle\phi|\hat{A}^\dagger|\psi\rangle = a_\phi^*\langle\phi|\psi\rangle \tag{11.86}$$

from which follows, since $|\psi\rangle$ is arbitrary

$$\langle\phi|\hat{A}^\dagger = a_\phi^*\langle\phi| \tag{11.87}$$

a perhaps not surprising result.

It is interesting to note that an operator may have eigenkets, but have no eigenbras. Thus, we have seen that the annihilation operator $\hat{a}$ has a continuously infinite number of eigenkets, but it has no eigenbras, for the same reason that $\hat{a}^\dagger$ has no eigenkets.

11.4.2 Eigenstates and Eigenvalues of Hermitean Operators

If an operator is Hermitean then its eigenstates and eigenvalues are found to possess a number of mathematical properties that are of substantial significance in quantum mechanics. So, if we suppose that an operator $\hat{A}$ is Hermitean i.e. $\hat{A} = \hat{A}^\dagger$ then the following three properties hold true.

1. The eigenvalues of $\hat{A}$ are all real.

The proof is as follows. Since

$$\hat{A}|a\rangle = a|a\rangle$$

then

$$\langle a|\hat{A}|a\rangle = a\langle a|a\rangle.$$

Taking the complex conjugate then gives

$$\langle a|\hat{A}|a\rangle^* = a^*\langle a|a\rangle.$$

Now, using the facts that $\langle\phi|\hat{A}|\psi\rangle = \langle\psi|\hat{A}|\phi\rangle^*$ (Eq. (11.63)), and that $\langle a|a\rangle$ is real, we have

$$\langle a|\hat{A}^\dagger|a\rangle = a^*\langle a|a\rangle.$$

Since $\hat{A} = \hat{A}^\dagger$ this then gives

$$\langle a|\hat{A}|a\rangle = a^*\langle a|a\rangle = a\langle a|a\rangle$$

and hence

$$(a^* - a)\langle a|a\rangle = 0.$$

And so, finally, since $\langle a|a\rangle \neq 0$,

$$a^* = a.$$

This property is of central importance in the physical interpretation of quantum mechanics in that all physical observable properties of a system are represented by Hermitean operators, with the eigenvalues of the operators representing all the possible values that the physical property can be observed to have.

2. Eigenvectors belonging to different eigenvalues are orthogonal, i.e. if $\hat{A}|a\rangle = a|a\rangle$ and $\hat{A}|a'\rangle = a'|a'\rangle$ where $a \neq a'$, then $\langle a|a'\rangle = 0$.

The proof is as follows. Since

$$\hat{A}|a\rangle = a|a\rangle$$

then

$$\langle a'|\hat{A}|a\rangle = a\langle a'|a\rangle.$$

But

$$\hat{A}|a'\rangle = a'|a'\rangle$$

so that

$$\langle a|\hat{A}|a'\rangle = a'\langle a|a'\rangle$$

and hence on taking the complex conjugate

$$\langle a'|\hat{A}^\dagger|a\rangle = a'^*\langle a'|a\rangle = a'\langle a'|a\rangle$$

where we have used the fact that the eigenvalues of $\hat{A}$ are real, and hence $a' = a'^*$. Overall then,

$$\langle a'|\hat{A}|a\rangle = a'\langle a'|a\rangle = a\langle a'|a\rangle$$

and hence

$$(a' - a)\langle a'|a\rangle = 0$$

so finally, if $a' \neq a$, then

$$\langle a'|a\rangle = 0.$$

The importance of this result lies in the fact that it makes it possible to construct a set of orthonormal states that define a basis for the state space of the system. To do this, we need the next property of Hermitean operators.

3. The eigenstates form a complete set of basis states for the state space of the system.

This can be proven to be always true if the state space is of finite dimension. If the state space is of infinite dimension, then completeness of the eigenstates of a Hermitean operator is not guaranteed. As we will see later, this has some consequences for the physical interpretation of such operators in quantum mechanics.

We can also always assume, *if the eigenvalue spectrum is discrete*, that these eigenstates are normalized to unity. If we were to suppose that they were not so normalized, for instance if the eigenstate $|a\rangle$ of the operator $\hat{A}$ is such that $\langle a|a\rangle \neq 1$, then we simply define a new state vector by

$$\widetilde{|a\rangle} = \frac{|a\rangle}{\sqrt{\langle a|a\rangle}} \tag{11.88}$$

which is normalized to unity. This new state $\widetilde{|a\rangle}$ is still an eigenstate of $\hat{A}$ with eigenvalue a – in fact it represents the same physical state as $|a\rangle$ – so we might as well have assumed from the very start that $|a\rangle$ was normalized to unity. Thus, provided the eigenvalue spectrum is discrete, then as well as the eigenstates forming a complete set of basis states, they also form an orthonormal set. Thus, if the operator $\hat{A}$ is Hermitean, and has a complete set of eigenstates $\{|a_n\rangle; n = 1, 2, 3 \ldots\}$, then these eigenstates form an orthonormal basis for the system. This means that any arbitrary state $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_n |a_n\rangle\langle a_n|\psi\rangle. \tag{11.89}$$

If the eigenvalue spectrum of an operator is continuous, then it is not possible to assume that the eigenstates can be normalized to unity. A different normalization scheme is required, as will be discussed in the next section.

11.4.3 Continuous Eigenvalues

Far from being the exception, Hermitean operators with continuous eigenvalues are basic to quantum mechanics, and it is consequently necessary to come to some understanding of the way the continuous case is distinct from the discrete case, and where they are the same. So in the following, consider a Hermitean operator $\hat{A}$ with continuous eigenvalues a lying in some range, between α_1 and α_2 say:

$$\hat{A}|a\rangle = a|a\rangle \qquad \alpha_1 < a < \alpha_2. \tag{11.90}$$

That there is a difficulty in dealing with eigenstates associated with a continuous range of eigenvalues can be seen if we make use of the (assumed) completeness of the eigenstates of a Hermitean operator, Eq. (11.89). It seems reasonable to postulate that in the case of continuous eigenvalues, this completeness relation would become an integral over the continuous range of eigenvalues:

$$|\psi\rangle = \int_{\alpha_1}^{\alpha_2} |a\rangle\langle a|\psi\rangle da. \tag{11.91}$$

We have seen this situation before in the discussion in Section 10.2.3 of the basis states $|x\rangle$ for the position of a particle. There we argued that the above form of the completeness relation can be used, but doing so requires that the inner product $\langle a'|a\rangle$, must be interpreted as a delta function:

$$\langle a'|a\rangle = \delta(a - a'). \tag{11.92}$$

The states $|a\rangle$ are said to be delta function normalized, in contrast to the orthonormal property of discrete eigenstates. As pointed out in Section 10.2.3, the result of this is that states such as $|a\rangle$ are of infinite norm and so cannot be normalized to unity. Such states cannot represent possible physical states of a system, which is an awkward state of affairs if the state is supposed to represent hat appears to be a physically reasonable state of the system. Fortunately it is possible to think of such states as idealized limits, and to work with them as if they were physically realizable,

provided care is taken. Mathematical (and physical) paradoxes can arise otherwise. However, linear combinations of these states can be normalized to unity, as this following example illustrates. If we consider a state $|\psi\rangle$ given by

$$|\psi\rangle = \int_{\alpha_1}^{\alpha_2} |a\rangle\langle a|\psi\rangle da, \tag{11.93}$$

then

$$\langle\psi|\psi\rangle = \int_{\alpha_1}^{\alpha_2} \langle\psi|a\rangle\langle a|\psi\rangle da. \tag{11.94}$$

But $\langle a|\psi\rangle = \psi(a)$ and $\langle\psi|a\rangle = \psi(a)^*$, so that

$$\langle\psi|\psi\rangle = \int_{\alpha_1}^{\alpha_2} |\psi(a)|^2 da. \tag{11.95}$$

Provided $|\psi(a)|^2$ is a well behaved function, this integral will converge to a finite result, so that the state $|\psi\rangle$ can indeed be normalized to unity and thus represent physically realizable states.

11.5 Dirac Notation for Operators

The above discussion of the properties of operators was based on making direct use of the defining properties of an operator, that is, in terms of their actions on ket vectors, in particular the vectors belonging to a set of basis states. All of these properties can be represented in a very succinct way that makes explicit use of the Dirac notation. The essential idea is to give a meaning to the symbol $|\phi\rangle\langle\psi|$, and we can see how this meaning is arrived at by considering the following example.

Suppose we have a spin half system with basis states $\{|+\rangle, |-\rangle\}$ and we have an operator $\hat{A}$ defined such that

$$\left.\begin{aligned} \hat{A}|+\rangle &= a|+\rangle + b|-\rangle \\ \hat{A}|-\rangle &= c|+\rangle + d|-\rangle \end{aligned}\right\} \tag{11.96}$$

and we calculate the quantity $\langle\phi|\hat{A}|\psi\rangle$ where $|\phi\rangle$ and $|\psi\rangle$ are arbitrary states. This is given by

$$\begin{aligned} \langle\phi|\hat{A}|\psi\rangle &= \langle\phi|\Big\{\hat{A}(|+\rangle\langle+|\psi\rangle + |-\rangle\langle-|\psi\rangle)\Big\} \\ &= \langle\phi|\Big[(a|+\rangle + b|-\rangle)\langle+|\psi\rangle + (c|+\rangle + d|-\rangle)\langle-|\psi\rangle\Big] \\ &= \langle\phi|\Big[a|+\rangle\langle+|\psi\rangle + b|-\rangle\langle+|\psi\rangle + c|+\rangle\langle-|\psi\rangle + d|-\rangle\langle-|\psi\rangle\Big]. \end{aligned} \tag{11.97}$$

We note that the term enclosed within the square brackets contains, symbolically at least, a common 'factor' $|\psi\rangle$ which we will move outside the brackets to give

$$\langle\phi|\hat{A}|\psi\rangle = \langle\phi|\Big[a|+\rangle\langle+| + b|-\rangle\langle+| + c|+\rangle\langle-| + d|-\rangle\langle-|\Big]|\psi\rangle \tag{11.98}$$

It is now tempting to make the identification of the operator $\hat{A}$ appearing on the left hand side of this expression with the combination of symbols appearing between the square brackets on the right hand side of the equation, i.e.

$$\hat{A} \leftrightarrow a|+\rangle\langle+| + b|-\rangle\langle+| + c|+\rangle\langle-| + d|-\rangle\langle-|. \tag{11.99}$$

We can do so provided we give appropriate meanings to this combination of ket-bra symbols such that it behaves in exactly the same manner as the operator $\hat{A}$ itself. Thus if we require that the action of this combination on a ket be given by

$$\begin{aligned} &\Big[a|+\rangle\langle+| + b|-\rangle\langle+| + c|+\rangle\langle-| + d|-\rangle\langle-|\Big]|\psi\rangle \\ &\qquad = a|+\rangle\langle+|\psi\rangle + b|-\rangle\langle+|\psi\rangle + c|+\rangle\langle-|\psi\rangle + d|-\rangle\langle-|\psi\rangle \\ &\qquad = |+\rangle(a\langle+|\psi\rangle + c\langle-|\psi\rangle) + |-\rangle(b\langle+|\psi\rangle + d\langle-|\psi\rangle) \end{aligned} \tag{11.100}$$

we see that this gives the correct result for $\hat{A}$ acting on the ket $|\psi\rangle$. In particular, if $|\psi\rangle = |\pm\rangle$ we recover the defining equations for $\hat{A}$ given in Eq. (11.96). If we further require that the action of this combination on a bra be given by

$$\begin{aligned}\langle\phi|\Big[a|+\rangle\langle+| + b|-\rangle\langle+| + c|+\rangle\langle-| + d|-\rangle\langle-|\Big] \\ &= a\langle\phi|+\rangle\langle+| + b\langle\phi|-\rangle\langle+| + c\langle\phi|+\rangle\langle-| + d\langle\phi|-\rangle\langle-| \\ &= (a\langle\phi|+\rangle + b\langle\phi|-\rangle)\langle+| + (c\langle\phi|+\rangle + d\langle\phi|-\rangle)\langle-|\end{aligned} \tag{11.101}$$

we see that this gives the correct result for $\hat{A}$ acting on the bra $\langle\psi|$. In particular, if $\langle\psi| = \langle\pm|$, this gives

$$\left.\begin{aligned}\langle+|\hat{A} &= a\langle+| + c\langle-| \\ \langle-|\hat{A} &= b\langle+| + d\langle-|\end{aligned}\right\} \tag{11.102}$$

which can be checked, using the defining condition for the action of an operator on a bra vector, Eq. (11.40), to be the correct result.

Thus we see that we can indeed write

$$\hat{A} = a|+\rangle\langle+| + b|-\rangle\langle+| + c|+\rangle\langle-| + d|-\rangle\langle-|. \tag{11.103}$$

as an valid expression for the operator $\hat{A}$ in terms of bra and ket symbols, provided we interpret the symbols in the manner indicated above, and summarized in more detail below.

The interpretation that is given is defined as follows:

$$\begin{aligned}(|\phi\rangle\langle\psi|)|\alpha\rangle &= |\phi\rangle\langle\psi|\alpha\rangle \\ \langle\alpha|(|\phi\rangle\langle\psi|) &= \langle\alpha|\phi\rangle\langle\psi|\end{aligned} \tag{11.104}$$

i.e. it maps kets into kets and bras into bras, exactly as an operator is supposed to.

If we further require $|\phi\rangle\langle\psi|$ to have the linear property

$$\begin{aligned}|\phi\rangle\langle\psi|(c_1|\psi_1\rangle + c_2|\psi_2\rangle) &= c_1(|\phi\rangle\langle\psi|)|\psi_1\rangle + c_2(|\phi\rangle\langle\psi|)|\psi_2\rangle \\ &= |\phi\rangle(c_1\langle\psi|\psi_1\rangle + c_2\langle\psi|\psi_2\rangle)\end{aligned} \tag{11.105}$$

and similarly for the operator acting on bra vectors, we have given the symbol the properties of a *linear operator*.

We can further generalize this to include sums of such bra-ket combinations, e.g.

$$c_1|\phi_1\rangle\langle\psi_1| + c_2|\phi_2\rangle\langle\psi_2|$$

where c_1 and c_2 are complex numbers, is an operator such that

$$(c_1|\phi_1\rangle\langle\psi_1| + c_2|\phi_2\rangle\langle\psi_2|)|\xi\rangle = c_1|\phi_1\rangle\langle\psi_1|\xi\rangle + c_2|\phi_2\rangle\langle\psi_2|\xi\rangle \tag{11.106}$$

and similarly for the action on bra vectors.

Finally, we can define the product of bra-ket combinations in the obvious way, that is

$$(|\phi\rangle\langle\psi|)(|\alpha\rangle\langle\beta|) = |\phi\rangle\langle\psi|\alpha\rangle\langle\beta| = \langle\psi|\alpha\rangle|\phi\rangle\langle\beta|. \tag{11.107}$$

Below we describe a number of examples that illustrate the usefulness of this notation.

Ex 11.22 The three operators (the Pauli spin operators) for a spin half system whose state space is spanned by the usual basis states $\{|+\rangle, |-\rangle\}$ are given, in Dirac notation, by the expressions

$$\hat{\sigma}_x = |-\rangle\langle+| + |+\rangle\langle-|$$
$$\hat{\sigma}_y = i|-\rangle\langle+| - i|+\rangle\langle-|$$
$$\hat{\sigma}_z = |+\rangle\langle+| - |-\rangle\langle-|.$$

Determine the action of these operators on the basis states $|\pm\rangle$.

First we consider $\hat{\sigma}_x|+\rangle$ which can be written

$$\hat{\sigma}_x|+\rangle = \Big[|-\rangle\langle+| + |+\rangle\langle-|\Big]|+\rangle = |-\rangle\langle+|+\rangle + |+\rangle\langle-|+\rangle = |-\rangle.$$

Similarly, for instance

$$\hat{\sigma}_y|-\rangle = \Big[i|-\rangle\langle+| - i|+\rangle\langle-|\Big]|-\rangle = i|-\rangle\langle+|-\rangle - i|+\rangle\langle-|-\rangle = -i|+\rangle.$$

In each of the above examples, the orthonormality of the states $|\pm\rangle$ has been used.

Ex 11.23 For the Pauli spin operators defined above, determine the action of these operators on the bra vectors $\langle\pm|$.

We find, for instance

$$\langle-|\hat{\sigma}_z = \langle-|\Big[|+\rangle\langle+| - |-\rangle\langle-|\Big] = \langle-|+\rangle\langle+| - \langle-|-\rangle\langle-| = -\langle-|.$$

Ex 11.24 Calculate the product $\hat{\sigma}_x\hat{\sigma}_y$ and the commutator $[\hat{\sigma}_x, \hat{\sigma}_y]$.

This product is:

$$\begin{aligned}\hat{\sigma}_x\hat{\sigma}_y &= \Big[|-\rangle\langle+| + |+\rangle\langle-|\Big]\Big[i|-\rangle\langle+| - i|+\rangle\langle-|\Big] \\ &= i|-\rangle\langle+|-\rangle\langle+| - i|-\rangle\langle+|+\rangle\langle-| + i|+\rangle\langle-|-\rangle\langle+| - i|+\rangle\langle-|+\rangle\langle-| \\ &= -i|-\rangle\langle-| + i|+\rangle\langle+| \\ &= i\hat{\sigma}_z.\end{aligned}$$

In the same fashion, it can be shown that $\hat{\sigma}_y\hat{\sigma}_x = -i\hat{\sigma}_z$ so that we find that

$$[\hat{\sigma}_x, \hat{\sigma}_y] = 2i\hat{\sigma}_z.$$

There are further important properties of this Dirac notation for operators worth highlighting.

Projection Operators In this notation, a projection operator $\hat{P}$ will be simply given by

$$\hat{P} = |\psi\rangle\langle\psi| \tag{11.108}$$

provided $|\psi\rangle$ is normalized to unity, since we have

$$\hat{P}^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \hat{P} \tag{11.109}$$

as required for a projection operator.

Completeness Relation This new notation also makes it possible to express the completeness relation in a particularly compact form. Recall that if the set of ket vectors $\{|\varphi_n\rangle; n = 1, 2, 3 \dots\}$ is a complete set of orthonormal basis states for the state space of a system, then any state $|\psi\rangle$ can be written

$$|\psi\rangle = \sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle \tag{11.110}$$

which in our new notation can be written

$$|\psi\rangle = \Big(\sum_n |\varphi_n\rangle\langle\varphi_n|\Big)|\psi\rangle \tag{11.111}$$

so that we must conclude that

$$\sum_n |\varphi_n\rangle\langle\varphi_n| = \hat{1} \tag{11.112}$$

where $\hat{1}$ is the unit operator. It is often referred to as a *decomposition of unity*.

In the case of continuous eigenvalues, the same argument as above can be followed through. Thus, if we suppose that a Hermitean operator $\hat{A}$ has a set of eigenstates $\{|a\rangle; \alpha_1 < a < \alpha_2\}$, then we can readily show that

$$\int_{\alpha_1}^{\alpha_2} |a\rangle\langle a| da = \hat{1}. \tag{11.113}$$

Note that, in practice, it is often the case that an operator can have both discrete and continuous eigenvalues, in which case the completeness relation can be written

$$\sum_n |\varphi_n\rangle\langle\varphi_n| + \int_{\alpha_1}^{\alpha_2} |a\rangle\langle a| da = \hat{1} \tag{11.114}$$

The completeness relation expressed in this fashion (in both the discrete and continuous cases) is extremely important and has widespread use in calculational work, as illustrated in the following examples.

Ex 11.25 Show that any operator can be expressed in terms of this Dirac notation. We can see this for an operator A by writing

$$\hat{A} = \hat{1}\hat{A}\,\hat{1} \tag{11.115}$$

and using the decomposition of unity twice over to give

$$\begin{aligned}\hat{A} &= \sum_m \sum_n |\varphi_m\rangle\langle\varphi_m|\hat{A}|\varphi_n\rangle\langle\varphi_n| \\ &= \sum_m \sum_n |\varphi_m\rangle\langle\varphi_n| A_{mn}\end{aligned} \tag{11.116}$$

where $A_{mn} = \langle\varphi_m|\hat{A}|\varphi_n\rangle$.

Ex 11.26 Using the decomposition of unity in terms of the basis states $\{|\varphi_n\rangle; n = 1, 2, 3 \dots\}$, expand $\hat{A}|\varphi_m\rangle$ in terms of these basis states. This calculation proceeds by inserting the unit operator in a convenient place:

$$\begin{aligned}\hat{A}|\varphi_m\rangle = \hat{1}\hat{A}|\varphi_m\rangle &= \Big(\sum_n |\varphi_n\rangle\langle\varphi_n|\Big)\hat{A}|\varphi_m\rangle \\ &= \sum_n |\varphi_n\rangle\langle\varphi_n|\hat{A}|\varphi_m\rangle \\ &= \sum_n A_{nm}|\varphi_n\rangle\end{aligned} \tag{11.117}$$

where $A_{nm} = \langle\varphi_n|\hat{A}|\varphi_m\rangle$.

Ex 11.27 Using the decomposition of unity, we can insert the unit operator between the two operators in the quantity $\langle\psi|\hat{A}\hat{B}|\phi\rangle$ to give

$$\langle\psi|\hat{A}\hat{B}|\phi\rangle = \langle\psi|\hat{A}\,\hat{1}\hat{B}|\phi\rangle = \sum_n \langle\psi|\hat{A}|\varphi_n\rangle\langle\varphi_n|\hat{B}|\phi\rangle. \tag{11.118}$$

Hermitean conjugate of an operator It is straightforward to write down the Hermitean conjugate of an operator. Thus, for the operator $\hat{A}$ given by

$$\hat{A} = \sum_n c_n|\phi_n\rangle\langle\psi_n| \tag{11.119}$$

we have

$$\langle\phi|\hat{A}|\psi\rangle = \sum_n c_n\langle\phi|\phi_n\rangle\langle\psi_n|\psi\rangle \tag{11.120}$$

so that taking the complex conjugate we get

$$\langle\psi|\hat{A}^\dagger|\phi\rangle = \sum_n c_n^*\langle\psi|\psi_n\rangle\langle\phi_n|\phi\rangle = \langle\psi|\Big(\sum_n c_n^*|\psi_n\rangle\langle\phi_n|\Big)|\phi\rangle. \tag{11.121}$$

We can then extract from this the result

$$\hat{A}^\dagger = \sum_n c_n^*|\psi_n\rangle\langle\phi_n|. \tag{11.122}$$

Spectral decomposition of an operator As a final important result, we can look at the case of expressing an Hermitean operator in terms of projectors onto its basis states. Thus, if we suppose that $\hat{A}$ has the eigenstates $\{|a_n\rangle; n = 1, 2, 3\ldots\}$ and associated eigenvalues $a_n, n = 1, 2, 3\ldots$, so that

$$\hat{A}|a_n\rangle = a_n|a_n\rangle \tag{11.123}$$

then by noting that the eigenstates of $\hat{A}$ form a complete orthonormal set of basis states we can write the decomposition of unity in terms of the eigenstates of $\hat{A}$ as

$$\sum_n |a_n\rangle\langle a_n| = \hat{1}. \tag{11.124}$$

Thus we find that

$$\hat{A} = \hat{A}\,\hat{1} = \hat{A}\sum_n |a_n\rangle\langle a_n| = \sum_n \hat{A}|a_n\rangle\langle a_n| = \sum_n a_n|a_n\rangle\langle a_n|. \tag{11.125}$$

so that

$$\hat{A} = \sum_n a_n|a_n\rangle\langle a_n|. \tag{11.126}$$

The analogous result for continuous eigenstates is then

$$\hat{A} = \int_{\alpha_1}^{\alpha_2} a|a\rangle\langle a|da \tag{11.127}$$

while if the operator has both continuous and discrete eigenvalues, the result is

$$\hat{A} = \sum_n a_n|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2} a|a\rangle\langle a|da. \tag{11.128}$$

This is known as the spectral decomposition of the operator $\hat{A}$, the name coming, in part, from the fact that the collection of eigenvalues of an operator is known as its eigenvalue spectrum.

Ex 11.28 Calculate the spectral decomposition of the operator $\hat{A}^2$, where $\hat{A}$ is as given in Eq. (11.128).

We can write $\hat{A}^2$ as

$$\begin{aligned}\hat{A}^2 &= \hat{A}\sum_n a_n|a_n\rangle\langle a_n| + \hat{A}\int_{\alpha_1}^{\alpha_2} a|a\rangle\langle a|da \\ &= \sum_n a_n\hat{A}|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2} a\hat{A}|a\rangle\langle a|da \\ &= \sum_n a_n^2|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2} a^2|a\rangle\langle a|da.\end{aligned}$$

Ex 11.29 Express $f(\hat{A})$ in Dirac notation, where $\hat{A}$ is a Hermitean operator given by Eq. (11.128) and where $f(x)$ can be expanded as a power series in x.

From the preceding exercise, it is straightforward to show that

$$\hat{A}^k = \sum_n a_n^k|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2} a^k|a\rangle\langle a|da.$$

Since $f(x)$ can be expanded as a power series in x, we have

$$f(\hat{A}) = \sum_{k=0}^{\infty} c_k\hat{A}^k$$

so that

$$\begin{aligned}f(\hat{A}) &= \sum_{k=0}^{\infty} c_k\left\{\sum_n a_n^k|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2} a^k|a\rangle\langle a|da\right\} \\ &= \sum_n\left\{\sum_{k=0}^{\infty} c_k a_n^k\right\}|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2}\left\{\sum_{k=0}^{\infty} c_k a^k\right\}|a\rangle\langle a|da \\ &= \sum_n f(a_n)|a_n\rangle\langle a_n| + \int_{\alpha_1}^{\alpha_2} f(a)|a\rangle\langle a|da\end{aligned}$$

Ex 11.30 Determine $f(\hat{A})|a_n\rangle$ where $\hat{A} = \sum_n a_n|a_n\rangle\langle a_n|$.

