Notes on wavefunctions

The double slit experiment

In the double slit experiment, a beam of light is send through a pair of slits, and then observed on a screen behind the slits. At first, we might expect to see a pair of bright lines on the screen, behind the two slits. In reality (assuming the separation between the slits is similar to the wavelength of the light), the brightest region is exactly in the middle, between the two slits. Surrounding this bright region are a series of bright and dark areas. If we define θ to be the angle of deflection of the light (see the diagram in the notes), the bright spots occur for $\sin(\theta) = n\lambda/d$ with $(n = 0, 1, 2, \dots)$ and the dark regions occur for $\sin(\theta) = (n + \frac{1}{2})\lambda/d$ with $(n = 0, 1, 2, \dots)$.

The explanation for this pattern of bright and dark spots is based on the interference of light from the two slits. At the location of the slits, the light waves are in phase with each other: the electric field oscillates up at one slit when it oscillates up at the other slit. The light waves then propagate outward from these two slits toward the screen. In the middle of the screen between the two slits, the distance to each slit is the same, so the light waves from these two slits are still in phase with each other. The two light waves constructively interfere, and we see a bright spot. As we move away from the center, we come to a place where the distance between the two slits differs by half a wavelength. Here, the two light waves are out of phase with each other; thus, we have destructive interference (the waves cancel each other out) and we have a dark spot.

The points on the screen at angles $\sin(\theta) = n\lambda/d$ are exactly the ones for which the distance to the two slits differs by an integer number of wavelengths (so that we get constructive interference) while the angles $\sin(\theta) = (n+\frac{1}{2})\lambda/d$ are the ones for which the distance to the two slits differs by $n+\frac{1}{2}$ wavelengths (so that we get constructive interference). The details of this may be found in the text, but for our purposes, there are two essential points to take away:

- The pattern on the screen is explained by understanding how light from the two slits combines.

- The pattern on the screen is explained using the fact that light is a wave with a particular wavelength.

The double slit experiment: photon explanation

What happens if we reduce the light intensity so much that only one photon at a time passes through the apparatus? In this case, we'll need to replace the screen with a more sensitive detector that can tell when and where a single photon hits (such a detector could be built from an array of photomultipliers or with a device similar to the "film" in digital cameras). In this case, we find that each photon hits a specific spot on the screen, but after many photons have hit, the distribution of hits reproduces the pattern of bright and dark regions that we had before.

How could we possibly get the same pattern as before if the pattern was caused by light from the two slits combining and interfering? If we just send individual photons through, don't they have to go through just one slit or the other? If so, how can there be anything like interference? The answer is that the individual photons don't just go through one slit or the other. Somehow, *each photon is able to go through both slits and interfere with itself*! To convince ourselves that this is true, we only need to repeat the experiment, but randomly cover up one slit or the other each time we send a photon through. If individual photons can only go through one slit or the other, this should not affect the pattern of hits, except that we should have half as many photons hitting the screen. But in fact, when we cover slits in this way, the pattern of hits completely changes. We conclude that the single photons must know about both slits; rather than being pointlike particles, the photons must be spread out in some way, enough so that they can pass through both of the slits.

But wait! didn't we say that each photon hits a specific spot on the screen? Now it sounds like we have a contradiction: we said each photon must be spread out enough to go through both slits, but also that each photon is detected at a single location on the screen.

Quantum superpositions

To resolve this, we need to bring in the idea of a QUANTUM SUPERPOSITION. Here are the basic rules:

- We can have photons with specific positions, but these are special states, which we call POSITION EIGENSTATES.

- General states are combinations, or "quantum superpositions" of these

position eigenstates. They don't have a definite location.

- When a measurement of position is made, the photon becomes one of the eigenstates. It randomly "chooses" a position, with the relative probability for each possible position related to the amount of that eigenstate present in the quantum superposition.

