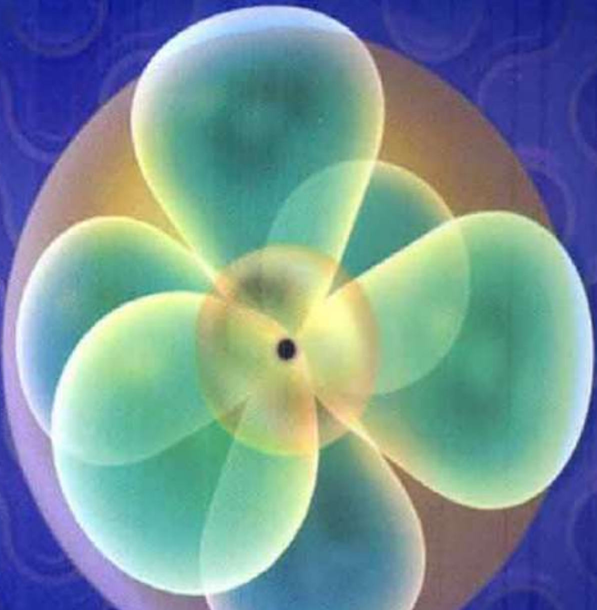
The Wonderland of Quantum Theory

Muhammad Ibrahim



The Wonderland of Quantum Theory

Muhammad Ibrahim

Introduction

There is a touch of magic in the word ‘Quantum’, as it creates an aura of mystery among all. Unfortunately, the science of quantum theory is hardly presented to any one but the students of higher physics and chemistry, in spite of its being the frame-work theory of modern physical sciences. This deprives the others of the joy of having at least a glimpse of this wonderland, making them understand what it is all about. This also makes it easier for others to misuse the word quantum just to give a color of mystery to anything they are doing, usually far from science.

It is true that quantum theory is mostly based on higher mathematics, and a full understanding can not be achieved without pursuing it mathematically. But it is still possible to get the essence of the theory from a simple description in ordinary language. That is what we would try to do in this book. Occasionally in this book, there would be a bit of math here and there, but these are basically nothing but simple arithmetic or graph drawing. Even the starting points of some known basic physics from school days have been briefly repeated here to refresh the memories– things like laws of motion, laws of electromagnetism, waves, radiations etc. One has to start from here, because the quantum theory has fundamentally changed the very notions of this long-established science that constituted the all-powerful edifice of what is called classical physics. It is from the understanding of this revolutionary change from the classical physics to the quantum physics that brings us to the ABC of the latter.

The last decade of the 19th century and the beginning one of the 20th saw several crises in the classical physics, in each of which only a totally new quantum idea had to come to the rescue. The rescue process itself in each successive case advanced the quantum idea a bit further, until it became an independent, self-contained, revolutionary theory in itself. The crises were about the radiation of the electromagnetic waves, the behavior of electrons coming out of metal surfaces when light is shone on these, and the correct structure of an atom, in that order. In each case the very foundation of physics was shaken, and only the prescriptions of a quantum nature could save it. We tried to follow this development historically to get a glimpse of the theory.

The final theory has a lot of bizarre content which go against our classical way of thinking, and hence the wonderland. Here the radiation has to come in discrete packets (quanta) of energy which can not be divided; waves behave like particles, and particles behave like waves, giving rise to a strange wave-particle duality; all knowledge of the states of a system come only as probabilities, not as a definite descriptions of motions of particles; there are pairs of parameters such as the position and the speed of a particle, in which if one is very certainly known that the other will be extremely uncertain; and so on. All these tend to make the physical world very fuzzy and uncertain, but in fact it was found that these give the best interpretation of the nature possible, and the results come in very precise correct form.

The quantum theory, however, is only applicable to the microscopic things of atomic dimensions– electron and other such particles, atoms, molecules etc. For anything bigger than that our good old classical physics is quite sufficient, as the quantum nature is negligible in these macroscopic dimensions we usually deal with. If it had not been so, and the quantum theory would be applicable here too, then all those bizarre things would face us in reality in the day to day life too. For example, than we would find a tennis ball which is confined in a very small box violently vibrating there and frantically shooting out of the box at a tremendous speed in any arbitrary direction, without leaving any hole or mark on the box. In another scene, a football dribbling gently on the ground may suddenly make a huge leap up, so much so that we would find it strangely on the other side of a 100 meter solid wall. In yet another scene, we may have two identical spinning tops made together from the same source and find that if one starts rotating in clockwise direction the other rotates in the counter clockwise direction; if the former take a new direction of its axis, the other will automatically take an opposite direction, and this would happen exactly the same way even if both are separated by thousands of miles from each other. Similar things do happen in the wonderland of the quantum theory; we do not experience these because the quantum theory is not applicable on bigger things.

Though we do not deal with things of atomic dimensions directly, yet everything we deal with are made of particles, atoms and molecules which are subject to quantum theory. So many of the bigger things and events of our physical world can only be explained in terms of quantum theory, nevertheless. That is how common things like emission of light, chemistry, crystals etc. can only be explained in terms of quantum theory. And certain things such as laser, semiconductors, superconductors would not even exist without the benefit of the quantum principles. We have tried in the book to have a glimpse of the process through which all these phenomena depend on the quantum theory.

But in spite of all these spectacular successes, the quantum theory goes against our normal way of thinking, because our brain is wired to think classically, not quantum theoretically. Therefore, everything has to be made classical ultimately for our own comprehension and measurement, all our measurement instruments and our way of sensing these being classical things. In the ultimate analysis, the quantum theory remains a wonderland, even for the scientists.

Muhammad Ibrahim
February 2021

Content

Physics as we knew it

This book intends to tell the story of quantum theory in physics. The story may give you a feeling as if you are in a wonderland, because it changed some of our well-established notions about nature. In many ways the theory takes us out of our comfort zone, as our thinking process itself matches better with the older way of doing science. However, we can still do a lot of physics in the old way, the quantum effects being negligible in that part. But the quantum theory has fundamentally changed physics forever, and now dominates the way we describe nature. This is inevitable because we have gradually realized that the nature itself wants to be revealed in the language of quantum. We will, however, start with a glimpse into our old physics, the one within our comfort zone.

The classical physics, that suites us well

First a preview of the wonderland:

In the old physics as well as in our common sense world, it was possible to find out where a particle is at this moment and then calculate from its speed where it will go the next moment. But in the quantum world we can not do that. All we can do is to calculate each of the various probabilities of the particle's existence in various points in an infinite space. When we actually measure the position of the particle, it will match the probabilities very precisely. This matching is true only for the totality of the final result. If we want to follow one particle to see how it goes from one point to another, we will be disappointed. In the quantum world there will also be some fundamental uncertainty that prevents us to measure a particle's position and velocity (speed and directions) precisely at the same time. If we make the position very certain, then the velocity will be infinitely uncertain – it can have any velocity you can think of.

This type of uncertainties can give rise to some very strange phenomena. For example, it will enable an electron to cross a voltage barrier many times greater than its own energy, which is like a boy or a girl suddenly going over a 30 meter high wall and ending up on the other side of it. In another example of such uncertainties, a perfect vacuum devoid of any matter and energy can happily bring into existence numerous virtual particles to vanish the next moment.

In the quantum world, something may exist both as a particle and a wave, and may demonstrate the properties of both. So an electron or an atom which we know to be material particles can go through two holes at the same time just by taking recourse to its wave nature, and then get superimposed on itself to create effects which we expect only from a wave. But all these strange behaviors are really the behaviors of the nature. Many of the mysteries of nature that defied the old physics could be explained only by quantum physics – very essential things like material properties, chemical bonds, semiconductors, superconductors, crystals etc. New phenomena such as laser, Bose-Einstein condensation etc. could happen only because nature behaves as in that wonderland.

Classical notions about motion:

The physics we have been familiar with till the year 1900 is called classical physics, in the sense as we refer to classical music or classical paintings. When Newton established the basics of motion and made it capable of predicting the motion of everything with a clockwork precision, the classical physics was born. From then on scientists had full confidence on this and developed it further into a beautiful edifice which reached its best at the end of the 19th century. It was this long period of inspiring success that made the later physicists to call it ‘classical’, which is also a convenient term to distinguish it from the revolutionary and bizarre theories that came after. As human beings we tend to think in that classical way; and conveniently it had been the framework of physics till the year 1900, roughly speaking. It still remains so if we confine ourselves to the discussion of familiar things – things of our size, things we can see, touch, or handle. Let us see some of the basic ideas of classical physics that got shaken up with the coming of quantum physics.

Newton’s three laws of motion and his gravitational theory form the backbone of the theory of motion in the classical physics. The laws of motion simply say that all objects are in a uniform motion in a straight line unless it is acted upon by a force. A force will bring a change to the motion either in speed or direction, the change being called an acceleration. It also introduced a property of the object called mass– which resists that change, also the concept of action-reaction– every action has an equal and opposite reaction. The theory of gravitation introduced a universal attraction between any two masses, the first of the four such forces of nature eventually discovered. Two similar forces with similar laws were discovered between electric charges and magnetic poles – these two however include a repulsive force in case the charges or the poles are similar. These two forces were later combined into one called electromagnetic force.

The theories of motion enable us to calculate precisely from the present position and velocity any future position that particle would take. It is customary sometimes to use momentum in place of velocity in such problems, the momentum being nothing other than a product of mass and velocity. If ‘m’ is mass and ‘v’ is velocity then the momentum is mv. The impact that happens during the collision of two objects is better described by momentum than just by velocity. This we realize when comparing the result of the collision between two light cars (low mass), with that between two heavy trucks (high mass) though speeds are kept same in both cases. Momentum has another special property – it is conserved within a closed system. This means if the system is not acted upon by any outside force, any number of objects that are within the system can interact and collide among themselves without affecting the total momentum of the system. The total momentum will always be the same – it will be conserved.

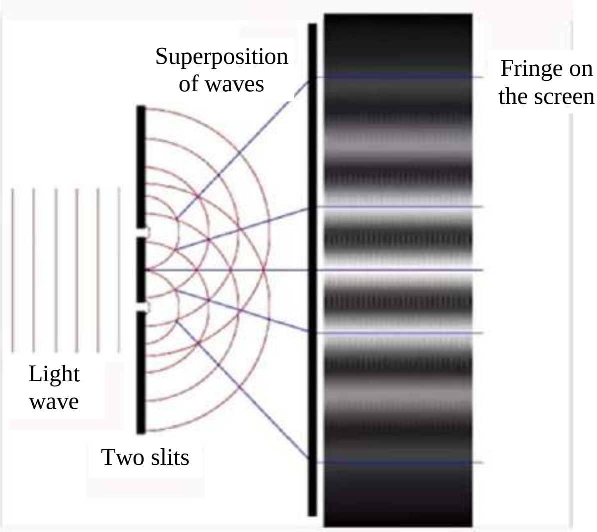
Another thing concerned with motion that is conserved in a similar way is angular momentum. This one is the momentum demonstrated in circular motion in contrast to the momentum or linear moment mentioned above, which is demonstrated in straight line motion. Angular momentum has to include the radius of the circle of motion along with the mass and velocity in the product that defines it. If ‘r’ is that radius, the angular momentum will be mvr. For our purpose we have to mention one more thing which is conserved – energy. Energy is the amount of work done. Motion may manifest energy, then we call it kinetic energy, which can be calculated if we know the mass and the velocity of the object carrying energy. But there is another kind of energy which is manifested by the particular circumstance or configuration of the object; for example, an object suspended above the ground, or an elastic band kept in a an extended state, or a charged battery – these are not showing any motion now, but can do so once they are released, or in the case of the battery, is put in a circuit. In these conditions all these are said to possess a potential energy, which can be converted into kinetic energy. Similarly, an electric charge in a position near another charge also has a potential energy. We get the total energy of a system if we add up all the kinetic energies and potential energies.

We have been talking about the energy being demonstrated by an object, which we can represent by a particle– the minuscule form of it. But in classical physics there is another entity called wave, which too can carry energy. Particles themselves move when they show kinetic energy. But in the case of wave, it is the propagation of an oscillation of the medium that carries the energy, not the propagation of the medium itself. In case the wave does not have a medium, situation is even more clear. In classical physics a carrier of energy is either a particle or a wave, it can not be both. To understand the basics of a wave we can look at one example such as water wave – which is quite visible. Just put your hand in water and take it out, you already have created a wave. Continue to do the in and out motion with your hand you have a continuous wave because of that oscillatory movement of the hand. Every wave needs an oscillation. Everywhere in the progression of the wave there is the up and down movement of water, the medium of the wave in this case. But while moving up and down the local water is not progressing with the wave, but remains in the same location. Now this periodic motion of water goes up to a peak of the wave and then comes down to a trough, above and below the normal surface of the water. We can represent this in a graph showing the peaks and troughs above the zero line (normal surface). The height of the peak above this zero line is called amplitude, which represents the intensity of the wave, showing how energetic it is. We can see the wave-shapes coming from of the source of oscillation (in this case the hand), following each other. The number of peaks passing any point in a given time (usually a second) is called frequency. The distance between the two successive peaks is called the wavelength. Obviously, the longer is the wavelength, the lower is the frequency. The wave has a speed – the length it reaches during its progression in a unit of time (usually a second).

In the case of the water wave, the direction of oscillation (up-down) is perpendicular to the direction of the progression of the wave; such waves are called transverse waves. Some other waves such as the wave that passes through a spring as successive contraction and elongation, are different in that the directions of the oscillation and the progression are the same. These are called longitudinal waves. Sound wave is another example of this. These too, however, can be represented in a graph exactly in the same way as the transverse waves. In the case of water wave, things are so visible and so slow that one can measure the amplitude, the wavelength, the frequency, and the speed easily; maybe even using very simple instruments like a ruler and a watch. This may not be the case with other waves which are not so visible or so easy to measure. We have then to depend on indirect methods of measuring these based on the universal properties of waves, which we examine next.

Some of the properties of waves such as reflection and refraction are well known; the former says when a wave reaches a barrier it comes back, and the latter says when a wave goes from one medium to another it bends. There are simple rules that these follow. Another property diffraction also makes a wave to bend when it passes through a narrow slit or passes by an edge. But the most important property of wave we will consider here is interference, which is best used to distinguish the waves from particles and to measure the various parameters of the wave by indirect method. We can understand this better if we arrange an experiment called ‘double-slit experiment’ with water wave.

Let us create an water wave in a pond and take it through two vertical slits cut on a plank. The two slits act as two new sources of two water waves. After going some distance from the plank these two waves are superimposed (falls on) each other. Now depending on the distance from the slits the situations of superimposition will somewhat vary at various points along a line parallel to the plank. The place where the peak of one wave falls on the peak of the other, the two peaks get added up and a maximum height of the water is achieved. But just by the side of it on the line the two superimposing waves have the peak of one falling on the other’s less than the peak stage. So the sum here will be smaller, and the height of water less high. There will thus be less and less heights as we move along the line. At one point one wave which is going up will fall on the other which is going down, so somewhat cancelling each other and making the sum even less. The next still the peak of one will fall on the trough of the other cancelling this completely and producing a zero height at the place. Thus between the maximum height and the zero height there will be various intermediate heights, and the next to the zero height intermediate ones again, and then the maximum again. Thus as we see along that line we find various heights – maximum, intermediate, zero, intermediate, another maximum, intermediate, another zero and so on. It will be as if a comb-like fringe is there and we can call it a fringe created by the interference of waves. If we get a fringe like that in a double-slit experiment we know that it is a wave. And if we measure the distance between the two slits, the distance between the slits and the fringe, the distance between the two consecutive maxima in the fringe, we can, from the theory of wave, easily calculate the wavelength and the frequency of the wave.



Double-slit experiment for light

This is how it was confirmed that light is a wave – from a double-slit experiment with light. The basic principle of the experiment remains the same, only the slits have to be much narrower because of the much shorter wavelength of light. If we use a light source which gives only one frequency (color) of light then we get a simple fringe on a screen placed at a distance from the slits. Though in this case waves can not be seen but the result of the interference of the waves can be seen as the fringe – a maximum light (brightest) in the middle and then on each side next are intermediate light and the zero light (dark), then intermediates again, and maximum again, and so on. Exactly similar sequence is obtained on the other side of the middle. All these happened exactly for the same reasons as we have seen in water wave, the superposition and interference between waves. The existence of this fringe proves that light is a wave, and the measurements of distances like before gives us the wavelength and frequency of light. Then various methods were invented to measure the speed of light which was found to be 300,000 kilometers per second.

Going back to our double-slit experiment with water, let us imagine we are not using water in it, rather we are using ping pong balls. The balls comes in a stream from a source and go through the two slits cut on the plank. Let us place a screen parallel to the plank at a distance from it. Would we see a fringe of balls with maximum number of balls striking in the middle, then intermediate number next, then no ball next, again intermediate and again a maximum? No, all we will see is that all the balls are striking the screen in only two places directly opposite to the slits, and that is all. This is because ping pong balls are not waves. They are particle-like. If light would be particle-like as Newton used to wrongly claim, then we would have got the same thing with light too, according to the classical physics – two bright vertical lines opposite the slits, and that is all.

Electromagnetic wave:

The theories of electricity and magnetism in the classical physics were fully developed in the 19th century, and became as important as Newton's achievements with the theory of motion. We have seen the two new forces electrical and magnetic were described in somewhat similar manner as gravitational force, these being later combined into electromagnetic force. Clerk Maxwell, the British scientist combined all the findings on electricity and magnetism and beautifully put them in four equations famously called Maxwell equations. He took up a concept of 'field' already discovered by another British scientist

Faraday – a kind of space of influence created by a mass (then it is a gravitational field), or by a charge (then it is an electrical field), or by a magnetic pole (then it is a magnetic field). If another mass, another charge, or another pole enters the field as the case may be, it will experience a force according to the point of the field at which it is. The force will be attractive in case of gravitational field, but attractive or repulsive according to the sign of the incoming charge or pole and the direction of the field. The direction depends on the sign of the first charge or the pole. According to Maxwell equations, we can create a magnetic field by changing an electrical field, the intensity of the field will depend on how rapidly the change takes place – the rate of change. Similarly we can create an electrical field by changing the magnetic field. The first of these principles we use when we design an electric motor, and the second principle we use when designing an electric generator. The two principles (the Maxwell equations) taken together can take the shape of a wave equation describing a wave where electrical field is oscillating and creating a wave, and that oscillation is also giving rise to a magnetic field that too is oscillating, and both propagating as a combined electromagnetic wave. Here in the electromagnetic wave, it is the electrical field and magnetic field that undulate everywhere in its progression. The plane of undulation of the two remain perpendicular to each other and also to the line of progression of the wave. As it is the fields that are oscillating (undulating) here, which can exist in vacuum too, electromagnetic wave does not need a medium.

The wave equation of the electromagnetic wave that Maxwell formulated indicates a speed of this wave in terms of two universal constants – originally part of the equations of electrical force and magnetic force (as the universal constant of gravitation is the part of the gravitational equation). From the precisely known values of these two constants the speed could be calculated, and what a surprise, the value of speed came exactly equal to the speed of light – 300,000 kilometers per second! This can not just be a coincidence. Maxwell had to conclude that this happened because light itself is nothing but an electromagnetic wave, not just in the visible part of the light spectrum consisting of colors we can see, but also in the extended part of it with shorter or longer wavelengths which we can not see. All these were done by Maxwell only theoretically. But soon German physicist Heinrich Hertz actually demonstrated this practically by creating electromagnetic wave with much longer wavelength than light (later called radio wave) and actually measured its speed, which came to exactly the speed of light. Soon a whole spectrum of electromagnetic waves of various wavelengths was discovered. The short wavelength end of it begins with gamma ray, then x-ray, then ultraviolet ray in order of longer and longer wavelength (all invisible). At this point we get the visible light, than waves of even longer wavelength Infrared, microwave, radio waves all invisible again. All have the same velocity – the velocity of light.

Probability in classical physics:

The thing we call probability is a concept of classical physics, that of mathematics as well as of everyday life. This is a way of predicting some result without actually performing the actions for it. It is based on two different methods. The first one maybe called a principle of equivalence. If there are several different alternative turn ups of an event, and there is an equivalence between all these alternatives without any bias towards any particular alternative, then the probability of any particular alternative to occur is the fraction in which 1 is divided by the number of alternatives available. To take a simple example, with an unbiased coin, a coin toss can yield 2 alternatives – head or tail. So the probability of getting a head will be $\frac{1}{2}$. The probability of getting a tail is also $\frac{1}{2}$. Therefore without tossing a coin even for once, we know that in 100 tosses we are very likely to get head 50 time, just by taking $\frac{1}{2}$ of 100.

The expectation by the probability is usually fulfilled, particularly if the number of events is a big one. Similarly in the case of a perfectly unbiased cube-shaped dice, out of the six surfaces the probability of getting the surface marked three is $\frac{1}{6}$, as the number of alternatives is 6. In practical term if we throw the dice 60 times we should get the surface marked three 10 times.