In this case, we can use the expansion for $f(\hat{A})$ as obtained in the previous example, that is

$$f(\hat{A}) = \sum_n f(a_n)|a_n\rangle\langle a_n|$$

so that

$$f(\hat{A})|a_k\rangle = \sum_n f(a_n)|a_n\rangle\langle a_n|a_k\rangle = \sum_n f(a_n)|a_n\rangle\delta_{nk} = f(a_k)|a_k\rangle.$$

Since k is a dummy index here, we can write this as $f(\hat{A})|a_n\rangle = f(a_n)|a_n\rangle$. Thus, the effect is simply to replace the operator in $f(\hat{A})$ by its eigenvalue, i.e. $f(a_n)$.

This last result can be readily shown to hold true in the case of continuous eigenvalues. It is a very important result that finds very frequent application in practice.

Chapter 12

Matrix Representations of State Vectors and Operators

In the preceding Chapters, the mathematical ideas underpinning the quantum theory have been developed in a fairly general (though, admittedly, not a mathematically rigorous) fashion. However, much that has been presented, particularly the concept of an operator, can be developed in another way that is, in some respects, less abstract than what has been used so far. This alternate form of presentation involves working with the *components* of the state vectors and operators, leading to their being represented by column and row vectors, and matrices. This development of the theory is completely analogous to the way in which this is done when dealing with the position vector in ordinary three dimensional space. Below, we will look at how the idea of writing the position vectors in two dimensional space can be written in terms of column and row vectors. We will then use the ideas developed there to show how state vectors and operators can be expressed in a similar fashion. This alternate route offers another way of introducing such concepts as adding, multiplying and taking the inverse of operators through their representations as matrices, and further provides another way to introduce the idea of the Hermitean adjoint of an operator, and of a Hermitian operator.

12.1 Representation of Vectors In Euclidean Space as Column and Row Vectors

When writing down a vector, we have so far made explicit the basis vectors when writing an expression such as $\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}}$ for a position vector, or $|S\rangle = a|+\rangle + b|-\rangle$ for the state of a spin half system. But the choice of basis vectors is not unique, i.e. we could use any other pair of orthonormal unit vectors $\hat{\mathbf{i}}'$ and $\hat{\mathbf{j}}'$, and express the vector $\mathbf{r}$ in terms of these new basis vectors, though of course the components of $\mathbf{r}$ will change. The same is true for the spin basis vectors $|\pm\rangle$, i.e. we can express the state $|S\rangle$ in terms of some other basis vectors, such as the states for which the x component of spin has the values $S_x = \frac{1}{2}\hbar$, though once again, the components of $|S\rangle$ will now be different. But it is typically the case that once the choice of basis vectors have been decided on, it should not be necessary to always write them down when writing down a vector, i.e. it would be just as useful to just write down the components of a vector. Doing this leads to a convenient notation in which vectors are written in terms of column and row vectors. It also provides a direct route to some of the important mathematical entities encountered in quantum mechanics such as bra vectors and operators that are more rigorously introduced in a more abstract way.

12.1.1 Column Vectors

To illustrate the ideas, we will use the example of a position vector in two dimensional space. The point that is to be established here is that a position vector is an independently existing geometrical

object 'suspended' in space, much as a pencil held in the air with a steady position and orientation has a fixed length and orientation. One end of the pencil, say where the eraser is, can be taken to be the origin O, and the other end (the sharp end) the position of a point P. Then the position and orientation of the pencil defines a position vector $\mathbf{r}$ of P with respect to the origin O. This vector can be represented by a single arrow joining O to P whose length and orientation specify the position of P with respect to O. As such, the vector $\mathbf{r}$ also has an independent existence as a geometrical object sitting in space, and we can work with this vector $\mathbf{r}$ and others like it by, for instance, performing vector additions by using the triangle law of vector addition as illustrated in Fig. (8.1), or performing scalar products by use of the definition Eq. (8.3).

In what was just described, we work only with the whole vector itself. This is in contrast with a very useful alternate way of working with vectors, that is to express any vector as a linear combination of a pair of basis vectors (in two dimensions), which amounts to building around these vectors some sort of 'scaffolding', a coordinate system such as a pair of X and Y axes, and describe the vector in terms of its components with respect to these axes. More to the point, what is provided is a pair of basis vectors such as the familiar unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ and write $\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}}$. We see that *any* position vector can be written in this way, i.e. the unit vectors constitute a pair of orthonormal basis vectors, and x and y are known as the components of $\mathbf{r}$ with respect to the basis vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. We can then work out how to add vectors, calculate scalar products and so on working solely with these components. For instance, if we have two vectors $\hat{\mathbf{r}_1}$ and $\hat{\mathbf{r}_2}$ given by $\hat{\mathbf{r}_1} = x_1\hat{\mathbf{i}} + y_1\hat{\mathbf{j}}$ and $\hat{\mathbf{r}_2} = x_2\hat{\mathbf{i}} + y_2\hat{\mathbf{j}}$ then $\hat{\mathbf{r}_1} + \hat{\mathbf{r}_2} = (x_1 + x_2)\hat{\mathbf{i}} + (y_1 + y_2)\hat{\mathbf{j}}$.

It is important to note that while a vector $\hat{\mathbf{r}}$ is a unique geometrical object, there is no unique choice of basis vectors, and correspondingly the components of the vector will change depending on the choice of basis vectors. Thus we could equally well have chosen the basis vectors $\hat{\mathbf{i}}'$ and $\hat{\mathbf{j}}'$, as illustrated in Fig. (12.1) so that the *same* vector $\mathbf{r}$ can be written

$$\mathbf{r} = x'\,\hat{\mathbf{i}}' + y'\,\hat{\mathbf{j}}' = x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}} \qquad (12.1)$$

with $x' \neq x$ and $y' \neq y$.

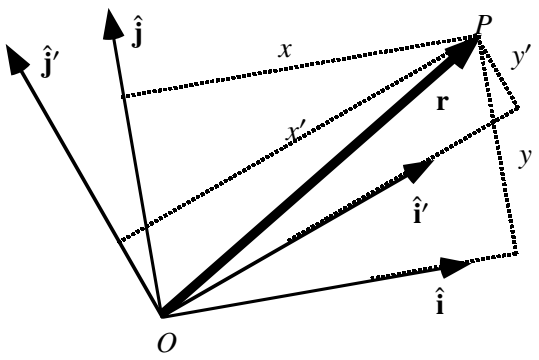


Figure 12.1: The position vector $\mathbf{r}$ written as a linear combination of two different pairs of orthonormal basis vectors. Although the coordinates of $\mathbf{r}$ are different with respect to the two sets of basis vectors, the vector $\mathbf{r}$ remains the same.

Once a choice of basis vectors has been made, it proves to be very convenient to work solely with the coordinates. There is a useful notation by which this can be done. In this notation, the vector $\mathbf{r}$ is written as

$$\mathbf{r} \doteq \begin{pmatrix} x \\ y \end{pmatrix} \qquad (12.2)$$

what is known as a *column vector*. We then say that this column vector is a *representation* of the vector $\mathbf{r}$ with respect to the basis vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. It is important to note that we do not say that $\mathbf{r}$ equals the column vector, in fact it is not an equal sign that is used in Eq. (12.2), rather the symbol '$\doteq$' is used, which is to be read as 'is represented by'. The reason for this is that, as mentioned above, while the vector $\mathbf{r}$ is a unique geometrical object, its components are not – they depend on the choice of basis vectors. We could have equally chosen basis vectors $\hat{\mathbf{i}}'$ and $\hat{\mathbf{j}}'$, and since the components x' and y' will be, in general, different from x and y, we end up with a different column

vector representing the same vector:

$$\mathbf{r} \doteq \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{12.3}$$

i.e. two apparently different column vectors representing the *same* vector $\mathbf{r}$. Equivalently, if we had two column vectors with exactly the same numbers in the two positions, we could not conclude that they represent the same vector unless we were told that the basis vectors were the same in each case. Thus if there is any chance of ambiguity, we need to make it clear when we use the column vector notation, exactly what the basis vectors are. The terminology then is to say that the vector $\mathbf{r}$ is given by the column vector in Eq. (12.2) in the $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$ representation.

Once a choice of basis vectors has been settled on, and consistently used, we can proceed with vector calculations using the new notation. Thus, for instance, we can add two vectors:

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2 \tag{12.4}$$

which becomes, using the $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$ representation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}. \tag{12.5}$$

12.1.2 Row Vectors

The scalar product $(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{r}_1 \cdot \mathbf{r}_2$ can be calculated by using the usual rule $\mathbf{r}_1 \cdot \mathbf{r}_2 = r_1 r_2 \cos\theta$, but it can also be expressed in terms of the components of $\mathbf{r}_1$ and $\mathbf{r_2}$ in, say, the $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$ representation, though note that the same numerical result is obtained whatever representation is used. The result is simply

$$(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{r}_1 \cdot \mathbf{r}_2 = x_1 x_2 + y_1 y_2. \tag{12.6}$$

At this point we note that, if we use the rules of matrix multiplication, this last result can be written

$$(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{r}_1 \cdot \mathbf{r}_2 = \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \tag{12.7}$$

where we note the appearance of the column vector representing the vector $\mathbf{r}_2$, but $\mathbf{r}_1$, the first factor in the scalar product, has been represented by a *row vector*. If the components of $\mathbf{r}_1$ and $\mathbf{r}_2$ were complex, then we would write the inner product as

$$(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{r}_1^* \cdot \mathbf{r}_2 = x_1^* x_2 + y_1^* y_2 = \begin{pmatrix} x_1^* & y_1^* \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \tag{12.8}$$

The use of a row vector to represent $\mathbf{r}_1$ can be looked on here as a convenience so that the rules of matrix multiplication can be applied, but there is a deeper significance to its use[1] that will become apparent when we look at the column and row vector representations of ket and bra vectors.

12.2 Representations of State Vectors and Operators

The procedure here is identical to that which was followed in the case of the position vector, i.e. we introduce a complete set of orthonormal basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$ that span the state space of the quantum system, and then work with the components of the ket and bra vectors, and the operators. Of course, we now do not have the luxury of interpreting these basis vectors as representing physical directions in real space – rather they are abstract vectors in a multi-dimensional complex vector space, but much of what has been said above in connection with vectors in ordinary Euclidean space can be carried over to this more abstract situation.

[1] Effectively, what is going on is that corresponding to any vector $\mathbf{r}$ represented by a column vector, there corresponds another vector $\mathbf{r}_*$ known as its dual which is represented by a row vector. The original vector is the 'physical' vector while its dual is an abstract mathematical companion. The original vector and its dual belong to two different vector spaces.

12.2.1 Row and Column Vector Representations for Spin Half State Vectors

To set the scene, we will look at the particular case of spin half state vectors for which, as we have seen earlier, Sec. 8.3, an arbitrary state $|S\rangle$ can be written

$$|S\rangle = |-\rangle\langle -|S\rangle + |+\rangle\langle +|S\rangle,$$

i.e the state $|S\rangle$ is expressed as a linear combination of the two basis states $|\pm\rangle$. We further saw that the ket vectors as $|+\rangle$, $|-\rangle$ could be put into direct correspondence with the (complex) unit vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ respectively, and that the probability amplitudes $\langle \pm|S\rangle$ are the components of $|S\rangle$ in the 'direction' of the basis states $|\pm\rangle$. We can complete the analogy with the case of ordinary vectors by noting that we could then write this ket vector as a column vector, i.e.

$$|S\rangle \doteq \begin{pmatrix} \langle -|S\rangle \\ \langle +|S\rangle \end{pmatrix}. \tag{12.9}$$

If we pursue this line further, we can get an idea of how to interpret bra vectors. To do this, consider the more general probability amplitudes $\langle S'|S\rangle$. This we can write as

$$\langle S'|S\rangle = \langle S'|-\rangle\langle -|S\rangle + \langle S'|+\rangle\langle +|S\rangle. \tag{12.10}$$

If we now use

$$\langle \pm|S\rangle = \langle S|\pm\rangle^* \tag{12.11}$$

this becomes

$$\langle S'|S\rangle = \langle -|S'\rangle^*\langle -|S\rangle + \langle +|S'\rangle^*\langle +|S\rangle \tag{12.12}$$

which we can write as

$$\langle S'|S\rangle = \begin{pmatrix} \langle -|S'\rangle^* & \langle +|S'\rangle^* \end{pmatrix} \begin{pmatrix} \langle -|S\rangle \\ \langle +|S\rangle \end{pmatrix}. \tag{12.13}$$

In other words, the bra vector $\langle S'|$ is represented by the row vector

$$\langle S'| \doteq \begin{pmatrix} \langle -|S'\rangle^* & \langle +|S'\rangle^* \end{pmatrix}. \tag{12.14}$$

This shows that a bra vector is more than just the 'complex conjugate' of a ket vector, since a row vector is not the same as a column vector.

We can now extend the idea to the more general situation of a state space of dimension $n > 2$.

12.2.2 Representation of Ket and Bra Vectors

In terms of the basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$, an arbitrary state vector $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle. \tag{12.15}$$

Let us now write

$$\langle\varphi_n|\psi\rangle = \psi_n. \tag{12.16}$$

We then have, by analogy with the position vector:

$$|\psi\rangle \doteq \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{pmatrix}. \tag{12.17}$$

This is a representation of $|\psi\rangle$ as a column vector with respect to the set of basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$. In particular, the basis state $|\varphi_m\rangle$ will have components

$$(\varphi_m)_n = \langle\varphi_n|\varphi_m\rangle = \delta_{nm} \tag{12.18}$$

and so they will be represented by column vectors of the form

$$|\varphi_1\rangle \doteq \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \qquad |\varphi_2\rangle \doteq \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \qquad \dots \tag{12.19}$$

i.e. the m^{th} component φ_{nm} of $|\varphi_n\rangle$ is zero except in the m^{th} position where $\varphi_{mm} = 1$.

Now form the inner product $\langle\chi|\psi\rangle$:

$$\langle\chi|\psi\rangle = \sum_n \langle\chi|\varphi_n\rangle\langle\varphi_n|\psi\rangle. \tag{12.20}$$

We know that $\langle\chi|\varphi_n\rangle = (\langle\varphi_n|\chi\rangle)^*$, and following on from the notation introduce above, we write $\chi_n = \langle\varphi_n|\chi\rangle$ so that

$$\langle\chi|\varphi_n\rangle = \chi_n^* \tag{12.21}$$

and hence

$$\langle\chi|\psi\rangle = \sum_n \chi_n^* \psi_n \tag{12.22}$$

which we can write as

$$\langle\chi|\psi\rangle = \begin{pmatrix} \chi_1^* & \chi_2^* & \chi_3^* & \dots \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{pmatrix} \tag{12.23}$$

which we evaluate by the usual rules of matrix multiplication. Note that here we have made the identification of the bra vector $\langle\chi|$ as a row vector:

$$\langle\chi| \doteq \begin{pmatrix} \chi_1^* & \chi_2^* & \chi_3^* & \dots \end{pmatrix} \tag{12.24}$$

with respect to the set of basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$. This can be compared with the representation of the ket vector $|\chi\rangle$ as a column vector:

$$|\chi\rangle \doteq \begin{pmatrix} \chi_1^* \\ \chi_2^* \\ \chi_3^* \\ \vdots \end{pmatrix}. \tag{12.25}$$

This difference in appearance of the representation of a bra and a ket vector, the first as a row vector, the second as a column vector, perhaps emphasizes the point made in Section 9.2.1 that the bra vectors form a vector space, the dual Hilbert space $\mathcal{H}^*$ related to, but distinct from, the Hilbert space $\mathcal{H}$ of the ket vectors. In a sense, a bra vector can be thought of as something akin to being the 'complex conjugate' of its corresponding ket vector.

12.2.3 Representation of Operators

Now turn to the operator equation

$$\hat{A}|\psi\rangle = |\phi\rangle \tag{12.26}$$

which we can write as

$$|\phi\rangle = \hat{A}|\psi\rangle = \hat{A}\sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle = \sum_n \hat{A}|\varphi_n\rangle\langle\varphi_n|\psi\rangle. \tag{12.27}$$

Then

$$\langle\varphi_m|\phi\rangle = \sum_n \langle\varphi_m|\hat{A}|\varphi_n\rangle\langle\varphi_n|\psi\rangle \tag{12.28}$$

which we can write as

$$\phi_m = \sum_n A_{mn}\psi_n \tag{12.29}$$

where

$$A_{mn} = \langle\varphi_m|\hat{A}|\varphi_n\rangle. \tag{12.30}$$

We can write this as a matrix equation:

$$\begin{pmatrix}\phi_1\\ \phi_2\\ \phi_3\\ \vdots\end{pmatrix} = \begin{pmatrix}A_{11} & A_{12} & A_{13} & \dots\\ A_{21} & A_{22} & A_{23} & \dots\\ A_{31} & A_{32} & A_{33} & \dots\\ \vdots & \vdots & \vdots & \end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \vdots\end{pmatrix} \tag{12.31}$$

where the operator $\hat{A}$ is represented by a *matrix*:

$$\hat{A} \doteq \begin{pmatrix}A_{11} & A_{12} & A_{13} & \dots\\ A_{21} & A_{22} & A_{23} & \dots\\ A_{31} & A_{32} & A_{33} & \dots\\ \vdots & \vdots & \vdots & \end{pmatrix}. \tag{12.32}$$

The quantities A_{mn} are known as the matrix elements of the operator $\hat{A}$ with respect to the basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$.

It is important to keep in mind that the column vectors, row vectors, and matrices above are constructed with respect to a particular set of basis states. If a different set of basis states are used, then the state vectors and operators remain the same, but the column or row vector, or matrix representing the state vector or operator respectively will change. Thus, to give any meaning to a row vector, or a column vector, or a matrix, it is essential that the basis states be known. An important part of quantum mechanics is the mathematical formalism that deals with transforming between different sets of basis states. However, we will not be looking at transformation theory here.

Ex 12.1 Consider two state vectors

$$|1\rangle = \tfrac{1}{\sqrt{2}}[|-\rangle - i|+\rangle] \qquad |2\rangle = \tfrac{1}{\sqrt{2}}[|-\rangle + i|+\rangle]$$

where $|\pm\rangle$ are the usual base states for a spin half system. We want to represent these ket vectors as column vectors with respect to the set of basis states $\{|+\rangle, |-\rangle\}$. Firstly, we note that in the general development described above, we assumed that the basis states

were named $|\varphi_1\rangle$, $|\varphi_2\rangle$ and so on. But here we are using a different way of labelling the basis states, which means we have a choice as to which of $|\pm\rangle$ we identify with $|\varphi_1\rangle$ and $|\varphi_2\rangle$. If makes no difference what we choose: we make the choice to suit ourselves, and provided we use it consistently then no problems should arise. Thus, here, we will choose $|\varphi_1\rangle = |+\rangle$ and $|\varphi_2\rangle = |-\rangle$. Thus we can write write

$$|+\rangle \doteq \begin{pmatrix}1\\0\end{pmatrix} \quad \text{and} \quad |-\rangle \doteq \begin{pmatrix}0\\1\end{pmatrix}.$$

We can then express the states $|1\rangle$ and $|2\rangle$ in column vector notation as

$$|1\rangle \doteq \tfrac{1}{\sqrt{2}}\begin{pmatrix}-i\\1\end{pmatrix}$$

which can also be written as

$$|1\rangle \doteq -\tfrac{i}{\sqrt{2}}\begin{pmatrix}1\\0\end{pmatrix} + \tfrac{1}{\sqrt{2}}\begin{pmatrix}0\\1\end{pmatrix}.$$

The corresponding bra vectors are

$$\langle 1| = \tfrac{1}{\sqrt{2}}[\langle -| + i\langle +|] \qquad \langle 2| = \tfrac{1}{\sqrt{2}}[\langle -| - i\langle +|]$$

or, as row vectors

$$\langle 1| \doteq (i \quad 1) \quad \text{and} \quad \langle 2| \doteq (-i \quad 1).$$

We can calculate inner products as follows:

$$\begin{aligned}\langle 1|2\rangle &= \tfrac{1}{2}(i \quad 1)\begin{pmatrix}i\\1\end{pmatrix}\\ &= 0,\\ \langle 1|1\rangle &= \tfrac{1}{2}(i \quad 1)\begin{pmatrix}-i\\1\end{pmatrix}\\ &= 1.\end{aligned}$$

and so on.

Ex 12.2 We can also look at the operator $\hat{A}$ defined by

$$\hat{A}|\pm\rangle = \pm\tfrac{1}{2}i\hbar|\mp\rangle$$

which can be written out in matrix form as

$$\hat{A} \doteq \begin{pmatrix}\langle +|\hat{A}|+\rangle & \langle +|\hat{A}|-\rangle\\ \langle -|\hat{A}|+\rangle & \langle -|\hat{A}|-\rangle\end{pmatrix} = \begin{pmatrix}0 & -\frac{1}{2}i\hbar\\ \frac{1}{2}i\hbar & 0\end{pmatrix}$$

so that, for instance

$$\begin{aligned}\hat{A}|1\rangle &\doteq \begin{pmatrix}0 & -\frac{1}{2}i\hbar\\ \frac{1}{2}i\hbar & 0\end{pmatrix}\begin{pmatrix}-\frac{i}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}\\ &= -\tfrac{1}{2}i\hbar\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{i}{\sqrt{2}}\end{pmatrix}\\ &= \tfrac{1}{2}\hbar\begin{pmatrix}-\frac{i}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}.\end{aligned}$$

Thus we have $\hat{A}|1\rangle = \frac{1}{2}\hbar|1\rangle$, which incidentally shows that $|1\rangle$ is an eigenstate of $\hat{A}$.

Using the representations of bra vectors and operators, it is straightforward to see what the action of an operator on a bra vector is given by. Thus, we have:

$$\langle\psi|\hat{A} \doteq (\psi_1 \quad \psi_2 \quad \psi_3 \quad \ldots)\begin{pmatrix} A_{11} & A_{12} & A_{13} & \ldots \\ A_{21} & A_{22} & A_{23} & \ldots \\ A_{31} & A_{32} & A_{33} & \ldots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

$$= (\psi_1 A_{11} + \psi_2 A_{21} + \ldots \quad \psi_2 A_{12} + \psi_2 A_{22} + \ldots \quad \psi_1 A_{13} + \psi_2 A_{23} + \ldots \quad \ldots). \quad (12.33)$$

The final result can then be written in the corresponding bra vector notation if desired. This can be illustrated by example.

Ex 12.3 Evaluate $\langle 2|\hat{A}$ using the representations of the bra vector $|2\rangle$ and operator $\hat{A}$:

$$\begin{aligned}\langle 2|\hat{A} &\doteq \tfrac{1}{\sqrt{2}}(-i \quad 1)\begin{pmatrix} 0 & -\frac{1}{2}i\hbar \\ \frac{1}{2}i\hbar & \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}(\tfrac{1}{2}i\hbar \quad -\tfrac{1}{2}\hbar) \\ &= -\tfrac{1}{2}\hbar \cdot \tfrac{1}{\sqrt{2}}(-i \quad 1)\end{aligned}$$

which can be written as $\langle 2|\hat{A} = -\frac{1}{2}\hbar\langle 2|$.

12.2.4 Properties of Matrix Representations of Operators

Many of the properties of operators can be expressed in terms of the properties of their representative matrices. Most of these properties are straightforward, and will be presented below without comment.

Equality

Two operators are equal if their corresponding operator matrix elements are equal, i.e. $\hat{A} = \hat{B}$ if $A_{mn} = B_{mn}$.

Unit and Zero Operator

The unit operator $\hat{1}$ is the operator such that $\hat{1}|\psi\rangle = |\psi\rangle$ for all states $|\psi\rangle$. It has the the matrix elements $\hat{1}_{mn} = \delta_{mn}$, i.e. the diagonal elements are all unity, and the off-diagonal elements are all zero. The unit operator has the same form in all representations, i.e. irrespective of the choice of basis states. The zero operator $\hat{0}$ is the operator such that $\hat{0}|\psi\rangle = 0$ for all states $|\psi\rangle$. Its matrix elements are all zero.

Addition of Operators

Given two operators $\hat{A}$ and $\hat{B}$ with matrix elements A_{mn} and B_{mn}, then the matrix elements of their sum $\hat{S} = \hat{A} + \hat{B}$ are given by

$$S_{mn} = A_{mn} + B_{mn}. \quad (12.34)$$

Multiplication by a Complex Number

If λ is a complex number, then the matrix elements of the operator $\hat{C} = \lambda\hat{A}$ are given by

$$C_{mn} = \lambda A_{mn}. \tag{12.35}$$

Product of Operators

Given two operators $\hat{A}$ and $\hat{B}$ with matrix elements A_{mn} and B_{mn}, then the matrix elements of their product $\hat{P} = \hat{A}\hat{B}$ are given by

$$P_{mn} = \sum_k A_{mk}B_{kn} \tag{12.36}$$

i.e. the usual rule for the multiplication of two matrices. Matrix multiplication, and hence operator multiplication, is not commutative, i.e. in general $\hat{A}\hat{B} \neq \hat{B}\hat{A}$. The difference, $\hat{A}\hat{B} - \hat{B}\hat{A}$, known as the commutator of $\hat{A}$ and $\hat{B}$ and written $[\hat{A}, \hat{B}]$, can be readily evaluated using the matrix representations of $\hat{A}$ and $\hat{B}$.