How does this apply to our photons? We can say that the photons approaching the double slits are in spread-out quantum superpositions involving many different position eigenstates. It is possible for the photon to go through both slits, since it can be in a state which is the superposition of being at one slit and being at the other slit. As the photon approaches the screen, it is still in a superposition, so there is no predetermined location where it will hit. But hitting the screen is like making a measurement of the photon's position, and at that time (according to the third rule) the photon will turn into one of the position eigenstates, and therefore hit the screen at a definite location.
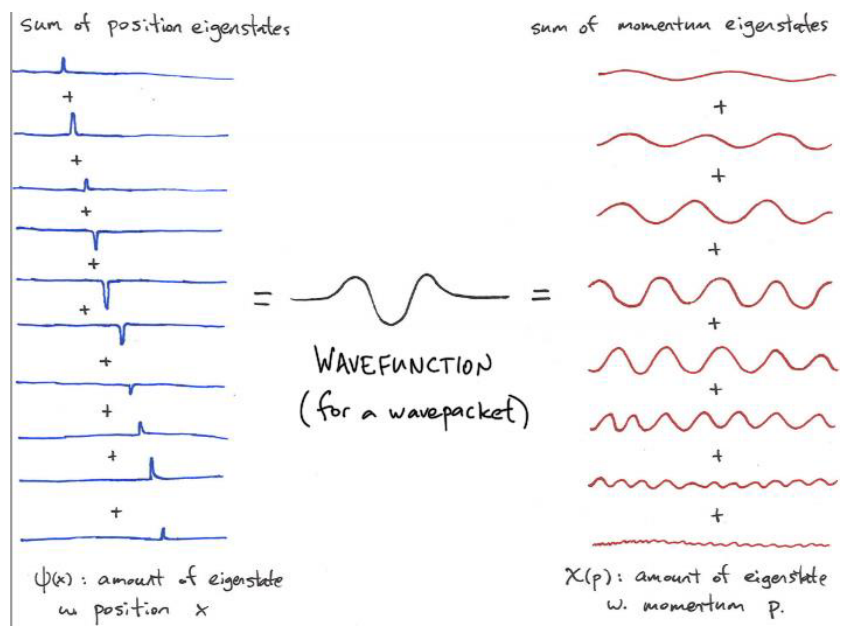
Same story for electrons and other particles

It turns out (in accord with de Broglie's prediction that particles like electrons will exhibit a wavelength) that the double slit experiment produces exactly the same results for electrons and other particles, with the wavelength for a particle of momentum p being given by

$$\lambda = \frac{h}{p} \ .$$

Explaining this requires exactly the same type of explanation in terms of quantum superposition. Thus, going forward, the discussion will apply equally well to electrons, neutrons and other particles as it does to photons.

Mathematical description of quantum superpositions

How do we describe these quantum superpositions precisely? When we make a measurement of the particle's position, the possible results are the various points in space, labeled by coordinates $\mathbf{x}$. Since we want to assign a probability (or more precisely a probability density) to each of these possible outcomes, we'll need one number for each possible location. This gives us a function $\psi(\mathbf{x})$. In the language of quantum superposition, we say that a

sum of position eigenstates

sum of momentum eigenstates

WAVEFUNCTION
(for a wavepacket)

$\psi(x)$: amount of eigenstate w. position x

$\chi(p)$: amount of eigenstate w. momentum p.

general state is a quantum superposition of position eigenstates (labeled by $\mathbf{x}$), and $\psi(\mathbf{x})$ tells us how much of this eigenstate is in the superposition. This is shown in figure ().

In practice, the function $\psi(\mathbf{x})$ tells us everything we can know about the state of the particle. This is known as the WAVEFUNCTION for the particle. If we make a measurement of position, $|\psi(\mathbf{x})|^2$ tells us the probability density for finding the particle at location $\mathbf{x}$. For example, to find the probability that the electron will be measured in a spatial region V, we integrate $|\psi(\mathbf{x})|^2$ over this region. Since the net probability of finding the electron somewhere in space must be 1, we require that any physical wavefunction must have $|\psi(\mathbf{x})|^2$ integrate to 1 over all space.