The second method of calculating probability which is more used in physics is the method of frequency. This makes a prediction on the basis of the past experience. The probability here is given by the number of times a certain alternative occurred in the past divided by the total number of events – which is the frequency of the occurrence of that alternative. Thus, for example, if any horse had previously won 3 out of 10 races running against the same competitors, the probability that this horse will win next time in the race against them is $\frac{3}{10}$. In practical terms this means in the next 10 races this horse is expected to win in any 3, or in the next 20 races should win in any 6, and so on. This is not unfamiliar to those who bet on horses, but this is also an important principle of classical physics. Here too, the probability is a simple fraction. Obviously we use probability to be able to say something about things we are ignorant of, in this case about the horse races that have not yet happened.

Maxwell of Maxwell equation also had the theory of the motion of gas molecules to his credit. It was already established that a gas in a container of a certain volume consists of innumerable molecules always in motion. The more is the temperature of the gas the faster is the average motion. It was important to know the nature of the motion of the molecules to predict the various measurable properties of gas such as pressure, density etc. In principle, the motion of each molecule as it moves and collides with other molecules can be precisely calculated using Newton’s laws. But in practice, it is impossible to do this on all these innumerable molecules. So Maxwell tried another way by finding a probable distribution curve for the velocities of the molecules. He had to make certain assumptions in doing this, assumptions which are reasonable under the circumstances being considered. According to these assumptions– molecules are colliding with each other in a perfectly elastic manner; the motion between two consecutive collisions involves uniform velocity in straight line; the distances between the molecules are far greater than the size of the molecules, minimizing any interaction between them; and most importantly, the initial motion of the molecules are completely random, each one having a random speed and direction.

The distribution depends only on the temperature of the gas – the higher is the temperature, the faster molecules are more represented in the random population of the molecules. The distribution is a graph showing the proportion of the population within a certain velocity range against that range. This makes roughly a bell-shaped curve, the top of the bell indicating the velocities of the largest proportion of the molecules. The further away the velocities are from these velocities, the less is the proportion. This does not tell about the velocities of an individual molecule, but can give an overall statistical picture which is valuable in calculating the measurable parameters. The whole thing depends on the concept of probability and this approach to study motion is called statistical dynamics.

Classical physics revealed a fundamental law of nature that in all conversions of heat to work, some heat is lost and be unavailable for useful work. This happens because a quantity known as entropy always increases within a close system whatever we do. Entropy can roughly be described as a measure of chaos or lack of discipline. Austrian scientist Boltzmann gave a beautiful method to explain the increase of entropy in the case of a gas. He showed that there is a finite number of ways in which molecules of a gas can be arranged in terms of their energy content. He could formulate an equation showing that the more diverse ways they can be arranged, the higher will be the entropy. The equation is true about any system, not just a gas.

To estimate the number of possible arrangements Boltzmann took the help of an artificial assumption. In classical physics energy content can be anything from zero to any quantity – because energy is not granular like particles, it is a continuous entity. Even so, Boltzmann assumed for the sake of ease of calculation that there is a minimum quantity of energy possible which he called ϵ. All other energies can come as multiples of ϵ, such as 2ϵ, 3ϵ, etc. He imagined that the total energy of the gas is distributed in various containers capable of having energies that can be counted in a certain unit. Thus the first container can have it only in the unit of ϵ, the second one in the units of 2ϵ, and the third one in the units of 3ϵ etc. Now the various number of molecules can belong to these containers. Those in the first container all have an energy ϵ, those in the second all have energy 2ϵ, and so on. This gives a simple method of calculating the ways the molecules can be arranged. To take a very simple example, let us say the total energy of the gas is 4ϵ, and there are only 3 molecules in the system. Now how many ways we can distribute energy 4ϵ among the 3 molecules? One way will be to have the first 2 molecules with energy ϵ and the third one with energy 2ϵ. Another way will be to have second and third molecule with ϵ, and the first with 2ϵ. A third way will be to have first and third molecule with ϵ, and the second with 2ϵ. There are only 3 ways and no more. Entropy can be calculated on the basis of this. Maxwell's and Boltzmann's methods were all within the classical physics – but these gave important clues to the scientists who came out of the classical physics.

The statistical dynamics by Maxwell and Boltzmann made possible another important principle called the principle of equipartition of energy. In a closed system there are certain 'degrees of freedom' for the molecules within it. If we consider only one molecule it is free to move in the 3 dimensions of the space – sidewise, up and down, and front and rear; so it has 3 degrees of freedom. If there are 'n' number of molecules the number of degrees of freedom will be 3n. Now the equipartition of energy says that for a definite temperature of a gas each degree of freedom must contain a definite and equal amount of energy. This energy multiplied by the number of degrees of freedom constitutes the total energy. In other words the total energy is distributed equally among all degrees of freedom. This principle too played a part in the first thoughts about the quantum theory.

Three crises in the classical physics

The basic concepts of the quantum theory had to be mooted first in response to some serious crises that arose when the 19th century was turning into the 20th century. There were three such crises each of which was a blow to the well-established classical physics. The way out of all these could be achieved only by conceiving new concepts, revolutionary departures from the classical physics – all indicating towards the quantum theory. Actually the solution of these three crises laid the foundation on which the elaborate edifice of the quantum theory could be built later. Let us, therefore, look at these crises and their solutions, one at a time.

The crisis of the blackbody radiation

An oven with an infinite energy!

The first crisis arose when some well-founded classical equations were predicting a strange result – whatever is its temperatures any oven will have an infinite amount of energy! Obviously this is not the real case, as otherwise we would be burned to oblivion every time we go near an oven. Actually the predictions those theories gave were not only about an ordinary oven, it applies to any hot cavity with a small hole through which radiation can come out. Such radiation is referred to as blackbody radiation, because when all the radiations are absorbed the thing is black, and in a cavity almost all the radiation is absorbed by its walls. The radiation coming out of this is a mix of all frequencies of electromagnetic waves including the visible light. Usually at the temperatures the ovens are used, the predominant radiations are in the infrared (heat) range. If now the temperature is increased, the predominant radiation will move towards the visible range starting with the low frequency color dull red, then increasing in frequency to orange, yellow and white (a mixture of all) as the oven temperature is increased. If we look through the hole we can see the color changes. The blacksmiths often estimate the temperature of their furnaces just by looking at these colors. With the precise measurements of the frequency and the intensity of radiation at that frequency, scientists could make a graph of intensities against frequencies, which would give roughly a bell-shaped curve similar to that of Maxwell's distribution. One such curve can be drawn for each temperature of the oven. With the increase of temperature the peak of the bell curve goes higher and moves towards a higher frequency. Now this curve being based on actual measurement, what would be the prediction from a theory?

Following Maxwell's statistical approach in a somewhat similar problem of proportion of molecules with a certain velocity against the velocities of molecules, a theoretical calculation of the blackbody radiation was made – this was called Reighley-Jean's law, using the names of its two discoverers. The graph drawn from this law coincides nicely with the measured graph in the low frequency regions, but sharply deviates from it in the higher frequencies. Instead of peaking and coming down on the higher frequency side, the intensity continues to go up and tends towards infinity! This is how the theory predicts an infinite energy for the blackbody radiation. This is where the classical physics faces a crisis – a theory based on the essential concepts of classical physics failed miserably to predict the reality. As this gross deviation takes place in the ultraviolet region of the black body frequencies this was called ultraviolet catastrophy. Another scientist Wien came up with classically sensible formula (not quite a theory) which fitted with the higher frequencies of the curve from the measurement, but failed in the lower frequencies. So the crisis remains.

One may consider the issue of the infinite energy of an oven from another point of view. For simplicity let us consider the oven to be a rectangular box. We may imagine that electromagnetic waves are being generated from the numerous little oscillators of one wall of the box and then either get absorbed on the opposite wall or get reflected from that. This goes on between all the three pairs of walls filling up the oven with all frequencies of electromagnetic waves – every frequency of which can be considered as an independent degree of freedom. According to the principle of equipartition of energy, that we came across in the last chapter, every degree of freedom will have a definite energy determined only by the temperature of the oven. Whatever this definite energy is, the total energy is bound to be infinite because of the infinite number of different frequencies. This is the ultraviolet catastrophy, we have already seen predicted by the classical laws.

The idea of quantum rescues the situation:

One of the scientists working on the ultraviolet catastrophy was Max Planck of Germany. He started from a picture of the oven we gave above – innumerable oscillators producing and absorbing electromagnetic waves. In formulating a theory for this he followed the statistical method by Maxwell and Boltzmann dealing with the distribution of molecules of a gas according to their velocities. In Planck’s case the problem changed into the distribution of total energy among those oscillators. He needed to know in how many ways such a distribution is possible, and he decided to use a similar mathematical trick that Boltzmann used. Whereas Boltzmann imagined a minimum amount of energy ϵ, and energy of various molecules to be as ϵ, 2ϵ, 3ϵ etc. Planck portioned the energy radiated by the oscillators into slices (packets) of fixed quantities of energy. Then he considered in how many ways these can be distributed among the oscillators, and as a part of the trick he had to assume that the amount of energy in a packet will be proportional to the frequency of the electromagnetic wave it is connected to. The energy of the packet thus can be expressed as $E = hf$, where E is the energy, f is the frequency and h is the proportionality constant to be determined, which later came to be known as Planck's constant. The word quantum means a quantity of energy, in this case the energy of the packet. Thus in this assumption energy was not treated as a continuous entity that can take any amount, but discrete quanta or packets of the amount hf; no fraction of those packets were possible.

This picture of discrete energy packets could solve the ultraviolet catastrophy. According to the principle of the equipartition of energy every frequency of the blackbody radiation should have an equal definite energy – not more, not less. Now that the energy has to come in the shape of discrete quantum (packet) each equaling hf, that definite energy has to be obtained in terms of certain number of these quanta. At low frequencies hf being very small, a certain number of these can fulfill the energy requirement of that degree of freedom (frequency). But when the frequency reaches a high value so much so that if the equipartitioned energy for that is to be fulfilled with by now large quanta (bigger hf) there may easily be a short fall or an excess which has to be adjusted by using a fraction of a quantum. But this is not allowed. As a result the radiation of such very high frequency would not be possible, nor will be anything above that frequency, for the same reason. Therefore, though the classical theory predicted infinite number of frequencies Plank's idea of quantum makes it finite. an oven does not radiate infinite energy after all; the crisis of blackbody radiation is solved. But it had to involve the idea of quantum, not quite approved by classical physics.

Planck used the quantum concept of energy along with the established classical concepts to give a new formula for the relation between radiation intensity and frequency, to replace the Reighleiy-Jeans theory and and Wien's formula. Unlike these, Planck's formula could be made to match the measured curve in all frequencies. To do this Planck had to try out various values for h, the Planck's constant, in the formula until he achieved a perfect match of the formula with the measured curve for all temperatures of the blackbody. This he realized is the correct value of h, a universal constant of nature. The value of h, thus found, is extremely small – only a billionth of a billionth of a billionth of one ordinary unit! This extremely small value of h show the very fine-grained nature of the energy quantum (as this energy is $E = hf$), but nevertheless even this miniscule packet of energy made such a difference to the classical concept of continuous energy, and much else in the classical physics (as we will see), that it had to give way to the quantum theory, remaining relevant only for things which are much bigger than h.

Plank introduced this very innovative concept of quantum in the year 1900. This gave rise to other discoveries ultimately resulting in the revolutionary quantum theory in about three decades. Planck himself, however, did not regard his theory as a revolutionary one. He tried to do everything within classical physics and did not mean to overturn it at all – perhaps not more than what Boltzmann thought when the latter introduced a minimum amount of energy as a trick to calculate entropy.

The crisis of behavior of photoelectrons

Intensity does not determine the energy of light, frequency does:

The scientist who demonstrated practically the electromagnetic wave as radio wave, Heinrich Hertz, made another important discovery– that of photoelectric effect. He showed that when light is shone on a metal plate a negatively charged ray called cathode ray came out of the metal (called a cathode plate). Another positive charged plate (called anode plate) placed near it would attract this cathode ray and an electric current would result if the circuit is completed. Here the cathode plate is the emitter of the ray and the anode plate is the collector. Later it was found that the cathode ray consists of negatively charged particles called electrons, which are ejected out of the cathode plate by the energy of the light. This is why these electrons are called photoelectrons, photo meaning light.

Further research made it clear that the energy of the light is transferred to the electrons at the surface of the cathode plate, increasing their motion and allowing some of these to come out in the space above, cutting off their bond with the metal. The remaining energy of these ejected electrons give them some velocity contributing to the current mentioned above. In one experiment with a fixed positive voltage on the anode if the intensity of light falling on the cathode is increased, there would be an increase in the current. If we make the anode plate negative but less negative than the cathode, then still there will be a current, though reduced. If the negative voltage at the anode is increased to such an extent that the current stopped altogether at that point, we can assume that this voltage is just enough to repel the electrons to stop them from arriving, exactly cancelling the effect of their own energy of motion. Thus the negative voltage required to stop the current will give a measure of the electron energy. Now if we repeat this latter experiment using different frequencies (colors) of light each time, it is found that as the frequency increases we need a higher negative voltage to stop the current, which means it is the frequency which determines the energy of the electrons. If we repeat this experiment with one frequency, and increase the intensity of light, the negative voltage to stop the current remains the same, showing that the intensity of light has no effect on the energy of the electrons. The previous experiment which gave higher current for higher intensity of light (with anode plate positive) only showed that a higher intensity of light ejects more electrons, thus increasing current.

This is a completely strange result from the point of view of the classical physics, and thus gave rise to another crisis in it. According to classical physics, light being a wave its intensity determines its energy. Therefore, at a higher intensity of light more energy should be transferred to the electrons at the surface of the cathode plate. This is what we find with other waves. For example, when the water waves break on the sea shore, striking and ejecting small pebbles, we see that higher waves make the pebble jump more energetically. But the reality of the photoelectrons is different; it is the frequency of the light that solely determines the energy of the photoelectron. There lies the second crisis of the classical physics, this too at the beginning of the 20th century.

Accept the quantum as a particle, and the crisis is over:

The crisis of photoelectron was solved by the young Einstein in 1905. He did it by going all the way – recognizing the Planck's quantum as a full fledged particle! He argued that the behavior of the photoelectron seemed strange because we regard light as a wave. The behavior will seem perfectly normal if we think that light is constituted of little particles. Then we can imagine a light particle coming down on the cathode plate where it is colliding with another particle, an electron, and transferring energy to it the way a marble coming down and striking another marble transfers its energy to the latter and ejects it out of the ground. If we combine this picture with Planck's concept of quantum and his relation E = hf, then we can think the quantum to be the light particle, with higher energy at higher frequency. In that case the energy of the photoelectron will naturally depend on the frequency of the light, and not on the intensity of the light. The photoelectron crisis will then be over. Later, the light particle was given the name photon.

However, while the photoelectron experiment shows light as particles there are other experiments known for long, which shows light as wave, the best of all – the double-slit experiment which we have discussed in details. Thus Einstein introduced the concept of wave-particle duality – that light is both a wave and a particle, but not both at the same time. In certain circumstances it behaves as a wave, and in others as particles. This is against the edict of classical physics which says an entity can either be wave or particle, but never both. Photoelectron theory was another crisis, the solution of which made the quantum theory inevitable.

The crisis of spectral line and atom model

Bright lines and dark lines in a spectrum:

Long ago in 1752 a Scottish scientist Thomas Melville noticed that if some hot gases are produced on a flame, the flame takes a characteristic color of that gas. For example, if some common salt is thrown on a colorless flame it takes yellow color, which is the characteristic color of sodium vapor from the sodium chloride (common salt). Sodium lamps that lit our streets also get their yellow light from the sodium vapor. Neon lamps, on the other hand, give red light, characteristic color of neon vapor. Many more things were discovered on these characteristic colors since Melville's time. It had been known from Newton's experiment that white light can be split into a spectrum of colors – violet, indigo, blue, green, yellow, orange and red (VIBGYOR) in increasing wavelength (decreasing frequency). In the more modern instrument called spectroscope a prism is placed at the centre of a round platform, it splits the light coming in the shape of a vertical line through a slit and the various colors go into various angles, which can be observed through a telescope that can be rotated around that prism. The telescope thus can be focused on any part of the spectrum. In the case of white light at the source, the spectrum comes continuous– the violet getting blended into the indigo next, and the indigo into the blue and so on; without any gap anywhere. But when we examine the light of a single color (single frequency) created by a hot gas, things are different.

In that case the spectrum will give us some bright lines characteristic of the hot gas which is giving the single color light. For example, if the source is a sodium lamp we get two bright yellow lines with a little gap in between, these in the region where the yellow should be in the white light spectrum. Apart from these two lines all else is dark. To take another example, the spectrum of the light coming from hot hydrogen gas give several lines one each of red, green, blue and violet, each in its appropriate place. A single such line means a specific frequency of light that can be found out from the calibration of the spectroscope. This means hot hydrogen gas is emitting four different frequencies of light represented by those four bright lines. These are called emission lines, and the spectrum an emission spectrum. The set of lines in each case form a pattern that is unique to a hot gas of an element; no two elements will have the same pattern. Thus the pattern serves as a sort of fingerprint for the element.

We can get another type of spectrum when white light is taken through a long container full of cold gas and then is examined in a spectroscope, we get the usual continuous white light spectrum VIBGYOR, but superimposed on this are dark lines characteristic of the element of the gas, in the appropriate regions of the white light spectrum. These dark lines correspond exactly to the bright lines in the emission spectrum of the same element. This can be explained by assuming that while coming through the cold gas certain characteristic frequencies of the white light have been absorbed by the gas and the light of those frequencies are therefore absent in the spectrum, creating dark lines there. The black lines are called absorption lines and the spectrum is called absorption spectrum. The fact that for the same element the pattern of bright lines in its emission spectrum corresponds exactly to the pattern of dark lines in tits absorption spectrum shows that the process of emission of light and the process of absorption of light are corresponding processes.

The absorption lines in the spectrum of light coming from the sun or any other star can be used to identify the elements which are there. For example, two black lines near each other in the yellow zone of sun’s spectrum corresponding to the bright lines from sodium vapor in the laboratory, indicates that the white light coming from the very hot central part of the sun got two yellow frequencies absorbed while coming through the comparatively cold gases in the outer zone of the sun.

Now this pattern of the two yellow lines is a sort of fingerprint for sodium, and that tells us that the sun has got sodium in it. In a similar way other elements were found in the sun just by looking at the dark lines in the spectrum of the light from it. In one case the scientists got a black line pattern which did not match with the emission spectrum of any known element. They concluded that it is coming from an element in the sun which has not been discovered in the earth as yet. This confidence on their spectral lines was soon vindicated when a new element with exactly the same emission and absorption spectrum was discovered here on the earth too. This was named helium to remind that it was first found in the sun, Helios being the Greek name for the sun.

The question relevant to current discussion was why does an element emit a very unique pattern of frequencies of light, and why does it absorb exactly the same frequencies. It was known that the one element differs from another basically in its atom, so the uniqueness must be coming from the differences in their atoms. This difference could be understood only if there were a knowledge about the structure of the atom. This was another big problem for the classical physics at the turn of the century, and we will see soon that the third crisis in the classical physics came through this quest for the structure of atom and for discovering the process of light emission and absorption by the atom. But before going into that let us mention how already a formula for predicting the emission and absorption lines was found out in a manner of solving a puzzle.

A Swiss school teacher Balmer did just that in 1885. He just thought out a formula by trial and error method which was successful in predicting the various spectral lines for hydrogen, the simplest element. These included the four visible lines for hydrogen mentioned before, and some more lines of it in the invisible ultraviolet region later discovered. It is not a small matter to fit one formula with all these lines irregularly arranged through the spectrum. It would be extremely unlikely that such a formula could exist if there was no physical basis underneath it. But Balmer arrived at the formula without any knowledge of that basis, purely by a puzzle-solving exercise.

His formula, known as Balmer's formula is the following

$f = R\,(\frac{1}{n_f^2} - \frac{1}{n_i^2})$

Where f is the frequency of the line, R is a constant that Balmer found out by trial and error, and n_f and n_i can be given small whole number values such as 1, 2, 3, 4 etc. n_f standing for 'final n' and n_i for 'initial n'. For calculating the four visible lines Balmer chose 2 as the value of n_f in every case, and 3, 4, 5, 6 as the value of n_i for the lines in that order. The frequencies calculated by this formula matched perfectly with the actual frequencies of the lines measured in the spectrum of hydrogen.

Inner world of the atom:

For thousands of years when philosophers thought about the atom, it was regarded as something like a little marble without any structure, and the idea persisted in the modern chemistry and physics too, at the beginning. But soon the cathode ray coming out of a hot metal plate and finding out that its constituents are lighter than atom and have electrical charge (atom themselves do not have charge) showed that the atom does have a structure, and there is a challenge for the scientists to find out is the insides of an atom. The first guess was made by the British scientist J J Thompson during the last years of the 19th century. He showed that the cathode ray is a flow of little negative charges coming out of the atom (later found out to be electrons with a mass about $\frac{1}{2000}$th of a hydrogen atom). To be charge-neutral atom must have a positive part too, which seemed to comprise most of the mass of the atom. So Thompson assumed that the positive part is continuous throughout the atom and the negative electrons are distributed in it. On a bigger scale, one could compare it with a plum pudding where the body of the pudding is the positive part and the plums spread within the pudding resemble the negative electrons. This made it popular as the plum pudding model of atom. Thompson tried to explain the various properties of atoms in terms of this model including the emission and absorption of light, but the model itself did not last long under further investigations.