Ex 12.4 Three operators $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\sigma}_3$, known as the Pauli spin matrices, that occur in the theory of spin half systems (and elsewhere) have the matrix representations with respect to the $\{|+\rangle, |-\rangle\}$ basis given by

$$\hat{\sigma}_1 \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \hat{\sigma}_2 \doteq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \hat{\sigma}_3 \doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The commutator $[\hat{\sigma}_1, \hat{\sigma}_2]$ can be readily evaluated using these matrices:

$$\begin{aligned} [\hat{\sigma}_1, \hat{\sigma}_2] &= \hat{\sigma}_1\hat{\sigma}_2 - \hat{\sigma}_2\hat{\sigma}_1 \\ &\doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} - \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \\ &= 2i\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

The final matrix can be recognized as the representation of $\hat{\sigma}_3$, so that overall we have shown that

$$[\hat{\sigma}_1, \hat{\sigma}_2] = 2i\hat{\sigma}_3.$$

Cyclic permutation of the subscripts then gives the other two commutators.

Functions of Operators

If we have a function $f(x)$ which we can expand as a power series in x:

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots = \sum_{n=0}^{\infty} a_nx^n \tag{12.37}$$

then we define $f(\hat{A})$, a function of the operator $\hat{A}$, to be also given by the same power series, i.e.

$$f(\hat{A}) = a_0 + a_1\hat{A} + a_2\hat{A}^2 + \cdots = \sum_{n=0}^{\infty} a_n\hat{A}^n. \tag{12.38}$$

Once again, using the matrix representation of $\hat{A}$, it is possible, in certain cases, to work out what the matrix representation is of $f(\hat{A})$.

Ex 12.5 One of the most important functions of an operator that is encountered is the exponential function. To illustrate what this means, we will evaluate here the exponential function $\exp(i\phi\hat{\sigma}_3)$ where $\hat{\sigma}_1$ is one of the Pauli spin matrices introduced above, for which

$$\hat{\sigma}_1 \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and ϕ is a real number. Using the power series expansion of the exponential function, we have

$$e^{i\phi\hat{\sigma}_1} = \sum_{n=0}^{\infty} \frac{i\phi^n}{n!}\hat{\sigma}_1^n.$$

It is useful to note that

$$\hat{\sigma}_3^2 \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

i.e. $\hat{\sigma}_1^2 = \hat{1}$, the identity operator. Thus we can always write

$$\hat{\sigma}_1^{2n} = \hat{1} \qquad \hat{\sigma}_1^{2n+1} = \hat{\sigma}_1.$$

Thus, if we separate the infinite sum into two parts:

$$e^{i\phi\hat{\sigma}_1} = \sum_{n=0}^{\infty} \frac{(i\phi)^{2n}}{(2n)!}\hat{\sigma}_1^{2n} + \sum_{n=0}^{\infty} \frac{(i\phi)^{2n+1}}{(2n+1)!}\hat{\sigma}_1^{2n+1}$$

where the first sum is over all the even integers, and the second over all the odd integers, we get

$$\begin{aligned} e^{i\phi\hat{\sigma}_1} &= \hat{1}\sum_{n=0}^{\infty}(-1)^n\frac{\phi^{2n}}{(2n)!} + i\hat{\sigma}_1\sum_{n=0}^{\infty}(-1)^n\frac{\phi^{2n+1}}{(2n+1)!} \\ &= \cos\phi + i\hat{\sigma}_1\sin\phi \\ &\doteq \begin{pmatrix} \cos\phi & 0 \\ 0 & \cos\phi \end{pmatrix} + \begin{pmatrix} 0 & i\sin\phi \\ i\sin\phi & 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos\phi & i\sin\phi \\ i\sin\phi & \cos\phi \end{pmatrix}. \end{aligned}$$

Inverse of an Operator

Finding the inverse of an operator, given its matrix representation, amounts to finding the inverse of the matrix, provided, of course, that the matrix has an inverse.

Ex 12.6 The inverse of $\exp(i\phi\hat{\sigma}_1)$ can be found by taking the inverse of its representative matrix:

$$(e^{i\phi\hat{\sigma}_1})^{-1} \doteq \begin{pmatrix} \cos\phi & i\sin\phi \\ i\sin\phi & \cos\phi \end{pmatrix}^{-1} = \begin{pmatrix} \cos\phi & -i\sin\phi \\ -i\sin\phi & \cos\phi \end{pmatrix}.$$

This inverse can be recognized as being just

$$\begin{pmatrix} \cos\phi & -i\sin\phi \\ -i\sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \cos(-\phi) & i\sin(-\phi) \\ i\sin(-\phi) & \cos(-\phi) \end{pmatrix}$$

which means that

$$(e^{i\phi\hat{\sigma}_1})^{-1} = e^{-i\phi\hat{\sigma}_1}$$

a perhaps unsurprising result in that it is a particular case of the fact that the inverse of $\exp(\hat{A})$ is just $\exp(-\hat{A})$, exactly as is the case for the exponential function of a complex variable.

12.2.5 Eigenvectors and Eigenvalues

Operators act on states to map them into other states. Amongst the possible outcomes of the action of an operator on a state is to map the state into a multiple of itself:

$$\hat{A}|\phi\rangle = a_\phi|\phi\rangle \tag{12.39}$$

where $|\phi\rangle$ is, in general, a complex number. The state $|\phi\rangle$ is then said to be an eigenstate or eigenket of the operator $\hat{A}$ with a_ϕ the associated eigenvalue. The fact that operators can possess eigenstates might be thought of as a mathematical fact incidental to the physical content of quantum mechanics, but it turns out that the opposite is the case: the eigenstates and eigenvalues of various kinds of operators are essential parts of the physical interpretation of the quantum theory, and hence warrant close study. Notationally, is is often useful to use the eigenvalue associated with an eigenstate to label the eigenvector, i.e. the notation

$$\hat{A}|a\rangle = a|a\rangle. \tag{12.40}$$

This notation, or minor variations of it, will be used almost exclusively here.

Determining the eigenvalues and eigenvectors of a given operator $\hat{A}$, occasionally referred to as solving the eigenvalue problem for the operator, amounts to finding solutions to the eigenvalue equation $\hat{A}|\phi\rangle = a_\phi|\phi\rangle$. Written out in terms of the matrix representations of the operator with respect to some set of orthonormal basis vectors $\{|\varphi\rangle; n = 1, 2, \dots\}$, this eigenvalue equation is

$$\begin{pmatrix} A_{11} & A_{12} & \dots \\ A_{21} & A_{22} & \dots \\ \vdots & \vdots & \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \end{pmatrix} = a\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \end{pmatrix}. \tag{12.41}$$

This expression is equivalent to a set of simultaneous, homogeneous, linear equations:

$$\begin{pmatrix} A_{11}-a & A_{12} & \dots \\ A_{21} & A_{22}-a & \dots \\ \vdots & \vdots & \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \end{pmatrix} = 0 \tag{12.42}$$

which have to be solved for the possible values for a, and the associated values for the components $\phi_1, \phi_2, \dots$ of the eigenvectors. The procedure is standard. The determinant of coefficients must vanish in order to get non-trivial solutions for the components $\phi_1, \phi_2, \dots$:

$$\begin{vmatrix} A_{11}-a & A_{12} & \dots \\ A_{21} & A_{22}-a & \dots \\ \vdots & \vdots & \end{vmatrix} = 0 \tag{12.43}$$

which yields an equation known as the secular equation, or characteristic equation, that has to be solved to give the possible values of the eigenvalues a. Once these are known, they have to be resubstituted into Eq. (12.41) and the components ϕ_1, ϕ_2, ... of the associated eigenvectors determined. The details of how this is done properly belongs to a text on linear algebra and will not be considered any further here, except to say that the eigenvectors are typically determined up to an unknown multiplicative constant. This constant is usually fixed by the requirement that these eigenvectors be normalized to unity. In the case of repeated eigenvalues, i.e. when the characteristic polynomial has multiple roots (otherwise known as degenerate eigenvalues), the determination of the eigenvectors is made more complicated still. Once again, issues connected with these kinds of situations will not be considered here.

In general, for a state space of finite dimension, it is found that the operator $\hat{A}$ will have one or more discrete eigenvalues a_1, a_2, ... and associated eigenvectors $|a_1\rangle$, $|a_2\rangle$, ... The collection of all the eigenvalues of an operator is called the *eigenvalue spectrum* of the operator. Note also that more than one eigenvector can have the same eigenvalue. Such an eigenvalue is said to be *degenerate*.

For the present we will be confining our attention to operators that have discrete eigenvalue spectra. Modifications needed to handle continuous eigenvalues will be introduced later.

12.2.6 Hermitean Operators

Apart from certain calculational advantages, the representation of operators as matrices makes it possible to introduce in a direct fashion Hermitean operators, already considered in a more abstract way in Section 11.3.1, which have a central role to play in the physical interpretation of quantum mechanics.

To begin with, suppose we have an operator $\hat{A}$ with matrix elements A_{mn} with respect to a set of orthonormal basis states $\{|\varphi_n\rangle; n = 1, 2, \dots\}$. From the matrix representing this operator, we can construct a new operator by taking the transpose and complex conjugate of the original matrix:

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & \dots \\ A_{21} & A_{22} & A_{23} & \dots \\ A_{31} & A_{32} & A_{33} & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix} \longrightarrow \begin{pmatrix} A_{11}^* & A_{21}^* & A_{31}^* & \dots \\ A_{12}^* & A_{22}^* & A_{32}^* & \dots \\ A_{13}^* & A_{23}^* & A_{33}^* & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix}. \tag{12.44}$$

The new matrix will represent a new operator which is obviously related to $\hat{A}$, which we will call $\hat{A}^\dagger$, i.e.

$$\hat{A}^\dagger \doteq \begin{pmatrix} (A^\dagger)_{11} & (A^\dagger)_{12} & (A^\dagger)_{13} & \dots \\ (A^\dagger)_{21} & (A^\dagger)_{22} & (A^\dagger)_{23} & \dots \\ (A^\dagger)_{31} & (A^\dagger)_{32} & (A^\dagger)_{33} & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \begin{pmatrix} A_{11}^* & A_{21}^* & A_{31}^* & \dots \\ A_{12}^* & A_{22}^* & A_{32}^* & \dots \\ A_{13}^* & A_{23}^* & A_{33}^* & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix}. \tag{12.45}$$

i.e.

$$\langle\varphi_m|\hat{A}^\dagger|\varphi_n\rangle = (\langle\varphi_n|\hat{A}|\varphi_m\rangle)^*. \tag{12.46}$$

The new operator that we have created, $\hat{A}^\dagger$, can be recognized as the Hermitean adjoint of $\hat{A}$. The Hermitean adjoint has a useful property which we can most readily see if we use the matrix representation of the operator equation

$$\hat{A}|\psi\rangle = |\phi\rangle \tag{12.47}$$

which we earlier showed could be written as

$$\phi_m = \sum_n \langle\varphi_m|\hat{A}|\varphi_n\rangle\psi_n. \tag{12.48}$$

If we now take the complex conjugate of this expression we find

$$\phi_m^* = \sum_n \psi_n^* \langle \varphi_n | \hat{A}^\dagger | \varphi_m \rangle \tag{12.49}$$

which we can write in row vector form as

$$\begin{pmatrix} \phi_1^* & \phi_2^* & \dots \end{pmatrix} = \begin{pmatrix} \psi_1^* & \psi_2^* & \dots \end{pmatrix} \begin{pmatrix} (A^\dagger)_{11} & (A^\dagger)_{12} & \dots \\ (A^\dagger)_{21} & (A^\dagger)_{22} & \dots \\ \vdots & \vdots & \end{pmatrix} \tag{12.50}$$

which is the matrix version of

$$\langle \phi | = \langle \psi | \hat{A}^\dagger. \tag{12.51}$$

In other words we have shown that if $\hat{A}|\psi\rangle = |\phi\rangle$, then $\langle\psi|\hat{A}^\dagger = \langle\phi|$, a result that we used earlier to motivate the definition of the Hermitean adjoint in the first place. Thus, there are two ways to approach this concept: either through a general abstract argument, or in terms of matrix representations of an operator.

Ex 12.7 Consider the operator $\hat{A}$ which has the representation in some basis

$$\hat{A} \doteq \begin{pmatrix} 1 & i \\ 0 & -1 \end{pmatrix}.$$

Then

$$\hat{A}^\dagger \doteq \begin{pmatrix} 1 & 0 \\ -i & -1 \end{pmatrix}.$$

To be noticed in this example is that $\hat{A} \neq \hat{A}^\dagger$.

Ex 12.8 Now consider the operator

$$\hat{A} \doteq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Then

$$\hat{A}^\dagger \doteq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

i.e. $\hat{A} = \hat{A}^\dagger$.

This is an example of a situation in which $\hat{A}$ and $\hat{A}^\dagger$ are identical. In this case, the operator is said to be selfadjoint, or Hermitean. Once again, we encountered this idea in a more general context in Section 11.4.2, where it was remarked that Hermitean operators have a number of very important properties which leads to their playing a central role in the physical interpretation of quantum mechanics. These properties of Hermitean operators lead to the identification of Hermitean operators as representing the physically observable properties of a physical system, in a way that will be discussed in the next Chapter.

Chapter 13

Observables and Measurements in Quantum Mechanics

Till now, almost all attention has been focussed on discussing the *state* of a quantum system. As we have seen, this is most succinctly done by treating the package of information that defines a state as if it were a vector in an abstract Hilbert space. Doing so provides the mathematical machinery that is needed to capture the physically observed properties of quantum systems. A method by which the state space of a physical system can be set up was described in Section 8.4.2 wherein an essential step was to associate a set of basis states of the system with the exhaustive collection of results obtained when measuring some physical property, or observable, of the system. This linking of particular states with particular measured results provides a way that the observable properties of a quantum system can be described in quantum mechanics, that is in terms of Hermitean operators. It is the way in which this is done that is the main subject of this Chapter.

13.1 Measurements in Quantum Mechanics

One of the most difficult and controversial problems in quantum mechanics is the so-called measurement problem. Opinions on the significance of this problem vary widely. At one extreme the attitude is that there is in fact no problem at all, while at the other extreme the view is that the measurement problem is one of the great unsolved puzzles of quantum mechanics. The issue is that quantum mechanics only provides probabilities for the different possible outcomes in an experiment – it provides no mechanism by which the actual, finally observed result, comes about. Of course, probabilistic outcomes feature in many areas of classical physics as well, but in that case, probability enters the picture simply because there is insufficient information to make a definite prediction. In principle, that missing information is there to be found, it is just that accessing it may be a practical impossibility. In contrast, there is no 'missing information' for a quantum system, what we see is all that we can get, even in principle, though there are theories that say that this missing information resides in so-called 'hidden variables'. But in spite of these concerns about the measurement problem, there are some features of the measurement process that are commonly accepted as being essential parts of the final story. What is clear is that performing a measurement always involves a piece of equipment that

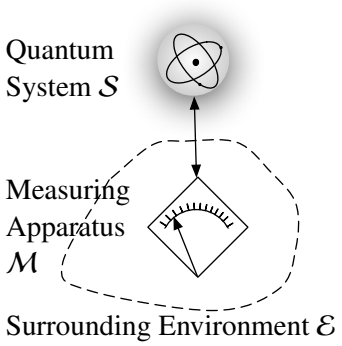


Figure 13.1: System $\mathcal{S}$ interacting with measuring apparatus $\mathcal{M}$ in the presence of the surrounding environment $\mathcal{E}$. The outcome of the measurement is registered on the dial on the measuring apparatus.

is macroscopic in size, and behaves according to the laws of classical physics. In Section 8.5, the process of decoherence was mentioned as playing a crucial role in giving rise to the observed classical behaviour of macroscopic systems, and so it is not surprising to find that decoherence plays an important role in the formulation of most modern theories of quantum measurement. Any quantum measurement then appears to require three components: the system, typically a microscopic system, whose properties are to be measured, the measuring apparatus itself, which interacts with the system under observation, and the environment surrounding the apparatus whose presence supplies the decoherence needed so that, 'for all practical purposes (FAPP)', the apparatus behaves like a classical system, whose output can be, for instance, a pointer on the dial on the measuring apparatus coming to rest, pointing at the final result of the measurement, that is, a number on the dial. Of course, the apparatus could produce an electrical signal registered on an oscilloscope, or bit of data stored in a computer memory, or a flash of light seen by the experimenter as an atom strikes a fluorescent screen, but it is often convenient to use the simple picture of a pointer.

The experimental apparatus would be designed according to what physical property it is of the quantum system that is to be measured. Thus, if the system were a single particle, the apparatus could be designed to measure its energy, or its position, or its momentum or its spin, or some other property. These measurable properties are known as *observables*, a concept that we have already encountered in Section 8.4.1. But how do we know what it is that a particular experimental setup would be measuring? The design would be ultimately based on classical physics principles, i.e., if the apparatus were intended to measure the energy of a quantum system, then it would also measure the energy of a classical system if a classical system were substituted for the quantum system. In this way, the macroscopic concepts of classical physics can be transferred to quantum systems. We will not be examining the details of the measurement process in any great depth here. Rather, we will be more concerned with some of the general characteristics of the outputs of a measurement procedure and how these general features can be incorporated into the mathematical formulation of the quantum theory.

13.2 Observables and Hermitean Operators

So far we have consistently made use of the idea that if we know something definite about the state of a physical system, say that we know the z component of the spin of a spin half particle is $S_z = \frac{1}{2}\hbar$, then we assign to the system the state $|S_z = \frac{1}{2}\hbar\rangle$, or, more simply, $|+\rangle$. It is at this point that we need to look a little more closely at this idea, as it will lead us to associating an operator with the physical concept of an observable. Recall that an observable is, roughly speaking, any measurable property of a physical system: position, spin, energy, momentum .... Thus, we talk about the position x of a particle as an observable for the particle, or the z component of spin, S_z as a further observable and so on.

When we say that we 'know' the value of some physical observable of a quantum system, we are presumably implying that some kind of measurement has been made that provided us with this knowledge. It is furthermore assumed that in the process of acquiring this knowledge, the system, after the measurement has been performed, survives the measurement, and moreover if we were to immediately remeasure the same quantity, we would get the same result. This is certainly the situation with the measurement of spin in a Stern-Gerlach experiment. If an atom emerges from one such set of apparatus in a beam that indicates that $S_z = \frac{1}{2}\hbar$ for that atom, and we were to pass the atom through a second apparatus, also with its magnetic field oriented in the z direction, we would find the atom emerging in the $S_z = \frac{1}{2}\hbar$ beam once again. Under such circumstances, we would be justified in saying that the atom has been prepared in the state $|S_z = \frac{1}{2}\hbar\rangle$. However, the reality is that few measurements are of this kind, i.e. the system being subject to measurement is physically modified, if not destroyed, by the measurement process. An extreme example is a measurement designed to count the number of photons in a single mode

cavity field. Photons are typically counted by photodetectors whose mode of operation is to absorb a photon and create a pulse of current. So we may well be able to count the number of photons in the field, but in doing so, there is no field left behind after the counting is completed. All that we can conclude, regarding the state of the cavity field, is that it is left in the vacuum state $|0\rangle$ after the measurement is completed, but we can say nothing for certain about the state of the field before the measurement was undertaken. However, all is not lost. If we fiddle around with the process by which we put photons in the cavity in the first place, it will hopefully be the case that amongst all the experimental procedures that could be followed, there are some that result in the cavity field being in a state for which every time we then measure the number of photons in the cavity, we always get the result n. It is then not unreasonable to claim that the experimental procedure has prepared the cavity field in a state which the number of photons in the cavity is n, and we can assign the state $|n\rangle$ to the cavity field.

This procedure can be equally well applied to the spin half example above. The preparation procedure here consists of putting atoms through a Stern-Gerlach apparatus with the field oriented in the z direction, and picking out those atoms that emerge in the beam for which $S_z = \frac{1}{2}\hbar$. This has the result of preparing the atom in a state for which the z component of spin would always be measured to have the value $\frac{1}{2}\hbar$. Accordingly, the state of the system is identified as $|S_z = \frac{1}{2}\hbar\rangle$, i.e. $|+\rangle$. In a similar way, we can associate the state $|-\rangle$ with the atom being in a state for which the z component of spin is always measured to be $-\frac{1}{2}\hbar$. We can also note that these two states are mutually exclusive, i.e. if in the state $|+\rangle$, then the result $S_z = -\frac{1}{2}\hbar$ is never observed, and furthermore, we note that the two states cover all possible values for S_z. Finally, the fact that observation of the behaviour of atomic spin show evidence of both randomness and interference lead us to conclude that if an atom is prepared in an arbitrary initial state $|S\rangle$, then the probability amplitude of finding it in some other state $|S'\rangle$ is given by

$$\langle S'|S\rangle = \langle S'|+\rangle\langle +|S\rangle + \langle S'|-\rangle\langle -|S\rangle$$

which leads, by the cancellation trick to

$$|S\rangle = |+\rangle\langle +|S\rangle + |-\rangle\langle -|S\rangle$$

which tells us that *any* spin state of the atom is to be interpreted as a vector expressed as a linear combination of the states $|\pm\rangle$. The states $|\pm\rangle$ constitute a complete set of orthonormal basis states for the state space of the system. We therefore have at hand just the situation that applies to the eigenstates and eigenvectors of a Hermitean operator as summarized in the following table:

Properties of a Hermitean Operator	Properties of Observable S_z
The eigenvalues of a Hermitean operator are all real.	Value of observable S_z measured to be real numbers $\pm\frac{1}{2}\hbar$.
Eigenvectors belonging to different eigenvalues are orthogonal.	States $\|\pm\rangle$ associated with different values of the observable are mutually exclusive.
The eigenstates form a complete set of basis states for the state space of the system.	The states $\|\pm\rangle$ associated with all the possible values of observable S_z form a complete set of basis states for the state space of the system.

It is therefore natural to associate with the observable S_z, a Hermitean operator which we will write as $\hat{S}_z$ such that $\hat{S}_z$ has eigenstates $|\pm\rangle$ and associate eigenvalues $\pm\frac{1}{2}\hbar$, i.e.

$$\hat{S}_z|\pm\rangle = \pm\tfrac{1}{2}\hbar|\pm\rangle \tag{13.1}$$

so that, in the $\{|-\rangle, |+\rangle\}$ basis

$$\hat{S}_z \doteq \begin{pmatrix} \langle +|\hat{S}_z|+\rangle & \langle +|\hat{S}_z|-\rangle \\ \langle -|\hat{S}_z|+\rangle & \langle -|\hat{S}_z|-\rangle \end{pmatrix} \tag{13.2}$$

$$= \tfrac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{13.3}$$

So, in this way, we actually *construct* a Hermitean operator to represent a particular measurable property of a physical system.

The term 'observable', while originally applied to the physical quantity of interest, is also applied to the associated Hermitean operator. Thus we talk, for instance, about the observable $\hat{S}_z$. To a certain extent we have used the mathematical construct of a Hermitean operator to draw together in a compact fashion ideas that we have been freely using in previous Chapters.

It is useful to note the distinction between a quantum mechanical observable and the corresponding classical quantity. The latter quantity, say the position x of a particle, represents a single possible value for that observable – though it might not be known, it in principle has a definite, single value at any instant in time. In contrast, a quantum observable such as S_z is an operator which, through its eigenvalues, carries with it *all* the values that the corresponding physical quantity could possibly have. In a certain sense, this is a reflection of the physical state of affairs that pertains to quantum systems, namely that when a measurement is made of a particular physical property of a quantum systems, the outcome can, in principle, be *any* of the possible values that can be associated with the observable, even if the experiment is repeated under identical conditions.

This procedure of associating a Hermitean operator with every observable property of a quantum system can be readily generalized. The generalization takes a slightly different form if the observable has a continuous range of possible values, such as position and momentum, as against an observable with only discrete possible results. We will consider the discrete case first.

13.3 Observables with Discrete Values

The discussion presented in the preceding Section can be generalized into a collection of postulates that are intended to describe the concept of an observable. So, to begin, suppose, through an exhaustive series of measurements, we find that a particular observable, call it Q, of a physical system, is found to have the values — all real numbers — $q_1, q_2, \ldots$. Alternatively, we may have sound theoretical arguments that inform us as to what the possible values could be. For instance, we might be interested in the position of a particle free to move in one dimension, in which case the observable Q is just the position of the particle, which would be expected to have any value in the range $-\infty$ to $+\infty$. We now introduce the states $|q_1\rangle$, $|q_2\rangle$, ... these being states for which the observable Q definitely has the value $q_1, q_2, \ldots$ respectively. In other words, these are the states for which, if we were to measure Q, we would be guaranteed to get the results $q_1, q_2, \ldots$ respectively. We now have an interesting state of affairs summarized below.

1. We have an observable Q which, when measured, is found to have the values $q_1, q_2, \ldots$ that are all real numbers.

2. For each possible value of Q the system can be prepared in a corresponding state $|q_1\rangle$, $|q_2\rangle$, ... for which the values $q_1, q_2, \ldots$ will be obtained with certainty in any measurement of Q.