Measurement

The wavefunctions for particle states are generally spread out (non-zero for a wide range of positions). However, when we measure the location of a particle, we always find it in one place. Furthermore, an immediately repeated measurement of the particle's location will give the same result (otherwise, it would be somewhat meaningless to say that the electron is at the location where we measure it). This implies that the wavefunction generally changes

in a measurement: after the location is measured, the particle does have a definite location so the state is now (approximately) one of the position eigenstates. The wavefunction changes from being spread out (non-zero for a wide range of positions) to being localized (non-zero only for one location or a small range of locations). This is often referred to as the COLLAPSE of the wavefunction.

Why should a measurement have any effect on the state of the particle? Can't we measure the particle without disturbing it? The answer is that in order to learn something about the particle, we have to interact with it somehow (e.g. by scattering a photon off it or having it hit a detector). In doing this, we always affect the state of the particle, apparently in such a way that the wavefunction collapses to the wavefunction for an eigenstate.

Classical versus quantum

The wavefunction (or the idea of quantum superposition) represents an entirely new description of physical systems. In the classical (pre-quantum) picture, the state of a particle at a given time is completely specified by giving its position and velocity. From these, the position and velocity at any later time can be predicted using Newton's Laws. Now, instead of these six numbers (three position coordinates and three velocity components), we need to give a whole function $\psi(\mathbf{x})$ to describe the state of a particle at some time.

Just as position and velocity evolve with time in the classical picture, the wavefunction for a particle evolves with time (as we saw in the double-slit simulation). We therefore need something to replace Newton's laws; something that will allow us to predict the wavefunction at some later time given the wavefunction at some initial time. This evolution equation for wavefunctions is known as the SCHRÖDINGER EQUATION. Our next task will be to "derive" this based on some simple observations and principles.

Momentum wavefunctions and uncertainty

The state of a particle at any time is described by a wavefunction $\psi(x)$. These wavefunction must change with time, since we know that particles can move. For a traveling particle, we might not be able to say exactly where the particle is at any given time, but the region where we are most likely to find

it changes with time as it travels along. Eventually, we would like a general equation that determines exactly how wavefunctions change with time. To start, we'll focus on the simple case of the wavefunction for an electron with some specific momentum p.

Wavefunctions for momentum states

Let's first ask the question: what does the wavefunction for a traveling electron look like? Here, we can use some of our experimental observations. Electrons with momentum p produce a pattern on the screen in a double slit experiment just like the pattern from a wave with wavelength $\lambda = h/p$. For all other types of waves (water waves, sound waves, classical light waves), such a pattern is explained by saying that before the slits we just have a pure sinusoidal wave with some wavelength traveling towards the apparatus, and that this wave goes through the slits and spreads out again, with waves from the two slits recombining at the screen to produce constructive and deconstructive interference. Since we get the same result for electrons, it makes sense that the mathematical explanation for the pattern should be the same: the traveling electrons with momentum p should be described by pure sinusoidal waves with wavelength $\lambda = h/p$.

Now that we understand that electron states in general are described by wavefunctions, we can make a more precise proposal: *the wavefunction $\psi(x)$ for a traveling electron with momentum p should look like a pure sinusoidal wave with wavelength h/p.* This proposal turns out to be exactly right.

Wavepackets

You might be bothered by the fact that (as we saw above) $|\psi_p(x)|^2$ is infinitely spread out. Our only real constraint on possible wavefunctions so far is that their probability densities should integrate to one, and this function clearly violates this, since the integral will be infinity. Since the wavefunction is infinitely spread out, the interpretation is that the electron is equally likely to be found anywhere in the universe. This doesn't sound very reasonable, especially since we started out trying to describe electrons inside our double slit experiment.

The resolution of this puzzle isn't really anything complicated to do with wavefunctions or quantum mechanics. It is simply that real waves in nature are never actually described by perfect sinusoidal functions like $cos(kx)$.

Physical waves always have finite extent, like this (just the real part is shown):

Here, there is still a clear wavelength, but the wave doesn't go on forever. A wavefunction like this can clearly be normalized so that the probability density integrates to one.