Ernst Rutherford, an ex-student of Thompson conducted an experiment in 1908, which became famous by the name gold-leaf experiment. By then it was known that some atoms (radioactive atoms) spontaneously emit energetic large positive particles called alpha particles. Rutherford arranged to bombard a very thin foil of gold by these alpha particles (hence the name of the experiment). He expected that if positive charge in the atom is continuous all over the atom as claimed by Thompson's model, the alpha particles will simply pass through the gold atom without deviating much, as their positive charge can not be repelled much by the low density of the positive charge of the atom. He arranged defectors for alpha particles around the gold foil to confirm this. But what happened was quite the opposite. Many alpha particles did pass through, but many others deviated sharply, and some even came straight back towards the source of the alpha particles. This disproved the Thompson model. Instead, Rutherford concluded from the results that most of the atom is empty space, that is why many alpha particles could go through. At the centre there is a small concentrated positive part, which was called nucleus. It was this which made the other alpha particles to deviate so sharply from their straight path. The electrons in the atom would be going round the nucleus to avoid getting attracted and falling down on the nucleus. This very much resembled a scaled down solar system where the nucleus is the sun and the electrons are planets going round – thus Rutherford’s atom model also acquired the name solar system model.

But this model had a serious defect in the eyes of the classical physics. If a negative charge moves in a circle, according to Maxwell’s electromagnetic theory this oscillation will produce electromagnetic waves, thus radiating energy. So electrons will soon lose all of their energy and will spiral down to the nucleus, thus collapsing the atom. This failure of the Rutherford’s model of atom was really the third crisis in classical physics. If classical physics prevents an atom model which is otherwise so effective, how could the spectral lines be explained?

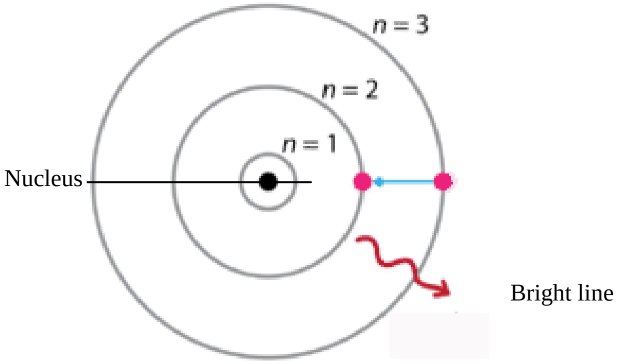
Quantum concept saves the atom model:

A kind of solution for the atom model crisis came from the Danish scientist Niel Bohr in 1912. He brought some alterations in the Rutherford atom model. According to him orbits of electrons in the atom can not be anywhere with any energy. There can only be certain ‘allowed’ orbits with fixed allowed energies for the electrons. The business of allowing arises from those quantum rules that prohibits fractional energy. Any orbit in between two allowed ones would be a violation of the quantum rule. So the electron remains in that allowed orbit; there is no question of spiraling down through those prohibited zones. We have seen that an object in a circular motion has an angular momentum given by mvr where m is the mass, v is the velocity and r is the radius of the circular path. Now Bohr brought this angular momentum from the realm of classical physics to that of quantum physics – allowing only definite discreet values to them which are the angular momenta of the various allowed orbits. This he did by making angular momentum take values only as multiples of $\frac{h}{2\pi}$, so the allowed values will be $\frac{h}{2\pi}$, $2\frac{h}{2\pi}$, $3\frac{h}{2\pi}$ and so on. To write these more easily usually a symbol ħ (pronounced as h-bar) is used instead of $\frac{h}{2\pi}$. Therefore the allowed angular momenta would be ħ, 2ħ, 3ħ etc. This too is an act of quantization, like the one done by Planck who made energy into discrete packets. In Bohr's case it is the angular momentum which is quantized. Thus angular momentum = mvr. = nħ, where n = 1, 2, 3, etc. The various values of n are called quantum numbers – each of the orbits is assigned a quantum number, the further is an orbit from the nucleus the higher is the quantum number.

Niel Bohr also gave an explanation of the emission and absorption of light from an atom using his altered model which came to be known as Bohr atom model. The explanation involves a very strange thing – the transfer of energy from one fixed orbit to another without going through the intermediate region. This is something inconceivable in classical physics. Such transfers are called 'quantum jumps'. According to this process when an atom is excited from its normal equilibrium state, called zero state, to a higher state the electron jumps from a lower orbit to a higher orbit. But this being a disturbance in equilibrium, the electron would soon go back to its equilibrium state by jumping back to a lower orbit by radiating energy as light. The energy thus radiating will be the difference of energy of the higher initial orbit and the lower final orbit – the final orbit. Thus according to the Planck's relation:

Energy radiated = $E_i - E_f = hf$

The frequency of the radiation thus depends on the energy difference between the two orbits. The light thus emitted forms a bright line in spectrum with this frequency. The dark line is formed by a reverse process in which the atom absorbs the light of a certain frequency by making a lower orbit electron to jump to a higher orbit, the frequency related to the difference of the energy of the two orbits matches with the frequency that has been absorbed. Thus that frequency is taken away from the white light spectrum leaving a dark line in its place.



Bohr's atom model: Orbits designated by various n (quantum numbers).

Bohr used the equations $E_i - E_f = hf$, $mvr = n\hbar$, along with some other laws of classical physics to arrive at an equation surprisingly similar to the Balmer's formula. This equation of Bohr's is

$$f = R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

where f is the frequency, n_f and n_i are small whole numbers indicating various orbits, R is a constant which Bohr arrived at combining to the relations the values of various parameters and constants, namely mass of electron, charge of electron, h, and a constant from the electromagnetic theory. From this equation the frequency of the spectrum lines are theoretically obtained through a combination of classical physics and quantum physics. The frequencies calculated from this matched perfectly with the measured frequencies of the lines of the spectrum giving a great victory to Bohr's atom model, and solving the crisis of the atom model. The Balmer's formula too had been similarly successful, but that was not from a theory or an atom model, it was found out by a sheer puzzle-solving method.

n_i and n_f in Bohr's formula are the two values of the quantum number n indicating two allowed orbits, each gives the electron a specific angular momentum $n\hbar$. The same quantum number n gives a specific energy of an orbit in terms of $\frac{1}{n^2}$. The exact match of the frequencies calculated from Bohr's formula with the measured lines of hydrogen was a great success for Bohr's atom model, and for the process suggested by him for the emission and absorption of light by an atom. This had to involve the ideas of quantum theory; but many other things in it followed the classical physics. So at best it was a mixture of classical and quantum considerations. This remained a weakness in Bohr's atom model, but, nevertheless, this was a great step towards a pure quantum theoretical model of atom, which eventually came within two decades.

Matter wave

If wave can be particle, particle can be wave too

The calculated wavelength of possible matter wave: We have seen how Einstein had to accept that light consists of particles – photons, to avoid the photoelectron crisis in classical physics. This moved the quantum theory further forward by admitting a wave-particle duality, unthinkable in classical physics. One general principle in physics is to expect symmetrical phenomena; for example, if something happens in left to right direction, the same thing is usually expected to happen in right to left direction too. Such a symmetry is not universally found, but it is natural to expect. Thus it occurred to some scientists that if, as Einstein showed, a light wave can be a particle, why can't be a matter-particle be a wave?

For De Broglie, a young French scientist in 1923, it did not remain just a thought; he elaborated a whole possible theory in his Ph.D. thesis based on such an idea of matter wave. He even deduced the wavelength of such a wave in terms of the momentum of the particle. To do this he utilized several established equations not only from classical physics and the budding quantum theory, but also from the theory of relativity, another new revolutionary theory. He used photon, the particle of light, for his deduction; but claimed that the result would be applicable to any particle including the matter particle.

The equations De Broglie used were $E = hf$ Planck's equation for the energy of a quantum and $E = mc^2$ the famous energy-mass equivalence equation from Einstein's relativity theory. In these E is energy, h is Planck's constant, f is frequency, m is mass and c is the velocity of light. De Broglie also used the well-known equation relating velocity, frequency and wavelength of any wave. As frequency is the number of waves passing a point per second, and the wavelength is the length of one wave, the velocity, the total distance covered per second, should be given by the product of the wavelength and the frequency. That is

$v = \lambda f$

where v is the velocity, f is the frequency, and the Greek letter λ (pronounced lambda) is the wavelength, λ being the usual symbol used for wavelength.

Now combining these three equations De Broglie could write

$E = mc^2 = hf$, so hf can be written as

$hf = (mc).c$

But c being the velocity of light, that is the velocity of photon, mc is the momentum of photon.

Also the equation $v = \lambda f$, in this case becomes

$c = \lambda f$

This makes $hf = (mc)\, \lambda .f$ giving

$\lambda = \frac{h}{mc}$

Thus De Broglie got the wavelength for photon as a ratio of Planck’s constant h and the momentum of photon – showing that a particle has wavelength (a wave property) which can be expressed in terms of its momentum (a particle property).

$\text{wavelength} = \frac{h}{\text{momentum}}$

This was shown for photon, but then De Broglie made a very bold prediction that it will be true for any particle; and any particle including that of matter, is also a wave. Most of the scientists dismissed this idea as an impossible generalization, but he struck to his gun, and asserted that a double-slit experiment done with such particle will vindicate his prediction, demonstrating their wave nature.

Matter wave is proved by experiment:

De Broglie’s prediction was proved to be true for electrons within a few years of the presentation of his theory. The experiment which proved it is quite similar in principle to the double-slit experiment we discussed before, though it differs from the latter in details. To begin with the arrangement for proper slits would not be easy. Even in the case of electron, the smallest matter particle, that was used in the experiment, the momentum will be much bigger than the value of h, so the wavelength would be extremely small, much smaller than that of light. As the slit-widths should be comparable to this very small wavelength, such slits would be very difficult to make. But such extremely narrow slits were found in nature in the shape of the very narrow gaps between the planes of molecules that comprise the solid crystals – a regular arrangement of such planes formed by arrays of molecules. Two American scientists Davison and Germer took advantage of this and used nickel crystals to provide the slits in a kind of double-slit experiment to test De-Broglie’s prediction. Though somewhat different in details we would describe their experiment as a double-slit experiment.

A flow of electrons from a source is allowed to pass through two slits and the beams coming through the two slits strike on the same place on a vertical screen. The screen is already covered with a sensitive material that gives rise to a tiny speck of light wherever the electron strikes it. Now if electrons behave like particles we would expect two vertical bright lines opposite to the two vertical slits or more practically a patch of uniform light there. But if electrons are waves they would behave exactly like light (photons) in a double-slit experiment with light, we discussed before. Electrons would form a fringe and we would be able to see that as variations in light – the highest brightness in the middle, then intermediate brightness, then dark, then intermediate again, then brightest again and so on at both sides of the middle. The light here actually represents the electron densities falling on the sensitive screen. The reason for forming such a fringe is exactly the same as that of light – the superposition of waves resulting in their constructive and destructive interference. So this proves that electrons are waves after all. The wavelength for it measured from the experiment matched exactly with the wavelength predicted by De Broglie

$$\text{wavelength of electron} = \frac{h}{\text{momentum of electron}}$$

Perhaps one could still interpret the result as a combined wave-effect when a huge number of electrons act in unison, which does not make an individual electron to be wave. This argument would be comparable to the behavior of water in water wave, which does not make individual water molecule to be wave. To rule out this interpretation the experiment was repeated using a very low density of electrons from a source, so much so that we could consider electrons to be coming to the slit one by one and reaching the screen one by one. But the result still remains the same – it is as if an individual electron is going through both the slits and interfering with itself to give the fringe effect. This shows that each of the electron is demonstrating a wave property. Later, similar experiments were done with bigger particles such as atoms and molecules, and they too demonstrated wave property. In proportion of their higher masses their wavelength were found to be even smaller than that of electrons.

Where does this wave property stop? If we use bigger and bigger things we may come up to a very big molecule, or even everyday thing of our visible world such as a marble or a tennis ball – then what happens? If we perform a double-slit experiment with tennis balls, for example, with appropriate size of slits, would we still find a fringe on an appropriate screen? the answer is obviously no. All the balls would be striking the screen opposite to the slits, and that is all. The experiment will not demonstrate a wave property for a tennis ball, and for that matter anything beyond the size of molecular dimensions. The masses, and therefore the momenta, would be so large that the wavelength and hence the wave nature would be negligible and of no consequence. This shows, once again, that the quantum theory is applicable to only things of the atomic or molecular dimensions.

Matter wave within an atom

So far, we have been describing particles which can move everywhere freely (free particles) and have found that they have wave properties. Electrons within an atom also have wave properties, but they are not free particles. We have to distinguish between the two different types of waves – waves which are free to travel, and waves which remain stationary because they are subject to some constraints. The former is called travelling wave, and the latter stationary wave. The simplest example of a stationary wave would be be the wave in a string kept fixed at both ends. If we excite this string to a wave, the wave will travel to one end where it will be reflected and travel towards the other end. The returning wave is superimposed on the original wave and create a stationary wave in which the points of maximum and minimum vibrations remain fixed in a static wave form on the string. Depending on the length of the string and the wavelength of the original exciting wave, there can be 1 or 2 or 3 or other whole numbers of segments, but not include a fraction. A string, having only one dimension– that of length, the stationary wave here is a one-dimensional (1D) one. But there can be two dimensional (2D) stationary waves as in the membrane of a drum when it is struck; and of course a three dimensional one (3D) which is usual in a confined space with length, breadth, and height. But in the case of 2D and 3D ones it is not as easy as the string wave to visualize the stationary wave. Electrons are in the three dimensional space of the atom constrained by the electric field created by the positive nucleus at the centre of the atom; somewhat like a string which is constrained by being fixed on both ends. Therefore each electron forms a 3D stationary matter wave.

De Broglie’s matter wave was able to remove one weakness of Bohr’s allowed orbits of electrons in an atom. Bohr could not give a reason why certain orbits are allowed while others are not. Another unexplained part of Bohr’s atom model was the use of $\frac{h}{2\barwedge}$ instead of just h in the quantization of angular momentum of electron in the atom. $\frac{h}{2\barwedge}$ did not come in a natural way as h did. De Broglie showed that the allowed orbits are allowed because only those can accommodate a certain whole number of stationary waves without including a fraction. These numbers of wavelength exactly fit within the circumference of these orbits. Let us see how. Let us imagine in a simplified picture that the electron orbit is a string in the shape of a circle, so its length is the circumference of the circle. In our string example above, the length of the string can contain the whole numbers of wavelengths λ, 2λ, 3λ, ... $n\lambda$, where n is a whole number. Now if r is the radius of the circle–

Circumference = $2\barwedge r = n\lambda$

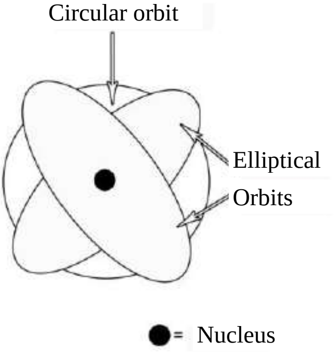
But De Broglie’s equation gives us $\lambda = \frac{h}{mv}$. So the above equation becomes $2\barwedge r = n \frac{h}{mv}$ which allows us to express angular momentum as

Angular momentum = $mvr = n \frac{h}{2\barwedge}$.

This is exactly what Bohr assumed while he quantized the angular momentum with n as his quantum number. But he only assumed the quantization using $\frac{h}{2\barwedge}$. De Broglie’s theory of matter wave made $\frac{h}{2\barwedge}$ (which we also write as $\hbar$) come naturally. This was an improvement .

The atom model gets more details

We have seen how Bohr’s atom model was thought out to explain the bright emission lines and dark absorption lines in the spectrum, in the latter case in a white light spectrum. Soon, however, it was revealed from more rigorous experiments that each of these lines have finer satellite lines accompanying them on both sides. Bohr’s model had no provision to explain these. But Sommerfield, an Austrian scientist offered an explanation by assuming that electron orbits are elliptical with various ellipticities as well as being circular. Those ellipses are allowed which are quantized with an additional quantum number k. Bohr’s quantum number n was now called the principal quantum number while the new one k was called the secondary quantum number. As before $n\hbar$ gives the various orbits with allowed angular momenta and thereby also the average distance of the orbit from the nucleus, and $k\hbar$ now gives the ellipticities of the various allowed orbits. The slight energy difference between the transfer from one eliptic orbit to the ground state and the transfer from another elliptical orbit (with a different k) to the ground state gives rise to two slightly different frequencies instead of one – thus creating two finer lines, along with the main line.



Bohr-Sommerfield atom model

Sometime later it was found that when a hot gas was placed within a very strong magnetic field, even those fine satellite lines get splitted further creating even finer lines very close to each other. Sommerfield explained these lines in term of the orientations his elliptical orbits take in a magnetic field. In a 3D space, the planes of the elliptical orbit may take various orientations letting the perpendicular on it make various angles with the direction of the magnetic field. This is because the movement of the electron along the orbit like any electric current makes a magnetic field of its own and interacts with the applied magnetic field. The orientations thus taken is quantized and is determined by one more quantum number m, called magnetic quantum number. The orientation is expressed as $m\hbar$. Again the light generation in the quantum jumps involving two different m will give two frequencies very slightly different from each other, thus explaining those finer splitting of lines. So the Bohr-Sommerfield atom metal was now described by three quantum numbers n, k, m.

But that was not the end of the story. By 1924-25 using very strong magnetic field additional splitting of spectrum of lines were discovered. Wolfgang Pauli explained it by assuming that an electron has a spinning rotation around its own axis which makes it like a small magnet, as any spinning charge would produce a magnetic field. Two different spins are possible for the electron – left spinning and right spinning. This way of looking at a spinning electron is a classical picture just like that of a spinning top. But Pauli looked at it as a quantum property and duly quantized it with another quantum number m_s called spin magnetic quantum number. The two spins are thus determined by $m_s\hbar$ with the two values of m_s. One special feature of this quantum number is that while all other quantum numbers take only whole number values this one takes two half values $+\frac{1}{2}$ and $-\frac{1}{2}$. By adding this new quantum number with the other three, Pauli arrived at a very important conclusion. Every state (orbit) of electron in an atom must have unique combination of values of n, k, m. Therefore, this state can accommodate a maximum of only two electrons, one with $m_s = +\frac{1}{2}$ and another with $m_s= -\frac{1}{2}$. If the atom has one more than two electrons the third one must go to the next orbit of higher energy, and so on. This conclusion of excluding in an orbit any electron in excess of two is called Pauli Exclusion Principle. This principle gives structured distribution of the electrons in the various energy states (orbits). This structure and the whole atom model using the quantum concept at every stage was a great success in physics and chemistry.

But this Bohr-Sommerfield model of atom was not a purely quantum theoretical one; concepts from both classical physics and quantum physics were used in it whenever required to serve the purpose of the model.

Even when De Broglie could give a reason for allowing certain orbits as they accommodate the whole numbers of wavelength of his stationary waves, the ideas was a mixture of both classical and quantum theory. One could say that both the theories are complementing each other here, but such a hybrid theory is not scientifically very elegant.

Moreover, the mystery of the quantum jump was still there. How could the electron go from one allowed orbit to another without passing through the space in between of intermediate energies? The photon emitted through such a quantum jump also remained somewhat mysterious to the extent that its directions etc. could not be predicted. What was really needed was a pure and self-contained quantum theory that can address these questions. Although De Broglie's matter wave could not provide such a theory, it provided the main clue towards that.

The matter wave tells it all

How is this wave related to the particle?

Pilot wave:
Every wave has an oscillation of something. In case of water wave it is the water which oscillates, in case of electromagnetic wave it is the electric field or the magnetic field that oscillates. But what oscillates in the matter wave? There was no easy answer to this question. De Broglie, the discoverer of the matter wave understood that there is no obvious ‘thing’ here that can oscillate, nor could he accept that somehow the electrons themselves are getting powdered into a dust which are oscillating into a wave! But the double-slit experiment was forcing a wave behavior on an electron, that superimposes and interferes just like any other wave. One way out was to assume some ‘internal wave property’ for our electron particle, as Einstein assumed for a photon particle, when he ascribed wave properties such as frequency to it. But De Broglie would rather go for a ‘pilot wave’ outside the electron particle though in contact with those internal wave property envisaged by Einstein. The pilot wave, as imagined by De Broglie, is a kind of guiding wave to the particle, that remains external to the particle.