At this stage we are still not necessarily dealing with a quantum system. We therefore assume that this system exhibits the properties of intrinsic randomness and interference that characterizes quantum systems, and which allows the state of the system to be identified as vectors belonging to the state space of the system. This leads to the next property:

3. If prepared in this state $|q_n\rangle$, and we measure Q, we only ever get the result q_n, i.e. we never observe the result q_m with $q_m \neq q_n$. Thus we conclude $\langle q_n|q_m\rangle = \delta_{mn}$. The states $\{|q_n\rangle; n = 1, 2, 3, \ldots\}$ are orthonormal.

4. The states $|q_1\rangle$, $|q_2\rangle$, ... cover all the possibilities for the system and so these states form a complete set of orthonormal basis states for the state space of the system.

That the states form a complete set of basis states means that any state $|\psi\rangle$ of the system can be expressed as

$$|\psi\rangle = \sum_n c_n|q_n\rangle \tag{13.4}$$

while orthonormality means that $\langle q_n|q_m\rangle = \delta_{nm}$ from which follows $c_n = \langle q_n|\psi\rangle$. The completeness condition can then be written as

$$\sum_n |q_n\rangle\langle q_n| = \hat{1} \tag{13.5}$$

5. For the system in state $|\psi\rangle$, the probability of obtaining the result q_n on measuring Q is $|\langle q_n|\psi\rangle|^2$ provided $\langle\psi|\psi\rangle = 1$.

The completeness of the states $|q_1\rangle$, $|q_2\rangle$, ... means that there is *no* state $|\psi\rangle$ of the system for which $\langle q_n|\psi\rangle = 0$ for *every* state $|q_n\rangle$. In other words, we must have

$$\sum_n |\langle q_n|\psi\rangle|^2 \neq 0. \tag{13.6}$$

Thus there is a non-zero probability for at least *one* of the results q_1, q_2, ... to be observed – if a measurement is made of Q, a result has to be obtained!

6. The observable Q is represented by a Hermitean operator $\hat{Q}$ whose eigenvalues are the possible results q_1, q_2, ... of a measurement of Q, and the associated eigenstates are the states $|q_1\rangle$, $|q_2\rangle$, ..., i.e. $\hat{Q}|q_n\rangle = q_n|q_n\rangle$. The name 'observable' is often applied to the operator $\hat{Q}$ itself.

The spectral decomposition of the observable $\hat{Q}$ is then

$$\hat{Q} = \sum_n q_n|q_n\rangle\langle q_n|. \tag{13.7}$$

Apart from anything else, the eigenvectors of an observable constitute a set of basis states for the state space of the associated quantum system.

For state spaces of finite dimension, the eigenvalues of any Hermitean operator are discrete, and the eigenvectors form a complete set of basis states. For state spaces of infinite dimension, it is possible for a Hermitean operator not to have a complete set of eigenvectors, so that it is possible for a system to be in a state which cannot be represented as a linear combination of the eigenstates of such an operator. In this case, the operator cannot be understood as being an observable as it would appear to be the case that the system could be placed in a state for which a measurement of the associated observable yielded no value! To put it another way, if a Hermitean operator could be constructed whose eigenstates did not form a complete set, then we can rightfully claim that such an operator cannot represent an observable property of the system.

It should also be pointed out that it is quite possible to construct all manner of Hermitean operators to be associated with any given physical system. Such operators would have all the mathematical

properties to be associated with their being Hermitean, but it is not necessarily the case that these represent either any readily identifiable physical feature of the system at least in part because it might not be at all appparent how such 'observables' could be measured. The same is at least partially true classically — the quantity px^2 where p is the momentum and x the position of a particle does not immediately suggest a useful, familiar or fundamental property of a single particle.

13.3.1 The Von Neumann Measurement Postulate

Finally, we add a further postulate concerning the state of the system immediately after a measurement is made. This is the von Neumann projection postulate:

7. If on measuring Q for a system in state $|\psi\rangle$, a result q_n is obtained, then the state of the system immediately after the measurement is $|q_n\rangle$.

   This postulate can be rewritten in a different way by making use of the projection operators introduced in Section 11.1.3. Thus, if we write

   $$\hat{P}_n = |q_n\rangle\langle q_n| \tag{13.8}$$

   then the state of the system after the measurement, for which the result q_n was obtained, is

   $$\frac{\hat{P}_n|\psi\rangle}{\sqrt{\langle\psi|\hat{P}_n|\psi\rangle}} = \frac{\hat{P}_n|\psi\rangle}{\sqrt{|\langle q_n|\psi\rangle|^2}} \tag{13.9}$$

   where the term in the denominator is there to guarantee that the state after the measurement is normalized to unity.

This postulate is almost stating the obvious in that we *name* a state according to the information that we obtain about it as a result of a measurement. But it can also be argued that if, after performing a measurement that yields a particular result, we immediately repeat the measurement, it is reasonable to expect that there is a 100% chance that the same result be regained, which tells us that the system must have been in the associated eigenstate. This was, in fact, the main argument given by von Neumann to support this postulate. Thus, von Neumann argued that the fact that the value has a stable result upon repeated measurement indicates that the system really has that value after measurement.

This postulate regarding the effects of measurement has always been a source of discussion and disagreement. This postulate is satisfactory in that it is consistent with the manner in which the idea of an observable was introduced above, but it is not totally clear that it is a postulate that can be applied to all measurement processes. The kind of measurements wherein this postulate is satisfactory are those for which the system 'survives' the measuring process, which is certainly the case in the Stern-Gerlach experiments considered here. But this is not at all what is usually encountered in practice. For instance, measuring the number of photons in an electromagnetic field inevitably involves detecting the photons by absorbing them, i.e. the photons are destroyed. Thus we may find that if n photons are absorbed, then we can say that there *were* n photons in the cavity, i.e. the photon field *was* in state $|n\rangle$, but after the measuring process is over, it is in the state $|0\rangle$. To cope with this fairly typical state of affairs it is necessary to generalize the notion of measurement to allow for this – so-called generalized measurement theory. But even here, it is found that the generalized measurement process being described can be understood as a von Neumann-type projection made on a larger system of which the system of interest is a part. This larger system could include, for instance, the measuring apparatus itself, so that instead of making a projective measurement on the system itself, one is made on the measuring apparatus. We will not be discussing these aspects of measurement theory here.

13.4 The Collapse of the State Vector

The von Neumann postulate is quite clearly stating that as a consequence of a measurement, the state of the system undergoes a discontinuous change of state, i.e. $|\psi\rangle \rightarrow |q_n\rangle$ if the result q_n is obtained on performing a measurement of an observable Q. This instantaneous change in state is known as 'the collapse of the state vector'. This conjures up the impression that the process of measurement necessarily involves a physical interaction with the system, that moreover, results in a major physical disruption of the state of a system – one moment the system is in a state $|\psi\rangle$, the next it is forced into a state $|q_n\rangle$. However, if we return to the quantum eraser example considered in Section 4.3.2 we see that there need not be any actual physical interaction with a system at all in order to obtain information about it. The picture that emerges is that the change of state is nothing more benign than being an updating, through observation, of the knowledge we have of the state of a system as a consequence of the outcome of a measurement. While obtaining this information must necessarily involve some kind of physical interaction involving a measuring apparatus, this interaction may or may not be associated with any physical disruption to the system of interest itself. This emphasizes the notion that quantum states are as much states of knowledge as they are physical states.

13.4.1 Sequences of measurements

Having on hand a prescription by which we can specify the state of a system after a measurement has been performed makes it possible to study the outcome of alternating measurements of different observables being performed on a system. We have already seen an indication of the sort of result to be found in the study of the measurement of different spin components of a spin half particle in Section 6.4.3. For instance, if a measurement of, say, S_x is made, giving a result $\frac{1}{2}\hbar$, and then S_z is measured, giving, say, $\frac{1}{2}\hbar$, and then S_x remeasured, there is an equal chance that either of the results $\pm\frac{1}{2}\hbar$ will be obtained, i.e. the formerly precisely known value of S_x is 'randomly scrambled' by the intervening measurement of S_z. The two observables are said to be incompatible: it is not possible to have exact knowledge of both S_x and S_z at the same time. This behaviour was presented in Section 6.4.3 as an experimentally observed fact, but we can now see how this kind of behaviour comes about within the formalism of the theory.

If we let $\hat{S}_x$ and $\hat{S}_z$ be the associated Hermitean operators, we can analyze the above observed behaviour as follows. The first measurement of S_x, which yielded the outcome $\frac{1}{2}\hbar$, results in the spin half system ending up in the state $|S_x = \frac{1}{2}\hbar\rangle$, an eigenstate of $\hat{S}_x$ with eigenvalue $\frac{1}{2}\hbar$. The second measurement, of S_z, results in the system ending up in the state $|S_z = \frac{1}{2}\hbar\rangle$, the eigenstate of $\hat{S}_z$ with eigenvalue $\frac{1}{2}\hbar$. However, this latter state cannot be an eigenstate of $\hat{S}_x$. If it were, we would not get the observed outcome, that is, on the remeasurement of S_x, we would not get a random scatter of results (i.e. the two results $S_x = \pm\frac{1}{2}\hbar$ occurring randomly but equally likely). In the same way we can conclude that $|S_x = -\frac{1}{2}\hbar\rangle$ is also not an eigenstate of S_z, and likewise, the eigenstates $|S_z = \pm\frac{1}{2}\hbar\rangle$ of $\hat{S}_z$ cannot be eigenstates of $\hat{S}_x$. Thus we see that the two incompatible observables S_x and S_z do not share the same eigenstates.

There is a more succinct way by which to determine whether two observables are incompatible or not. This involves making use of the concept of the commutator of two operators, $\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A}$ as discussed in Section 11.1.3. To this end, consider the commutator $\left[\hat{S}_x, \hat{S}_z\right]$ and let it act on the eigenstate $|S_z = \frac{1}{2}\hbar\rangle = |+\rangle$:

$$\left[\hat{S}_x, \hat{S}_z\right]|+\rangle = \left(\hat{S}_x\hat{S}_z - \hat{S}_z\hat{S}_x\right)|+\rangle = \hat{S}_x\left(\tfrac{1}{2}\hbar|+\rangle\right) - \hat{S}_z\left(\hat{S}_x|+\rangle\right) = \left(\tfrac{1}{2}\hbar - \hat{S}_z\right)\left(\hat{S}_x|+\rangle\right). \tag{13.10}$$

Now let $\hat{S}_x|+\rangle = |\psi\rangle$. Then we see that in order for this expression to vanish, we must have

$$\hat{S}_z|\psi\rangle = \tfrac{1}{2}\hbar|\psi\rangle. \tag{13.11}$$

In other words, $|\psi\rangle$ would have to be the eigenstate of $\hat{S}_z$ with eigenvalue $\frac{1}{2}\hbar$, i.e. $|\psi\rangle \propto |+\rangle$, or

$$\hat{S}_x|+\rangle = \text{constant} \times |+\rangle \tag{13.12}$$

But we have just pointed out that this cannot be the case, so the expression $\left[\hat{S}_x, \hat{S}_z\right]|+\rangle$ cannot be zero, i.e. we must have

$$\left[\hat{S}_x, \hat{S}_z\right] \neq 0. \tag{13.13}$$

Thus, the operators $\hat{S}_x$ and $\hat{S}_z$ do not commute.

The commutator of two observables serves as a means by which it can be determined whether or not the observables are compatible. If they do not commute, then they are incompatible: the measurement of one of the observables will randomly scramble any preceding result known for the other. In contrast, if they do commute, then it is possible to know precisely the value of both observables at the same time. An illustration of this is given later in this Chapter (Section 13.5.4), while the importance of compatibility is examined in more detail later in Chapter 14.

13.5 Examples of Discrete Valued Observables

There are many observables of importance for all manner of quantum systems. Below, some of the important observables for a single particle system are described. As the eigenstates of any observable constitutes a set of basis states for the state space of the system, these basis states can be used to set up representations of the state vectors and operators as column vectors and matrices. These representations are named according to the observable which defines the basis states used. Moreover, since there are in general many observables associated with a system, there are correspondingly many possible basis states that can be so constructed. Of course, there are an infinite number of possible choices for the basis states of a vector space, but what this procedure does is pick out those basis states which are of most immediate physical significance.

The different possible representations are useful in different kinds of problems, as discussed briefly below. It is to be noted that the term 'observable' is used both to describe the physical quantity being measured as well as the operator itself that corresponds to the physical quantity.

13.5.1 Position of a particle (in one dimension)

The position x of a particle is a continuous quantity, with values ranging over the real numbers, and a proper treatment of such an observable raises mathematical and interpretational issues that are dealt with elsewhere. But for the present, it is very convenient to introduce an 'approximate' position operator via models of quantum systems in which the particle, typically an electron, can only be found to be at certain discrete positions. The simplest example of this is the O_2^- ion discussed in Section 8.4.2.

This system can be found in two states $|\pm a\rangle$, where $\pm a$ are the positions of the electron on one or the other of the oxygen atoms. Thus these states form a pair of basis states for the state space of the system, which hence has a dimension 2. The position operator $\hat{x}$ of the electron is such that

$$\hat{x}|\pm a\rangle = \pm a|\pm a\rangle \tag{13.14}$$

which can be written in the position representation as a matrix:

$$\hat{x} \doteq \begin{pmatrix} \langle +a|\hat{x}|+a\rangle & \langle +a|\hat{x}|-a\rangle \\ \langle -a|\hat{x}|+a\rangle & \langle -a|\hat{x}|-a\rangle \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}. \tag{13.15}$$

The completeness relation for these basis states reads

$$|+a\rangle\langle +a| + |-a\rangle\langle -a| = \hat{1}. \tag{13.16}$$

which leads to

$$\hat{x} = a|+a\rangle\langle+a| - a|-a\rangle\langle-a|. \tag{13.17}$$

The state space of the system has been established as having dimension 2, so any other observable of the system can be represented as a 2×2 matrix. We can use this to construct the possible forms of other observables for this system, such as the momentum operator and the Hamiltonian.

This approach can be readily generalized to e.g. a CO_2^- ion, in which case there are three possible positions for the electron, say $x = \pm a, 0$ where $\pm a$ are the positions of the electron when on the oxygen atoms and 0 is the position of the electron when on the carbon atom. The position operator $\hat{x}$ will then be such that

$$\hat{x}|\pm a\rangle = \pm a|\pm a\rangle \quad \hat{x}|0\rangle = 0|0\rangle. \tag{13.18}$$

13.5.2 Momentum of a particle (in one dimension)

As is the case of position, the momentum of a particle can have a continuous range of values, which raises certain mathematical issues that are discussed later. But we can consider the notion of momentum for our approximate position models in which the position can take only discrete values. We do this through the observation that the matrix representing the momentum will be a $N\times N$ matrix, where N is the dimension of the state space of the system. Thus, for the O_2^- ion, the momentum operator would be represented by a two by two matrix

$$\hat{p} \doteq \begin{pmatrix} \langle+a|\hat{p}|+a\rangle & \langle+a|\hat{p}|-a\rangle \\ \langle-a|\hat{p}|+a\rangle & \langle-a|\hat{p}|-a\rangle \end{pmatrix} \tag{13.19}$$

though, at this stage, it is not obvious what values can be assigned to the matrix elements appearing here. Nevertheless, we can see that, as $\hat{p}$ must be a Hermitean operator, and as this is a 2×2 matrix, $\hat{p}$ will have only 2 real eigenvalues: the momentum of the electron can be measured to have only two possible values, at least within the constraints of the model we are using.

13.5.3 Energy of a Particle (in one dimension)

According to classical physics, the energy of a particle is given by

$$E = \frac{p^2}{2m} + V(x) \tag{13.20}$$

where the first term on the RHS is the kinetic energy, and the second term $V(x)$ is the potential energy of the particle. In quantum mechanics, it can be shown, by a procedure known as canonical quantization, that the energy of a particle is represented by a Hermitean operator known as the Hamiltonian, written $\hat{H}$, which can be expressed as

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x}) \tag{13.21}$$

where the classical quantities p and x have been replaced by the corresponding quantum operators. The term Hamiltonian is derived from the name of the mathematician Rowan Hamilton who made profoundly significant contributions to the theory of mechanics. Although the Hamiltonian can be identified here as being the total energy E, the term Hamiltonian is usually applied in mechanics if this total energy is expressed in terms of momentum and position variables, as here, as against say position and velocity.

That Eq. (13.21) is 'quantum mechanical' is not totally apparent. Dressing the variables up as operators by putting hats on them is not really saying that much. Perhaps most significantly there is no $\hbar$ in this expression for $\hat{H}$, so it is not obvious how this expression can be 'quantum mechanical'.

For instance, we have seen, at least for a particle in an infinite potential well (see Section 5.3), that the energies of a particle depend on $\hbar$. The quantum mechanics (and the $\hbar$) is to be found in the properties of the operators so created that distinguish them from the classical variables that they replace. Specifically, the two operators $\hat{x}$ and $\hat{p}$ do not commute, in fact, $[\hat{x}, \hat{p}] = i\hbar$ as shown later, Eq. (13.133), and it is this failure to commute by an amount proportional to $\hbar$ that injects 'quantum mechanics' into the operator associated with the energy of a particle.

As the eigenvalues of the Hamiltonian are the possible energies of the particle, the eigenvalue is usually written E and the eigenvalue equation is

$$\hat{H}|E\rangle = E|E\rangle. \tag{13.22}$$

In the position representation, this equation becomes

$$\langle x|\hat{H}|E\rangle = E\langle x|E\rangle. \tag{13.23}$$

It is shown later that, from the expression Eq. (13.21) for $\hat{H}$, that this eigenvalue equation can be written as a differential equation for the wave function $\psi_E(x) = \langle x|E\rangle$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_E(x)}{dx^2} + V(x)\psi_E(x) = E\psi_E(x). \tag{13.24}$$

This is just the time independent Schrödinger equation.

Depending on the form of $V(\hat{x})$, this equation will have different possible solutions, and the Hamiltonian will have various possible eigenvalues. For instance, if $V(\hat{x}) = 0$ for $0 < x < L$ and is infinite otherwise, then we have the Hamiltonian of a particle in an infinitely deep potential well, or equivalently, a particle in a (one-dimensional) box with impenetrable walls. This problem was dealt with in Section 5.3 using the methods of wave mechanics, where it was found that the energy of the particle was limited to the values

$$E_n = \frac{n^2\hbar^2\pi^2}{2mL^2}, \qquad n = 1, 2, \ldots.$$

Thus, in this case, the Hamiltonian has discrete eigenvalues E_n given by Eq. (13.5.3). If we write the associated energy eigenstates as $|E_n\rangle$, the eigenvalue equation is then

$$\hat{H}|E_n\rangle = E_n|E_n\rangle. \tag{13.25}$$

The wave function $\langle x|E_n\rangle$ associated with the energy eigenstate $|E_n\rangle$ was also derived in Section 5.3 and is given by

$$\begin{aligned}\psi_n(x) = \langle x|E_n\rangle &= \sqrt{\frac{2}{L}}\sin(n\pi x/L) \quad 0 < x < L \\ &= 0 \qquad\qquad\qquad\qquad x < 0, \quad x > L.\end{aligned} \tag{13.26}$$

Another example is that for which $V(\hat{x}) = \frac{1}{2}k\hat{x}^2$, i.e. the simple harmonic oscillator potential. In this case, we find that the eigenvalues of $\hat{H}$ are

$$E_n = (n + \tfrac{1}{2})\hbar\omega, \qquad n = 0, 1, 2, \ldots \tag{13.27}$$

where $\omega = \sqrt{k/m}$ is the natural frequency of the oscillator. The Hamiltonian is an observable of particular importance in quantum mechanics. As will be discussed in the next Chapter, it is the Hamiltonian which determines how a system evolves in time, i.e. the equation of motion of a quantum system is expressly written in terms of the Hamiltonian. In the position representation, this equation is just the time dependent Schrödinger equation.

The Energy Representation If the state of the particle is represented in component form with respect to the energy eigenstates as basis states, then this is said to be the *energy representation*. In contrast to the position and momentum representations, the components are often discrete. The energy representation is useful when the system under study can be found in states with different energies, e.g. an atom absorbing or emitting photons, and consequently making transitions to higher or lower energy states. The energy representation is also very important when it is the evolution in time of a system that is of interest.

13.5.4 The O_2^- Ion: An Example of a Two-State System

In order to illustrate the ideas developed in the preceding sections, we will see how it is possible, firstly, how to 'construct' the Hamiltonian of a simple system using simple arguments, then to look at the consequences of performing measurements of two observables for this system.

Constructing the Hamiltonian

The Hamiltonian of the ion in the position representation will be

$$\hat{H} \doteq \begin{pmatrix} \langle +a|\hat{H}|+a\rangle & \langle +a|\hat{H}|-a\rangle \\ \langle -a|\hat{H}|+a\rangle & \langle -a|\hat{H}|-a\rangle \end{pmatrix}. \tag{13.28}$$

Since there is perfect symmetry between the two oxygen atoms, we must conclude that the diagonal elements of this matrix must be equal i.e.

$$\langle +a|\hat{H}|+a\rangle = \langle -a|\hat{H}|-a\rangle = E_0. \tag{13.29}$$

We further know that the Hamiltonian must be Hermitean, so the off-diagonal elements are complex conjugates of each other. Hence we have

$$\hat{H} \doteq \begin{pmatrix} E_0 & V \\ V^* & E_0 \end{pmatrix} \tag{13.30}$$

or, equivalently

$$\hat{H} = E_0|+a\rangle\langle +a| + V|+a\rangle\langle -a| + V^*|-a\rangle\langle +a| + E_0|-a\rangle\langle -a|. \tag{13.31}$$

Rather remarkably, we have at hand the Hamiltonian for the system with the barest of physical information about the system.

In the following we shall assume $V = -A$ and that A is a real number so that the Hamiltonian matrix becomes

$$\hat{H} \doteq \begin{pmatrix} E_0 & -A \\ -A & E_0 \end{pmatrix} \tag{13.32}$$

The physical content of the results are not changed by doing this, and the results are a little easier to write down. First we can determine the eigenvalues of $\hat{H}$ by the usual method. If we write $\hat{H}|E\rangle = E|E\rangle$, and put $|E\rangle = \alpha|+a\rangle + \beta|-a\rangle$, this becomes, in matrix form

$$\begin{pmatrix} E_0 - E & -A \\ -A & E_0 - E \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0. \tag{13.33}$$

The characteristic equation yielding the eigenvalues is then

$$\begin{vmatrix} E_0 - E & -A \\ -A & E_0 - E \end{vmatrix} = 0. \tag{13.34}$$

Expanding the determinant this becomes

$$(E_0 - E)^2 - A^2 = 0 \tag{13.35}$$

with solutions

$$E_1 = E_0 + A \qquad E_2 = E_0 - A. \tag{13.36}$$

Substituting each of these two values back into the original eigenvalue equation then gives the equations for the eigenstates. We find that

$$|E_1\rangle = \frac{1}{\sqrt{2}}(|+a\rangle - |-a\rangle) \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} \tag{13.37}$$

$$|E_2\rangle = \frac{1}{\sqrt{2}}(|+a\rangle + |-a\rangle) \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \tag{13.38}$$

where each eigenvector has been normalized to unity. Thus we have constructed the eigenstates and eigenvalues of the Hamiltonian of this system. We can therefore write the Hamiltonian as

$$\hat{H} = E_1|E_1\rangle\langle E_1| + E_2|E_2\rangle\langle E_2| \tag{13.39}$$

which is just the spectral decomposition of the Hamiltonian.

We have available two useful sets of basis states: the basis states for the position representation $\{|+a\rangle, |-a\rangle\}$ and the basis states for the energy representation, $\{|E_1\rangle, |E_2\rangle\}$. Any state of the system can be expressed as linear combinations of either of these sets of basis states.

Measurements of Energy and Position

Suppose we prepare the O_2^- ion in the state

$$\begin{aligned}|\psi\rangle &= \frac{1}{5}[3|+a\rangle + 4|-a\rangle] \\ &\doteq \frac{1}{5}\begin{pmatrix}3\\4\end{pmatrix}\end{aligned} \tag{13.40}$$

and we measure the energy of the ion. We can get two possible results of this measurement: E_1 or E_2. We will get the result E_1 with probability $|\langle E_1|\psi\rangle|^2$, i.e.

$$\langle E_1|\psi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\end{pmatrix} \cdot \frac{1}{5}\begin{pmatrix}3\\4\end{pmatrix} = -\frac{1}{5\sqrt{2}} \tag{13.41}$$

so that

$$|\langle E_1|\psi\rangle|^2 = 0.02 \tag{13.42}$$

and similarly

$$\langle E_2|\psi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\end{pmatrix} \cdot \frac{1}{5}\begin{pmatrix}3\\4\end{pmatrix} = \frac{7}{5\sqrt{2}} \tag{13.43}$$

so that

$$|\langle E_2|\psi\rangle|^2 = 0.98. \tag{13.44}$$

It is important to note that if we get the result E_1, then according to the von Neumann postulate, the system ends up in the state $|E_1\rangle$, whereas if we got the result E_2, then the new state is $|E_2\rangle$.

Of course we could have measured the position of the electron, with the two possible outcomes $\pm a$. In fact, the result $+a$ will occur with probability

$$|\langle +a|\psi\rangle|^2 = 0.36 \tag{13.45}$$

and the result $-a$ with probability

$$|\langle -a|\psi\rangle|^2 = 0.64. \tag{13.46}$$

Once again, if the outcome is $+a$ then the state of the system after the measurement is $|+a\rangle$, and if the result $-a$ is obtained, then the state after the measurement is $|-a\rangle$.