For a wavepacket, the wavelength λ and the extend of the packet, which we'll call Δx are two independent quantities. There is one obvious restriction: that the extent of the wavepacket must be at least as big as a wavelength. This gives us our first hint of the HEISENBERG UNCERTAINTY PRINCIPLE: for a wave with a well-defined wavelength (or equivalently momentum), the wavefunction must be nonzero over a range of values as least as large as this wavelength. Thus, knowing something definite about the momentum of a particle restricts how certain we can be about its position.

Wavepackets as sums of pure waves.

We now come to a mathematical fact that is crucial for our understanding of wavefunctions in quantum mechanics. We have argued that physical traveling electrons should be described by wavepackets rather than pure (infinitely extended) waves. But it is a fact that any such wavepacket – in fact any wavefunction at all – can be written as a sum (i.e. a superposition) of pure waves with various wavelengths.[1]

For different wavefunctions, the amount of each pure wave will be different, so we need to introduce a function $A(p)$ that tells us how much of the pure wave with momentum p we have (i.e. with what amplitude do we add the wave with momentum p into our superposition). In terms of $A(p)$ our claim is that for any wavefunction $\psi(x)$, we can write

$$\psi(x) = \sum_p A(p)\psi_p(x) \ . \tag{1}$$

The physical interpretation of this expression is that a general state can be thought of as a quantum superposition of momentum eigenstates. This

---

[1]This is covered in greater detail in the text, chapter 40. The "Fourier: making waves" PHET is also very helpful for understanding this.

means that generally states do not have a definite value for momentum, but if we measure the momentum, the probability density for finding the value p is $|A(p)|^2$. So $A(p)$ acts just like a wavefunction for momentum. Again, to be precise, we should use an integral here instead of a sum, since there is actually a continuous range of possible momenta.

How can we find the momentum wavefunction $A(p)$ if we know the wavefunction $\psi(x)$? It turns out there is a nifty formula for doing this (thanks, mathematicians!) known as a Fourier transform, but we'll skip the details for now. For us, the most important point is that this formula exists: given any wavefunction, we can in principle figure out what superposition of momentum eigenstates that wavefunction corresponds to. Once we know this, we can figure out the probabilities for finding different momenta if we measure the momentum.

Just like for position, immediately repeated measurements of momentum should give the same result. Thus, after a measurement of the momentum, the state of the particle will be approximately a momentum eigenstate (i.e. we will have a very spread out wavepacket with a sharply defined wavelength), since only in a momentum eigenstate can we be guaranteed that the next measurement of momentum will have a definite value (equal to the previously measured value). Since measuring momentum results in a spread-out wavefunction, it tends to decrease out knowledge about the location of the particle. This is another hint of the Heisenberg Uncertainty Principle, which we will now make more precise.

The Heisenberg Uncertainty Principle

We can now be more precise about the connection between position uncertainty and momentum uncertainty. A property of Fourier transformations is that in order to describe a position wavefunction that is very localized (e.g. a very short wavepacket), we must add up pure waves with a very broad range of momenta. Conversely, if we combine pure waves with a very narrow range of momenta, the resulting wavefunction will necessarily be very spread out. Quantitatively, we can prove that

$$\Delta x \Delta p \geq \frac{h}{4\pi} \ , \tag{2}$$

where Δx represents the "uncertainty in position", i.e. the range of values of x over which there is a significant change to find the particle, and Δp

represents the "uncertainty in momentum, i.e. the range of values of p over which we are likely to find the momentum. A simple way to give the precise definitions of Δx and Δp is that if we think of $|\psi(x)|^2$ or $|A(p)|^2$ as a histogram for the possible values of position/momenta, then Δx and Δp are the standard deviation.

Equation (3) is the precise form of the Heisenberg Uncertainty Principle. This result has a dramatic physical interpretation: it says that *it is impossible to know both the position and the momentum of a particle at the same time.* The more accurately we know the position, the larger the uncertainty in momentum much be. Similarly, the more accurately we know the momentum, the less certain we can be about position. The uncertainty relation (3) has an inequality sign rather than an equality since for some wavefunctions, both position and momentum are very uncertain (can you think of an example?).