One way to elaborate the pilot wave is to use that in describing the double slit experiment for electron. When an electron arrives at the two slits it can not go through both the slits at the same time, but its pilot wave can. It is this pilot wave which after going through the two slits takes the shape of two waves which superimpose on one another, interfere, and create the fringe on the screen. All the while the electron itself remains aloof from these activities maintaining its integrity. The fringe thus is not created by the action of the electron, but rather by that of the pilot wave. Once the highs and lows of this wave at the fringe on the screen are created, the densities of the electrons arriving there at every position are determined by those highs and lows. In this way De Broglie was able to keep the electrons intact with its particle nature, while leaving all those wave-like interference work to the pilot wave. Though this pilot wave concept was later abandoned by the scientists, it was a good stepping stone to the ultimate shape of the quantum theory.

We have seen that De Broglie depicted electron waves within an atom as stationary waves. But outside of an atom, outside of any constrained space, the electron wave has to be a travelling wave. Even in the case of a travelling wave the electron can be shown as a localized entity. When waves of many different frequencies superimpose on one another and produce a region with maximum amplitude, on both sides of which amplitude gradually reducing to zero – forming a distinct wave packet or pulse. The travelling wave has a velocity, but then the wave packet itself has a velocity of its own different than the travelling wave – this wave packet velocity or group velocity being assumed to be the velocity of electron. This is as far as De Broglie could go in associating the electron with the wave. If a question is made about what the wave is made of, De Broglie's answer stops here.

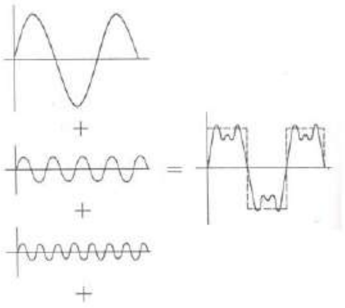
The wave is all that an electron has:

The Austrian scientist Schrödinger had more elaborate ideas in answering the question – what is the thing the electron wave is made of. He accepted the electron wave packet concept, but took it much further. While De Broglie thought that it somehow represents the electron, Schrödinger said it is the electron. He did not find any necessity for the existence of our electron particle separate from this wave packet; and the same he thought for any particle. He ascribed a lot of meaning to the shape of the electron wave that undulates up and down – the height of the wave at every point of space gives all the information there is for the ‘state’ of the particle at that point, such as position, momentum, energy etc. Whatever that thing is which is so undulating or oscillating, that can be expressed as a mathematical function of the position in space.

In mathematics function expresses the mathematical dependence of one quantity on another. For example, if a quantity ‘b’ depends on another quantity ‘a’ by any mathematical relation then ‘b’ is said to be a function of ‘a’, and maybe expressed as b(a). The relationship can be something simple such as $b = a^2$ or $b = a+3$ or some more complicated relationship involving higher algebra or trigonometry etc. In the cases mentioned above the functions are $b(a) = a^2$ or $b(a) = a+3$. In representing the function as a graph we may take the horizontal axis as ‘a’ and the vertical axis as ‘b’, the two axes defining a two dimensional graph; then for every value of ‘a’ we can calculate the value of ‘b’ and then plot a point in the graph paper going up to that value of vertically above the value of a in the horizontal axis. Joining all such points would give us the graphical or visual representation of the function. A function can be that of more than one quantity (variable), then relationship of ‘b’ may be not with just ‘a’, but with additional quantities such as ‘c’ and ‘d’ and others, for example such a function can be $b = a^2+c$, and can be indicated as $b = b\ (a,c)$. To make a graphical representation of this we would need three axes one each for a, b and c. Visualizing a three dimensional graph and the plots of points would give us not a simple curve but a more complicated three dimensional figure.

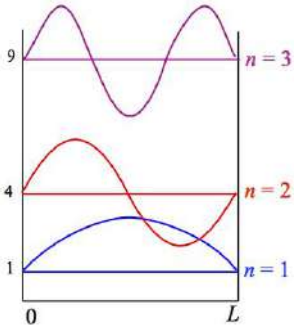
Schrödinger represented the amplitude values of the matter wave as a function of position, or as a function of time, or as a function of both position and time. This function was called wave function. He made the wave function as the most important thing in his quantum theory containing all the information of a system – one particle constituting a system, or more than one particle doing so. Our natural space being three dimensional, we often refer to the three dimensions as left to right, back to front, down to up – customarily designating them as x axis (giving various values of x increasing left to right), y axis (giving various values of y increasing from back to front), and z-axis (giving various values of z increasing from down to up). It is also a convention to designate the wave functions as the Greek letter Ψ (pronounced as psai). Thus as a function of the natural three dimensional (3D) space it is written as $\Psi = \Psi$ (x, y, z). If it depends on time, the time function is written as $\Psi = \Psi$ (t), where t stands for the value of time. It can be a function of space and time simultaneously and then can be written as $\Psi = \Psi$ (x, y, z, t).

We have already seen how Schrödinger, following De Broglie, had conceived the particle as a wave packet created by the superimposition of many waves of different frequencies. For this he invoked a famous theorem from mathematics known as Fourier's theorem. This theorem states that any function can be expressed as the sum of an infinite number of regular-wavelike periodic functions each with a different frequency. Thus even if the graphical depiction of the original function is not like a wave at all, yet it can be expressed as the sum of those periodic waves the graph of each of which will be like a regular wave of a specific frequency. The mathematical step of reaching such an expression of sum is called Fourier analysis. This way Schrödinger could express his total wave function as a sum of regular periodic wave functions each with one specific frequency through Fourier analysis. Each of these is called a harmonic with a particular frequency. All the harmonics superimposed on one another produce the total wave function. The amplitude of each harmonic represents the proportional contribution of that harmonic to the total amplitude of the original wave function at that point of space.



Superposition of various harmonics, each with a distinct frequency, result in the total wave function

To simplify all these, let us consider a one dimensional (1D) space having only x axis, and take as a simile a taut string tied at both ends. When plucked, the string gives rise to infinite number of waves (harmonics). In one of these the string vibrates in one segment giving the largest wavelength. In another it vibrates in two segments giving half the previous wavelength; and in yet another vibrates in three segments, and so on. giving smaller and smaller wavelength, that is larger and larger frequency. Every one of these pure harmonics has one frequency, called natural frequencies of the string. The totality of the sound after that plucking of the string that we hear is the sum of all the harmonics. The properties of the string, its tightness etc. will make a different contribution of every harmonic in the sound heard, proportional to the amplitude of that harmonic. These waves in the strings are actually stationary waves, and are good similes to the stationary electron waves in the atom, only simplified to 1D space in place of the actual 3D space. Every harmonic will represent an allowed electron orbit of Bohr, and it is the wavelength of this harmonic that would exactly fit the orbit when multiplied by an appropriate whole number, as De Broglie showed. But Schrödinger needed neither the allowed orbit assumed by Bohr nor that fitting of wavelengths within the length of the orbit shown by De Broglie, requiring a mix of classical and quantum concepts. For him the allowed orbits follow naturally from the wave function that gives them as harmonics.



The various harmonics of the electron wave function in an atom (stationary wave), shown here in one dimension. Every harmonic is related to an orbit (designated by a quantum number n)

The challenge Schrödinger took was to find out the wave function for a system. He did it by devising an equation involving the wave function that can represent any system when some parameters from the system can be put into that equation. The solution of that equation then provides the wave function holding all the information of states of the system. This was a great leap forward to a pure quantum theory without any assumption or any mix of classical physics.

Schrödinger equation

Why did it involve complex numbers:

In physics any wave can be expressed as a wave equation – a mathematical description of how the undulation in the wave takes place. Naturally, the main unknown quantity in the equation is the thing whose ups and downs form the wave. Therefore the equation would involve the rates of changes, the mathematics of which would usually get into differential calculus, the branch of mathematics that deal with rates of changes. The equations themselves are of the type called differential equation. A solution gives the function describing the up-down movement of that main unknown quantity. Schrödinger equation is a wave equation like any other. While in other waves we have something physical that goes up and down such as water in the case of water wave, or electrical and magnetic field in the case of electromagnetic wave, in the case Schrödinger equation it had to be just a mathematical function Ψ, the wave function. The circumstances of the system concerned had to be brought into the equation in the shape of the total energy experienced – the sum of kinetic and potential energy, as the case may be in each system. The other aspects of the equation have to conform to the already available quantum rules and the De Broglie equation. Schrödinger modeled his equation on a well known equation of the classical physics called diffusion equation which describes how something that flows diffuses through space (changes its density from point to point in space) as the quantity of it within the system increases. The diffusion equation is derived from another more fundamental principle saying that things can neither be created nor destroyed – the principle of continuity. Thus the left hand side of the diffusion equation (left side of the equality sign) is the rate of change of density with respect to time and the right side consists of the rate of change of the density with respect to space multiplied by a constant called diffusion constant, that depends on the thing being diffused.

Now in Schrödinger’s case he considered the energy as that something which is neither created, nor destroyed and also can diffuse. So he associated the total energy with the rate of change of Ψ with respect to space, but here the rate is taken doubly – as the rate of change, of the rate of change. In place of the diffusion constant that multiples the time rate of change, Schrödinger equation has a very different kind of number symbolized by i called imaginary number. So the other side of the Schrödinger equation consists of the rate of change of Ψ with respect to time, multiplied by i and also from the quantum theory the familiar ħ. We will not go into the exact Schrödinger equation because our scope in mathematics here is limited, but we would try to explain why such an unusual number i comes into the equation.

i is just another way of writing $\sqrt{-1}$, the square root of number -1. In mathematics i or any number multiplied by i is called an imaginary number, because no usual number, positive or negative, when squared can give a negative number (for example, square of both +2 and -2 is +4). Nevertheless, mathematics allows it as a number and use it in the mathematical operations. While combined with the usual numbers, called real numbers, the combination is called a complex number. For example (2+i3) is a complex number with 2 as its real part and i3 as the imaginary part. Similarly any complex number can be divided into a real and imaginary part in whatever way it turns up. In classical physics the imaginary part has no physical meaning and is usually rejected from the final physical result. But this is not the case with the quantum physics, where it is an essential part of physics itself, not just of the mathematics.

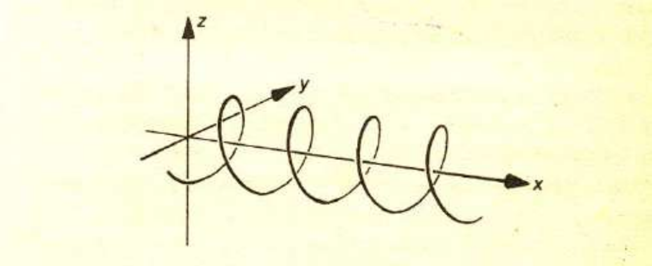
Let us now go back to the wave function and take one that describes position states of a system – describing positions of a particle in space. If the particle has a high tendency to exist at a particular point of space the wave function Ψ will be quite high at that point. To simplify matters, let us assume a 1D space having only x axis in the graph, then the graphical representation of Ψ will be just a curve varying in height as x is changed. In this case the curve will look like a spike at the point where the particle has a high tendency to exist, and zero in all other points. It will not look like an undulating wave form that we expect from a wave function. How can we make it look like a wave? Multiplying this by any usual real number does not help, as it only scale the spike up or down. But multiplying by an imaginary number i does the trick. To have a sense of it let us multiply i by i, and again multiply the product by i, and go on repeating this continuously. Thus $i \times i = -1$, $i \times i \times i = -1i = -i$, $-i \times i = +1$. As we go on multiplying by i, we keep on getting i, -1, -i, +1 again and again, in the same sequence. As is usually done on a graph paper, on the horizontal axis, we take one unit right to the centre point as +1, and one unit left of the centre point as -1; now on the vertical axis we take one unit above the zero point us +i, and one unit below the zero point as -i. We can then graphically visualize complex numbers with their real part (on horizontal axis) and imaginary part (on vertical axis). The continuous multiplication by i will now have the visual effect on graph similar as a ring going round and round – another way of describing a regularly undulating wave. Thus Ψ of any shape can be made to undulate by multiplying it with i; and that is the reason Schrödinger had to include an imaginary number number in his equation, which then seeps into the wave function as a necessity in quantum theory, becoming a part and parcel of the physics – not to be conveniently jettisoned once the mathematical operations are over. This is another bizarre aspect of quantum theory.

When the Schrödinger equation has to represent the stationary version of the matter wave, as it is in the atom, Ψ does not change with respect to time. So in this time-independent Schrödinger equation the rate of change of Ψ with respect to time is not there; that side of the equation now consists simply of a parameter (of energy) multiplied by Ψ. This parameter is called eigenvalue which means proper value or characteristic value. It is for these values of energy, that solutions exist of the equation. These are actually a measure of the allowed electron orbits in the atom, in the atomic case; and whole numbers– the quantum numbers– are associated with them.

Success in predicting spectral lines:

We have seen that considering the graphical view of a wave function for a position-state, the greater is the height of it at a particular point of space there is a greater possibility of a particle's position to be there. But motion does not involve only the position, it also involves the velocity – where would it be the next moment. Usually velocity is considered through momentum, the product of the mass and the velocity. Let us remember the harmonics of the wave function, each with a particular frequency. The frequency of the matter wave is related to the momentum, and therefore while wave function Ψ itself is giving us the position, its constituent harmonics are giving us the momenta. So even when we are talking about a position, it is already linked to a momentum.

We can write the simplified 1D version of wave function as Ψ(x). Let us visualize it on a graph where only the x axis is required for Ψ(x), but we would still keep y and z axes for other purposes. Those two can be used to visualize a complex number, as we did when talking about multiplying with i. For this, let the real part be along y axis and the imaginary part along the z axis. As before, the effect of i will be to create a ring-like rotation of Ψ on the y-z plane. Thus the complex wave function would show an undulating wavy motion moving in all three axes, though only one of them (x) is representing space. We can always do exactly this not with the whole Ψ, but with only one of its harmonics. It is only in case of a harmonic that a regular wave with a single fixed frequency is always expected. So the wavy motion now would be circular ring-like ones on the y-z plane, but also advancing along the x axis as we change position in space. Thus the ring now takes the shape of a spiral along the x axis. The distance between two subsequent ring here will give the frequency. This being only one frequency (in one harmonic) this distance remains constant. This distance, therefore, gives a measure of the momentum in this harmonic. The height of the ring at any point will give the amplitude which is proportional to the contribution of this harmonic at this point of space to the total height of Ψ. Such a visualization is possible because we are considering a 1D space. For the real 3D space it is not possible to provide the so many extra axes required for visualizing complex quantities in all the dimensions, and therefore an easy visualization is not possible. Nevertheless we can imagine a complicated shape of Ψ from our visualization of simple 1D case – the mathematics will work for the 3D space all the same, and can calculate in all the three dimensions. Schrödinger did exactly this for the hydrogen atom. Let us see the steps in doing this.



One harmonic of the wave function (ψ) giving the position state. This also gives the momentum-state (from the distance between rings) and contribution of this harmonic in the total wave function (from the radius of the ring)

The first thing to do was to formulate the Schrödinger equation for the specific requirement of the hydrogen atom. As it would be stationary wave, the equation will not have the time-dependent part. As required in Schrödinger equation, the kinetic energy and potential energy of an electron created by the electric field due to the nucleus within the atom, have to be put there. Schrödinger could solve this equation for the various harmonics. As we saw each harmonic is associated with an allowed electron orbit given by the quantum numbers n, k and m. While Schrödinger solved the equation in 3D, each dimension yielded one of the three quantum numbers.

Visually speaking, this comes more easily if considering the spherical shape of the atom, we use a different system of coordinate as did Schrödinger instead of the usual x, y, z system for rectangular shape of space. In this new system of coordinates, one dimension is in the radial direction of the sphere – from the centre to the surface, designated by r, the value of which increases as we go further from the centre. Another dimension is in terms of the angle that the radial line creates with the equatorial plane when it moves toward the north pole or the south pole (speaking in terms of the earth’s sphere). The third dimension is another angle created by the radial line on the equatorial plane when it moves around the axis of the sphere. The solution of the Schrödinger equation in the first dimension gives naturally the quantum number n, the various whole number values of which are at various radial distances from the centre of the atom. Similarly the other two quantum numbers k and m follow naturally from the solutions in the other two dimensions in the order these have been discussed above. They represent the ellipticity of the orbits and the orientation of the orbits respectively. Once the orbits with all the subtleties were obtained from the solutions of the equations, the the energy differences between the orbits gave all the lines of the spectrum in their calculated frequencies. And these lines were a total match with the real lines of hydrogen atoms as have been long known. This was a huge success of Schrödinger’s concepts taking everything as waves and wave functions, purely from quantum theory.

But soon Schrödinger suffered from a great disappointment too. It was revealed by others that the physical meaning that Schrödinger wanted to ascribe to his wave function can not actually materialize. To Schrödinger the wave function was a reality in the classical sense, but us we will see in the next chapter, it was not to be. The meaning that ultimately could be ascribed to the wave function was not satisfactory to Schrödinger; but it did wonders all the same.

Quantum predictions are probabilities

Every point is a probable position

The wave packet concept in trouble:
Schrödinger’s particle was nothing but a wave packet – a puffed up part in the wave created by the superposition of many waves. By treating everything as a wave he was trying to avoid the wave-particle duality, and if possible, to keep everything within the classical physics. He was also trying to explain other strange things such as a quantum jump which had chancy elements and take place without crossing the intermediate space. For these he was banking on a classical wave phenomenon known as beats. If two waves with a very little difference in frequency are superimposed on one another the interference builds up a wave packet gradually in time, takes it to a maximum, and then, gradually lets it collapse, repeating the process again and again. These are called beat waves. In the case of sound waves these cause successive waxing and waning of the sound; and it is more familiar that way. The number of times the cycle of the waxing and waning happen per second is called the beat frequency, which is found to be equal to the difference between the frequencies of the two original waves. Schrödinger believed that he could associate the two frequencies with the two electron orbits in an atom between which the quantum jumps take place, then he could regard the frequency of the emitted light as the beat frequency. He thought perhaps this would make the quantum jump a more natural process, avoiding the associated mysteries in it.

But the wave packet which was central to these concepts was itself in trouble. Other scientists showed that it will not be stable enough to represent a particle. It can be demonstrated that for free electrons it gradually spreads out and collapses, in a process similar to the diffusion where the density comes down and diffuses away. Besides if such a free electron collides with an atom, it can be considered as a collision between a wave packet (the electron) and a vibrating system (the atom), any atom being considered as a system vibrating at a particular frequency. In the case of collision between two marbles, the states of the two marbles after the collision can be predicted by classical physics from their states before the collision. But for electron and atom in their pulsating picture above, it is not to be. In the wave picture of the quantum theory one can not predict which direction and what velocity the electron would take after the collision. The best one can say is that it has the probability of going in all direction with all sorts of speeds. It is as if the electron has been smashed into a powder, and the latter is scattered in every way possible. There goes the integrity that Schrödinger wanted to give to his particle.

All these show that Schrödinger's attempts to regard electron as a wave packet is untenable. His wave function can not thus be identified with the particle itself. If the wave function is not the particle, then what is it, what is its relationship with the particle?

Superposition of probabilities:

It was Max Born, an Austrian scientist, who demonstrated that it is not possible to predict which states they would assume after a collision of an electron and an atom. This led to the demise of the Schrödinger's interpretation of the wave function. It was Max Born again who offered an alternative interpretation. He proposed that the wave function is actually indicating the probability of the state, rather than state itself. Then in answer to the question what the wave is made of, one could only say that it is just an abstract mathematical wave that tells us the probability of a particular occurrence, and there is no need to mix up the particle with what goes up and down in the wave. Thus in the graphical representation of the wave function, the height at every point represents just the probability of an event's occurrence at that point. Of course we have to remember that it is not the same probability as we know from classical physics, which has to be a fraction of real numbers. To keep that distinction, let us call this 'probability amplitude' instead of just probability. Later some ways have to be devised to obtain the classical probability from this, so that we can put a meaning to it that makes a practical sense as the latter.

Why did Born bring probability into the picture, why not something more certain? Actually this became clear in the research that Born had conducted on the collision between electrons and atoms. The findings were, as we have put above, one can not predict the direction and the velocity in the electron's scattering by the atom, and the best one can do is to try to find out the probability of a particular occurrence. The higher is the probability, we may expect to see that particular event on more occasions. More specifically, this is the probability for a system A to go to system B, or to system C, or to system D etc. That is what the height of the wave function at a particular point indicates – we called it the probability amplitude of an event to occur; it could be the probability of the position of the particle to be there, or it could be the probability the particle at that point to have a certain momentum, or something else as the case may be. The fact that the theoretical results come in terms of probability does not mean that the end-result obtained is an uncertain thing – not at all. The result provides very definite knowledge about nature that can very accurately predict what we would get when a measurement is made.

The spread of the wave function over all points of space may give an impression, for example, that the particle exists to some extent at every point. But that is not what the wave function tells. What is spread out is the probability amplitude of the position of the particle (in this case) – ultimately the particle would be found in one point only. If we designate the probability amplitude at point 1 as c_1, the same at point 2 as c_2, and so on, then c_1, c_2, c_3 etc. gives the heights of the wave function at various points. Thus the wave function Ψ itself can be expressed as a sum of these c_1, c_2 etc. The sum here means a sort of superposition of all these values.