Finally, we can consider what happens if we were to do a sequence of measurements, first of energy, then position, and then energy again. Suppose the system is initially in the state $|\psi\rangle$, as above, and the measurement of energy gives the result E_1. The system is now in the state $|E_1\rangle$. If we now perform a measurement of the position of the electron, we can get either of the two results $\pm a$ with equal probability:

$$|\langle \pm a|E_1\rangle|^2 = 0.5. \tag{13.47}$$

Suppose we get the result $+a$, so the system is now in the state $|+a\rangle$ and we remeasure the energy. We find that now it is not guaranteed that we will regain the result E_1 obtained in the first measurement. In fact, we find that there is an equal chance of getting either E_1 or E_2:

$$|\langle E_1|+a\rangle|^2 = |\langle E_2|+a\rangle|^2 = 0.5. \tag{13.48}$$

Thus we must conclude that the intervening measurement of the position of the electron has scrambled the energy of the system. In fact, if we suppose that we get the result E_2 for this second energy measurement, thereby placing the system in the state $|E_2\rangle$ and we measure the position of the electron again, we find that we will get either result $\pm a$ with equal probability again! The measurement of energy and electron position for this system clearly interfere with one another. It is not possible to have a precisely defined value for *both* the energy of the system and the position of the electron: they are incompatible observables. We can apply the test discussed in Section 13.4.1 for incompatibility of $\hat{x}$ and $\hat{H}$ here by evaluating the commutator $\left[\hat{x}, \hat{H}\right]$ using their representative matrices:

$$\begin{aligned}\left[\hat{x}, \hat{H}\right] =& \hat{x}\hat{H} - \hat{H}\hat{x} \\ =& a\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}E_0 & -A\\ -A & E_0\end{pmatrix} - a\begin{pmatrix}E_0 & -A\\ -A & E_0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \\ =& -2aA\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \neq 0.\end{aligned} \tag{13.49}$$

13.5.5 Observables for a Single Mode EM Field

A somewhat different example from those presented above is that of the field inside a single mode cavity (see pp 132, 150). In this case, the basis states of the electromagnetic field are the number states $\{|n\rangle, n = 0, 1, 2, \ldots\}$ where the state $|n\rangle$ is the state of the field in which there are n photons present.

Number Operator From the annihilation operator $\hat{a}$ (Eq. (11.57)) and creation operator $\hat{a}^\dagger$ (Eq. (11.72)) for this field, defined such that

$$\begin{aligned}\hat{a}|n\rangle &= \sqrt{n}|n-1\rangle, \qquad \hat{a}|0\rangle = 0 \\ \hat{a}^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle\end{aligned}$$

we can construct a Hermitean operator $\hat{N}$ defined by

$$\hat{N} = \hat{a}^\dagger\hat{a} \tag{13.50}$$

which can be readily shown to be such that

$$\hat{N}|n\rangle = n|n\rangle. \tag{13.51}$$

This operator is an observable of the system of photons. Its eigenvalues are the integers $n = 0, 1, 2, \ldots$ which correspond to the possible results obtained when the number of photons in the cavity are measured, and $|n\rangle$ are the corresponding eigenstates, the number states, representing the state in which there are exactly n photons in the cavity. This observable has the spectral decomposition

$$\hat{N} = \sum_{n=0}^{\infty} n|n\rangle\langle n|. \tag{13.52}$$

Hamiltonian If the cavity is designed to support a field of frequency ω, then each photon would have the energy $\hbar\omega$, so that the energy of the field when in the state $|n\rangle$ would be $n\hbar\omega$. From this information we can construct the Hamiltonian for the cavity field. It will be

$$\hat{H} = \hbar\omega\hat{N}. \tag{13.53}$$

A more rigorous analysis based on 'quantizing' the electromagnetic field yields an expression $\hat{H} = \hbar\omega(\hat{N} + \frac{1}{2})$ for the Hamiltonian. The additional term $\frac{1}{2}\hbar\omega$ is known as the zero point energy of the field. Its presence is required by the uncertainty principle, though it apparently plays no role in the dynamical behaviour of the cavity field as it merely represents a shift in the zero of energy of the field.

13.6 Observables with Continuous Values

In the case of measurements being made of an observable with a continuous range of possible values such as position or momentum, or in some cases, energy, the above postulates need to modified somewhat. The modifications arise first, from the fact that the eigenvalues are continuous, but also because the state space of the system will be of infinite dimension.

To see why there is an issue here in the first place, we need to see where any of the statements made in the case of an observable with discrete values comes unstuck. This can best be seen if we consider a particular example, that of the position of a particle.

13.6.1 Measurement of Particle Position

If we are to suppose that a particle at a definite position x is to be assigned a state vector $|x\rangle$, and if further we are to suppose that the possible positions are continuous over the range $(-\infty, +\infty)$ and that the associated states are complete, then we are lead to requiring that any state $|\psi\rangle$ of the particle must be expressible as

$$|\psi\rangle = \int_{-\infty}^{\infty} |x\rangle\langle x|\psi\rangle\, dx \tag{13.54}$$

with the states $|x\rangle$ δ-function normalised, i.e.

$$\langle x|x'\rangle = \delta(x - x'). \tag{13.55}$$

The difficulty with this is that the state $|x\rangle$ has infinite norm: it cannot be normalized to unity and hence cannot represent a possible physical state of the system. This makes it problematical to introduce the idea of an observable – the position of the particle – that can have definite values x associated with unphysical states $|x\rangle$. There is a further argument about the viability of this idea, at least in the context of measuring the position of a particle, which is to say that if the position were to be precisely defined at a particular value, this would mean, by the uncertainty principle $\Delta x \Delta p \geq \frac{1}{2}\hbar$ that the momentum of the particle would have infinite uncertainty, i.e. it could have any value from $-\infty$ to ∞. It is a not very difficult exercise to show that to localize a particle to a region of infinitesimal size would require an infinite amount of work to be done, so the notion of preparing a particle in a state $|x\rangle$ does not even make physical sense.

The resolution of this impasse involves recognizing that the measurement of the position of a particle is, in practice, only ever done to within the accuracy, δx say, of the measuring apparatus. In other words, rather than measuring the precise position of a particle, what is measured is its position as lying somewhere in a range $(x - \frac{1}{2}\delta x, x + \frac{1}{2}\delta x)$. We can accommodate this situation within the theory by defining a new set of states that takes this into account. This could be done in a number of ways, but the simplest is to suppose we divide the continuous range of values of x into intervals of length δx, so that the n^{th} segment is the interval $((n-1)\delta x, n\delta x)$ and let x_n be a point within the n^{th} interval. This could be any point within this interval but it is simplest to take it to be the midpoint of the interval, i.e. $x_n = (n - \frac{1}{2})\delta x$. We then say that the particle is in the state $|x_n\rangle$ if the measuring apparatus indicates that the position of the particle is in the n^{th} segment.

In this manner we have replaced the continuous case by the discrete case, and we can now proceed along the lines of what was presented in the preceding Section. Thus we can introduce an observable $x_{\delta x}$ that can be measured to have the values $\{x_n; n = 0, \pm 1, \pm 2 \ldots\}$, with $|x_n\rangle$ being the state of the particle for which $x_{\delta x}$ has the value x_n. We can then construct a Hermitean operator $\hat{x}_{\delta x}$ with eigenvalues $\{x_n; n = 0, \pm 1, \pm 2 \ldots\}$ and associated eigenvectors $\{|x_n\rangle; n = 0, \pm 1, \pm 2, \ldots\}$ such that

$$\hat{x}_{\delta x}|x_n\rangle = x_n|x_n\rangle. \tag{13.56}$$

The states $\{|x_n; n = 0, \pm 1, \pm 2, \ldots\rangle\}$ will form a complete set of orthonormal basis states for the particle, so that any state of the particle can be written

$$|\psi\rangle = \sum_n |x_n\rangle\langle x_n|\psi\rangle \tag{13.57}$$

with $\langle x_n|x_m\rangle = \delta_{nm}$. The observable $\hat{x}_{\delta x}$ would then be given by

$$\hat{x}_{\delta x} = \sum_n x_n|x_n\rangle\langle x_n|. \tag{13.58}$$

Finally, if a measurement of $x_{\delta x}$ is made and the result x_n is observed, then the immediate post-measurement state of the particle will be

$$\frac{\hat{P}_n|\psi\rangle}{\sqrt{\langle\psi|\hat{P}_n|\psi\rangle}} \tag{13.59}$$

where $\hat{P}_n$ is the projection operator

$$\hat{P}_n = |x_n\rangle\langle x_n|. \tag{13.60}$$

To relate all this back to the continuous case, it is then necessary to take the limit, in some sense, of $\delta x \to 0$. This limiting process has already been discussed in Section 10.2.2, in an equivalent but slightly different model of the continuous limit. The essential points will be repeated here.

Returning to Eq. (13.57), we can define a new, unnormalized state vector $\widetilde{|x_n\rangle}$ by

$$\widetilde{|x_n\rangle} = \frac{|x_n\rangle}{\sqrt{\delta x}} \tag{13.61}$$

The states $\widetilde{|x_n\rangle}$ continue to be eigenstates of $\hat{x}_{\delta x}$, i.e.

$$\hat{x}_{\delta x}\widetilde{|x_n\rangle} = x_n\widetilde{|x_n\rangle} \tag{13.62}$$

as the factor $1/\sqrt{\delta x}$ merely renormalizes the length of the vectors. Thus these states $\widetilde{|x_n\rangle}$ continue to represent the same physical state of affairs as the normalized state, namely that when in this state, the particle is in the interval $(x_n - \frac{1}{2}\delta x, x_n + \frac{1}{2}\delta x)$.

In terms of these unnormalized states, Eq. (13.57) becomes

$$|\psi\rangle = \sum_n \widetilde{|x_n\rangle}\,\widetilde{\langle x_n|\psi\rangle}\delta x. \tag{13.63}$$

If we let $\delta x \to 0$, then, in this limit the sum in Eq. (13.63) will define an integral with respect to x:

$$|\psi\rangle = \int_{-\infty}^{\infty} |x\rangle\langle x|\psi\rangle\, dx \tag{13.64}$$

where we have introduced the symbol $|x\rangle$ to represent the $\delta x \to 0$ limit of $\widetilde{|x_n\rangle}$ i.e.

$$|x\rangle = \lim_{\delta x\to 0} \frac{|x_n\rangle}{\sqrt{\delta x}}. \tag{13.65}$$

This then is the idealized state of the particle for which its position is specified to within a vanishingly small interval around x as δx approaches zero. From Eq. (13.64) we can extract the completeness relation for these states

$$\int_{-\infty}^{\infty} |x\rangle\langle x|\, dx = \hat{1}. \tag{13.66}$$

This is done at a cost, of course. By the same arguments as presented in Section 10.2.2, the new states $|x\rangle$ are δ-function normalized, i.e.

$$\langle x|x'\rangle = \delta(x - x') \tag{13.67}$$

and, in particular, are of infinite norm, that is, they cannot be normalized to unity and so do not represent physical states of the particle.

Having introduced these idealized states, we can investigate some of their further properties and uses. The first and probably the most important is that it gives us the means to write down the probability of finding a particle in any small region in space. Thus, provided the state $|\psi\rangle$ is normalized to unity, Eq. (13.64) leads to

$$\langle\psi|\psi\rangle = 1 = \int_{-\infty}^{\infty} |\langle x|\psi\rangle|^2\, dx \tag{13.68}$$

which can be interpreted as saying that the total probability of finding the particle somewhere in space is unity. More particularly, we also conclude that $|\langle x|\psi\rangle|^2\, dx$ is the probability of finding the position of the particle to be in the range $(x, x + dx)$.

If we now turn to Eq. (13.58) and rewrite it in terms of the unnormalized states we have

$$\hat{x}_{\delta x} = \sum_n x_n \widetilde{|x_n\rangle}\,\widetilde{\langle x_n|}\delta x \tag{13.69}$$

so that in a similar way to the derivation of Eq. (13.64) this gives, in the limit of $\delta x \to 0$, the new operator $\hat{x}$, i.e.

$$\hat{x} = \int_{-\infty}^{\infty} x|x\rangle\langle x|\, dx. \tag{13.70}$$

This then leads to the $\delta x \to 0$ limit of the eigenvalue equation for $\hat{x}_{\delta x}$, Eq. (13.62) i.e.

$$\hat{x}|x\rangle = x|x\rangle \tag{13.71}$$

a result that also follows from Eq. (13.70) on using the δ-function normalization condition. This operator $\hat{x}$ therefore has as eigenstates the complete set of δ-function normalized states $\{|x\rangle; -\infty <$

$x < \infty\}$ with associated eigenvalues x and can be looked on as being the observable corresponding to an idealized, precise measurement of the position of a particle.

While these states $|x\rangle$ can be considered idealized limits of the normalizable states $|x_n\rangle$ it must always be borne in mind that these are not physically realizable states – they are not normalizable, and hence are not vectors in the state space of the system. They are best looked on as a convenient fiction with which to describe idealized situations, and under most circumstances these states can be used in much the same way as discrete eigenstates. Indeed it is one of the benefits of the Dirac notation that a common mathematical language can be used to cover both the discrete and continuous cases. But situations can and do arise in which the cavalier use of these states can lead to incorrect or paradoxical results. We will not be considering such cases here.

The final point to be considered is the projection postulate. We could, of course, idealize this by saying that if a result x is obtained on measuring $\hat{x}$, then the state of the system after the measurement is $|x\rangle$. But given that the best we can do in practice is to measure the position of the particle to within the accuracy of the measuring apparatus, we cannot really go beyond the discrete case prescription given in Eq. (13.59) except to express it in terms of the idealized basis states $|x\rangle$. So, if the particle is in some state $|\psi\rangle$, we can recognize that the probability of getting a result x with an accuracy of δx will be given by

$$\int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} |\langle x'|\psi\rangle|^2 dx' = \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} \langle\psi|x'\rangle\langle x'|\psi\rangle dx'$$
$$= \langle\psi|\left[\int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} |x'\rangle\langle x'|dx'\right]|\psi\rangle = \langle\psi|\hat{P}(x,\delta x)|\psi\rangle \quad (13.72)$$

where we have introduced an operator $\hat{P}(x,\delta x)$ defined by

$$\hat{P}(x,\delta x) = \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} |x'\rangle\langle x'|dx'. \quad (13.73)$$

We can readily show that this operator is in fact a projection operator since

$$\begin{aligned}\left[\hat{P}(x,\delta x)\right]^2 &= \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} dx' \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} dx''|x'\rangle\langle x'|x''\rangle\langle x''| \\ &= \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} dx' \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} dx''|x'\rangle\delta(x'-x'')\langle x''| \\ &= \int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} dx'|x'\rangle\langle x'| \\ &= \hat{P}(x,\delta x). \end{aligned} \quad (13.74)$$

This suggests, by comparison with the corresponding postulate in the case of discrete eigenvalues, that if the particle is initially in the state $|\psi\rangle$, then the state of the particle immediately after measurement be given by

$$\frac{\hat{P}(x,\delta x)|\psi\rangle}{\sqrt{\langle\psi|\hat{P}(x,\delta x)|\psi\rangle}} = \frac{\int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} |x'\rangle\langle x'|\psi\rangle dx'}{\sqrt{\int_{x-\frac{1}{2}\delta x}^{x+\frac{1}{2}\delta x} |\langle x'|\psi\rangle|^2 dx'}} \quad (13.75)$$

It is this state that is taken to be the state of the particle immediately after the measurement has been performed, with the result x being obtained to within an accuracy δx.

Further development of these ideas is best done in the language of generalized measurements where the projection operator is replaced by an operator that more realistically represents the outcome of the measurement process. We will not be pursuing this any further here.

At this point, we can take the ideas developed for the particular case of the measurement of position and generalize them to apply to the measurement of any observable quantity with a continuous range of possible values. The way in which this is done is presented in the following Section.

13.6.2 General Postulates for Continuous Valued Observables

Suppose we have an observable Q of a system that is found, for instance through an exhaustive series of measurements, to have a continuous range of values $\theta_1 < q < \theta_2$. In practice, it is not the observable Q that is measured, but rather a discretized version in which Q is measured to an accuracy δq determined by the measuring device. If we represent by $|q\rangle$ the idealized state of the system in the limit $\delta q \to 0$, for which the observable definitely has the value q, then we claim the following:

1. The states $\{|q\rangle; \theta_1 < q < \theta_2\}$ form a complete set of δ-function normalized basis states for the state space of the system.

That the states form a complete set of basis states means that any state $|\psi\rangle$ of the system can be expressed as

$$|\psi\rangle = \int_{\theta_1}^{\theta_2} c(q)|q\rangle \tag{13.76}$$

while δ-function normalized means that $\langle q|q'\rangle = \delta(q - q')$ from which follows $c(q) = \langle q|\psi\rangle$ so that

$$|\psi\rangle = \int_{\theta_1}^{\theta_2} |q\rangle\langle q|\psi\rangle \, dq. \tag{13.77}$$

The completeness condition can then be written as

$$\int_{\theta_1}^{\theta_2} |q\rangle\langle q| \, dq = \hat{1} \tag{13.78}$$

2. For the system in state $|\psi\rangle$, the probability of obtaining the result q lying in the range $(q, q + dq)$ on measuring Q is $|\langle q|\psi\rangle|^2 dq$ provided $\langle\psi|\psi\rangle = 1$.

Completeness means that for any state $|\psi\rangle$ it must be the case that

$$\int_{\theta_1}^{\theta_2} |\langle q|\psi\rangle|^2 dq \neq 0 \tag{13.79}$$

i.e. there must be a non-zero probability to get some result on measuring Q.

3. The observable Q is represented by a Hermitean operator $\hat{Q}$ whose eigenvalues are the possible results $\{q; \theta_1 < q < \theta_2\}$, of a measurement of Q, and the associated eigenstates are the states $\{|q\rangle; \theta_1 < q < \theta_2\}$, i.e. $\hat{Q}|q\rangle = q|q\rangle$. The name 'observable' is often applied to the operator $\hat{Q}$ itself.

The spectral decomposition of the observable $\hat{Q}$ is then

$$\hat{Q} = \int_{\theta_1}^{\theta_2} q|q\rangle\langle q|\, dq. \tag{13.80}$$

As in the discrete case, the eigenvectors of an observable constitute a set of basis states for the state space of the associated quantum system.

A more subtle difficulty is now encountered if we turn to the von Neumann postulate concerning the state of the system after a measurement is made. If we were to transfer the discrete state postulate directly to the continuous case, we would be looking at proposing that obtaining the result q in a measurement of $\hat{Q}$ would mean that the state after the measurement is $|q\rangle$. This is a state that is not permitted as it cannot be normalized to unity. Thus we need to take account of the way a measurement is carried out in practice when considering the state of the system after the measurement. Following on from the particular case of position measurement presented above, we will suppose that Q is measured with a device of accuracy δq. This leads to the following general statement of the von Neumann measurement postulate for continuous eigenvalues:

4. If on performing a measurement of Q with an accuracy δq, the result is obtained in the range $(q - \frac{1}{2}\delta q, q + \frac{1}{2}\delta q)$, then the system will end up in the state

$$\frac{\hat{P}(q, \delta q)|\psi\rangle}{\sqrt{\langle\psi|\hat{P}(q, \delta q)|\psi\rangle}} \tag{13.81}$$

where

$$\hat{P}(q, \delta q) = \int_{q-\frac{1}{2}\delta q}^{q+\frac{1}{2}\delta q} |q'\rangle\langle q'|dq'. \tag{13.82}$$

Even though there exists this precise statement of the projection postulate for continuous eigenvalues, it is nevertheless a convenient fiction to assume that the measurement of an observable Q with a continuous set of eigenvalues will yield one of the results q with the system ending up in the state $|q\rangle$ immediately afterwards. While this is, strictly speaking, not really correct, it can be used as a convenient shorthand for the more precise statement given above.

As mentioned earlier, further development of these ideas is best done in the language of generalized measurements.

13.7 Examples of Continuous Valued Observables

13.7.1 Position and momentum of a particle (in one dimension)

These two observables are those which are most commonly encountered in wave mechanics. In the case of position, we are already able to say a considerable amount about the properties of this observable. Some further development is required in order to be able to deal with momentum.

Position observable (in one dimension)

In one dimension, the position x of a particle can range over the values $-\infty < x < \infty$. Thus the Hermitean operator $\hat{x}$ corresponding to this observable will have eigenstates $|x\rangle$ and associated eigenvalues x such that

$$\hat{x}|x\rangle = x|x\rangle, \qquad -\infty < x < \infty. \tag{13.83}$$

As the eigenvalues cover a continuous range of values, the completeness relation will be expressed as an integral:

$$|\psi\rangle = \int_{-\infty}^{\infty} |x\rangle\langle x|\psi\rangle \tag{13.84}$$

where $\langle x|\psi\rangle = \psi(x)$ is the wave function associated with the particle. Since there is a continuously infinite number of basis states $|x\rangle$, these states are delta-function normalized:

$$\langle x|x'\rangle = \delta(x - x'). \tag{13.85}$$

The operator itself can be expressed as

$$\hat{x} = \int_{-\infty}^{\infty} x|x\rangle\langle x|\, dx. \tag{13.86}$$

The Position Representation The wave function is, of course, just the components of the state vector $|\psi\rangle$ with respect to the position eigenstates as basis vectors. Hence, the wave function is often referred to as being the state of the system in the *position representation*. The probability amplitude $\langle x|\psi\rangle$ is just the wave function, written $\psi(x)$ and is such that $|\psi(x)|^2 dx$ is the probability of the particle being observed to have a momentum in the range x to $x + dx$.

The one big difference here as compared to the discussion in Chapter 12 is that the basis vectors here are continuous, rather than discrete, so that the representation of the state vector is not a simple column vector with discrete entries, but rather a function of the continuous variable x. Likewise, the operator $\hat{x}$ will not be represented by a matrix with discrete entries labelled, for instance, by pairs of integers, but rather it will be a function of two continuous variables:

$$\langle x|\hat{x}|x'\rangle = x\delta(x - x'). \tag{13.87}$$

The position representation is used in quantum mechanical problems where it is the position of the particle in space that is of primary interest. For instance, when trying to determine the chemical properties of atoms and molecules, it is important to know how the electrons in each atom tend to distribute themselves in space in the various kinds of orbitals as this will play an important role in determining the kinds of chemical bonds that will form. For this reason, the position representation, or the wave function, is the preferred choice of representation. When working in the position representation, the wave function for the particle is found by solving the Schrödinger equation for the particle.

Momentum of a particle (in one dimension)

As for position, the momentum p is an observable which can have any value in the range $-\infty < p < \infty$ (this is non-relativistic momentum). Thus the Hermitean operator $\hat{p}$ will have eigenstates $|p\rangle$ and associated eigenvalues p:

$$\hat{p}|p\rangle = p|p\rangle, \qquad -\infty < p < \infty. \tag{13.88}$$

As the eigenvalues cover a continuous range of values, the completeness relation will also be expressed as an integral:

$$|\psi\rangle = \int_{-\infty}^{+\infty} |p\rangle\langle p|\psi\rangle\, dp \tag{13.89}$$

where the basis states are delta-function normalized:

$$\langle p|p'\rangle = \delta(p - p'). \tag{13.90}$$

The operator itself can be expressed as

$$\hat{p} = \int_{-\infty}^{+\infty} p|p\rangle\langle p|\, dp. \tag{13.91}$$

Momentum Representation If the state vector is represented in component form with respect to the momentum eigenstates as basis vectors, then this is said to be the *momentum representation*. The probability amplitude $\langle p|\psi\rangle$ is sometimes referred to as the momentum wave function, written $\tilde{\psi}(p)$ and is such that $|\tilde{\psi}(p)|^2 dp$ is the probability of the particle being observed to have a momentum in the range p to $p + dp$. It turns out that the momentum wave function and the position wave function are Fourier transform pairs, a result that is shown below.

The momentum representation is preferred in problems in which it is not so much where a particle might be in space that is of interest, but rather how fast it is going and in what direction. Thus, the momentum representation is often to be found when dealing with scattering problems in which a particle of well defined momentum is directed towards a scattering centre, e.g. an atomic nucleus, and the direction in which the particle is scattered, and the momentum and/or energy of the scattered particle are measured, though even here, the position representation is used more often than not as it provides a mental image of the scattering process as waves scattering off an obstacle. Finally, we can also add that there is an equation for the momentum representation wave function which is equivalent to the Schrödinger equation.

Properties of the Momentum Operator

The momentum operator can be introduced into quantum mechanics by a general approach based on the space displacement operator. But at this stage it is nevertheless possible to draw some conclusions about the properties of the momentum operator based on the de Broglie hypothesis concerning the wave function of a particle of precisely known momentum p and energy E.