In three dimensions, there are similar uncertainty relations that apply to the y and z directions, so the complete set of relations is:

$$\Delta x \Delta p_x \geq \frac{h}{4\pi} \qquad \Delta y \Delta p_y \geq \frac{h}{4\pi} \qquad \Delta z \Delta p_z \geq \frac{h}{4\pi} \; . \tag{3}$$

Here, the three position uncertainties are all independent quantities. For example, if we have a wavefunction that is very spread out in x but localized in y, that would correspond to a case with large Δx and small Δy.

Time dependence and the Schrödinger equation

We now understand that the wavefunctions for traveling particles with momentum p look like wavepackets with wavelength $\lambda = h/p$, but we haven't really said anything about the time-dependence. Our goal for this part of the notes is to learn how to predict what a wavefunction will look like in the future if we know the wavefunction now.

Time dependence of wavepackets

Suppose we have a wavepacket describing a particle with momentum p. Let's assume that it is a very broad wavepacket with a small uncertainty in momentum. How do we expect this to change with time?

One thing is fairly certain: since we know the particle is moving with momentum p, it should have a velocity $v = p/m$ (assuming $p \ll m$), so the wavepacket should move along at this velocity. If the wavepacket is centered at x_0 at time $t = 0$, then at a later time $t = T$, it should be centered somewhere near $x = x_0 + vT$.

In terms of wavelength, we can say that wavepackets with wavelength λ should travel at speed

$$v_{packet} = \frac{p}{m} = \frac{h}{m\lambda} \ . \tag{4}$$

It is interesting that the speed should depend on the wavelength. This does not happen for light in a vacuum or for waves on a string (assuming they obey the simplest wave equation) where waves with all possible wavelengths travel at one velocity. However, most realistic waves (e.g. water waves or light waves in air/water/glass) do have wavelength-dependent velocities.[2]

Time dependence of pure waves

For wavepackets of finite length, there isn't a single definite momentum, so the wavepacket spreads out as it moves along (just like a group of people that are walking along with a range of different speeds). However, for pure waves that describe momentum eigenstates, the time dependence is really just like a traveling wave: if the initial wavefunction is

$$\psi(x) = cos(\frac{2\pi}{\lambda}x)$$

then the wavefunction at later time is

$$\psi(x, t) = cos(\frac{2\pi}{\lambda}(x - vt))$$

where the wavelength here is equal to h/p, and the velocity v here is proportional to the momentum (and inversely proportional to the wavelength). If we calculate the frequency f of this wave, we find that

$$hf = \frac{p^2}{2m} = \frac{1}{2}mv^2 \ . \tag{5}$$

_____

[2]For example, the index of refraction of a material is the speed of light in the vacuum divided by the speed of light in the material. A prism makes a rainbow because the index of refraction is different for the different colors; in other words, the light velocity in the glass is different for different wavelengths.

There is a really simple interpretation of this result. Since $p^2/(2m)$ is just the energy of the electron (so far, we're talking about "free" electrons traveling in regions without potential energy) what we have found is that

$$hf = E .$$

Thus, pure-wave wavefunctions with frequency f describe particles with energy $E = hf$, exactly the same relation between energy and frequency as we have for photons!

   To summarize, our final result is the the time-dependent wavefunctions for particles with momentum p are traveling sinusoidal waves with wavelength $\lambda = h/p$ and frequency given by $hf = p^2/(2m)$ (or equivalently with velocity proportional to p).