$\Psi = c_1 + c_2 + c_3 + ... + c_n +....$

Any particular value c_n is actually expressing the 'weight' within Ψ for the probability that the particle's position is at the point n. This will be the case when the state concerned would be about the position of the particle. Every point has a weight like that – large or small – giving the probability at that point. The sum tells the whole story. But because c_1, c_2, c_3 etc. (and also Ψ) are expressed in complex numbers, the summing process is not an ordinary one, it will have to follow the mathematics of the complex numbers.

In the sum c_1, c_2, c_3 etc. can be regarded as the various alternative states that the system may take. In the case above, it was the question of position, which may be called position-state; hence the various alternatives means various positions. But in other cases it could be momentum-state, energy-state etc. the alternatives being momentum or energy instead. Similar sums maybe formed when the sum is not of various alternative states, rather it is of say two alternative systems – then it would be a superposition of two systems A and B, each giving its own wave function Ψ_A or Ψ_B. Then the combined result would be a third wave function which could the sum of these two – following the complex number mathematics here too.

$\Psi = \Psi_A + \Psi_B$

From probability amplitude to our familiar probability:
Our familiar probability both in classical physics and ordinary life has to be a fraction involving real numbers only. On the other hand, the probability amplitude of the wave function is a complex number. This does not show the number of times an alternative would occur out of a certain total – the information we get from the fraction in our familiar probability. Rather in its simplified graphical representation, we see the probability amplitude going up and down undulating from negative to positive and negative again, caused by the rotating effect of the complex number itself. How can we find from it a real positive probability as a fraction which is the only expression of probability that has any practical sense to us.

Max Born, while giving his interpretation for the wave function had an answer to this too. He got it from the mathematics of complex number. Any number, real or complex, has an absolute value which is also called modulus. Thus, for example, 3 and i3 are two different numbers one real, the other imaginary; but they have the same absolute value (modulus) of 3. Usually the modulus of a complex number c is written as $|c|$. For a complex number with both the real and imaginary parts the modulus $|c|$ has to be found out through a prescribed mathematical method. Now Born showed that the correct method of converting a probability amplitude c_n into a real probability is to take the square of its modulus, which is $|c_n|^2$. Being a square it must be a positive real number and can represent a positive fraction, giving the familiar probability – the type of probability that tells us how many times of the next 10 horse races, my horse is expected to win.

Considering the position-state, the real probability of the particle to be at the point n is $|c_n|^2$, but what is the real probability of the position-state of the whole system? That will be given by $|\Psi|^2$ which can be expressed as

$|\Psi|^2 = |c_1 + c_2 + c_3 + ... + c_n + ... |^2$

The actual calculation of this combined probability has some more mathematical process to follow. To give a simple description of it let us assume that there are only two probability amplitudes c_1 and c_2. Then

$|\Psi|^2 = |c_1 + c_2|^2$

If we calculate it we would find this as a sum of $|c_1|^2$ and $|c_2|^2$, as well as one mixed quantity which is a kind of a superposition of $|c_1|^2$ and $|c_2|^2$.

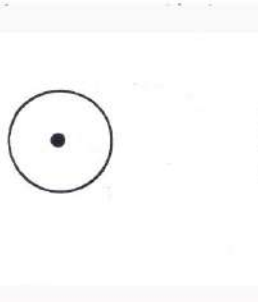
So, $|\Psi|^2 = |c_1|^2 + |c_2|^2$ + a mixed quantity

This mixed quantity, as a result of the superposition, can take values ranging over negative, positive and zero according to the circumstances. This could be compared to the superposition of two waves creating an interference fringe – the intensity values at the screen giving high, low and zero repeating the sequence as the various geometric lengths vary. In the $|\Psi|^2$ expression when the mixed quantity takes a positive value the sum becomes high, when it takes zero value the sum becomes low, and when it takes a negative value the sum at one stage would become zero. In all circumstances $|\Psi|^2$ remains positive giving the real probability in the classical sense, but even then there is a quantum character in the shape of the mixed quantity which makes the probability high, low or zero according to the circumstances.

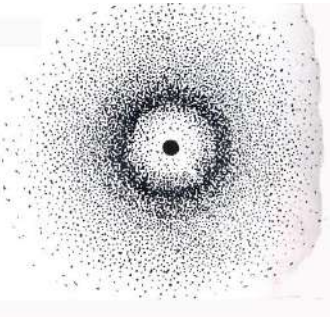
This provides us with another reason why a big thing like a tennis ball does not need quantum theory. In a minute particle such as an electron the mixed quantity plays a role giving it the quantum character. But a macroscopic object like a tennis ball is composed of innumerable such minute particles, and the calculation of their combined probability would involve innumerable such mixed quantities. Their random values would make the contribution of the mixed quantities zero as the positives cancel out the negatives. Thus anything macroscopic will not have the quantum characters; those are demonstrated only by the microscopic particles.

A purely quantum model of atom:

Now that we have a method of calculating the real probability at every point, we can make a probability picture of a system such as the atom, depicting every point with a shade whose depth of shade would be proportional to the probability density there. The higher probability density would mean there is higher probability of electrons being there in our picture – a greater depth of the shade would show this. Thus the whole picture of the atom would display shades of various depths. It is clear by now that the electron would not be confined to an orbit as given in the Bohr-Sommerfield model. All we would know is that there is some probability of its being positioned at every point within the space of the atom – in some regions with very low probably density. But there would, on the other hand, be regions where the probability density is very high, say more than 95%. These regions would be the place where it is very likely to find the electrons, and these would be the nearest equivalent to the orbits in the purely quantum model of atom, and are called orbitals. But orbitals are not orbits. An orbit is a real path where electrons actually reside. But the orbital is a three dimensional region within the atom which has high probabilities of electron being there – the probabilities, not the actual electrons. In a 2D depiction of the 3D sphere of atom we can draw the atom as a circle on the plain of the paper, and in that case the lowest orbital (quantum number n = 1) as calculated from the solution of the Schrödinger equation is a broad ring of the deepest shade surrounding and very near the nucleus. The actual shape in 3D, of course, would be a doughnut shaped space. There would be bigger rings for higher values of n. But, at the same time, comparable to the elliptical orbits in the Bohr-Sommerfield model, of various values of k, there would be dense (high probability) areas of strange shapes, such as dumbbell shapes. The whole atom outside the nucleus would seem to be constituted of cloudy regions of various densities and shapes. This is why sometimes this model is referred to as the cloud model of atom. This model is based purely on the quantum theory, everything comes from the wave function and the Born interpretation of that. The orbitals can be used to calculate the frequency of the emitted light (as Schrödinger did with orbits), and many other properties of the atom as well as of molecules and other combinations made of atom – all very precisely. This has been the most successful model in the long interesting history of atom models.



The lowest ‘orbit’ of hydrogen atom in the Bohr-Sommerfield atom model (shows actual electron positions)



The lowest ‘orbital’ of the hydrogen atom in the quantum theoretical atom model (shows probabilities only).

Before the measurement, and after it

Probabilistic calculations become real:
The things, we get from calculations with the wave function are the probabilities. But science is not about probability, those probabilities have to be real information about nature. The reliability of the information would depend on matching with the actual measurements. To conduct such measurements we have to set up instruments and would need human observers to sense those instruments by looking at the dials, listening to sounds, or comprehending the outputs of a computer. All these work in a classical manner. So when it comes to measurement, the domain of the quantum theory ends, and that of classical theory begins. But useful results had already been predicted before that, and the predictions now come true, only the quantum theory can do it. The wonderland of the quantum theory can be very practical at the same time.

In that quantum world we got things in superimposed states which give rise to mysterious things such as those mixed quantities in calculating $|\Psi|^2$. This shows that probabilities of the existence of the particle are everywhere, but a particle can not really be everywhere simultaneously, what it actually says is that if we would make a number of measurements, in proportion to the probability a certain number of the measurements will find the particle at one point, certain others would yield the particle at other points. But it would not be able to predict what one single measurement at a point would give – whether the particle would be there or not; there can be no prediction for that. Once that single measurement is made, no probability, no superposed states, are relevant any more – it is the old classical result, either the particle is there or it is not there. This is called the collapse of the superposition, and that of the whole quantum edifice, which so precisely had predicted all the results statistically. Those predictions tell us most of the things we need to know about nature. There may be chancy things on the way such as in the quantum jumps, but there is nothing chancy in the results given – such as the frequency of emitted light.

Another visit to the double-slit experiment:

It will be quite instructive to make one last visit to the double slit experiment, this time to illustrate Max Born's interpretation of the path followed by electrons in it. Let us assume that the electrons are coming from the source one by one, and that one of these electrons is approaching the two slits. Now it has two alternative paths – one to go through the slit 1, and the other to go through slit 2. In both cases it proceeds to the screen after going through the slit. There is a scale extending from left to right and the electron would ultimately strike at one position of the scale. Let us choose a point p on the scale and assume that the probability amplitude of the electron going through slit 1 and striking at p is $c_{1,}$ and that to go through slit 2 and strike at p is c_2. Therefore the total real probability of the electron reaching point p is $|c_1 + c_2|^2$, which can be expressed as

Real probability = $|c_1 + c_2|^2 = |c_1|^2 + |c_2|^2$ + a mixed quantity

Depending on the distance between the electron source and the point p (which will vary with the position of p on the scale), the mixed quantity will have a positive, or a negative, or a zero value making the real probability maximum, intermediate, or zero. This means an individual electron may have the probability to reach a particular point in the scale amounting to high, low or zero. Probability being what it is, only a probability, it may still reach that point in any case, or may not reach it in any case. But considering a great number of electrons, most electrons would reach where the maximum probability was, less number would reach where the low probability was, and hardly any would reach where the zero probability was. We would find this when actual measurements are made, for example, by making the scale sensitive to electron in which each electron striking would give a speck of light. This visualizes the result of measurement as an interference fringe. But up to the moment of the measurement there were only probabilities created by superposition. The moment the measurement was done, those collapsed and the certainty emerged. Similarly, it would not be possible to predict whether an electron passes through slit 1 or slit 2. Of course one could say that the probability of going through slit 1 is given by $|c_1|^2$. But it would not work, because there is then a chance that it goes through slit 1 and being independent can strike at a point which in reality is supposed to be dark. Therefore, it is not possible to consider $|c_1|^2$ or $|c_2|^2$ separately, nor is it possible to measure the electron's presence in slit 1 by putting a particle detector there. Only the combined $|c_1 + c_2|^2$ can yield the probabilities. It is as if the quantum theory is implying that an electron goes through both the slits and creates the superposition. If a detector could detect an electron in slit 1, that would be the collapse of the quantum effect right then and there.

Schrödinger's cat:

Successful as it was, Schrödinger could never reconcile with the interpretation of wave function given by Born and accepted by most of the scientists. Schrödinger, like some other scientists including Einstein, wanted to keep the issue of interpretation open, hoping that someday a more deterministic interpretation would appear. To give expression to his dissatisfaction with the probabilistic interpretation, Schrödinger imagined a thought-experiment that depicts this as a paradox, something which is obviously impossible. But this has rather become an interesting way to look at the quantum theory with Born interpretation as it is widely accepted. That thought-experiment is usually referred to as Schrödinger cat.

In it Schrödinger imagined that a cat is confined within a box that allows the cat to live, but keeps everything inside the box hidden from outside observers. There is a radioactive material within the box which emits charged particles in a random fashion. The time gap between emissions may vary very widely – there maybe an emission next second, or there may not be one for hours. Once the next emission takes place the released charged particle will create an electrical signal that, after due amplification, will switch on a motor to raise a hammer which will break a small vial full of a poison gas. Therefore, the moment the radioactive material emits, a sequence of activities would start to kill the cat by poisoning. Without that emission, the cat continues to live. For the sake of the thought-experiment, Schrödinger disregarded the fact that the cat and some other content of the box are macroscopic and therefore do not constitute a quantum theoretical system. He made them subject to the quantum theory, and there comes the paradox.

He raised the question – what would be the ‘state’ of the cat after say an hour, will it be in dead state or living state? How would quantum theory describe the state? Obviously there is a probability for the cat to be alive, and also a probability for it to be dead. The theory can ascribe a probability amplitude to each of these options. One could calculate the real probability from the wave function of the system that comes from the superposition of the two states – living and dead. The best one can say is that the cat is in a superpositioned state between being living and dead! We would not get rid of this ridiculous state till we open the box and see (measure) whether the cat is alive or dead. Schrödinger assumed that if the superpositioned state of the electron does not shock us enough, the same between life and death of the cat will.

But in a way Schrödinger's attempt to depict the Born interpretation as a paradox was counterproductive. More than 100 years have elapsed since the interpretation started to work for the quantum theory, and during this time it went from strength to strength in explaining nature. Schrodinger's cat has become a popular thought-experiment with the students of quantum theory as a means of understanding the Born interpretation, rather than revealing its weakness, if any.

An inbuilt uncertainty giving certain results

The meaning of uncertainty

Nothing is uncertain in classical physics:
We have already seen that the theories of classical physics can predict everything given the initial conditions. It can also potentially allow the measurement of measurable quantities to any degree of precision, subject to the availability of the proper measurement technology. There is nothing essentially inherent which may prevent the attainment of such precisions. Yes, there might be practical problems in both theoretical calculations and measurements. For example, it is not practically possible to calculate the future motion of every molecule of a gas even with a computer, because of the time required. That is why the use of statistical physics had to come to the rescue. But there is nothing to hinder the calculation in principle. Similarly, we may not be able to see and measure something directly, because these are too small even for our most powerful microscope today. But that too is not inherently impossible, a better microscope in future will solve the problem.

Some situations, however, do arise in measuring small things. If the probe we are using for the measurement is bigger than the thing measured, that may create an uncertainty. It is somewhat like the loss of precision when we try to push the keys of a very small phone keyboard with our fingers. We might need a narrow stylus to do the job precisely. In a microscope if the intensity of light used is too high, it may shove and displace the very fine thing we are trying to see and measure. This may create uncertainty in the measurement. But we can always reduce the intensity of light to the extent that the uncertainty is reduced to our desired level. And in principle there is no limit how far down the level of uncertainty can be pushed to. Too little light used for that may create practical problems which can be technically solved. There is another way out in the shape of the calculation of corrections to be added to compensate for any uncertainty created. Such corrections can always be theoretically devised in the classical physics.

A dilemma introduced by quantum theory:

Let us now consider our measurement in a microscope from the point of view of the quantum theory. Let us imagine a microscope that can show and measure an electron. In order to do that we must use a light of extremely small wavelength comparable to the size of the electron. We have already seen that the probe has to be comparable in size to the thing being probed. So our ‘light’ has to be a kind of gamma ray of extremely small wavelength. With sufficiently small wavelength of gamma ray, we should be able to measure the position of the electron with certainty. But the small wavelength means very high frequency, and from the Planck’s relation $E= hf$, a very high energy of the light photon. When such a high energy gamma photon collides with the electron to be measured, it will be like the electron-atom collision described by Born, that resulted in an uncertain scattering of the electron. This means any attempt to measure the momentum of the electron will give uncertain result, and the uncertainty will increase with the decrease of the wavelength of the gamma light. So while we try to achieve a certainty in the position of the electron by reducing the light wavelength, the certainty in the measurement of the momentum is decreased. The opposite will happen when we try to increase the certainty in the measurement of the momentum of electron by reducing the photon energy that would require to increase the wavelength of light thus making the position measurement less certain. So here is a dilemma. We can not make the both the position measurement and the momentum measurement certain at the same time. One of these has to be achieved at the cost of the other. There is an inherent uncertainty involved.

However much we improve our measurement technology, we can not avoid this inherent uncertainty. Here the quantum theory drastically differs from the classical theory.

Uncertainty principle

The minimum inbuilt uncertainty:
Werner Heisenberg, the German scientist, was one of the major creators of quantum theory. He developed a quantum theory based on a branch of mathematics called matrix algebra, which gave perfect results. Later Schrödinger gave an alternative way to the theory based on his wave function concept. It is the latter we have chosen to discuss because it is easier to understand. Schrödinger proved that both these formulations of quantum theory are equivalent. Schrödinger’s formulation, however is much more used in practice. It was Heisenberg who established a principle of uncertainty in quantum theory which showed that there is a minimum inherent uncertainty which we can not avoid. The amount of that uncertainty is defined by the Planck constant h. This is called Heisenberg’s uncertainty principle which work in the measurement of pairs of parameters such as the position and the momentum coming as a pair, the one we have already discussed.

To make our discussion more precise in terms of Heisenberg’s uncertainty principle, we can adopt a simple method to prove the principle on this pair. Let us reintroduce our imaginary gamma light microscope to see an electron. The gamma light here has to go through a tiny opening, as light has to do in all microscopes. As a result, it will show the property of diffraction. One part of the diffraction phenomenon is the self-superposition of light creating in this case circular interference fringes like those we saw in a double-slit experiment. This forms the successive bright and dark rings at the edge of the spot where the light is focused. In our case, the gamma light is focused on an electron, with its concentric bright and dark rings. Now the position of the electron has to be visualized from this image in the field of view of the microscope. But the rings will introduce an uncertainty in doing so. The theory of diffraction says that the minimum in such uncertainty will be equal to the wavelength of the light. Thus–

the uncertainty in the position $\geq \lambda$ where λ is the wavelength of the light used. Confined to a single dimension to keep it simple, the position may be designated by x and its uncertainty as Δx. This makes

$\Delta x \geq \lambda$

Now let us see again what happens when the light photon collides with the electron in the process of measuring it. The electron’s momentum will be disturbed by the collision at least to the extent of photon’s own momentum which can be calculated from De Broglie’s equation:

$\lambda = \frac{h}{\text{momentum}}$; Therefore, momentum $= \frac{h}{\lambda}$

If we designate momentum as p, and the uncertainty in momentum as Δp, them

$\Delta p \geq \frac{h}{\lambda}$

The product of the two uncertainties is then

$\Delta x. \Delta p \geq \lambda. \frac{h}{\lambda}$

or $\Delta x. \Delta p \geq h$

This is Heisenberg's uncertainty principle. It fixes up the minimum inherent uncertainty one encounters when trying to measure both the position and the momentum of a particle. The amount of uncertainty being of the order of h, it is applicable only to the very small things of atomic dimensions, and is negligible by far when it comes to bigger things such as a tennis ball, a grain of sand, or things even smaller.

The product of the two uncertainties being fixed, the decrease of the uncertainty in position automatically increases the uncertainty in momentum, and vice versa. If the former becomes zero, making position certain, the momentum will be infinitely uncertain, which is impossible. Even if we make Δx extremely small the momentum will then be extremely uncertain meaning it can be very large without any bound, shooting out with any velocity in any direction. On the other hand, if we make Δp very small, then Δx becoming very big– we would hardly know the position of the particle, which could be a very fuzzy thing existing everywhere in the space.

Some strange results of the position- momentum uncertainty:

The uncertainty principle gives some very strange behaviors of small particles such as electron. Suppose we confine an electron in a small container which can be squeezed further to a smaller and smaller size. The smaller is the size of the container, more certain becomes the position of the electron, which means the uncertainty in position would decrease. As the process continues, the uncertainty in the position may be reduced to the order of h, then according to the uncertainty principle, the uncertainty in the momentum of the electron will be very high, and the electron may shoot out of the container with unimaginable speeds in any direction. When the position gets almost certain, the momentum will get infinitely uncertain, which is impossible. This means an electron can never have a certain position. This is the reason that we could not expect an electron to actually go through one of the slits in the double-slit experiment, as in that case it would have an almost certain position while within the slit (slit being extremely narrow), which is impossible. So quantum theory could only state that the electron goes through both the slits in some mysterious way.

It is this fuzziness of its position that prevents an electron to be on an orbit given precisely by the Bohr-Sommerfield atom model. Also it is this fuzziness that never allows an atom or molecule to be perfectly still within a material. The classical theory predicts that at the lowest temperature possible, the absolute zero temperature, the molecules will be standing still; but not so the quantum theory. A molecule still has some vibration left even at this absolute zero temperature according to the uncertainty principle of the quantum theory.

Time-energy uncertainty

Another pair of parameters bound by uncertainty:
Energy and the time over which the energy is emitted form another pair of parameters which have uncertainty relation with each other, exactly similar to that in position and momentum. Let us take the example of the light emission from an atom. When an atom is excited to a higher energy an electron from the ground state goes up to a higher state. But this is only a transient unstable situation. Therefore, the atom has to regain the

equilibrium by the quantum jump of the excited electron back to the ground state. In the process, light is emitted with the frequency corresponding to the difference of energy between the two states, creating a bright line in the spectrum.

We may compare this process to a simile in which I am trying to balance a pencil on its pointed tip, putting the tip on my finger. So long as I keep everything very still the pencil will maintain an equilibrium and remain in that position. But soon even at the slightest deviation from the temporary equilibrium energy, the pencil will fall down in its attempt to go to a lower permanent equilibrium energy. The more uncertain the energy of the earlier equilibrium of the balanced pencil becomes, it will fall down within a smaller and more certain time. A similar thing happens when an atom comes back to its lowest ground state energy from the unstable equilibrium of the excited state. The more uncertain is the unstable state, this quantum jump will take place in a more small and certain time. Thus such a more certain time corresponds to a more uncertain frequency (more uncertain energy difference of states) of the light emitted giving a spread in the frequency of the emitted bright line. On the other hand, a more uncertain longer time of emission gives a line with a more precise frequency (more certain energy difference of the two states).