The momentum operator in the position representation From the de Broglie relation and Einstein's formula, the wave function $\Psi(x,t)$ to be associated with a particle of momentum p and energy E will have a wave number k and angular frequency ω given by $p = \hbar k$ and $E = \hbar\omega$. We can then guess what this wave function would look like:

$$\Psi(x,t) = \langle x|\Psi(t)\rangle = Ae^{i(kx-\omega t)} + Be^{-i(kx-\omega t)} + Ce^{i(kx+\omega t)} + De^{-i(kx+\omega t)}. \tag{13.92}$$

The expectation is that the wave will travel in the same direction as the particle, i.e. if $p > 0$, then the wave should travel in the direction of positive x. Thus we must reject those terms with the argument $(kx + \omega t)$ and so we are left with

$$\langle x|\Psi(t)\rangle = Ae^{i(px/\hbar-\omega t)} + Be^{-i(px/\hbar-\omega t)} \tag{13.93}$$

where we have substituted for k in terms of p. The claim then is that the state $|\Psi(t)\rangle$ is a state for which the particle definitely has momentum p, and hence it must be an eigenstate of the momentum operator $\hat{p}$, i.e.

$$\hat{p}|\Psi(t)\rangle = p|\Psi(t)\rangle \tag{13.94}$$

which becomes, in the position representation

$$\begin{aligned}\langle x|\hat{p}|\Psi(t)\rangle =& p\langle x|\Psi(t)\rangle \\ =& p\left(Ae^{i(px/\hbar-\omega t)} + Be^{-i(px/\hbar-\omega t)}\right).\end{aligned} \tag{13.95}$$

The only simple way of obtaining the factor p is by taking the derivative of the wave function with respect to x, though this does not immediately give us what we want, i.e.

$$\langle x|\hat{p}|\Psi(t)\rangle = -i\hbar\frac{\partial}{\partial x}\left(Ae^{i(px/\hbar-\omega t)} - Be^{-i(px/\hbar-\omega t)}\right) \neq p\langle x|\Psi(t)\rangle \tag{13.96}$$

which tells us that the state $|\Psi(t)\rangle$ is not an eigenstate of $\hat{p}$, at least if we proceed along the lines of introducing the derivative with respect to x. However, all is not lost. If we choose one or the other of the two terms in the expression for $\langle x|\Psi(t)\rangle$ e.g.

$$\langle x|\Psi(t)\rangle = Ae^{i(px/\hbar-\omega t)} \tag{13.97}$$

we find that

$$\langle x|\hat{p}|\Psi(t)\rangle = -i\hbar\frac{\partial}{\partial x}\langle x|\Psi(t)\rangle = p\langle x|\Psi(t)\rangle \tag{13.98}$$

as required. This suggests that we have arrived at a candidate for the wave function for a particle of definite momentum p. But we could have chosen the other term with coefficient B. However, this other choice amounts to reversing our choice of the direction of positive x and positive t — its exponent can be written $i(p(-x)/\hbar - \omega(-t))$. This is itself a convention, so we can in fact use either possibility as the required momentum wave function without affecting the physics. To be in keeping with the convention that is usually adopted, the choice Eq. (13.97) is made here.

Thus, by this process of elimination, we have arrived at an expression for the wave function of a particle with definite momentum p. Moreover, we have extracted an expression for the momentum operator $\hat{p}$ in that, if $|p\rangle$ is an eigenstate of $\hat{p}$, then, in the position representation

$$\langle x|\hat{p}|p\rangle = -i\hbar\frac{d}{dx}\langle x|p\rangle. \tag{13.99}$$

This is a result that can be readily generalized to apply to any state of the particle. By making use of the fact that the momentum eigenstate form a complete set of basis states, we can write any state $|\psi\rangle$ as

$$|\psi\rangle = \int_{-\infty}^{+\infty} |p\rangle\langle p|\psi\rangle\, dp \tag{13.100}$$

so that

$$\begin{aligned}\langle x|\hat{p}|\psi\rangle &= \int_{-\infty}^{+\infty} \langle x|\hat{p}|p\rangle\langle p|\psi\rangle\, dp \\ &= -i\hbar\frac{d}{dx}\int_{-\infty}^{+\infty} \langle x|p\rangle\langle p|\psi\rangle\, dp \\ &= -i\hbar\frac{d}{dx}\langle x|\psi\rangle\end{aligned}$$

or

$$\langle x|\hat{p}|\psi\rangle = -i\hbar\frac{d}{dx}\psi(x). \tag{13.101}$$

From this result it is straightforward to show that

$$\langle x|\hat{p}^n|\psi\rangle = (-i\hbar)^n\frac{d^n}{dx^n}\psi(x). \tag{13.102}$$

For instance,

$$\langle x|\hat{p}^2|\psi\rangle = \langle x|\hat{p}|\phi\rangle \tag{13.103}$$

where $|\phi\rangle = \hat{p}|\psi\rangle$. Thus

$$\langle x|\hat{p}^2|\psi\rangle = -i\hbar\frac{d}{dx}\phi(x). \tag{13.104}$$

But

$$\phi(x) = \langle x|\phi\rangle = \langle x|\hat{p}|\psi\rangle = -i\hbar\frac{d}{dx}\psi(x). \tag{13.105}$$

Using this and Eq. (13.104), we get

$$\langle x|\hat{p}^2|\psi\rangle = -i\hbar\frac{d}{dx}\left(-i\hbar\frac{d}{dx}\right)\psi(x) = -\hbar^2\frac{d^2}{dx^2}\psi(x) \tag{13.106}$$

In general we see that, when working in the position representation, the substitution

$$\hat{p} \longrightarrow -i\hbar\frac{d}{dx} \tag{13.107}$$

can consistently be made. This is an exceedingly important result that plays a central role in wave mechanics, in particular in setting up the Schrödinger equation for the wave function.

One final result can be established using this correspondence. Consider the action of the operator

$$\hat{D}(a) = e^{i\hat{p}a/\hbar} \tag{13.108}$$

on an arbitrary state $|\psi\rangle$ i.e.

$$|\phi\rangle = \hat{D}(a)|\psi\rangle \tag{13.109}$$

which becomes, in the position representation

$$\langle x|\phi\rangle = \phi(x) = \langle x|\hat{D}(a)|\psi\rangle. \tag{13.110}$$

Expanding the exponential and making the replacement Eq. (13.107) we have

$$\hat{D}(a) = \hat{1} + i\hat{p}a/\hbar + (ia/\hbar)^2\frac{1}{2!}\hat{p}^2 + (ia/\hbar)^3\frac{1}{3!}\hat{p}^3 + \ldots \longrightarrow 1 + a\frac{d}{dx} + \frac{a^2}{2!}\frac{d^2}{dx^2} + \frac{a^3}{3!}\frac{d^3}{dx^3} + \ldots \tag{13.111}$$

we get

$$\begin{aligned}\phi(x) =& \left(1 + a\frac{d}{dx} + \frac{a^2}{2!}\frac{d^2}{dx^2} + \frac{a^3}{3!}\frac{d^3}{dx^3} + \ldots\right)\psi(x)\\ =& \psi(x) + a\psi'(x) + \frac{a^2}{2!}\psi''(x) + \frac{a^3}{3!}\psi'''(x) + \ldots\\ =& \psi(x + a)\end{aligned} \tag{13.112}$$

where the series appearing above is recognized as the Maclaurin series expansion about $x = a$. Thus we see that the state $|\phi\rangle$ obtained by the action of the operator $\hat{D}(a)$ on $|\psi\rangle$ is to diplace the wave function a distance a along the x axis. This result illustrates the deep connection between momentum and displacement in space, a relationship that is turned on its head in Chapter 16 where momentum is defined in terms of the displacement operator.

The normalized momentum eigenfunction Returning to the differential equation Eq. (13.99), we can readily obtain the solution

$$\langle x|p\rangle = Ae^{ipx/\hbar} \tag{13.113}$$

where A is a coefficient to be determined by requiring that the states $|p\rangle$ be delta function normalized. Note that our wave function Eq. (13.97) gives the time development of the eigenfunction $\langle x|p\rangle$, Eq. (13.113).

The normalization condition is that

$$\langle p|p'\rangle = \delta(p - p') \tag{13.114}$$

which can be written, on using the completeness relation for the position eigenstates

$$\begin{aligned}\delta(p - p') =& \int_{-\infty}^{+\infty}\langle p|x\rangle\langle x|p'\rangle\, dx\\ =& |A|^2\int_{-\infty}^{+\infty} e^{-i(p-p')x/\hbar}\, dp\\ =& |A|^2 2\pi\hbar\delta(p - p')\end{aligned} \tag{13.115}$$

where we have used the representation of the Dirac delta function given in Section 10.2.3. Thus we conclude

$$|A|^2 = \frac{1}{2\pi\hbar}. \tag{13.116}$$

It then follows that

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}. \tag{13.117}$$

This result can be used to relate the wave function for a particle in the momentum and position representations. Using the completeness of the momentum states we can write for any state $|\psi\rangle$ of the particle

$$|\psi\rangle = \int_{-\infty}^{+\infty} |p\rangle\langle p|\psi\rangle \, dp \tag{13.118}$$

which becomes, in the position representation

$$\langle x|\psi\rangle = \int_{-\infty}^{+\infty} \langle x|p\rangle\langle p|\psi\rangle \, dp \tag{13.119}$$

or, in terms of the wave functions:

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} e^{ipx/\hbar} \tilde{\psi}(p) \, dp \tag{13.120}$$

which immediately shows that the momentum and position representation wave functions are Fourier transform pairs. It is a straightforward procedure to then show that

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} e^{-ipx/\hbar} \psi(x) \, dx \tag{13.121}$$

either by simply inverting the Fourier transform, or by expanding the state vector $|\psi\rangle$ in terms of the position eigenstates.

That the position and momentum wave functions are related in this way has a very important consequence that follows from a fundamental property of Fourier transform pairs. Roughly speaking, there is an inverse relationship between the width of a function and its Fourier transformed companion. This is most easily seen if suppose, somewhat unrealistically, that $\psi(x)$ is of the form

$$\psi(x) = \begin{cases} \dfrac{1}{\sqrt{a}} & |x| \leq \frac{1}{2}a \\ 0 & |x| > \frac{1}{2}a \end{cases} \tag{13.122}$$

The full width of $\psi(x)$ is a. The momentum wave function is

$$\tilde{\psi}(p) = \sqrt{\frac{2\hbar}{\pi}} \frac{\sin(pa/\hbar)}{p}. \tag{13.123}$$

An estimate of the width of $\tilde{\psi}(p)$ is given by determining the positions of the first zeroes of $\tilde{\psi}(p)$ on either side of the central maximum at $p = 0$, that is at $p = \pm\pi\hbar/a$. The separation of these two zeroes, $2\pi\hbar/a$, is an overestimate of the width of the peak so we take this width to be half this separation, thus giving an estimate of $\pi\hbar/a$. Given that the square of the wave functions i.e. $|\psi(x)|^2$ and $|\tilde{\psi}(p)|^2$ give the probability distribution for position and momentum respectively, it is clearly the case that the wider the spread in the possible values of the position of the particle, i.e. the larger a is made, there is a narrowing of the spread in the range of values of momentum, and vice versa. This inverse relationship is just the Heisenberg uncertainty relation reappearing, and is more fully quantified in terms of the uncertainties in position and momentum defined in Chapter 14.

Position momentum commutation relation The final calculation in here is to determine the commutator $[\hat{x}, \hat{p}] = \hat{x}\hat{p} - \hat{p}\hat{x}$ of the two operators $\hat{x}$ and $\hat{p}$. This can be done most readily in the position representation. Thus we will consider

$$\langle x|[\hat{x}, \hat{p}]|\psi\rangle = \langle x|(\hat{x}\hat{p} - \hat{p}\hat{x})|\psi\rangle \tag{13.124}$$

where $|\psi\rangle$ is an arbitrary state of the particle. This becomes, on using the fact that the state $\langle x|$ is an eigenstate of $\hat{x}$ with eigenvalue x

$$\langle x|[\hat{x}, \hat{p}]|\psi\rangle = x\langle x|\hat{p}|\psi\rangle - \langle x|\hat{p}|\xi\rangle \tag{13.125}$$

where $|\xi\rangle = \hat{x}|\psi\rangle$. Expressed in terms of the differential operator, this becomes

$$\langle x|[\hat{x}, \hat{p}]|\psi\rangle = -i\hbar\left(x\frac{d}{dx}\psi(x) - \frac{d}{dx}\xi(x)\right). \tag{13.126}$$

But

$$\xi(x) = \langle x|\xi\rangle = \langle x|\hat{x}|\psi\rangle = x\langle x|\psi\rangle = x\psi(x) \tag{13.127}$$

so that

$$\frac{d}{dx}\xi(x) = x\frac{d}{dx}\psi(x) + \psi(x). \tag{13.128}$$

Combining this altogether then gives

$$\langle x|[\hat{x}, \hat{p}]|\psi\rangle = i\hbar\psi(x) = i\hbar\langle x|\psi\rangle. \tag{13.129}$$

The completeness of the position eigenstates can be used to write this as

$$\int_{-\infty}^{+\infty} |x\rangle\langle x|[\hat{x}, \hat{p}]|\psi\rangle = i\hbar\int_{-\infty}^{+\infty} |x\rangle\langle x|\psi\rangle \tag{13.130}$$

or

$$[\hat{x}, \hat{p}]|\psi\rangle = i\hbar\hat{1}|\psi\rangle. \tag{13.131}$$

Since the state $|\psi\rangle$ is arbitrary, we can conclude that

$$[\hat{x}, \hat{p}] = i\hbar\hat{1} \tag{13.132}$$

though the unit operator on the right hand side is usually understood, so the relation is written

$$[\hat{x}, \hat{p}] = i\hbar. \tag{13.133}$$

This is perhaps the most important result in non-relativistic quantum mechanics as it embodies much of what makes quantum mechanics 'different' from classical mechanics. For instance, if the position and momentum operators were classical quantities, then the commutator would vanish, or to put it another way, it is that fact that $\hbar \neq 0$ that gives us quantum mechanics. It turns out that, amongst other things, the fact that the commutator does not vanish implies that it is not possible to have precise information on the position and the momentum of a particle, i.e. position and momentum are *incompatible* observables.

13.7.2 Field operators for a single mode cavity

A physical meaning can be given to the annihilation and creation operators defined above in terms of observables of the field inside the cavity. This done here in a non-rigorous way, relying on a trick by which we relate the classical and quantum ways of specifying the energy of the field. We have earlier arrived at an expression for the quantum Hamiltonian of the cavity field as given in Eq. (13.53), i.e. $\hat{H} = \hbar\omega\hat{a}^\dagger\hat{a}$, which as we have already pointed out, is missing the zero-point

contribution $\frac{1}{2}\hbar\omega$ that is found in a full quantum theory of the electromagnetic field. However, assuming we had never heard of this quantum theory, then we could proceed by comparing the quantum expression we have derived with the classical expression for the energy of the single mode EM field inside the cavity.

A little more detail about the single mode field is required before the classical Hamiltonian can be obtained. This field will be assumed to be a plane polarized standing wave confined between two mirrors of area $\mathcal{A}$, separated by a distance L in the z direction. Variation of the field in the x and y directions will also be ignored, so the total energy of the field will be given by

$$H = \tfrac{1}{2}\int_{\mathcal{V}} \left(\epsilon_0 E^2(z,t) + B^2(z,t)/\mu_0\right) dxdydz \tag{13.134}$$

where the classical electric field is $E(z,t) = \mathrm{Re}[\mathcal{E}(t)]\sin(\omega z/c)$, i.e. of complex amplitude $\mathcal{E}(t) = \mathcal{E}e^{-i\omega t}$, the magnetic field is $B(z,t) = c^{-1}\mathrm{Im}[\mathcal{E}(t)]\cos(\omega z/c)$, and where $\mathcal{V} = \mathcal{A}L$ is the volume of the cavity.

The integral can be readily carried out to give

$$H = \tfrac{1}{4}\epsilon_0 \mathcal{E}^*\mathcal{E}\mathcal{V} \tag{13.135}$$

We want to establish a correspondence between the two expressions for the Hamiltonian, i.e.

$$\hbar\omega\hat{a}^\dagger\hat{a} \longleftrightarrow \tfrac{1}{4}\epsilon_0\mathcal{E}^*\mathcal{E}\mathcal{V}$$

in order to give some sort of physical interpretation of $\hat{a}$, apart from its interpretation as a photon annihilation operator. We can do this by reorganizing the various terms so that the correspondence looks like

$$\left(2e^{-i\phi}\sqrt{\frac{\hbar\omega}{\mathcal{V}\epsilon_0}}\hat{a}^\dagger\right)\left(2e^{i\phi}\sqrt{\frac{\hbar\omega}{\mathcal{V}\epsilon_0}}\hat{a}\right) \longleftrightarrow \mathcal{E}^*\mathcal{E}$$

where $\exp(i\phi)$ is an arbitrary phase factor, i.e. it could be chosen to have any value and the correspondence would still hold. A common choice is to take this phase factor to be i. The most obvious next step is to identify an operator $\hat{\mathcal{E}}$ closely related to the classical quantity $\mathcal{E}$ by

$$\hat{\mathcal{E}} = 2i\sqrt{\frac{\hbar\omega}{\mathcal{V}\epsilon_0}}\hat{a}$$

so that we get the correspondence $\hat{\mathcal{E}}^\dagger\hat{\mathcal{E}} \longleftrightarrow \mathcal{E}^*\mathcal{E}$.

We can note that the operator $\hat{\mathcal{E}}$ is still not Hermitean, but recall that the classical electric field was obtained from the real part of $\mathcal{E}$, so that we can define a Hermitean electric field operator by

$$\hat{E}(z) = \tfrac{1}{2}[\hat{\mathcal{E}} + \hat{\mathcal{E}}^\dagger]\sin(\omega z/c) = i\sqrt{\frac{\hbar\omega}{\mathcal{V}\epsilon_0}}\,[\hat{a} - \hat{a}^\dagger]\sin(\omega z/c)$$

to complete the picture. In this way we have identified a new observable for the field inside the cavity, the electric field operator. This operator is, in fact, an example of a quantum field operator. Of course, the derivation presented here is far from rigorous. That this operator is indeed the electric field operator can be shown to follow from the full analysis as given by the quantum theory of the electromagnetic field.

In the same way, an expression for the magnetic field operator can be determined from the expression for the classical magnetic field, with the result:

$$\hat{B}(z) = \sqrt{\frac{\mu_0\hbar\omega}{\mathcal{V}}}(\hat{a} + \hat{a}^\dagger)\cos(\omega z/c). \tag{13.136}$$

The above are two new observables for the system so it is natural to ask what is their eigenvalue spectrum. We can get at this in an indirect fashion by examining the properties of the two Hermitean operators $i(\hat{a} - \hat{a}^\dagger)$ and $\hat{a} + \hat{a}^\dagger$. For later purposes, it is convenient to rescale these two operators and define

$$\hat{X} = \sqrt{\frac{\hbar}{2\omega}}(\hat{a} + \hat{a}^\dagger) \quad \text{and} \quad \hat{P} = -i\sqrt{\frac{\hbar\omega}{2}}(\hat{a} - \hat{a}^\dagger). \tag{13.137}$$

As the choice of notation implies, the aim is to show that these new operators are, mathematically at least, closely related to the position and momentum operators $\hat{x}$ and $\hat{p}$ for a single particle. The distinctive quantum feature of these latter operators is their commutation relation, so the aim is to evaluate the commutator of $\hat{X}$ and $\hat{P}$. To do so, we need to know the value of the commutator $[\hat{a}, \hat{a}^\dagger]$. We can determine this by evaluating $[\hat{a}, \hat{a}^\dagger]|n\rangle$ where $|n\rangle$ is an arbitrary number state. Using the properties of the annihilation and creation operators $\hat{a}$ and $\hat{a}^\dagger$ given by

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle \quad \text{and} \quad \hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$$

we see that

$$\begin{aligned}
[\hat{a}, \hat{a}^\dagger]|n\rangle &= \hat{a}\hat{a}^\dagger|n\rangle - \hat{a}^\dagger\hat{a}|n\rangle \\
&= \hat{a}\sqrt{n+1}|n+1\rangle - \hat{a}^\dagger\sqrt{n}|n-1\rangle \\
&= (n+1)|n\rangle - n|n\rangle \\
&= |n\rangle
\end{aligned}$$

from which we conclude

$$[\hat{a}, \hat{a}^\dagger] = \hat{1}. \tag{13.138}$$

If we now make use of this result when evaluating $[\hat{X}, \hat{P}]$ we find

$$[\hat{X}, \hat{P}] = -i\tfrac{1}{2}\hbar[\hat{a} + \hat{a}^\dagger, \hat{a} - \hat{a}^\dagger] = i\hbar \tag{13.139}$$

where use has been made of the properties of the commutator as given in Eq. (11.24).

In other words, the operators $\hat{X}$ and $\hat{P}$ obey exactly the same commutation relation as position and momentum for a single particle, $\hat{x}$ and $\hat{p}$ respectively. This is of course a mathematical correspondence i.e. there is no massive particle 'behind the scenes' here, but the mathematical correspondence is one that is found to arise in the formulation of the quantum theory of the electromagnetic field. But what it means to us here is that the two observables $\hat{X}$ and $\hat{P}$ will, to all intents and purposes have the same properties as $\hat{x}$ and $\hat{p}$. In particular, the eigenvalues of $\hat{X}$ and $\hat{P}$ will be continuous, ranging from $-\infty$ to $+\infty$. Since we can write the electric field operator as

$$\hat{E}(z) = -\sqrt{\frac{2}{\mathcal{V}\epsilon_0}}\,\hat{P}\sin(\omega z/c) \tag{13.140}$$

we conclude that the electric field operator will also have a continuous range of eigenvalues from $-\infty$ to $+\infty$. This is in contrast to the number operator or the Hamiltonian which both have a discrete range of values. A similar conclusion applies for the magnetic field, which can be written in terms of $\hat{X}$:

$$\hat{B}(z) = \omega\sqrt{\frac{2\mu_0}{\mathcal{V}}}\hat{X}\cos(\omega z/c). \tag{13.141}$$

What remains is to check the form of the Hamiltonian of the field as found directly from the expressions for the electric and magnetic field operators. This can be calculated via the quantum version of the classical expression for the field energy, Eq. (13.134), i.e.

$$\hat{H} = \tfrac{1}{2}\int_{\mathcal{V}}\left[\epsilon_0\hat{E}^2(z) + \hat{B}^2(z)/\mu_0\right]dxdydz. \tag{13.142}$$

Substituting for the field operators and carrying out the spatial integrals gives

$$= \tfrac{1}{4}\hbar\omega\left[-(\hat{a}-\hat{a}^\dagger)^2 + (\hat{a}+\hat{a}^\dagger)^2\right] \tag{13.143}$$

$$= \tfrac{1}{2}\hbar\omega\left[\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}\right]. \tag{13.144}$$

Using the commutation rule $\left[\hat{a}, \hat{a}^\dagger\right] = \hat{1}$ we can write this as

$$\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \tfrac{1}{2}\right). \tag{13.145}$$

Thus, we recover the original expression for the Hamiltonian, but now with the additional zero-point energy contribution $\frac{1}{2}\hbar\omega$. That we do not recover the assumed starting point for the Hamiltonian is an indicator that the above derivation is not entirely rigorous. Nevertheless, it does achieve a useful purpose in that we have available expressions for the electric and magnetic field operators, and the Hamiltonian, for the a single mode electromagnetic field.

Chapter 14

Probability, Expectation Value and Uncertainty

We have seen that the physically observable properties of a quantum system are represented by Hermitean operators (also referred to as 'observables') such that the eigenvalues of the operator represents all the possible results that could be obtained if the associated physical observable were to be measured. The eigenstates of the operator are the states of the system for which the associated eigenvalue would be, with 100% certainty, the measured result, if the observable were measured. If the system were in any other state then the possible outcomes of the measurement cannot be predicted precisely – different possible results could be obtained, each one being an eigenvalue of the associated observable, but only the probabilities can be determined of these possible results. This physically observed state-of-affairs is reflected in the mathematical structure of quantum mechanics in that the eigenstates of the observable form a complete set of basis states for the state space of the system, and the components of the state of the system with respect to this set of basis states gives the probability amplitudes, and by squaring, the probabilities of the various outcomes being observed.

Given the probabilistic nature of the measurement outcomes, familiar tools used to analyze statistical data such as the mean value and standard deviation play a natural role in characterizing the results of such measurements. The expectation or mean value of an observable is a concept taken more or less directly from the classical theory of probability and statistics. Roughly speaking, it represents the (probability-weighted) average value of the results of many repetitions of the measurement of an observable, with each measurement carried out on one of very many identically prepared copies of the same system. The standard deviation of these results is then a measure of the spread of the results around the mean value and is known, in a quantum mechanical context, as the uncertainty.

14.1 Observables with Discrete Values

The probability interpretation of quantum mechanics plays a central, in fact a defining, role in quantum mechanics, but the precise meaning of this probability interpretation has as yet not been fully spelt out. However, for the purposes of defining the expectation value and the uncertainty it is necessary to look a little more closely at how this probability is *operationally* defined. The way this is done depends on whether the observable has a discrete or continuous set of eigenvalues, so each case is discussed separately. We begin here with the discrete case.

14.1.1 Probability

If A is an observable for a system with a discrete set of values $\{a_1, a_2, \ldots\}$, then this observable is represented by a Hermitean operator $\hat{A}$ that has these discrete values as its eigenvalues, and associated eigenstates $\{|a_n\rangle, n = 1, 2, 3, \ldots\}$ satisfying the eigenvalue equation $\hat{A}|a_n\rangle = a_n|a_n\rangle$.

These eigenstates form a complete, orthonormal set so that any state $|\psi\rangle$ of the system can be written

$$|\psi\rangle = \sum_n |a_n\rangle\langle a_n|\psi\rangle$$

where, provided $\langle\psi|\psi\rangle = 1$, then

$$|\langle a_n|\psi\rangle|^2 = \text{ probability of obtaining the result } a_n \text{ on measuring } \hat{A}.$$

To see what this means in an operational sense, we can suppose that we prepare N identical copies of the system all in exactly the same fashion so that each copy is presumably in the same state $|\psi\rangle$. This set of identical copies is sometimes called an *ensemble*. If we then perform a measurement of $\hat{A}$ for each member of the ensemble, i.e. for each copy, we find that we get randomly varying results e.g.