Time dependence of general wavefunctions

We have now solved the problem of time-dependence for the special case of momentum eigenstates. While this might seem like only a small step towards finding the time dependence of general wavefunctions, it actually gives us everything we need. The crucial point is that any state can be written as a superposition of momentum eigenstates (or equivalently, any wavefunction can be written as a superposition of pure waves). Suppose that our wavefunction at time $t = 0$ is $\psi(x)$. Then (from the previous notes) we know that we can write $\psi(x)$ as a superposition of pure waves:

$$\psi(x, t = 0) = \sum A(p)\psi_p(x) \tag{6}$$

where $A(p)$ can be determined using the Fourier transform formula. But we now know from equation (**??**) how all the individual pure waves evolve with time. Thus, to find out the wavefunction at a time t, we just take the same combination of pure waves, but evaluated at the later time:

$$\psi(x, t = 0) = \sum A(p)\psi_p(x - vt) \tag{7}$$

where the velocities are proportional to the momenta (i.e. the waves with a smaller wavelength travel faster). This gives a precise method to determine $\psi(x, t)$ for any wavefunction given $\psi(x, 0)$. While it may seem complicated to carry out in practice, these this is a relatively easy calculation for a computer to do (or we can do it by hand for many possible forms of the initial wavefunction).

Time dependence from the Schrödinger equation

There is a very convenient way to summarize everything we've learned so far about how wavefunctions change with time, and that is by using a differential equation known as the SCHRÖDINGER EQUATION. Let's start by writing it down; it says that for a wavefunction $\psi(x, t)$[3]

$$\frac{d\psi}{dt} = \frac{ih}{4\pi m}\frac{d^2\psi}{dx^2} \tag{8}$$

where the time derivative is taken with x fixed and the x derivatives are taken with t fixed. You may not have much experience with differential equations, let alone this more complicated variety with two different types of derivatives (known as a partial differential equation). But it's easy enough to figure out what this is telling us. First remember that the time derivative on the left hand side is defined to be

$$\frac{d\psi}{dt} \equiv \frac{\psi(x, t + \delta t) - \psi(x, t)}{\delta t}$$

in the limit where δt goes to zero. Rewriting our Schrödinger equation with this definition, we get:

$$\psi(x, t + \delta t) = \psi(x, t) + \delta t\frac{ih}{4\pi m}\frac{d^2\psi}{dx^2} \ . \tag{9}$$

In this way of writing things, we see that the left side is just the wavefunction evaluated at a time which is slightly later than t. The right hand side is just the wavefunction at time t plus a constant times some x derivatives of the wavefunction at time t. So the Schrödinger equation tells us what the wavefunction will be at a slightly later time based on what the wavefunction is now. So apparently it gives us a second way of finding the wavefunction $\psi(x, t)$ starting from the initial wavefunction $\psi(x, t = 0)$.[4] It is not too hard to show that the Schrödinger equation predicts exactly the same time-dependence that we found in our general formula (**??**).

---

[3]You may be wondering what "i" is. This is actually the complex number whose square is -1. If you are wondering what the heck this is doing in the Schrödinger equation, you can read the appendix. For now, it's safe to ignore it.

[4]The equation (9) only works if δt is infinitesimal, so we have to break up the time t into many parts and then apply (9) a bunch of times. This is something that a computer can do easily. There are also analytical (i.e. calculating by hand) techniques for solving equations like this.

Newton's First Law v2.0

We've learned so far that in quantum mechanics, the description of the state of a particle is given by a wavefunction $\psi(x)$ instead of a position and velocity (x, v). The Schrödinger equation (11) tells us how the wavefunction changes with time, but only in the special case where there are no forces on the particle. Thus, it is the quantum version of Newton's First Law ($m\frac{d^2x}{dt^2} = 0$ in the absence of forces). The solutions

$$\psi(x, t) = \sum A(p)\psi_p(x - vt) \ . \tag{10}$$

are the quantum version of constant velocity particles.

$$x = x_0 + vt \ .$$

An interesting difference between the classical picture and the quantum picture is that in classical physics, we need to know both the position and the time derivative of position (i.e. the velocity) at some time to determine what the position will be at later times. For the Schrödinger equation we only need to know the wavefunction, but not its time derivative to determine what the wavefunction will be at some later time. The reason for the difference is that the wavefunction contains all information about position *and* velocity (since we can figure out the momentum wavefunction directly from the position wavefunction).

Now that we know the quantum mechanical version of Newton's First Law, our next task is to understand what replaces Newton's Second Law; how do forces enter into the Schrödinger equation?