This relationship between the uncertainties in energy and time is an inevitable one and follows the same uncertainty principle as before:

$\Delta E . \Delta t \geq h$ where ΔE is uncertainty in energy and Δt is the uncertainty in time.

In this pair too, a decrease in ΔE will increase Δt, and vice versa. For example, if the time over which the energy is released becomes very certain, the uncertainty in the energy it releases will approach infinity and can assume any value, however large it is!

Some strange results of time-energy uncertainty:

In our classical world a particle with a certain energy can never cross an energy-barrier which is higher than that energy. This is for the same reason as a high-jumper with his/her limited energy can not jump over (say) a 50 meter high wall. But in the quantum world the uncertainty principle creates opportunities for particles to do exactly that – cross any barrier. For example, an electric current can be stopped by applying an opposite voltage, because classically the electrons can not overcome the barrier of a sufficient voltage. But according to the quantum theory an arrangement can be made which can take advantage of a very short and certain time for the emission of energy, and thus have a real probability to attain a very uncertain energy assuming any value – even a very high one enough to overcome the barrier. Thus a current can be created even against a very high voltage. This is as surprising as a high jumper jumping over a 50 meter wall, but it is quite possible in the quantum world!

One may understand it by imagining that the electron is borrowing the extra energy from outside within that very small and certain time, using that to cross the barrier and then paying back the energy borrowed within that same small time. So at the end of it, everything is same as before, but the electron is already on the other side of the barrier. It is as if the electron has tunneled through the barrier unnoticed by anyone. A classical simile will be that of a tennis ball with limited energy striking a 100 meter high 2 meter thick concrete wall again and again and suddenly find itself going through it to the other side without leaving any dent or any hole in it, nor any trace in any form.

These events may seem to be very strange, but these are real, so much so that we have used them in our technologies with excellent effect. One example is the very commonly used device in electronics known as tunnel diode. It works exactly as we described above by overcoming a voltage barrier taking advantage of the uncertainty principle, as if the electrons are tunneling through the barrier – hence the name.

Another remarkable use of it is in the Scanning Tunneling Microscope (STM). This can create a visible image of such minute thing as individual atoms in an array on a highly polished metal surface. Here an extremely sharp pointed needle tip scans over the surface remaining very close to it – almost touching it, but not quite. The needle and the surface are parts of an electric circuit and a current can jump over the miniscule gap between the needle tip and the atom on the surface, because it can tunnel through the barrier here. The current increases when the gap is smaller. As the gap is the lowest when the tip is near the top of an atom, but gets bigger when the tip moves toward the valley in-between two adjacent atoms, the current forms an image of the contour of the array of atoms. In effect, every atom is individually imaged by the microscope as the current fluctuates while passing one atom to the next. Without the uncertainty principle in the quantum theory there would be no STM.

These are some examples of the uses we make of the uncertainty principle for our technologies. But nature itself also makes many uses of the principle. For example, this is essential in the energy producing process inside the sun and other stars. Sun’s energy comes from nuclear fusion energy being produced within it – the same reaction which produces energy within a hydrogen bomb. For this two hydrogen nuclei have to come very near to each other and fuse together ultimately making a helium nucleus. A huge amount of energy is needed to overcome the repulsive energy of the two positively charged hydrogen nuclei. The energy available from the collapse of the great masses of the sun towards the centre is not sufficient for this. But the time involved being very small the uncertainty principle shoots up the energy as the latter becomes very uncertain, thus momentarily fuse the two hydrogen nuclei. Such fusion with a great amount of hydrogen creates the enormous energy of the sun, without which we would not have existed. There are many instances like this where the nature takes advantage of the uncertainty principle. The one described below has been instrumental in the creation of the universe.

The creation of the universe out of nothing:

It was with the help of the other great revolutionary theory at the beginning of the 20th century, the theory of relativity, that quantum uncertainty can explain the creation of the universe. According to the established theory of this creation, called Big Bang theory, the universe suddenly came into being as a little point of space and time, some 14 billion years ago. There was nothing before that – no space, no time; classically we can regard it as an absolute vacuum. But the quantum theory gives a different picture of the vacuum. If we could zoom into the details of this vacuum through an imaginary microscope we would see energy appearing there in the shape many particles – appearing now and disappearing immediately. This is uncertainty principle in action. There is always a possibility within a very small and certain window of time to borrow any amount of energy and then give it back. Thus energy in the form of particles can appear and disappear within that window of time without violating any law. Because the particles are so transient, they are called virtual particles. These have to come in pairs for the sake of obeying the laws of conservation.

Thus the initial total momentum being zero, it will still be zero if two particles of equal and opposite momenta are created simultaneously; similarly for the conservation of angular momentum, electrical charge etc. Thus for example, a particle and antiparticle pair, such as an electron and positron pair is created, and then is annihilated immediately.

How can such transient particles create a universe? For this we have to refer to the theory of relativity. According to this theory, when highly amplified by an imaginary microscope we would see bubbles of space-time frothing within the ‘vacuum’, a bubble coming up and then collapsing immediately. But the relativity theory showed how there is a probability of one of the bubbles to expand enormously very very rapidly before it could collapse. This unimaginably fast expansion is called inflation which within that miniscule time would expand it to a macroscopic size. Then it would continue to expand at a slower pace. Now the tiny bubble of space-time would have some virtual particles and would capture them before they had a chance to get annihilated. This is what happened in the creation of the universe. That tiny bubble of space-time with energy within to be expressed as matter and energy. Thus the uncertainty principle has a role in this great event too.

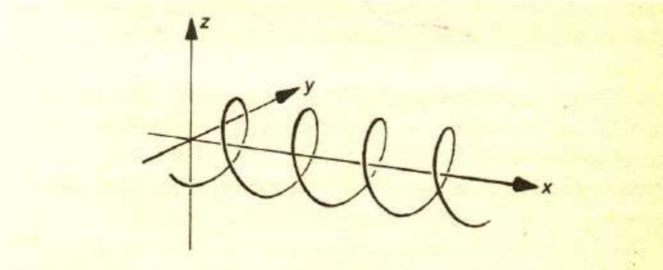
An alternative way to look at the uncertainty principle

Some scientists do not like to look at the uncertainty principle as an inevitable problem of measurement at the level of the things of atomic size, the way we have already introduced the subject. They would prefer it to come naturally from the wave function. Let us try to look at it also from that point of view in a rather simple manner.

To keep it simple, let us go back to our 1D graph of the wave function $\Psi(x)$. $\Psi(x)$ is a complex quantity whose height for every value of x gives the probability amplitude at that point of space. Now we recall the graph of one of the harmonics of $\Psi(x)$ which resembled a screw with rings spirally progressing along the x axis. The distance between the consecutive rings is constant and gives the momentum of the harmonic, while the height of the ring gives the contribution of this harmonic in $\Psi(x)$. Now let us suppose that the concerned particle has a certain fixed momentum, without any uncertainty. Then there will be only one harmonic for the $\Psi(x)$, the one with this momentum, and this is the only one contributing to $\Psi(x)$. So at every point x, the height of $\Psi(x)$ will be the same – equal to the height of the rings in this harmonic. $\Psi(x)$ giving the position-state of the particle, the probability amplitude of the position is the same everywhere, it has equal probability to be at every point. This means its position is totally uncertain. Thus we see that when momentum is very certain, the position is very uncertain. Here we find one extreme case of the position-momentum uncertainty principle, this time naturally from the wave function.

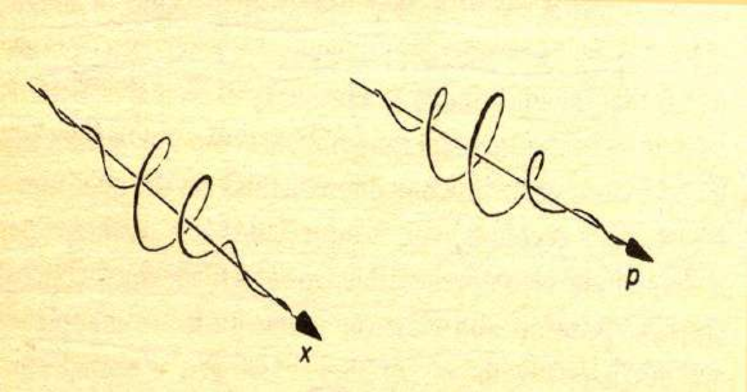
For the other extreme, we have to draw a 1D graph of Ψ, but this time Ψ will be a function of momentum (p) instead of being a function of position (x). So we have to write Ψ as $\Psi(p)$. To distinguish it from the previous one we use $\overline{\Psi}$ (pronounced psai bar) instead of Ψ. A graph of $\overline{\Psi}(p)$ against the values of p will give us the momentum probability amplitudes for various momenta.

Now its harmonic will be screw-like rings spirally going along the p axis. The distance between the consecutive rings will be same and will give the position of the particle and the height of the ring gives the contribution of the harmonic to $\overline{\Psi}(p)$. Let us suppose that $\overline{\Psi}(p)$ has only one harmonic, so this is the only one that contributes to $\overline{\Psi}$(p). The position of the particle is then related to this distance between rings, and there is no uncertainty in it. The contribution to $\overline{\Psi}$(p) being only from this harmonic $\overline{\Psi}$(p) will be same at all values of p. As a result it will be totally uncertain. So when position is extremely certain, the momentum is extremely uncertain giving the other extreme case of the uncertainty principle.



Let us look at a harmonic of the ψ(x) again. This time let us assume that this is they only harmonic ψ has, therefore, the momentum is fixed (certain) obtained from the distance between rings in this only harmonic. This only contribution of amplitude (radius of ring) makes the probabilities of position same everywhere, making position infinitely uncertain. We can also look at this as the only harmonic of $\overline{\Psi}$(p). In that case the x axis will be p axis instead. In this case the single harmonic makes the position fixed (certain) and probability of p (radius of ring) is same for all kinds of p's, making

As both these extreme situations are impossible, in reality there will be some uncertainty in both position and the momentum, and that would be reflected in the graphs. In the case of Ψ(x) the height will be maximum at one value of x, but there will be some heights at other values too. Similarly in the case the of $\overline{\Psi}$(p), the height will be the maximum for one value of p, but their will be some heights for other values of p too. There will be various harmonics in both the cases and their relative contributions will determine the heights. It can be shown that the product of the uncertainties in x and that in p will be equal or bigger than h. Thus the position-momentum uncertainty may follow from the purely theoretical considerations of the wave function, and so can the other pairs of uncertainties in a similar fashion.



ψ(x) in reality is the sum of many harmonics (various momenta). They (with various ring radi) contribute so that probability of position is high at one point of space, but lower in other points. Thus there are some uncertainty in both position and momentum.

Ψ(p) in reality is the sum of many harmonics (various probability of position). They (with various ring radi) contribute so that the probability of some momentum is high while others are low. Thus there are some uncertainties in both position and momentum. Both ψ(x) and Ψ(p) give the same conclusion about uncertainty.

The quantum world

The last three chapters of the book described how a pure quantum theory without the help of classical concepts emerged to give an account of the new facts in physics. Once the main frame-work of the quantum theory was in place, many problems of physics, new or old, could be solved by using it. Moreover, many new phenomena of nature were discovered which would not have been possible without this theory. Though only the very small things of atomic dimension are directly concerned with the quantum theory, these being the fundamental constituent of all physical entities, many macroscopic events too need this theory for their explanation. Thus the quantum theory has now occupied the centre-stage of the physical world when it comes to the intricate scientific explanations of its workings. We will see some areas of this world where the quantum theory has been on the driving seat of explaining and innovating.

Atoms, molecules, crystals

Quantum theory behind the intricacies of light:

We have already seen how much of the story of quantum theory was about light. It was through the discoveries of the process of emission and absorption of light by the atom and the correct predictions of the frequencies (colors) of the spectral lines that the quantum theory went from success to success. There were other phenomena concerning light which could attempt an explanation for the first time thanks to the discovery of the quantum theory. Fluorescence was one of those phenomena. Certain materials absorb light when they are exposed to it and can themselves become luminescent emitting light, usually of a lower frequency than the one absorbed, but sometimes of equal or even higher frequencies. This is called fluorescence. Quantum theory can explain it by showing how in a fluorescent material electrons get excited by the light around the material and go from the ground state to a higher state. When the atom then comes back to the equilibrium, the electron returns to the ground state by emitting the light of fluorescence. But usually a part of the energy gets scattered by the electron and is lost to the process, hence a lower frequency of the emission.

A very familiar example is the tube light (also called fluorescent light or the compact fluorescent light) we often use. Here electrons flow through a gas in the tube causing collisions between electrons and gas atoms, thus creating ultraviolet light (invisible). The inner surface of the tube is coated with a fluorescent material that absorbs the ultraviolet and emits a visible light of a lower frequency. Other interesting uses are in the identification of minerals and precious stones which give characteristic visible light when exposed to invisible ultraviolet, because of their fluorescent property. The technique is even used for the quality control of the manufacture of medicinal tablets which would fluoresce in a characteristic color if the right proportion of content is there. If the light to which the fluorescent material is exposed is very intense then two instead of one photons excite the electron which goes to a much higher state, as a result. Then the emitted light maybe of a higher frequency than the absorbed one. In case both the frequencies happen to be equal, then one can say the light shone on it is in resonance with the atom.

In the fluorescence process, the emission of light follows its absorption in almost no time, so we see the fluorescence only when the light shone on the material is still present. But there is a different process in which a material can emit light long after it had been exposed to the original light – the process is called phosphorescence. Quantum theory has at last revealed the mystery of this process too, a curious phenomenon in which we see things glowing softly in the dark – popularly in the beads of necklace, dials of table clocks and watches, and the toys. Here an electron excited by the light shone on it goes to a higher state as usual but can not come back to the ground state directly. It has to do that via an intermediate state which is not the usual way of doing it. The reason for this unusual path maybe a changed spin state of the electron and in this situation it may not be able to be accommodated at the ground state. Therefore, it has to seek another state which would accommodate its changed spin state and prepare it to go back to the ground state, emitting the light. All these take time, different time for different atoms of the material. As a result, one would see soft light coming in the dark from some or other such delayed emission.

What is true for light, is also true for other parts of electromagnetic wave spectrum – for x-rays for example. The x-ray frequency being very high, the energy difference between the two states involved in the quantum jump has to be very big, and therefore, states deep into the centre of the atom have to be involved. The x-ray is produced in an evacuated glass tube where a very high energy electron ray strikes a target made of a specific metal, that gives rise to the x-ray pattern characteristic to that metal. The electron ray is so energetic that it is capable of expelling electrons from the deepest levels where they are very strongly bound. The equilibrium demands that there can not be electrons in a higher state while a lower state has been emptied of its required number of electrons. This makes the various upper state electrons to quantum jump to refill the inner vacant states, thus emitting the characteristic x-ray frequency pattern of that target metal.

The fluorescence, the phosphorescence, x-rays – all these intricate phenomena of light had been discovered earlier. But all these needed the quantum theory for an explanation.

Quantum chemistry:

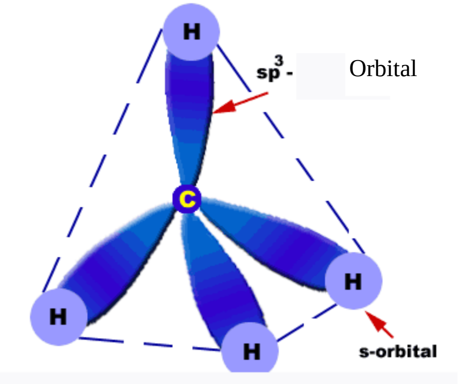
The tasks of chemistry include the creation of molecules out of atoms and allowing the molecules to react by breaking themselves or recombining themselves. So the application of quantum theory means that we should be able to write Schrödinger equations for molecules, and to solve them, so that the chemistry of it is explained. But all that the quantum theory could do is to solve the Schrödinger equation exactly for atoms only, and that too for the simplest atom, hydrogen. The hydrogen atom involves only one electron; any increase in the number of particles in the atom makes the mathematical problem too complex to solve exactly. But scientists were not daunted by this, they developed methods of simplification which would give results to the precisions that we need. These would enable the prediction of the states of more complex atoms, molecules, and even a solid made up of innumerable such molecules. This is how the quantum theory can solve many complex problems including those of chemistry, such as the explanation of the chemical bonds.

In the case of the hydrogen atom where an exact solution is available, the lowest state is designated as 1s, 1 indicating the lowest value of the quantum number n, and s indicating the lowest value of quantum number k. Electron would have the highest probability to be in the orbital defined by this. When excited, it goes to higher orbitals named as 2s, 2p, 2d etc. All s orbitals are spherical, but those with higher values of quantum number k, designated by symbols p, d, etc. are not spherical. For example, the orbital 2p is dumbbell shaped with a particular direction as the axis of the dumbbell. Thus to make simplification in a more complex atom, we may regard this similar to a hydrogen atom replacing the usual electron orbital of hydrogen with the excited orbitals of the same, for this new atom. Thus 2s, an excited orbital of hydrogen, may become the lowest orbital of this new complex atom. We can still use the same symbols 2s, 2p etc. because their shapes remain same as in the hydrogen atom. However as the wave function of the new atom will be slightly different, therefore these orbitals will occupy slightly different areas in the new atom, and will have slightly different energy. As the new atom would have more electron than one, the orbitals would tell us where the extra electron will go in the equilibrium condition. This is important, as for example, the bonds between atoms will depend on the number of electrons in the outermost orbitals of the atoms.

To illustrate the bond formation, let us take the example of the type of bond called covalent bond, very common in compounds. In this type of bond electrons in the outermost orbitals are shared between the two atoms so that both can fulfill their full quota of electrons, and be in equilibrium. As the two atoms remaining separate will have higher total energy, it is energetically favorable to form the bond. The quantum theoretical picture of the orbitals show that the place where the two outermost orbitals superimpose on each other they spread out to more space than they would if they remained separate. This spreading means more uncertainty in the position of electrons, making the uncertainty in momentum to decrease, according to the uncertainty principle. This means the reduction in the kinetic energy causing the total energy to reduce too, making the bond formation the preferable option. Also the quantum theoretical results show that the interference between two s orbitals and two p orbitals are more constructive than destructive. Being constructive means that the amplitudes of the two waves would add up and result in the increase of the amplitude of the wave function, which indicates a greater probability for the bonding than not bonding.

There is another way to consider the covalent bond without localizing it in the space in-between the two atoms, as we have just done. In the new approach we may regard the superimposed outer orbitals to be an orbital of the whole molecule consisting of the two atoms. As a result, instead of being localized, the bond encompasses the whole molecule that form two poles – one positive and the other negative. This consideration shows that the molecule can vibrate in some natural frequencies just like a spring. These natural frequencies are usually in the infrared region (heat). This is how carbon dioxide molecules in the atmosphere absorb heat released by the earth's surface, as the heat waves have exactly the same frequencies as the natural frequencies of the carbon dioxide molecules, and hence excite them into vibration. This prevents the heat to escape to the outer space, and thus creates the global warming, carbon dioxide acting as a 'green house gas'.

There are cases where quantum chemistry can go beyond our simplistic view of the molecule, and can give us some intricate details. We know that methane is a simple molecule with one carbon atom bonded to four hydrogen atoms. In a 3-dimensional model the carbon atom is at the centre of a regular tetrahedron and each hydrogen atom is at one of the four corners of the tetrahedron. This makes the four carbon-hydrogen bond spread out from the carbon atom at right angle to each other. All these bonds are of equal strength and are symmetric in every way. But later quantum theoretical considerations showed that things are not that simple. It shows that one of the carbon-hydrogen bonds is quite unique, and do not resemble the other three. An s orbital of carbon is involved in this bond; s orbital bond being spherical it does not have any directionality, and it would be rather weak. Each of the other three bonds use a p orbital of carbon atom; p orbital being dumbbell shaped it has a clear directionality along the axis of the dumbbell, which also makes it a stronger bond. But the practical chemical considerations do not show such an asymmetry among the four bonds. Quantum chemistry explains it by showing that the p bonds and the s bond exist in a state of quantum superposition giving four equal hybrid (mixed) bonds designated by sp^3. The 3 over the p indicates that one part of s bond is mixed with 3 parts of p bond in each of the bonds. This makes all the four sp^3 bonds equally strong and at right angles to each other, as demanded by the practical chemistry. But these quantum theoretical details provide much more information useful to chemistry.



The orbitals of a methane molecule consisting of one carbon atom and four hydrogen atoms.