Copy:	1	2	3	4	5	6	7	8	...	N
Result:	a_5	a_3	a_1	a_9	a_3	a_8	a_5	a_9	...	a_7

Now count up the number of times we get the result a_1 – call this N_1, and similarly for the number of times N_2 we get the result a_2 and so on. Then

$$\frac{N_n}{N} = \text{ fraction of times the result } a_n \text{ is obtained.} \tag{14.1}$$

What we are therefore saying is that if the experiment is performed on an increasingly large number of identical copies, i.e. as $N \to \infty$, then we find that

$$\lim_{N\to\infty} \frac{N_n}{N} = |\langle a_n|\psi\rangle|^2. \tag{14.2}$$

This is the so-called 'frequency' interpretation of probability in quantum mechanics, i.e. the probability predicted by quantum mechanics gives us the frequency with which an outcome will occur in the limit of a very large number of repetitions of an experiment.

There is another interpretation of this probability, more akin to 'betting odds' in the sense that in application, it is usually the case that the experiment is a 'one-off' affair, i.e. it is not possible to repeat the experiment many times over under identical conditions. It then makes more sense to try to understand this probability as a measure of the likelihood of a particular result being actually observed, in the same way that betting odds for a horse to win a race is a measure of the confidence there is in the horse actually winning. After all, a race is only run once, so while it is possible to imagine the race being run many times over under identical conditions, and that the odds would provide a measure of how many of these imagined races a particular horse would win, there appears to be some doubt of the practical value of this line of thinking. Nevertheless, the 'probability as frequency' interpretation of quantum probabilities is the interpretation that is still most commonly to be found in quantum mechanics.

14.1.2 Expectation Value

We can now use this result to arrive at an expression for the average or mean value of all these results. If the measurement is repeated N times, the average will be

$$\frac{N_1}{N}a_1 + \frac{N_2}{N}a_2 + \ldots = \sum_n \frac{N_n}{N}a_n. \tag{14.3}$$

Now take the limit of this result for $N \to \infty$, and call the result $\langle \hat{A} \rangle$:

$$\langle \hat{A} \rangle = \lim_{N\to\infty} \sum_n \frac{N_n}{N} a_n = \sum_n |\langle a_n|\psi\rangle|^2 a_n. \tag{14.4}$$

It is a simple matter to use the properties of the eigenstates of $\hat{A}$ to reduce this expression to a compact form:

$$\begin{aligned}
\langle \hat{A} \rangle &= \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle a_n \\
&= \sum_n \langle\psi|\hat{A}|a_n\rangle\langle a_n|\psi\rangle \\
&= \langle\psi|\hat{A}\left\{\sum_n |a_n\rangle\langle a_n|\right\}|\psi\rangle \\
&= \langle\psi|\hat{A}|\psi\rangle
\end{aligned} \tag{14.5}$$

where the completeness relation for the eigenstates of $\hat{A}$ has been used.

This result can be readily generalized to calculate the expectation values of any function of the observable $\hat{A}$. Thus, in general, we have that

$$\langle f(\hat{A}) \rangle = \sum_n |\langle a_n|\psi\rangle|^2 f(a_n) = \langle\psi|f(\hat{A})|\psi\rangle \tag{14.6}$$

a result that will find immediate use below when the uncertainty of an observable will be considered.

While the expectation value has been defined in terms of the outcomes of measurements of an observable, the same concept of an 'expectation value' is applied to non-Hermitean operators which do not represent observable properties of a system. Thus if $\hat{A}$ is not Hermitean and hence not an observable we cannot make use of the idea of a collection of results of experiments over which an average is taken to arrive at an expectation value. Nevertheless, we continue to make use of the notion of the expectation value $\langle \hat{A} \rangle = \langle\psi|\hat{A}|\psi\rangle$ even when $\hat{A}$ is not an observable.

14.1.3 Uncertainty

The uncertainty of the observable A is a measure of the spread of results around the mean $\langle \hat{A} \rangle$. It is defined in the usual way, that is the difference between each measured result and the mean is calculated, i.e. $a_n - \langle \hat{A} \rangle$, then the average taken of the *square* of all these differences. In the limit of $N \to \infty$ repetitions of the same experiment we arrive at an expression for the uncertainty $(\Delta A)^2$:

$$(\Delta A)^2 = \lim_{N\to\infty} \sum_n \frac{N_n}{N} (a_n - \langle \hat{A} \rangle)^2 = \sum_n |\langle a_n|\psi\rangle|^2 (a_n - \langle \hat{A} \rangle)^2. \tag{14.7}$$

As was the case for the expectation value, this can be written in a more compact form:

$$\begin{aligned}
(\Delta A)^2 &= \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle (a_n - \langle \hat{A} \rangle)^2 \\
&= \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle (a_n^2 - 2a_n\langle \hat{A} \rangle + \langle \hat{A} \rangle^2) \\
&= \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle a_n^2 - 2\langle \hat{A} \rangle \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle a_n + \langle \hat{A} \rangle^2 \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle
\end{aligned}$$

Using $f(\hat{A})|a_n\rangle = f(a_n)|a_n\rangle$ then gives

$$\begin{aligned}(\Delta A)^2 &= \sum_n \langle\psi|\hat{A}^2|a_n\rangle\langle a_n|\psi\rangle - 2\langle\hat{A}\rangle\sum_n\langle\psi|\hat{A}|a_n\rangle\langle a_n|\psi\rangle + \langle\hat{A}\rangle^2\langle\psi|\left\{\sum_n |a_n\rangle\langle a_n|\right\}|\psi\rangle \\ &= \langle\psi|\hat{A}^2\left\{\sum_n |a_n\rangle\langle a_n|\right\}|\psi\rangle - 2\langle\hat{A}\rangle\langle\psi|\hat{A}\left\{\sum_n |a_n\rangle\langle a_n|\right\}|\psi\rangle + \langle\hat{A}\rangle^2\langle\psi|\psi\rangle\end{aligned}$$

Assuming that the state $|\psi\rangle$ is normalized to unity, and making further use of the completeness relation for the basis states $|a_n\rangle$ this becomes

$$(\Delta A)^2 = \langle\psi|\hat{A}^2|\psi\rangle - 2\langle\hat{A}\rangle^2 + \langle\hat{A}\rangle^2 = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2. \tag{14.8}$$

We can get another useful expression for the uncertainty if we proceed in a slightly different fashion:

$$\begin{aligned}(\Delta\hat{A})^2 &= \sum_n \langle\psi|a_n\rangle\langle a_n|\psi\rangle(a_n - \langle\hat{A}\rangle)^2 \\ &= \sum_n \langle\psi|a_n\rangle(a_n - \langle\hat{A}\rangle)^2\langle a_n|\psi\rangle\end{aligned}$$

Now using $f(\hat{A})|a_n\rangle = f(a_n)|a_n\rangle$ then gives

$$\begin{aligned}(\Delta\hat{A})^2 &= \sum_n \langle\psi|(\hat{A} - \langle\hat{A}\rangle)^2|a_n\rangle\langle a_n|\psi\rangle \\ &= \langle\psi|(\hat{A} - \langle\hat{A}\rangle)^2\left\{\sum_n |a_n\rangle\langle a_n|\right\}|\psi\rangle \\ &= \langle\psi|(\hat{A} - \langle\hat{A}\rangle)^2|\psi\rangle \\ &= \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle.\end{aligned} \tag{14.9}$$

Thus we have, for the uncertainty $\Delta\hat{A}$, the two results

$$(\Delta\hat{A})^2 = \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2. \tag{14.10}$$

Just as was the case for the expectation value, we can also make use of this expression to formally calculate the uncertainty of some operator $\hat{A}$ even if it is not an observable, though the fact that the results cannot be interpreted as the standard deviation of a set of actually observed results means that the physical interpretation of the uncertainty so obtained is not particularly clear.

Ex 14.1 For the O_2^- ion (see p 117) calculate the expectation value and standard deviation of the position of the electron if the ion is prepared in the state $|\psi\rangle = \alpha|+a\rangle + \beta|-a\rangle$.

In this case, the position operator is given by

$$\hat{x} = a|+a\rangle\langle+a| - a|-a\rangle\langle-a|$$

so that

$$\hat{x}|\psi\rangle = a\alpha|+a\rangle - a\beta|-a\rangle$$

and hence the expectation value of the position of the electron is

$$\langle\psi|\hat{x}|\psi\rangle = \langle\hat{x}\rangle = a(|\alpha|^2 - |\beta|^2)$$

which can be recognized as being just

$$\langle\hat{x}\rangle = (+a)\times\text{probability } |\alpha|^2 \text{ of measuring electron position to be } +a$$
$$+ (-a)\times\text{probability } |\beta|^2 \text{ of measuring electron position to be } -a.$$

In particular, if the probability of the electron being found on either oxygen atom is equal, that is $|\alpha|^2 = |\beta|^2 = \frac{1}{2}$, then the expectation value of the position of the electron is $\langle\hat{x}\rangle = 0$.

The uncertainty in the position of the electron is given by

$$(\Delta\hat{x})^2 = \langle\hat{x}^2\rangle - \langle\hat{x}\rangle^2.$$

We already have $\langle\hat{x}\rangle$, and to calculate $\langle\hat{x}^2\rangle$ it is sufficient to make use of the expression above for $\hat{x}|\psi\rangle$ to give

$$\begin{aligned}\hat{x}^2|\psi\rangle &= \hat{x}\big[a\alpha|+a\rangle - a\beta|-a\rangle\big]\\ &= a^2\big[\alpha|+a\rangle + \beta|-a\rangle\big]\\ &= a^2|\psi\rangle\end{aligned}$$

so that $\langle\psi|\hat{x}^2|\psi\rangle = \langle\hat{x}^2\rangle = a^2$. Consequently, we find that

$$\begin{aligned}(\Delta\hat{x})^2 &= a^2 - a^2(|\alpha|^2 - |\beta|^2)^2\\ &= a^2\Big[1 - (|\alpha|^2 - |\beta|^2)^2\Big]\\ &= a^2(1 - |\alpha|^2 + |\beta|^2)(1 + |\alpha|^2 - |\beta|^2)\end{aligned}$$

Using $\langle\psi|\psi\rangle = |\alpha|^2 + |\beta|^2 = 1$ this becomes

$$(\Delta\hat{x})^2 = 4a^2|\alpha|^2|\beta|^2$$

from which we find that

$$\Delta\hat{x} = 2a|\alpha\beta|.$$

This result is particularly illuminating in the case of $|\alpha|^2 = |\beta|^2 = \frac{1}{2}$ where we find that $\Delta\hat{x} = a$. This is the symmetrical situation in which the mean position of the electron is $\langle\hat{x}\rangle = 0$, and the uncertainty is exactly equal to the displacement of the two oxygen atoms on either side of this mean.

Ex 14.2 As an example of an instance in which it is useful to deal with the expectation values of non-Hermitean operators, we can turn to the system consisting of identical photons in a single mode cavity (see pp 132, 150, 193). If the cavity field is prepared in the number state $|n\rangle$, an eigenstate of the number operator $\hat{N}$ defined in Eq. (13.50), then we immediately have $\langle\hat{N}\rangle = n$ and $\Delta\hat{N} = 0$.

However, if we consider a state such as $|\psi\rangle = (|n\rangle + |n+1\rangle)/\sqrt{2}$ we find that

$$\begin{aligned}\langle\hat{N}\rangle &= \tfrac{1}{2}(\langle n| + \langle n+1|)\hat{N}(|n\rangle + |n+1\rangle)\\ &= \tfrac{1}{2}(\langle n| + \langle n+1|)(n|n\rangle + (n+1)|n+1\rangle)\\ &= n + \tfrac{1}{2}.\end{aligned}$$

while

$$\begin{aligned}(\Delta\hat{N})^2 &= \langle\hat{N}^2\rangle - \langle\hat{N}\rangle^2\\ &= \tfrac{1}{2}(\langle n| + \langle n+1|)\hat{N}\,\hat{N}(|n\rangle + |n+1\rangle) - (n+\tfrac{1}{2})^2\\ &= \tfrac{1}{2}(n\langle n| + (n+1)\langle n+1|)(n|n\rangle + (n+1)|n+1\rangle) - (n+\tfrac{1}{2})^2\\ &= \tfrac{1}{2}(n^2 + (n+1)^2) - (n+\tfrac{1}{2})^2\\ &= \tfrac{1}{4}.\end{aligned}$$

Not unexpectedly,we find that the uncertainty in the photon number is now non-zero.

We can also evaluate the expectation value of the annihilation operator $\hat{a}$, Eq. (11.57) for the system in each of the above states. Thus, for the number state $|n\rangle$ we have

$$\langle\hat{a}\rangle = \langle n|\hat{a}|n\rangle = \sqrt{n}\langle n|n-1\rangle = 0$$

while for the state $|\psi\rangle = (|n\rangle + |n+1\rangle)/\sqrt{2}$ we find that

$$\langle\hat{a}\rangle = \tfrac{1}{2}\sqrt{n+1}.$$

It is an interesting exercise to try to give some kind of meaning to these expectation values for $\hat{a}$, particularly since the expectation value of a non-Hermitean operator does not, in general, represent the average of a collection of results obtained when measuring some observable. The meaning is provided via the demonstration in Section 13.5.5 that $\hat{a}$ is directly related to the electric field strength inside the cavity. From that demonstration, it can be seen that the imaginary part of $\langle\hat{a}\rangle$ is to be understood as being proportional to the average strength of the electric (or magnetic) field inside the cavity, i.e. $\mathrm{Im}\langle\hat{a}\rangle \propto \langle\hat{E}\rangle$ where $\hat{E}$ is the electric field operator. We can then interpret the expressions obtained above for the expectation value of the field as saying little more than that for the field in the number state $|n\rangle$, the electric field has an average value of zero. However, this is not to say that the field is not overall zero. Thus, we can calculate the uncertainty in the field strength $\Delta\hat{E}$. This entails calculating

$$\Delta\hat{E} = \sqrt{\langle n|\hat{E}^2|n\rangle - \langle n|\hat{E}|n\rangle^2} = \sqrt{\langle n|\hat{E}^2|n\rangle}.$$

Using the expression

$$\hat{E} = i\sqrt{\frac{\hbar\omega}{2\mathcal{V}\epsilon_0}}\,[\hat{a} - \hat{a}^\dagger]$$

we see that

$$\langle n|\hat{E}^2|n\rangle = \frac{\hbar\omega}{2\mathcal{V}\epsilon_0}\langle n|\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}|n\rangle$$

where the other contributions, $\langle n|\hat{a}\hat{a}|n\rangle = \langle n|\hat{a}^\dagger\hat{a}^\dagger|n\rangle = 0$. It then follows from the properties of the creation and annihilation operators that

$$\langle n|\hat{E}^2|n\rangle = \frac{\hbar\omega}{2\mathcal{V}\epsilon_0}(2n+1)$$

and hence

$$\Delta\hat{E} = \sqrt{\frac{\hbar\omega}{2\mathcal{V}\epsilon_0}}\,\sqrt{(2n+1)}.$$

So, while the field has an expectation value of zero, it has non-zero fluctuations about this mean. In fact, if the cavity contains no photons ($n = 0$), we see that even then there is a random electric field present:

$$\Delta\hat{E} = \sqrt{\frac{\hbar\omega}{2\mathcal{V}\epsilon_0}}.$$

This non-vanishing electric random electric field is known as the vacuum field fluctuations, and its presence shows that even the quantum vacuum is alive with energy!

14.2 Observables with Continuous Values

Much of what has been said in the preceding Section carries through with relatively minor changes in the case in which the observable under consideration has a continuous eigenvalue spectrum. However, rather than discussing the general case, we will consider only the particular instance of the measurement of the position represented by the position operator $\hat{x}$ of a particle.

14.2.1 Probability

Recall from Section 13.6.1, p 194, that in order to measure the position – an observable with a continuous range of possible values – it was necessary to suppose that the measuring apparatus had a finite resolution δx in which case we divide the continuous range of values of x into intervals of length δx, so that the n^{th} segment is the interval $((n-1)\delta x, n\delta x)$. If we let x_n be the point in the middle of the n^{th} interval, i.e. $x_n = (n - \frac{1}{2})\delta x$, we then say that the particle is in the state $|x_n\rangle$ if the measuring apparatus indicates that the position of the particle is in the n^{th} segment. The observable being measured is now represented by the Hermitean operator $\hat{x}_{\delta x}$ with discrete eigenvalues $\{x_n; n = 0, \pm 1, \pm 2 \ldots\}$ and associated eigenvectors $\{|x_n\rangle; n = 0, \pm 1, \pm 2, \ldots\}$ such that $\hat{x}_{\delta x}|x_n\rangle = x_n|x_n\rangle$.

We can now imagine setting up, as in the discrete case discussed above, an ensemble of N identical systems all prepared in the state $|\psi\rangle$, and the position of the particle measured with a resolution δx. We then take note of the number of times N_n in which the particle was observed to lie in the n^{th} interval and from this data construct a histogram of the frequency N_n/N of the results for each interval. An example of such a set of results is illustrated in Fig. 14.1.

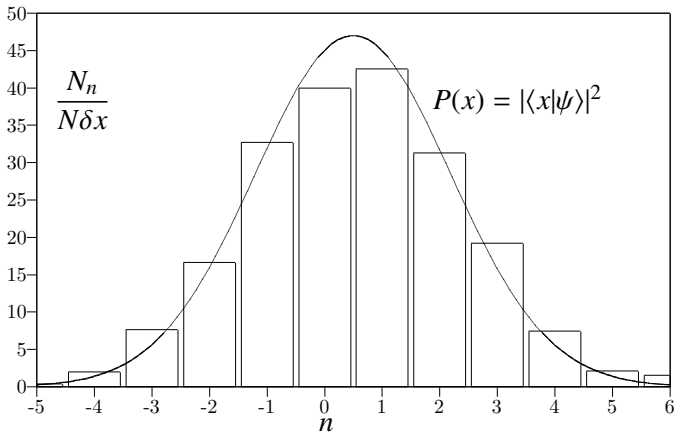


Figure 14.1: Histogram of the frequencies of detections in each interval $((n-1)\delta x, n\delta x), n = 0, \pm 1, \pm 2, \ldots$ for the measurement of the position of a particle for N identical copies all prepared in the state $|\psi\rangle$ and using apparatus with resolution δx. Also plotted is the probability distribution $|\langle x|\psi\rangle|^2$ expected in the limit of $N \to \infty$ and $\delta x \to 0$.

As was discussed earlier, the claim is now made that for N large, the frequency with which the particle is observed to lie in the n^{th} interval $((n-1)\delta x, n\delta x)$, that is N_n/N, should approximate to the probability predicted by quantum mechanics, i.e.

$$\frac{N_n}{N} \approx |\langle x|\psi\rangle|^2 \delta x \tag{14.11}$$

or, in other words

$$\frac{N_n}{N\delta x} \approx |\langle x|\psi\rangle|^2. \tag{14.12}$$

Thus, a histogram of the frequency N_n/N normalized per unit δx should approximate to the idealized result $|\langle x|\psi\rangle|^2$ expected in the limit of $N \to \infty$. This is also illustrated in Fig. 14.1. In the limit of the resolution $\delta x \to 0$, the tops of the histograms would approach the smooth curve in Fig. 14.1. All this amounts to saying that

$$\lim_{\delta x \to 0} \left(\lim_{N \to \infty} \frac{N_n}{N\delta x} \right) = |\langle x|\psi\rangle|^2. \tag{14.13}$$

The generalization of this result to other observables with continuous eigenvalues is essentially along the same lines as presented above for the position observable.

14.2.2 Expectation Values

We can now calculate the expectation value of an observable with continuous eigenvalues by making use of the limiting procedure introduced above. Rather than confining our attention to the particular case of particle position, we will work through the derivation of the expression for the expectation value of an arbitrary observable with a continuous eigenvalue spectrum. Thus, suppose we are measuring an observable $\hat{A}$ with a continuous range of eigenvalues $\alpha_1 < a < \alpha_2$. We do this, as indicated in the preceding Section by supposing that the range of possible values is divided up into small segments of length δa. Suppose the measurement is repeated N times on N identical copies of the same system, all prepared in the same state $|\psi\rangle$. In general, the same result will not be obtained every time, so suppose that on N_n occasions, a result a_n was obtained which lay in the n^{th} interval $((n-1)\delta a, n\delta a)$. The average of all these results will then be

$$\frac{N_1}{N}a_1 + \frac{N_2}{N}a_2 + \ldots. \tag{14.14}$$

If we now take the limit of this result for $N \to \infty$, then we can replace the ratios N_n/N by $|\langle a_n|\psi\rangle|^2\delta a$ to give

$$|\langle a_1|\psi\rangle|^2 a_1\delta a + |\langle a_2|\psi\rangle|^2 a_2\delta a + |\langle a_3|\psi\rangle|^2 a_3\delta a + \ldots = \sum_n |\langle a_n|\psi\rangle|^2 a_n\delta a. \tag{14.15}$$

Finally, if we let $\delta a \to 0$ the sum becomes an integral. Calling the result $\langle\hat{A}\rangle$, we get:

$$\langle\hat{A}\rangle = \int_{\alpha_1}^{\alpha_2} |\langle a|\psi\rangle|^2 a da. \tag{14.16}$$

It is a simple matter to use the properties of the eigenstates of $\hat{A}$ to reduce this expression to a compact form:

$$\begin{aligned}\langle\hat{A}\rangle &= \int_{\alpha_1}^{\alpha_2} \langle\psi|a\rangle a\langle a|\psi\rangle da \\ &= \int_{\alpha_1}^{\alpha_2} \langle\psi|\hat{A}|a\rangle\langle a|\psi\rangle da \\ &= \langle\psi|\hat{A}\left\{\int_{\alpha_1}^{\alpha_2} |a\rangle\langle a| da\right\}|\psi\rangle \\ &= \langle\psi|\hat{A}|\psi\rangle\end{aligned} \tag{14.17}$$

where the completeness relation for the eigenstates of $\hat{A}$ has been used. This is the same compact expression that was obtained in the discrete case earlier.

In the particular case in which the observable is the position of a particle, then the expectation value of position will be given by

$$\langle\hat{x}\rangle = \int_{-\infty}^{+\infty} |\langle x|\psi\rangle|^2 x\, dx = \int_{-\infty}^{+\infty} |\psi(x)|^2 x\, dx \tag{14.18}$$

Also, as in the discrete case, this result can be readily generalized to calculate the expectation values of any function of the observable $\hat{A}$. Thus, in general, we have that

$$\langle f(\hat{A})\rangle = \int_{\alpha_1}^{\alpha_2} |\langle a|\psi\rangle|^2 f(a)\, da = \langle\psi|f(\hat{A})|\psi\rangle \tag{14.19}$$

a result that will find immediate use below when the uncertainty of an observable with continuous eigenvalues will be considered.

Once again, as in the discrete case, if $\hat{A}$ is not Hermitean and hence not an observable we cannot make use of the idea of a collection of results of experiments over which an average is taken to arrive at an expectation value. Nevertheless, we continue to make use of the notion of the expectation value as defined by $\langle\hat{A}\rangle = \langle\psi|\hat{A}|\psi\rangle$ even when $\hat{A}$ is not an observable.

14.2.3 Uncertainty

We saw in the preceding subsection how the expectation value of an observabel with a continuous range of possible values can be obtained by first 'discretizing' the data, then making use of the ideas developed in the discrete case to set up the expectation value, and finally taking appropriate limits. The standard deviation or uncertainty in the outcomes of a collection of measurements of a continuously valued observable can be calculated in essentially the same fashion – after all it is nothing but another kind of expectation value, and not surprisingly, the same formal expression is obtained for the uncertainty. Thus, if $\hat{A}$ is an observable with a continuous range of possible eigenvalues $\alpha_1 < a < \alpha_2$, then the uncertainty $\Delta\hat{A}$ is given by $\langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle$

$$(\Delta A)^2 = \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2 \tag{14.20}$$

where

$$\langle\hat{A}^2\rangle = \int_{\alpha_1}^{\alpha_2} a^2 |\langle a|\psi\rangle|^2 \, da \qquad \text{and} \qquad \langle\hat{A}\rangle = \int_{\alpha_1}^{\alpha_2} a|\langle a|\psi\rangle|^2 \, da \tag{14.21}$$

In particular, the uncertainty in the position of a particle would be given by Δx where

$$(\Delta x)^2 = \langle\hat{x}^2\rangle - \langle\hat{x}\rangle^2 = \int_{-\infty}^{+\infty} x^2 |\psi(x)|^2 \, dx - \left(\int_{-\infty}^{+\infty} x|\psi(x)|^2 \, dx \right)^2 . \tag{14.22}$$

Ex 14.3 In the case of a particle trapped in the region $0 < x < L$ by infinitely high potential barriers – i.e. the particle is trapped in an infinitely deep potential well – the wave function for the energy eigenstates are given by

$$\psi_n(x) = \begin{cases} \sqrt{\dfrac{2}{L}} \sin(n\pi x/L) & 0 < x < L \\ 0 & x < 0 \quad x > L \end{cases}$$

The integrals are straightforward (see Eq. (5.39)) and yield

$$\Delta x = \frac{L}{2n\pi} \sqrt{\frac{n^2\pi^2 - 6}{3}}$$

which approaches, for $n \to \infty$

$$\Delta x \to \frac{L}{2\sqrt{3}}$$

which is the result that would be expected for a particle that was equally likely to be found anywhere within the potential well.

14.3 The Heisenberg Uncertainty Relation

14.4 Compatible and Incompatible Observables

In all the discussion so far regarding measurement of observables of a system, we have confined our attention to considering one observable only at a time. However, it is clear that physical systems will in general possess a number of observable properties.

14.4.1 Degeneracy

Chapter 15

Time Evolution in Quantum Mechanics

Physical systems are, in general, dynamical, i.e. they evolve in time. The dynamics of classical mechanical systems are described by Newton's laws of motion, while the dynamics of the classical electromagnetic field is determined by Maxwell's equations. Either Newton's equations or Maxwell's equations constitute laws of physics whose form is particular to the kind of system that is being described, though some unification is provided in terms of Hamilton's principle, a universal way of stating the laws of classical dynamics. What we require is a means by which the dynamical properties of quantum systems can be described, i.e. given the state $|\psi(0)\rangle$ of a quantum system at some initial time $t = 0$, what is required is a physical law that tells us what the state will be at some other time t, i.e. a quantum law of evolution. Given that a state vector is a repository of the information known about a system, what is required is a general physical law that tells us how this information evolves in time in response to the particular physical circumstances that the system of interest finds itself in. It is to be expected that the details of this law will vary from system to system, but it turns out that the law of evolution can be written in a way that holds true for *all* physical systems, in some sense similar to the way that Hamilton's principle provides a way of stating classical dynamical laws in one compact form. Of course, the quantum law of evolution is Schrödinger's equation.

15.1 Stationary States

The idea of 'stationary states' was first introduced by Bohr as a name given to those states of a hydrogen atom for which the orbits that the electron occupied were stable, i.e. the electron remained in the same orbital state for all time, in contrast to the classical physics prediction that no orbiting electron could remain in orbit forever: it would always radiate away its energy and spiral into the proton. Although Bohr's original concept is now no longer appropriate, the notion of a stationary state remains a valid one. It is a term that is now used to identify those states of a quantum system that do not change in time. This is not to say that a stationary state is one for which 'nothing happens' – there is still a rich collection of dynamical physics to be associated with such a state – but a stationary state turns out to be one for which the probabilities of outcomes of a measurement of any property of the system is the same no matter at what time the measurement is made. Our aim initial aim here is to try to learn what properties the state vector representing a stationary state might have.

The starting point is to let $|\psi(0)\rangle$ be the initial state of the system, and $|\psi(t)\rangle$ be the state at some other time t. As this state vector at time t is supposed to represent the system in the same physical state, the most that these two states can differ by is a multiplicative factor, $u(t)$ say, i.e.

$$|\psi(t)\rangle = u(t)|\psi(0)\rangle. \tag{15.1}$$

Since we would expect that if the initial state $|\psi(0)\rangle$ is normalized to unity, then so would the state $|\psi(t)\rangle$, we can write

$$\langle\psi(t)|\psi(t)\rangle = |u(t)|^2\langle\psi(0)|\psi(0)\rangle \tag{15.2}$$

and hence

$$|u(t)|^2 = 1 \tag{15.3}$$

so we must conclude that

$$u(t) = e^{-i\phi(t)} \tag{15.4}$$

where $\phi(t)$ is an as yet unknown time dependent phase. Thus we can write

$$|\psi(t)\rangle = e^{-i\phi(t)}|\psi(0)\rangle. \tag{15.5}$$

This result is very general, and perhaps not all that surprising, but it becomes more revealing if we focus on an important special case, that in which the system is assumed to be isolated i.e. that it has no interaction with any other system – the dynamics of the system are determined solely by its own internal interactions. What this means is that if, for such a system, we were initially to prepare the system in some state and allow it to evolve for, say, ten minutes, it should not matter when this 'initial' time might be. If we start the experiment at noon and allow it to run for ten minutes, or if we started it at midnight and allow it to run for ten minutes, the state of the system after the ten minutes has elapsed should be the same in either case, simply because the system is isolated: the way it behaves cannot be affected by anything external. In contrast, if the system were *open*, i.e. not isolated, then there is the prospect of time dependent external influences acting on the system – e.g. an externally produced laser pulse fired at an atom – so the evolution of the state of the atom would differ depending on the starting time. In such a situation the system might not have any stationary states at all as it is always potentially being forced to change because of these external influences.

What all this amounts to as that, for an isolated system, we can chose the initial time to be any arbitrary time t_0, and the evolution of $|\psi(t_0)\rangle$ up to the time t would be given by

$$|\psi(t)\rangle = e^{-i\phi(t-t_0)}|\psi(t_0)\rangle \tag{15.6}$$

i.e. what matters is how long the system evolves for, not what we choose as our starting time.

If we now consider the evolution of the system in a stationary state over a time interval (t, t'), then we have

$$|\psi(t')\rangle = e^{-i\phi(t'-t)}|\psi(t)\rangle = e^{-i\phi(t'-t)}e^{-i\phi(t-t_0)}|\psi(t_0)\rangle \tag{15.7}$$

but over the interval (t_0, t') we have

$$|\psi(t')\rangle = e^{-i\phi(t'-t_0)}|\psi(0)\rangle \tag{15.8}$$

so by comparing the last two equations we find that

$$e^{-i\phi(t'-t_0)} = e^{-i\phi(t'-t)}e^{-i\phi(t-t_0)} \tag{15.9}$$

or

$$\phi(t - t_0) = \phi(t' - t) + \phi(t - t_0), \tag{15.10}$$

an equation for ϕ with the solution

$$\phi(t) = \omega t \tag{15.11}$$

where ω is a constant. Thus we can conclude that

$$|\psi(t)\rangle = e^{-i\omega t}|\psi(0)\rangle \tag{15.12}$$

which is the required expression for the evolution in time of a stationary state of an isolated system.

15.2 The Schrödinger Equation – a 'Derivation'.

The expression Eq. (15.12) involves a quantity ω, a real number with the units of (time)$^{-1}$, i.e. it has the units of angular frequency. In order to determine the physical meaning to be given to this quantity, we can consider this expression in the particular case of a free particle of energy E and momentum p, for which the wave function is given by

$$\Psi(x,t) = Ae^{ikx}e^{-i\omega t} \tag{15.13}$$

where, according to the de Broglie-Bohr relations, $E = \hbar\omega$ and $p = \hbar k$. In this case, we see that ω is directly related to the energy of the particle. This is a relationship that we will assume applies to any physical system, so that we will rewrite Eq. (15.12) as

$$|\psi(t)\rangle = e^{-iEt/\hbar}|\psi(0)\rangle \tag{15.14}$$

and further identify the stationary state $|\psi(0)\rangle$ as being an eigenstate of the energy observable $\hat{H}$ for the system, otherwise known as the Hamiltonian.

We now make an inductive leap, and identify a stationary state of *any* physical system as being an eigenstate of the Hamiltonian of the system, with the associated eigenvalue being the energy to be associated with the stationary state.

$$|E(t)\rangle = e^{-iEt/\hbar}|E\rangle \tag{15.15}$$

with

$$\hat{H}|E\rangle = E|E\rangle \tag{15.16}$$

from which we get

$$\hat{H}|E(t)\rangle = E|E(t)\rangle. \tag{15.17}$$

If we differentiate Eq. (15.15) we find

$$i\hbar\frac{d|E(t)\rangle}{dt} = E|E(t)\rangle = \hat{H}|E(t)\rangle \tag{15.18}$$

which we can now use to obtain the equation satisfied by an arbitrary state vector $|\psi(t)\rangle$. First, since the eigenstates of $\hat{H}$, call them $\{|E_1\rangle, |E_2\rangle, \ldots, |E_N\rangle\}$, form a complete set of basis states, we can write for the initial state

$$|\psi(0)\rangle = \sum_{n=1}^{N} |E_n\rangle\langle E_n|\psi(0)\rangle. \tag{15.19}$$

The time evolved version is then

$$|\psi(t)\rangle = \sum_{n=1}^{N} e^{-iE_nt/\hbar}|E_n\rangle\langle E_n|\psi(0)\rangle \tag{15.20}$$

which, on differentiating, becomes

$$\begin{aligned} i\hbar\frac{d|\psi(t)\rangle}{dt} =& i\hbar\sum_{n=1}^{N} E_n e^{-iE_nt/\hbar}|E_n\rangle\langle E_n|\psi(t)\rangle \\ =& \sum_{n=1}^{N} \hat{H}e^{-iE_nt/\hbar}|E_n\rangle\langle E_n|\psi(t)\rangle \\ =& \hat{H}\sum_{n=1}^{N} e^{-iE_nt/\hbar}|E_n\rangle\langle E_n|\psi(t)\rangle \\ =& \hat{H}|\psi(t)\rangle \end{aligned} \tag{15.21}$$

so we have

$$\hat{H}|\psi(t)\rangle = i\hbar\frac{d|\psi(t)\rangle}{dt} \tag{15.22}$$

which is the celebrated Schrödinger equation in vector form.

Determining the solution of this equation is the essential task in determining the dynamical properties of a quantum system. If the eigenvectors and eigenvalues of the Hamiltonian can be readily determined, the solution can be written down directly, i.e. it is just

$$|\psi(t)\rangle = \sum_n e^{-iE_nt/\hbar}|E_n\rangle\langle E_n|\psi(0)\rangle. \tag{15.23}$$

However, it is not necessary to follow this procedure, at least explicitly, as there are in general many ways to solving the equations. In any case, the process of obtaining the eigenvalues and eigenvectors is often very difficult, if not impossible, for any but a small handful of problems, so approximate techniques have to be employed.

Typically information about the Hamiltonian is available as its components with respect to some set of basis vectors, i.e. the Hamiltonian is given as a matrix. In the next Section, an example of solving the Schrödinger equation when the Hamiltonian matrix is given.

15.2.1 Solving the Schrödinger equation: An illustrative example

The aim here is to present an illustrative solution of the Schrödinger equation using as the example the O_2^- ion studied in the preceding Chapter. There it was shown that the Hamiltonian for the ion takes the form, Eq. (13.32), in the position representation:

$$\hat{H} \doteq \begin{pmatrix} E_0 & -A \\ -A & E_0 \end{pmatrix} \tag{15.24}$$

where A is a real number. The eigenvalues of $\hat{H}$ are

$$E_1 = E_0 + A \qquad E_2 = E_0 - A.$$

and the associated eigenstates are

$$|E_1\rangle = \frac{1}{\sqrt{2}}(|+a\rangle - |-a\rangle) \doteq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$
$$|E_2\rangle = \frac{1}{\sqrt{2}}(|+a\rangle + |-a\rangle) \doteq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Now suppose the system is prepared in some initial state $|\psi(0)\rangle$. What is the state of the system at some other time t? There are a number of ways of doing this. The first is to solve the Schrödinger equation, with the Hamiltonian as given in Eq. (15.24). The second is to make use of the known information about the eigenvalues and eigenstates of the Hamiltonian.

Direct solution of the Schrödinger equation.

It is easiest to work using the vector/matrix notation. Thus, we will write the state of the ion at time t as

$$|\psi(t)\rangle = C_+(t)|+a\rangle + C_-(t)|-a\rangle \doteq \begin{pmatrix} C_+(t) \\ C_-(t) \end{pmatrix} \tag{15.25}$$

where $C_\pm(t)$ are the probability amplitudes $\langle \pm a|\psi(t)\rangle$.

The Schrödinger equation now becomes

$$i\hbar\frac{d}{dt}\begin{pmatrix}C_+(t)\\C_-(t)\end{pmatrix}=\begin{pmatrix}E_0 & -A\\-A & E_0\end{pmatrix}\begin{pmatrix}C_+(t)\\C_-(t)\end{pmatrix}. \tag{15.26}$$

In the following we will use the 'dot' notation for a derivative with respect to time, i.e. $\dot{C}_+ = dC_+/dt$. So, this last equation becomes

$$i\hbar\begin{pmatrix}\dot{C}_+(t)\\\dot{C}_-(t)\end{pmatrix}=\begin{pmatrix}E_0 & -A\\-A & E_0\end{pmatrix}\begin{pmatrix}C_+(t)\\C_-(t)\end{pmatrix}. \tag{15.27}$$

Expanding out this matrix equation we get two coupled first-order differential equations:

$$i\hbar\dot{C}_+(t) = E_0C_+(t) - AC_-(t) \tag{15.28}$$
$$i\hbar\dot{C}_-(t) = -AC_+(t) + E_0C_-(t). \tag{15.29}$$

There are many different ways of solving these equations (one of them, in fact, being to work with the original matrix equation!). Here we will exploit the symmetry of these equations by defining two new quantities

$$X = C_+ + C_- \quad Y = C_+ - C_- \tag{15.30}$$

in terms of which we find

$$C_+ = \tfrac{1}{2}(X+Y) \quad C_- = \tfrac{1}{2}(X-Y). \tag{15.31}$$

So, if we add the two equations Eq. (15.28) and Eq. (15.29) we get:

$$i\hbar\dot{X} = (E_0 - A)X \tag{15.32}$$

while if we subtract them we get

$$i\hbar\dot{Y} = (E_0 + A)Y. \tag{15.33}$$

Eq. (15.32) has the immediate solution

$$X(t) = X(0)e^{-i(E_0-A)t/\hbar} \tag{15.34}$$

while Eq. (15.33) has the solution

$$Y(t) = Y(0)e^{-i(E_0+A)t/\hbar}. \tag{15.35}$$

We can now reconstruct the solutions for $C_\pm(t)$ by use of Eq. (15.31):

$$C_+(t) = \tfrac{1}{2}(X+Y) = \tfrac{1}{2}e^{-iE_0t/\hbar}\left(X(0)e^{iAt/\hbar} + Y(0)e^{-iAt/\hbar}\right) \tag{15.36}$$
$$C_-(t) = \tfrac{1}{2}(X-Y) = \tfrac{1}{2}e^{-iE_0t/\hbar}\left(X(0)e^{iAt/\hbar} - Y(0)e^{-iAt/\hbar}\right). \tag{15.37}$$

To see what these solutions are telling us, let us suppose that at $t = 0$, the electron was on the oxygen atom at $x = -a$, i.e. the initial state was

$$|\psi(0)\rangle = |-a\rangle \tag{15.38}$$

so that

$$C_-(0) = 1 \quad C_+(0) = 0 \tag{15.39}$$

and hence

$$X(0) = 1 \quad Y(0) = -1. \tag{15.40}$$

We then have

$$\begin{aligned} C_+(t) =& \tfrac{1}{2} e^{-iE_0 t/\hbar} \left(e^{iAt/\hbar} - e^{-iAt/\hbar} \right) \\ =& i e^{-iE_0 t/\hbar} \sin(At/\hbar) \end{aligned} \tag{15.41}$$

$$\begin{aligned} \not{c} C_-(t) =& \tfrac{1}{2} e^{-iE_0 t/\hbar} \left(e^{iAt/\hbar} + e^{-iAt/\hbar} \right) \\ =& e^{-iE_0 t/\hbar} \cos(At/\hbar) \end{aligned} \tag{15.42}$$

where we have used the relations

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$$

Consequently, the probability of finding the ion on the oxygen atom at $x = -a$, i.e. the probability of finding it in the state $|-a\rangle$ will be

$$P_-(t) = |C_-(t)|^2 = \cos^2(At/\hbar) \tag{15.43}$$

while the probability of finding the electron on the oxygen atom at $x = +a$, i.e. the probability of finding it in the state $|+a\rangle$ will be

$$P_+(t) = |C_+(t)|^2 = \sin^2(At/\hbar). \tag{15.44}$$

These two probabilities add to unity for all time, as they should:

$$P_-(t) + P_+(t) = \cos^2(At/\hbar) + \sin^2(At/\hbar) = 1. \tag{15.45}$$

We can plot these two probabilities as a function of time:

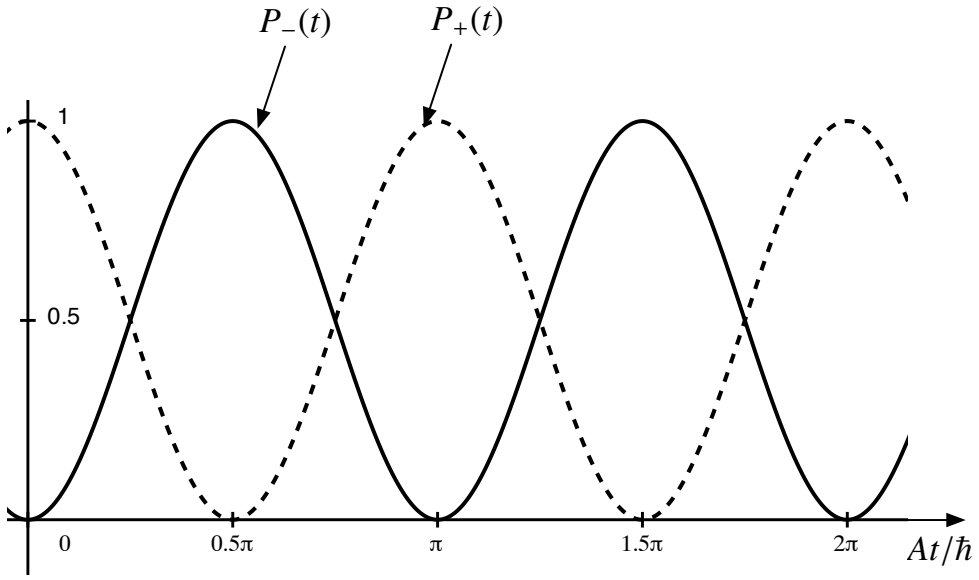


Figure 15.1: Probabilities $P_\pm(t)$ for the electron to be found on the oxygen atom at $x = \pm a$ in an O_2^- ion. The electron was initially on the atom at $x = +a$.

We can see from these curves that the probability oscillates back and forth with an angular frequency $A/\hbar$ and hence a frequency $A/\pi\hbar$, or a period $\pi\hbar/A$. Note that this oscillation is not to be interpreted as the electron being lost somewhere between the oxygen atoms between the peaks of the oscillations, rather this must be interpreted as the probability of the electron being on, say, the right hand atom, diminishing from a maximum of unity, while the probability of the atom being on the left hand atom increases from a minimum of zero, and vice versa. The electron is either more likely, or less likely, to be found at one atom or the other.

Solution using energy eigenvectors as basis vectors.

An alternate way of solving this problem is to recognize that the energy eigenstates can equally well be used as basis states for the state of the ion, i.e. we could write for the state of the ion at $t = 0$:

$$|\psi(0)\rangle = C_1|E_1\rangle + C_2|E_2\rangle \tag{15.46}$$

so that the time evolved state will be

$$|\psi(t)\rangle = C_1 e^{-i(E_0+A)t/\hbar}|E_1\rangle + C_2 e^{-i(E_0-A)t/\hbar}|E_2\rangle \tag{15.47}$$

where we have used the known energy eigenvalues $E_1 = E_0 + A$ and $E_2 = E_0 - A$ for the energy eigenstates $|E_1\rangle$ and $|E_2\rangle$ respectively. Thus, we already have the solution. But, if we want to make use of the initial condition that the electron was initially on the atom at $-a$, i.e. $|\psi(0)\rangle = |-a\rangle$, we have to use this information to determine the coefficients C_1 and C_2. Thus we have

$$|-a\rangle = C_1|E_1\rangle + C_2|E_2\rangle \tag{15.48}$$

and hence

$$C_1 = \langle E_1|-a\rangle = -\frac{1}{\sqrt{2}} \quad C_2 = \langle E_2|-a\rangle = \frac{1}{\sqrt{2}}. \tag{15.49}$$

where we have used

$$\begin{aligned}|E_1\rangle =& \frac{1}{\sqrt{2}}(|+a\rangle - |-a\rangle)\\ |E_2\rangle =& \frac{1}{\sqrt{2}}(|+a\rangle + |-a\rangle).\end{aligned}$$

Thus, the time evolved state is

$$|\psi(t)\rangle = -\frac{1}{\sqrt{2}}e^{-iE_0t/\hbar}\left(e^{-iAt/\hbar}|E_1\rangle - e^{iA)t/\hbar}|E_2\rangle\right). \tag{15.50}$$

The probability of finding the electron on the atom at $+a$ will then be

$$P_+(t) = |\langle +a|\psi(t)\rangle|^2 = |C_+(t)|^2 \tag{15.51}$$

which requires us to calculate

$$\begin{aligned}C_+(t) = \langle +a|\psi(t)\rangle =& -\frac{1}{\sqrt{2}}e^{-iE_0t/\hbar}\left(e^{-iAt/\hbar}\langle +a|E_1\rangle - e^{iA)t/\hbar}\langle +a|E_2\rangle\right) && (15.52)\\ =& -\tfrac{1}{2}e^{-iE_0t/\hbar}\left(e^{-iAt/\hbar} - e^{iA)t/\hbar}\right) && (15.53)\\ =& ie^{-iE_0t/\hbar}\sin(At/\hbar) && (15.54)\end{aligned}$$

as before. Similarly, we can determine $\langle -a|\psi(t)\rangle$, once again yielding the result for $C_-(t)$ obtained earlier.

15.2.2 The physical interpretation of the O_2^- Hamiltonian

The results just obtained for the time evolution of the state of the O_2^- ion makes it possible to give a deeper physical interpretation to the elements appearing in the Hamiltonian matrix, in particular the off-diagonal element A.

It is important to note that evolution of the system critically depends on the off-diagonal element A having a non-zero value. If $A = 0$, then the matrix is diagonal in the position representation, which means that the states $|\pm a\rangle$ are eigenstates of $\hat{H}$, and hence are stationary states. So, if $A = 0$, once

an electron is placed on one or the other of the oxygen atoms, it stays there. Thus A is somehow connected with the internal forces responsible for the electron being able to make its way from one oxygen atom to the other. To see what this might be, we need to take a closer look at the physical makeup of the O_2^- ion.

The electron in the ion can be shown to move in a double potential well with minima at $x = \pm a$, i.e. at the positions of the two oxygen atoms. The general situation is illustrated in Fig. (15.2)

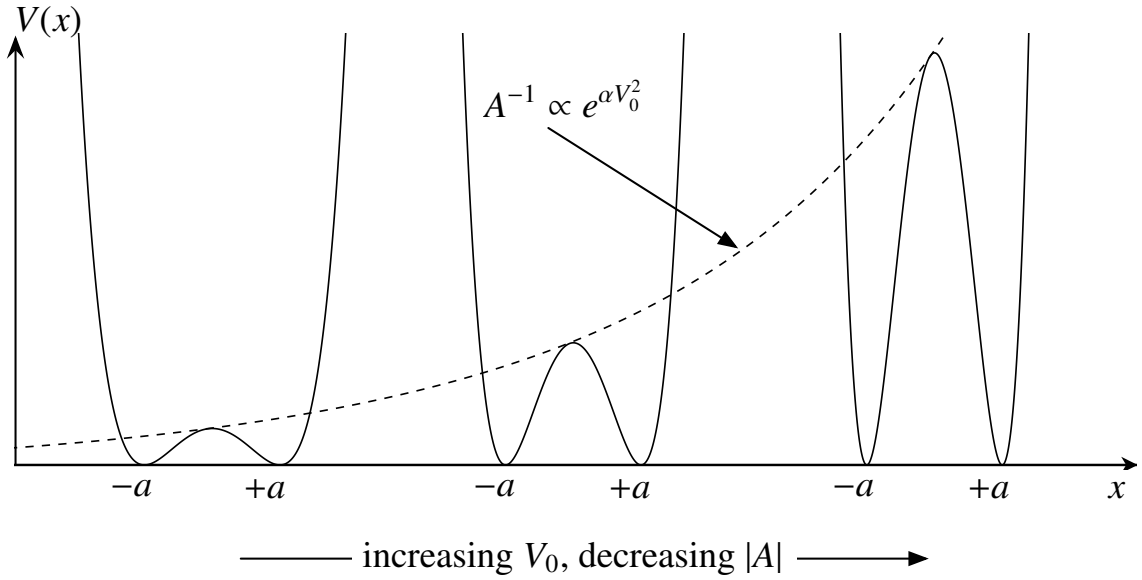


Figure 15.2: Potential experienced by electron in O_2^- ion. For increasing V_0 (the height of the barrier), the off-diagonal element of the Hamiltonian *decreases*, becoming zero when the barrier is infinitely high.

So, from a classical physics perspective, the electron would reside in the vicinity of the position of either of these two minima, unless it was provided with enough energy by some external force to cross over the intervening potential barrier of height V_0. Indeed, in the limit of $V_0 \to \infty$, the electron would never be able to cross this barrier, and so would be confined to either of the oxygen atoms to which it initially became attached. This is also true quantum mechanically – for V_0 infinitely large, the electron would remain in either of the states $\pm a$; these states would be stationary states, and hence eigenstates of the Hamiltonian. This in turn would mean that the matrix representing $\hat{H}$ would be diagonal in the position representation, which amounts to saying that $A = 0$. However, for a finite barrier height the electrons are able to 'tunnel' through the potential barrier separating the two oxygen atoms, so that the states $\pm a$ would no longer be stationary states, and the Hamiltonian would not be diagonal in the position representation, i.e. $A \neq 0$. In fact, it can be shown that the $A \propto \exp(-\alpha V_0^2)$ where α is a constant that depends on the detailed physical characteristics of the ion, so as V_0 increases, A decreases, the chances of the electron tunnelling through the barrier decreases, and the oscillation frequency of the probability will decrease.

15.3 The Time Evolution Operator

Chapter 16

Displacements in Space

Chapter 17

Rotations in Space

Chapter 18

Symmetry and Conservation Laws