Semiconductor explained:

Some material properties can not be explained without the use of quantum theory, semiconductor property is one of them. These are solids – chunk of materials constituted of innumerable atoms. We have seen how two is already too many when trying to solve a Schrödinger equation for particles. How could we then have such an equation for all these atoms and solve it? Again, the simplification by approximation does it. Semiconductors are usually crystals, that is a great help in the approximation as crystals are regular arrays of atoms at equal distances from each other – whichever direction of the solid we look into. So every atom's environment is like every other. If we have to formulate the Schrödinger equation for the crystal we have to know the ups and downs of the potential energy everywhere within it. To make this easier to understand, let us try to do it on an imaginary one dimensional (1D) crystal – just a row of atoms at a regular intervals. Of these the outer electrons are usually free to move about throughout the crystal, while the nuclei and the inner electrons form a positive core of the atoms which are basically static at those regular positions in the array. If we try to draw a graph of the potential energy at one of these cores, the positive charge will give a kind of well-like depression going down below the zero energy line. This may be called a potential well tapered downward, and flared upward like a funnel merging with the zero energy line there. While the energy thus goes down the well at the core, it remains zero in the space in-between two cores, till we come to the next core and the next potential well, and so on.

For further simplification, we may neglect the funnel shape of the potential wells and imagine them to be rectangular box-like wells. This would make calculations much easier while not affecting the quantum theoretical points to a significant extent. The variation of the potential energy being easy and regular, the formulation of the Schrödinger equation is straight forward. To try to solve this we could start with a trial solution as expected from the situation. The free electron part will give a wave function which should be a flow of continuous regular parallel waves – having the same energy everywhere in the crystal. But then there are wells at periodic distance and the wave function would be attracted down at every core – giving a periodic function. Therefore the continuous wave part will be modulated by this periodic part – the trial solution thus will be a product of the mathematical expressions of these two parts.

Putting the trial solution into the equation we get the real solution. The solution gives the picture of the allowed states. But whereas the allowed states in an atom are represented by thin lines, here we are dealing with many atoms and the lines here merge into wider bands, and we get the allowed energy bands from the solution, with non-allowed energies in-between. The highest of the energy band is called the conduction band. If some of the conduction band states are filled with electrons those would be in the lowest part of the band the higher part being empty. If then a voltage is applied the electrons at the lower part will gain energy and transfer to the upper part thus creating a conduction. So the materials which have a partially filled conduction band, are conductors. But in other materials the conduction band maybe completely empty. The band right below, however, will be completely filled, and is called the valence band.

For such materials two different cases may arise, the gap formed by the non-allowed region between the conduction band and the valence band is called the energy gap. The first type of materials have energy gaps which are too big, so that at ordinary temperatures even the topmost electrons can not have enough energy to cross this energy gap. As a result the conduction band remains empty as before. As there will not be electrons partially occupying the conduction band, no electrons are available to from a current on the application of a voltage, and no conduction is possible. These materials are, therefore, non-conductors or insulators. But there are the second type of materials for which the energy gap is small. So at various temperatures some electrons can gain enough energy to quantum jump over the energy gap and get into the conduction band.

The electron in the conduction band can now constitute a negatively charged carrier particle to conduct electricity. But while leaving a full valence band the electron had left a ‘hole’ in the valence band and when a voltage is applied, the hole, basically an empty place, can move within the valence band as if it is a positive charge. The hole, therefore, forms another carrier of current along with the electron that had jumped into the conduction band. Thus each such jump creates two carriers of current in the shape of an electron-hole pair – both contributing to the same current. The material behaves like a conductor, but in a limited sense because only a limited number of current carriers are available at a certain temperature. That is why these are called semiconductors.

The conductivity of the semiconductors can be much enhanced by introducing some impurities in them. One type of impurity can donate electrons in the conduction band. This makes negatively charged electrons the predominant carriers of the current, the reason why such a semiconductor is called n type (negative). The other type of impurities accept electrons from the valence band and thus introduce holes there. The positively charged holes become the predominant carriers of current in these semiconductors, and are, therefore, called p type (positive). A junction of p type and n type semiconductors (called p-n junction) is instrumental in the working of the most semiconductor electronic devices such as transistors. Without the application of the quantum theory the semiconductor science and technology would not be possible.

Laser made possible by the quantum theory:

We have seen how light is emitted by atoms. This emitted light always has a spread in its frequency; and also in a beam of such light, the rays are not exactly parallel but diverge with the distance. It is, therefore, not possible to have a very focused intense light of a precise frequency in all our common types of light. But scientists have discovered a special type of light with a special process of creating it which does not have these shortcomings.

When the electron of an excited atom makes a quantum jump to go back to the ground state it is a spontaneous affair; it does not need a cue from anywhere. This is, therefore, called a spontaneous emission. But the quantum theory has predicted another type of emission that can be called a stimulated emission. Suppose an atom has been excited by letting it absorb energy from outside. Then while it is still in the excited state another photon with exactly the same frequency and momentum involved in the excitation of the atom, strikes the atom. This will cause the immediate quantum jump of the original excited electron to the ground state emitting a photon exactly of the same frequency and momentum as the incoming photon, and starts to move along with it. This is the stimulated emission, as it is induced by the incoming photon. Now there are two exactly similar photons in place of one, and thus an ‘amplification’ has happened in the photon number. These are the fundamental processes behind the creation of LASER, and the word is the acronym for Light Amplification by Stimulated Emission of Radiation. The fact that the two photons move together in unison means that there is a fixed phase difference between the two – meaning that the peaks of their waves always maintains the same relationship in how much one is ahead or behind the other. In that case we can call them coherent. So at least these two photons have exactly the same of frequency; and being coherent they would hardly diverge from each other – the properties that we expect from laser. But only two photons will not do; the amplification must go on, and that requires another thing called ‘population inversion’.

In equilibrium, the lower states of an atom are all filled up. So when one of the electrons jump up to a higher state, still it is the lower states which are much more crowded than the upper ones – this is how the electron population is usually distributed. But if somehow the upper states maybe made more populated than the lower ones, we may call it a population inversion, and that would open up the prospect of a greater number of stimulated emissions and thus more amplification. One way to arrange this is to select certain materials where once the atom is excited, there is no immediate quantum jump to the ground state; instead the electron that went up comes down to an intermediate state, only a little below than where it was, radiating a little energy in the form of heat. Not only that, the electron can stay at that intermediate state for a long time; other electrons can do the same populating such intermediate states. Thus more electrons can be up there than the lower states, creating a population inversion. Stimulated emissions through the quantum jumps of these upper electrons to the ground state, can be induced by the photons already moving within the material; thus amplification will continue giving many photons of the same frequency and momentum, all coherent to each other.

Now let us go through the whole sequence how laser is produced. To begin with, an ordinary flash light is used which gives an intense light momentarily. Some laser, however, may use more common source of light giving not so intense or momentary light. This excites the atoms, and the special material used make the electrons thus excited to come down to intermediate states creating a population inversion there. The first photon with the characteristic of this laser light is then provided which causes the first stimulated emission of a second photon exactly with the same frequency and momentum, and coherent with the first. Two perfectly reflecting mirrors are placed face to face within the material, perfectly parallel to each other. The two photons already there would be repeatedly reflected back and forth the mirrors and would induce more stimulated emissions through the electron jumps from the crowded intermediate states. Thus two photon become four, and then eight, and so on, and the amplication quickly increases the number of photons to very high, providing intense rays of light of the same frequency, and perfectly coherent. Of the two mirrors, one is at the front through which laser is supposed to come out, and the other is at the back. The front one has the reflectivity of only 99%, while the back one has it 100%. The remaining 1% light from the front mirror is transmitted through it and goes out. Very soon that 1% makes a very intense beam of laser rays. Because of the perfect parallelism of the mirrors, the rays are exactly parallel; a ray which comes deviating from parallel even by the slightest amount, would be immediately lost to the beam. So we get a very intense coherent beam of light of a precise frequency which can go a huge distance without spreading out even by a little. It is these exceptional properties that make laser so useful – from melting metals to executing intricate medical operations, to optically carrying digital information to a great distance. The science that made laser possible is out and out quantum theoretical.

Quantum behavior in things of our size

The quantum world is the world of things of atomic dimension. However, we can experience this in the phenomena of light, chemistry, crystals etc. things which have expressions in things of our size. But there are a few phenomena where quantum theory is applied directly on things of our size, bringing the whole chunk of visible material comprehensively within the fold of quantum theory. Superconductivity and Bose-Einstein condensation are two examples.

The quantum theory of superconductivity:

The vibration of atoms offers the main resistance to the electric current by scattering the carrier electrons. As the temperature is lowered the vibration gets reduced, and so does the resistance, increasing the conductivity of a material. In 1911 a phenomenon was discovered in which the sharp reduction of the resistance could not be explained by the lowering of vibration alone. It was noticed that mercury when cooled down to a temperature of precisely 4.2^0 Kelvin (about 269 degrees below the temperature of the melting ice) its electrical resistance becomes zero! This unthinkable phenomenon was called superconductivity. Kept at that temperature (called transition temperature) or below, mercury would demonstrate a perpetual electric current without any further application of a voltage. Later, superconductivity was found in other materials all at a very low transition temperature comparable to that of mercury. Superconductivity gave rise to many other research results and had many applications, but it took until 1957 to have a theory explaining this phenomenon. That year three scientists Bardeen, Cooper and Schrieffer could give a successful theory, that came to be known as BCS theory using the first letter of their names.

The theory showed that the basic reason for superconductivity is a surprising quantum effect that forms pair bonding between electrons, which normally should repel each other because of the same negative charge they have. But under the conditions of superconductivity, an electron interacts with the rest of the crystal which as a whole consists of arrays of the positive cores (matrix). An electron will sort of attract the matrix towards itself, thus creating an undulation in it. The undulating matrix in its part will attract another electron towards it. As a result, the second electron comes close to the first through the intermediary of the matrix. The electrons thus form a pair known as Cooper Pair, overcoming the repulsive force between them. Now the undulating matrix may be regarded as a quantum of sound to be referred to as phonon. This is because any vibration is a sound, and ‘phon’ is the Greek word for sound. So the whole interaction may be represented as a electron-phonon-electron quantum interaction.

The cooper pair itself can be regarded as a new kind of particle which unlike electron’s $\frac{1}{2}$ spin, has a whole spin, and therefore is not subject to the Pauli exclusion principle. It is this exclusion principle which prevents more than two particle to be in one state. The cooper pairs, not subject to the exclusion principle, would all reside at the lowest state, as the natural tendency would be to be at the lowest energy possible. All the electrons now residing as cooper pairs within the lowest state, there is no way they can interact with the vibrating atoms; so they can remain to be the carriers of current without in any way creating a resistance. To have any such resistance-producing interaction, the cooper pairs will have to cross a small energy gap between the lowest band and the band above it – bands not unlike those in the semiconductor we saw. Such a crossing is not possible at such an extremely low temperature. Another way to look at it is that whenever one electron of the cooper pair tries to have a collision with a vibrating atom, the other electron of the pair will pull it away, maintaining a zero resistance. Here the totality of the chunk of the crystal is participating in the quantum process of the superconductivity , as a whole. In superconductivity we are experiencing a quantum effect involving a whole material of our size.

Bose-Einstein condensation:

In 1924 a young professor of Dhaka University in the British India (now in Bangladesh) Satyendranath Bose sent to Einstein the manuscript of a short scientific paper. Einstein arranged to get it published, and published a paper of his own on the same topic. Bose's paper gave a completely new insight into the behavior of photon, and Einstein's paper took it up further predicting a very exceptional new state of matter beyond the states already known – solids, liquids and gases. This new state was later known as Bose-Einstein condensed state. The two papers also gave rise to a new quantum statistical theory called Bose-Einstein statistics. All the particle that are subjected to this statistics are called boson, taking Bose's name, photon being an example. All other particles are called fermion, those subject to the statistics given by Enrico Fermi, those which come within the Pauli exclusion principle, electron being an example.

In his 1924 paper, Bose imagined photons as molecules of a gas on which the quantum theoretical constraints were applied. He thus showed that photons can all take the lowest energy state one falling on the other hardly distinguishable one from the other. On the basis of this, Einstein in his paper predicted that at a sufficiently low temperature bosons of any kind will condense at the lowest energy to form a completely new state of matter in which viscosity that gives liquids their stickiness, will become zero. This will result in an infinite fluidity for the material. The phenomenon was later named as Bose-Einstein condensation. Within ten years of the prediction, a sort of realization of it came in the shape of superfluid helium. When helium was cooled to the extreme law temperature of 2.17^0K the liquid helium lost all its viscosity, became a superfluid, and climbed up the wall of the test tube to overflow out. This phenomenon had some resemblance to the superconductivity. In the case of superfluid helium, each atom of helium acts as a boson, whereas in the case of superconductivity each cooper pair of electrons is a boson. Both phenomena are created by their boson characters.

But this superfluity was not the full manifestation of Einstein's prediction; that would need much lower temperature– extremely close to the absolute zero temperature. Perhaps Einstein did not conceive that such low temperature could ever be achieved. But it was achieved through some spectacular experiments in 1990s. The real Bose-Einstein condensation needed a temperature as low as a billionth of a degree above the absolute zero (– 273^0 C). Absolute zero is the never achievable lowest temperature; but one can approach very close to it. This was achieved on rubidium through two new techniques – one by bombarding the atoms with appropriate laser photons which took away the last beat of temperature from the atoms. The second one is to keep the atoms in a sort of magnetic trap where the warmest atoms are at the top; then lowering the trap wall slightly would allow those warmest (comparatively speaking) atoms to escape leaving only the very cold ones. As the required cooling was achieved Rubidium atoms acted as bosons condensed at the lowest state, one falling above the other as if to form a huge single atom. Though it grew into such a big ‘atom’ that we can call it our size, it still remained very much a quantum entity subject to all the quantum properties such as the uncertainty principle. At a temperature of almost absolute zero the momentum being almost zero and very certain, the position of this huge atom would be very uncertain; that makes it extremely inflated and fuzzy – a kind of willow the wisp! It is very much our size, and also a quantum entity. Already we could observe the quantum superposition among such huge atoms of our size – quantum theory in action at this size.

The quantum horizon of the existing technology

Quantum clock:

The precision time keeper we know as atomic clocks actually run in quantum principles. Here the precise frequency of the emission of light from the atom is used to keep the time. This way, time can be very accurately kept – in the best cases it may not lose or gain a second in 5 billion years! For our usual chores we do not need such accuracy. But this is very useful under certain circumstances. For example, one of the sophisticated systems very familiar to us by its everyday use is the GPS system, and this gives our global position by measuring the time the signals from various satellites take to reach us and to get back to them. Even a very little inaccuracy in the time would make the results seriously wrong. Many such applications and many scientific investigations need precision measurement of time by atomic clocks.

The effectiveness of an atomic clock depends on two factors – its stability and its accuracy. The frequencies emitted by atoms have a certain spread in them because the quantum jumps from various atoms are not exactly synchronized. It is only with the average frequency of all atoms that the theoretical frequency is expected to match. The less time over which the averaging has to be done to match the theoretical frequency, the better is the stability. Previously the time for averaging required used to be days and even months, now it is only seconds. The accuracy of the atomic clock is proportional to the square root of the number of atoms used. If the number of atoms is increased 100 times, the accuracy would increase only 10 times. The increase of the number of atoms also increases the difficulty of keeping all those atoms synchronized to keep the stability; so the accuracy may come at the cost of stability.

One interesting quantum property has been useful in increasing the accuracy of the clock without increasing the number of atoms. We will discuss that property in the next chapter. But it is sufficient to say here that if two particles are created in a common process, some of their characteristics remain connected to each other even if the particles are far away from each other. Thus some atoms can remain outside the clock but still can participate in the time keeping because they are quantum theoretically connected with the atoms within the clock. Thus we could increase the effective number of atoms without facing the hazard of reducing the stability. It has been possible to achieve records in the stability as well as accuracy in the latest atomic clocks, which have become quantum clocks in more senses than before.

Quantum microscope:

We have already met an exceptional microscope using quantum principles – Scanning Tunneling Microscope (STM). But there are other ways in which the microscope has been improved by the quantum theory. The electron microscope use electrons instead of light. This is possible because quantum theory has shown that electron also behave like waves, but has a much shorter wavelength than light, thus enabling much higher magnifying power of a microscope.

Another type of microscope use the interference of light for its use, but needs an exceptional quantum property to use it better. In such a microscope two rays of light from a single source is shone on a very small object which we want to see. Now the two rays will interfere with each other according to the path difference they have in reaching the object. The path difference is created because of the high and low bumps on the surface of the object. So the interference pattern will highly magnify the bumps and make the details of the object’s surface visible to us. Even if the bump is one billionth part of a millimeter, it would be shown in the image created by the microscope.

The problem is – this needs comparatively very intense light to work satisfactorily. Such intensity may not be suitable for certain subtle objects such as live biological tissues, the energy of the light itself damaging altering the object. On the other hand, a lower intensity will reduce the signal to noise ratio, and more ‘noise’ will be in the image than the actual signal information from the object. It is found that if the number of photons in the light is reduced N times than the signal to noise ratio will be reduced by $\sqrt{N}$ times. The same quantum property came to the rescue here as in the atomic clock mentioned above. Some of photons are kept outside the microscope but are connected with the photons which inside by that quantum property. These are not contributing to the intensity of light, but can provide more information about the interference and image, because of their connectedness with the photons inside. Then signal to noise ratio becomes much higher than $\sqrt{N}$, improving the image, while keeping the intensity low.

Quantum computer:

The computer we have been using so far has been based on semiconductor electronics. Its capacity has been increasing in leaps and bound as the density of the number of transistors (the basic logic elements of the computer) accommodated on an area of the semiconductor chip increased steadily. But now it seems to be approaching a limit when a greater density can not be achieved without hampering its functions. Therefore, radically different ideas are being contemplated for future computers; the quantum computer is one of these.

Our digital computer work based on a binary system (two digits) – all the information are coded in terms of two states designated by 0 and 1. It may be realized by the ‘off’ and ‘on’ state of a switch of electric current, for example. The processing takes place according to a logic determined by a binary algebra, with the ‘bits’ in the shape of 0 and 1 as its constituent. Binary processing needs many steps one after another because it is limited to only two states. But computer being fast, and its memory density being high, this did not matter for the type of work done so far. But now that a limit is being approached and the future computational problem would be more complex requiring faster and simultaneous processing, the binary processing may not be the best idea.

The quantum computers would use quantum bits (qubits) instead of bits, actually realized by the various superposed quantum states of 0 and 1 states in a particle which can be photon, electron, atom ion (charged atom) etc. A whole new algorithm is used to process these many qubits, whereas the binary algorithm was used for the processing of only the two bits. As qubits are many more than two in number, they can represent many more data and can process much faster, making simultaneous parallel processing also possible. A conventional computer based on two bits has to finish one step before it can take up the next one; but here processing on several problems can go in parallel, resulting in superfast computation.

But there is a basic problem, the qubits are quantum entities based on the superposition of states. The final result the processing would give, would also be a superposed state. The moment we would like to 'read' it or measure it, the superposition would collapse and we would see only a classical bit like 0 or 1, losing the intricate information that the output is supposed to give. We have already seen how the moment one measures a parameter it ceases to be of quantum character and only ordinary classical manifestation results – such as either the thing is there or it is not there (like 1 or 0), and no superposition between the two.

This is the ultimate problem, but before that we have to face another problem of similar nature – how do we contain those quantum particles, the qubits, in a container? The container necessarily has to be classical in nature, and will thus affect the quantum character of the particles. Several alternatives are being tried to avoid this problem. One of these is a magnetic trap which can confine the particles and is itself a quantum entity; another possibility is a light trap consisting of coherent laser which itself is a quantum entity too. Some very small specks of semiconductor can be kept very isolated from anything classical, and can serve the same purpose; any such speck is called a ‘quantum dot’. A similar arrangement can be made using a superconductive material kept an extremely low temperature below its transition temperature. Being superconductive it is a quantum entity and can act as a container of qubit particles. None of these are easy to implement and are offering difficult technological challenges. But to a limited extent, some of these have already been achieved.

Now back to the ultimate problem of reading the result of the quantum computer. The most feasible way out is a similar process we have already met in connection with the quantum clocks and quantum microscopes. It is to have pairs of quantum-connected (formally called quantum entangled, as we will see in the next chapter) particles– one member of the pair remaining within the computer, the other outside it. A measurement on the outside particle will tell us the result of the processing because this would reflect the state of its partner within the computer which has taken part in the processing. The measurement outside, however, will not cause the collapse the quantum states of the qubits in the computer. We would have some more details when we discuss the quantum entanglement in the next chapter. Some rudimentary quantum computers have already been successful, though we are yet to achieve those super-performance that we would expect from the future quantum computers.

Some more wonders

Quantum theory matched to relativity theory

The two revolutions:

The early 20th century saw two revolutions in physics – the quantum theory and the relativity theory, both deviating completely from the classical physics in their own ways. Though they negated many of the established theories of classical physics, both of these have been spectacularly successful in predicting the nature, and in causing paradigm shifts in physics. However, both have areas of applicability where their effects would be significant. In other areas the effects would be negligible and the classical physics would still prevail. Usually the domain of the quantum theory involves the tiny things whereas that of the relativity theory involves the very fast things and the massive things. Therefore, for most cases no conflict arises even if the two theories are not compatible to each other. But in certain cases their domains overlap, and the two theories have to match each other in such situations. For example, the world of photons and electrons is the world of both the small and the fast – and thus are subjects to both the quantum theory and the relativity theory. But as the two theories were developed independent of each other, one tends not to take the other into account, or even when they tried to do so, the task was found difficult.

The particular overlap of photons, electrons etc. happens between the quantum theory and one part of Einstein's relativity theory known as the special theory of relativity. This part of relativity theory tells about the relativity of space and time when looked at by two observers who are in uniform velocity in straight line with respect to each other. This uniform motion makes it special. Some of the conclusions it reaches are the following: nothing can go faster than light; objects shrink as its velocity increases (space contracts); time gets slower, and mass increases in the same situation. There is no absolute space or absolute time – these two are always relative to the motion of the observers; time and space are equivalent, one can be interchanged for the other in a 4-dimensional space-time (3 for space and 1 for time).

According to this, an object gets shrinked in space, and the time gets slowed down for it, both significantly, only if it moves at a velocity in the order of the velocity of light. Quantum particles such as photon and electron fulfill this criterion, and that is why they come within the jurisdiction of both the quantum theory and the special theory of relativity. Einstein later extended his theory to include accelerated motion too, and called it the general theory of relativity. This showed that the acceleration and gravitation are the same thing; the space-time get distorted near the mass and that is the reason for gravitation (a new theory for gravitation); the amount and the distribution of all the masses in the universe gives the latter its shape, and predicts its future. The effects of the general theory becomes significant at the level of massive things only.

As the quantum particles are very much within the domain of the special theory, the developers of the quantum theory tried to accommodate it within their quantum formulations right from the beginning. The early attempts did not succeed, but the success came ultimately, as we will see. The attempts to match the quantum theory and the general theory of relativity has not, however, succeeded as yet, and remains a challenge for future.

Dirac equation:

The lack of inclusion of the special theory within the quantum theory was a disappointment for the early quantum theorists. This disappointment was increased as the electron spin was later discovered – something whose classical equivalent was a spinning top. It was soon found, however, that the spin is a purely quantum mechanical phenomenon and can not be quite understood if the special theory is not incorporated within the quantum equation.

If we recall the Schrödinger equation we notice that on one side of it there is a two-fold rate of change, meaning the rate of change, of the rate of change of Ψ, with respect to the space. The other side has only a simple rate of change of Ψ with respect to time – an ordinary one-fold rate of change. So obviously, the equation does not put space and time on equal footing, which is a necessity in the special theory of relativity. Schrödinger had tried to use two-fold rate of change on both sides to put space and time on equal footing, and thus to accommodate the situation demanded by the special theory in his first attempts in an equation. But this created so much mathematical complications that he had to settle for his now famous equation disregarding the needs of the special theory. Later attempts by others to accommodate the special theory created equations which failed to yield some of the well-established results of the quantum theory given by the Schrödinger equation.

Finally, it was a British scientist Paul Dirac who succeeded in developing a relativistic quantum equation which achieved desired goal, including an explanation for the spin, without sacrificing any of the well established results of the quantum theory. Dirac could keep the equivalence of space and time by using one-fold rate of change with respect to both. Moreover, his equation was a

4-dimensional one as demanded by a 4-dimensional space-time of the special theory. The equation, therefore, has four components giving four different solutions. This is roughly in the same way as a 2-dimensional equation represented by $x^2 = 4$ has two mathematically perfectly valid solutions as $x = 2$ and $x = –2$. But being the equation of a physical phenomenon, Dirac had to have a physical meaning of each of its four solutions. Out of these four solutions, Dirac found two are positive and two are negative. The two positive solutions did not pose any problem in ascribing a physical meaning to them as the two spins – spin up and spin down. These involves the two quantum numbers $+\frac{1}{2}$ and $-\frac{1}{2}$, spin itself being a positive quantity in both the cases. Classical assumptions like the spins of a top are no longer required to explain the practical results of these spins – they simply follow from the solutions of the equation which itself in based on purely quantum theoretical and special relativistic considerations of the Dirac theory. Yes, there is a magnetic moment involved, associated with the quantum angular momentum of the spin, but this too is a purely quantum effect, and is correctly predicted by two positive solutions.

The problem was with the physical significance of the two negative solutions of the Dirac equation. They seem to represent two strange states of the electron where energy is negative! Negative energy does not make any sense in physics – it gives a scenario in which when an object is pushed in the direction it is going, instead of gaining speed it rather loses speed! If something absurd like this results from any solution of a physical equation, the practice of the scientists is to discard it and to keep only the meaningful ones. Dirac refused to do that. Rather, he tried alternative forms of his equation that can avoid negative solutions. He succeeded, but he did not accept those equations because they were not ‘beautiful’ as his original one. Dirac believed that the mathematical elegance (beauty) of an equation is a proof of its correctness, because the nature prefers beauty. So he stuck to his original equation and rather concentrated on showing that the two negative solutions are not absurd after all – the nature perhaps is trying to say something real through those solutions.

Dirac thought that perhaps in a special circumstance an abrupt quantum jump can take place from a negative energy state to a positive energy state. The circumstance he imagined was very unconventional – it is as if the whole universe is immersed in a negative energy sea. This sea of negative energy is filled up with electrons without any vacancy, and the whole thing is in perfect equilibrium. In this situation, there will not be any interaction between this sea and anything else, because like fish immersed in water electrons would not know of any other entity. But the moment an electron changes its state to a positive energy one, the situation changes. This electron can now very well be an ordinary electron. At the same time, there will be a vacancy or a ‘hole’ in the sea of negative energy states in the place which used to be occupied by the electron that changed its state. Within the otherwise completely filled sea this ‘hole’ can now move about, as if it is a particle with positive charge, otherwise similar to electron, which has an equal negative charge. We have already seen such a ‘hole’ in a different context when talking about semiconductors. Proton being the only positive particle known at that time, Dirac concluded that the negative energy solution represents a proton, the two solutions indicating two spin of the proton – spin up and spin down. But soon he realized that the ‘hole’ he thought about could not represent a particle which is 2000 times heavier than electron, as the mass of proton happens to be.

There was no particle known at that time which could foot the bill on this count too. But Dirac remained undaunted, he had unweavering trust on the correctness of his equation, and the physical interpretation of its solutions. He boldly predicted that such a particle will be discovered, and stuck to his theory. He even coined a name for the particle yet to be discovered – anti-electron. This would be exactly similar to electron except for its positive charge equal to the charge of negative electron. Within a few years of the prediction, the particle was discovered in 1932 by an American scientist Anderson in the cosmic rays, matching in everyway Dirac's predicted anti-electron. The name was later changed to positron.

The Dirac equation is applicable to any particle with spin $\frac{1}{2}$, same as the electron. Thus any such particle should have its own anti-particle, similar to the electron's positron. Such anti-proton, anti-neutron etc. have actually been discovered. As protons, neutrons and electrons can form an atom, so can anti-protons, anti-neutrons and positron form an anti-atom. It has been experimentally shown that a superposition of a particle on its anti-particle annihilate the both and create only energy. So it has been theorized that a contact between matter and anti-matter, the latter constituted of anti-atoms, would result in the annihilation of both. Such are the dramatic conjectures of theoretical physics that could follow from Dirac equation, and his bold predictions from it. Most importantly, Dirac equation made quantum theory compatible with the special theory of relativity.

Quantum field theory

Field is the main actor:

Quantum theory was still encountering problems while applied to certain events such as photon-photon interaction, and electron-positron pair creation from vacuum. When a photon bumps on another photon the quantum theory has little to say about it. As for the pair creation, it was found that an electron-positron pair can be spontaneously created out of nothing (vacuum), but the pair is very short lived as it gets annihilated immediately; there is hardly any quantum theoretical explanation what makes such pair creation inevitable. Then the nucleus is also full of mysteries – why do protons and neutrons crowd together in it, why should radioactive nuclei get chipped off spontaneously giving a part of its positive charge, how could the negative beta particles (electrons) defy the attraction of the positive nucleus and get out, wherefrom it comes in the nucleus consisting only of protons and neutrons anyway? Many of these events can be showed to be possible because of the uncertainty principle. But that does not explain why they should occur. It was felt that if one important component of the classical physics, the field, could have a quantum counterpart, some of these problems could be predicted by the quantum theory.

We have seen that the concept of field in the classical physics such as the electrical field and magnetic field have done wonders in explaining things like electromagnetic wave and interaction in classical electrodynamics (science of motion involving electromagnetism). A similar achievement was still absent in the quantum theory. Here is an irony. At the beginning, the scientists were treating quantum theory as a bizarre theory imposed upon them by the circumstances, which they wished would go away. But now they were so dependent of this theory that the absence of a quantum theoretical electrodynamics with fields and all, was making them sad.

Thus first came the quantum electrodynamics, a simple expression of quantum field theory. This is about electrons and electromagnetic field. But here field comes in a different shape than the classical field. Field is like a ‘thing’ in the classical physics (remember the lines of force), but in quantum theory it is nothing like a ‘thing’; rather it creates the ‘thing’ such as a particle. That is what field is for. Not being a thing does not take away all those forces which are expressed as a quantity at every point of the space. To understand that quantity (the intensity of the field) better, we may imagine a physical construction at every point, which we can later discard as a scaffold no longer necessary. Let us say there is ball at the top of a spring vibrating up and down, at every point of space; The height of the ball gives the intensity of the field at that point. Let us assume that every ball-spring is connected with every other with other springs. Now if I strike at any part of the system, a wave of vibration would flow through the system. From the quantum theoretical point of view we can describe it as the creation of a particle that passes through the field. In this picture the field is more fundamental, it creates the particle. Now striking again, we can create a second particle which may collide with the first – one may scatter the other. The characteristics of this interaction will depend on the properties of the field – in this picture those properties are determined by how hard are the springs of the ball-spring, how hard are the springs connecting various ball-springs, and the like. Though we could describe the vibration passing through the system as both waves and particles according to the wave-particle duality of the quantum theory, the quantum field theory always look at them as particles.

In this make-believe picture we created the particle by striking anywhere in the system; but in quantum theory we can not strike with any energy, it has to come in terms of quanta – no fraction is allowed. This imposes two conditions on the vibration of the ball-springs. First, they can never be still, because there always will be a minimum energy. The second, the balls can only be at certain allowed heights, not at all heights; this means the field is quantized. So now only the particles with certain allowed energies can be created. For an observer who has no relative motion with the particle, the allowed energy will be equivalent to a 'rest mass' of the particle. The mass, however, will increase according to the relative speed of the observer – all according to the special theory of relativity. It is the rest mass which we can talk about as the minimum mass of a particle.

It is the electron field that creates electron. The production process can be looked at as an interaction of electron field and the electromagnetic field. Because this interaction energy is very small, the rest mass of electron is very small. On the other hand, when the scientists were looking for another particle called Higg's Boson which has a very high rest mass they had to use a very high energy particle accelerator where comparatively heavy particles collided with very high energy. In this process the interaction energy with the Higg's field being very high, the particle found, the Higg's Boson, has a very high rest mass. On the other end of the scale a very little burst of electromagnetic wave creates a photon which has zero rest mass. It is the field which is important in every case. Ultimately it is the field, which through various interaction process, could solve the mysteries of the problems mentioned at the beginning of the section.

Feynman's version of the field theory:

Quantum field theory has been discovered independently by several scientists and each was very successful. One version of it due to the American scientist Richard Feynman has been quite illuminating in certain ways. To have a glimpse of this it would be helpful to look at the somewhat different interpretation Feynman gave to the quantum theory. Long ago in 1657 famous French scientist and mathematician Fermat gave a curious explanation of why light travels in a straight line. He said light takes a straight line to go from part A to point B because by doing so it can reach the destination in the shortest time – he called it the principle of least time. Feynman adapted the principle in a quantum theoretical way, and raised the obvious question – how does the light know beforehand that following this path it would need the least amount of time? Feynman's answer to this question is very striking; according to him the path the photon particles follow can be obtained from the sum of all the infinite number of paths possible between point A and point B, and in the case of photons that sum happens to be a straight line. This interpretation is called 'the sum of all paths'. To make this process of taking the sum possible, each possible path is assigned an amplitude according to the weight it carries within the sum. The sum, or rather the integration as in the integral calculus, is taken over all these amplitudes. We can, for example, apply the concept on the double-slit experiment for photons and look at the whole thing as an alternative interpretation, Feynman's interpretation. In this a photon may take all kind of paths from the source to a particular point on the screen via the two slits. The actual path will be determined by the sum of all these possible paths; and the sum of all the amplitudes of these paths will be equal to the probability calculated from the wave function method to reach that point on the scale. These sums of amplitudes to reach the different points of the screen will create the familiar interference fringe. This will work out like this for all quantum things – photon, electron all. But if we try it on tennis balls, for example, each of the balls will end up at the same point of the screen, because, the sum for reaching any other point will be zero – giving a classical result, because the tennis ball is not a quantum thing.

We have seen how in the quantum field theory a quantized field is creating a particle from its energy quantum and then interacting with that particle. For example, a quantum of electromagnetic field creates a photon during the interaction between two charged particles such as electron and electron, or electron and proton. The photon soon vanishes. This is the quantum theory of electromagnetic field. Similarly, if an electron appears and vanishes – that would be the quantum theory of electron field, and so on. In every case, the core of the theory is its field equation, the solution of which gives the practical results. But the exact solutions are never possible; all we can do is to solve approximately by a method called perturbation method. In this method the part of the equation which does not have complexities, and yields an exact solution, is separated out and is called the zeroth order equation. Now to approach the real case a perturbation or disturbance is added to it as a sort of correction term. This may come as a first step of correction which maybe comparatively a big one and is called the first order perturbation. This maybe further corrected by adding a second order perturbation, which is smaller. Thus finer and finer correction terms are added if one would like to take the equation closer and closer to the actual one.

In terms of his concept of the sum of all paths, the most direct path gives the zeroth order equation. But the other paths are not neglected; these come as perturbation of various orders. The actual calculation of the sum (integral) can be extremely complicated. To keep an easy track of the information about which field is interacting with which other field, Feynman expressed these as a diagrammatic system known as Feynman diagrams. These diagrams have been very useful and made his version of the quantum field theory quite popular. But all versions of the theory have achieved wonders by revealing intricate information about nature to an accuracy of unbelievable degree. In a way, this has been the most beautiful vindication of the quantum theory.

The quantum entangled states

The meaning of entanglement:

We had mentioned quantum particles which are related to each other because they were produced together, so that when one goes into one state, the other automatically goes into a similar or related state in spite of the two being separated by a long distance. We saw how this phenomenon has been used in several technological innovations. Here we describe this strange quantum phenomenon known as quantum entanglement, in more details. First a little history of its discovery.

In 1935 Einstein along with two other scientists, Podolosky and Rossen, expressed some apprehension that the quantum theory is not yet a complete theory and is waiting for a proper completion. The view expressed is usually referred to as EPR as the initial letters of the names of its authors go. The view resulted from a thought-experiment they did. A thought experiment consists of all the steps of an experiment as these are thought out in the experimenters' head, but nothing is done practically. Such execution of the steps of an experiment in mind only, can give useful results too. This thought-experiment showed that two particles can be related to each other quantum theoretically in such a way that some or other properties of theirs can be similar or related. Then if somebody measures the property of one of the pair, that of the other will be automatically known. If one is changed, the other will automatically change, even if the particles are separated by a long distance. This is what the quantum entanglement means. The initiators of EPR found it very puzzling, and soon we will see why.

But first let us give a concrete example of the entanglement. Two particles maybe created in such a way that their spins are opposite to each other – if one is spin up then the other is bound to be spin down. We may move one particle from the other miles away even million miles; still they would remain entangled so far as their spin is concerned. If the spin of one particle is placed in a particular direction, the other particle's spin will be automatically oriented to an opposite direction. And what is more, in spite of the huge distance between the two particles, the latter will follow the former's action instantaneously, without requiring any time for the information of the one's orientation to travel to the other! This information then is moving faster than light, which is a violation Einstein's special theory of relativity. EPR insisted that such impossible result coming out of this thought-experiment is a proof that the quantum theory is incomplete. The thought-experiment led to some real experiments, which proved in practice, what EPR only thought. It was clearly demonstrated that quantum entanglement is a natural result of the quantum theory. However, eventually it was found that there is nothing in it that would make quantum theory incomplete, because in the real sense there is no violation of the relativity theory – as we will see.

Let us look at one of the practical experiments that showed quantum entanglement in reality. This experiment was conducted by a scientist named Aspect in 1976. For this experiment photon particles were obtained by exciting a calcium atom electromagnetically. Having a common source of birth the photons would be entangled to each other with their state of polarization related. To get an idea of the polarization we should consider the photon as a wave where normally the undulation of the electromagnetic wave takes place in all planes perpendicular to the direction that the wave travels. When all these are made to undulate on a single plane, we call the wave to be polarized. When considered as a photon particle this property of polarization is still with it in a quantum theoretical sense. The relatedness of the polarization in the two entangled photon in this experiment was that if one of the photons has its polarization perpendicular to a certain plane, the other would be parallel to that plane. The way photons were produced as entangled pairs, the members of the pairs with perpendicular polarization are supposed to go to one direction, and the members which have parallel polarization to the opposite direction. While polarization of photons were measured in the beams going to the two directions, the relationship as expected in the entanglement was established statistically, thus proving the phenomenon of quantum entanglement. Later, similar experiments confirmed entanglement when the two members of the pair of photons were far apart. This showed that the entanglement, resulting from the quantum theory, is a law of nature; and there is no way someone can ‘correct’ the quantum theory by getting rid of this phenomenon.

Various ways of getting entangled particles:

In the practical experiment we described, the entangled particles are photons which got entangled because they were born out of the same calcium atom. Thus one way to get entangled particles is to have a common source of birth; and photons are the easiest particles to get so entangled. We can get entangled photons by other means too. For example, we can divide one high energy photon into two each having half of the energy – then both of these would remain entangled with each other. This can be done by shining a photon of violet color (high energy) on the face of a certain kind of crystal, resulting in two photons of another color (smaller energy) going out from the opposite sides of the crystal – remaining entangled.

Though photons are easier to produce, they are not the best entangled particles to use. One strategy is to start with entangled photons and then use them to produce second generation particles which too remain entangled. For example, we can arrange the two photons to be absorbed by two atoms. The state of each atom will depend on the polarization of the photon it absorbed. The two photons being entangled, this would now make the two atoms to be entangled. Unlike photon, atoms are particle of matter, and have slower motion. These two properties would make atoms much more useful in applications such as in a quantum computer. There in another way to create entangled particles, this time without needing any photon. Two atoms can be placed very near to each other and are excited to such a high extent that they would interact to get entangled. We see in all these means of creating entangled particles that the particle concerned have to share a common history. This history can never be erased from them; that is why they remain entangled forever, even if they are separated by a great distance.

The information flow is not instantaneous:

Let us now examine whether the quantum entangled particles really have a way to communicate information faster than light. For this we would imagine a situation found in a quantum computer. As we have seen before in the case of quantum computer, bits are not just 0 and 1, they are qubits – various quantum superpositions of 0 and 1 too. But we can not see the qubits in the final result, because the process of seeing (measuring) collapses them to classical state giving back only the bits 0 and 1. We avoid this situation by placing somewhere outside the computer, the quantum entangled particles of the particles acting as qubits within the computer. The particles inside the computer (qubits) giving the result, will have their counterparts outside automatically taking up to the related state. These can be measured without collapsing the computer itself. Let us call the particles within the computer inner particles. In an actual quantum computer the inner particles, serving as qubits, are never measured. But for our purpose let us assume that there is a mechanism to measure those too.

Let us measure one of the inner particles, then its superposed state collapses and suppose as a result, we get the classical bit 1 (it could be 0 too). Till this measurement of the inner particle, the corresponding outer particle was giving the opposite superposed state of the inner one. But now it would give 0, as the opposite of 1. Now, if the scientist who is with the outer particle measures it, he will see 0. Let us suppose that the outer particles are 10 miles away from the computer with its inner particles. Scientist with the outer particle would not know whether the 0 she is seeing now, is the result of a measurement of the inner particle creating the opposite 0 in the outer, or whether it is just giving a classical value 0 because it has collapsed due the measurement she has just made and the inner particle has not actually been measured as yet. To know which of these two alternatives is true she has to talk with her colleague who is with the inner particles. This she can do by using phone, or email or something else – all of which is a signal travelling at or lower than the speed of light. Now if the two scientists, 10 miles apart, expect that they can communicate with each other instantaneously just by measuring their part of the entangled particles, and thus knowing what coded signal the other scientist has sent her, they would be disappointed. Because without that phone call/email communication one would not know whether that is a coded message from the other or not. So instantaneous communication by the entangled particles are impossible. Thus the quantum entanglement does not violate the special theory of relativity, after all.

Quantum theory, which seemed like a wonderland to the scientists to begin with, has become a familiar landscape to them with the progress of science and technology that depend on that theory. But it is still a wonderland to our usual way of thinking about things around us; and will always remain so.