The Schrödinger equation in a potential and energy eigenstates.

We have now seen that the wavefunction for a free electron changes with time according to the Schrödinger Equation

$$i\hbar\frac{d\psi}{dt} = -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} \tag{11}$$

where we have written the previous equation in a somewhat more conventional form by defining $\hbar$ (pronounced "h-bar") to be $h/(2\pi)$.

The solutions to this equation include time-dependent pure waves that describe momentum eigenstates, and various other time-dependent wavefunctions that describe superpositions of momentum eigenstates. Because these general wavefunctions are superpositions of electrons with a range of possible velocities, they will all tend to spread out with time. So the position uncertainty always goes to infinity as t goes to infinity. Examples include wavepackets for traveling electrons which become more and more spread out, or highly localized wavefunctions (e.g. that we obtain after a measurement of position) that quickly spread out in all directions.[5]

In nature, it is clearly not true that all electron wavefunctions spread out more and more as time passes. If we have an electron in an atom, the electron stays in the atom unless disturbed by some outside influence. Similarly, for an electron in a metal, the wavefunction might spread out through the metal, but the electron will not escape without some outside influence. The physical difference in these situations from the case of free electrons is that electrons in an atom or in a metal have forces acting on them. Equivalently, the electrons have a potential energy $U(x)$ that is different for different locations (this leads to forces $F(x) = -U'(x)$). To learn how wavefunctions evolve in the presence of these forces or potentials, we need to understand how to modify the Schrödinger equation when they are present. As a specific example, we might keep in mind the case of a hydrogen atom, where the potential energy that an electron feels is the Coulomb potential from the proton:

$$U(x) = -\frac{ke^2}{|x|} \ . \tag{12}$$

Momentum eigenstates in a constant potential.

To understand how the potential $U(x)$ should affect the time evolution of wavefunctions, let's start by thinking about the simple situation where we have an electron with a definite momentum p traveling a region where the potential energy is constant but non-zero. In this case, the total energy of

---

[5]A good way to develop intuition for these time-dependent wavefunctions is by playing with the PHET "quantum tunneling" simulation, choosing a constant potential.

the electron is now
$$E = \frac{p^2}{2m} + U \ .$$

Since the electron still has momentum p, it should still be described as a pure wave with wavelength h/p. But according to the relation between energy and frequency, we expect that the frequency should now be

$$f = \frac{E}{h} = \frac{1}{h}(\frac{p^2}{2m} + U) \ .$$

Thus, the momentum eigenstate wavefunctions in a region of potential energy U should be pure waves with wavelength h/p and frequency $hf = \frac{p^2}{2m} + U$ instead of frequency $hf = \frac{p^2}{2m}$. Again, we can find the time-dependence of general wavefunctions in a region of non-zero potential energy by decomposing them as sums of these pure waves.

The Schrödinger equation with a potential

This change in the frequency for the pure waves in the presence of a potential is equivalent to a change in the Schrödinger equation where we add one extra piece:

$$i\hbar\frac{d\psi}{dt} = -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U\psi \tag{13}$$

We can at least roughly understand this as follows: the left side (a time derivative of the wavefunction) is related to the frequency. The higher the frequency, the faster the wavefunction changes with time, and so the bigger the time derivative. The equation tells us that this frequency will be bigger if either the second x derivative is bigger (this will happen when the function is more curvy, which will be true if we have a shorter wavelength/larger momentum) OR if the potential energy is bigger. So we can at least see that it's consistent with the basic relation $hf = \frac{p^2}{2m} + U$.

By this simple modification, we now have an equation that correctly gives the time dependence of momentum eigenstates in a potential, even if U is a function of space, as for example in an atom. This is the general Schrödinger equation for quantum mechanics in one dimension. It is the exact quantum mechanical analog of Newton's second law $\vec{F} = m\vec{a}$.

For three dimensions, the wavefunction can depend on x, y, and z, so the equation includes y and z derivatives also:

$$i\hbar\frac{d\psi}{dt} = -\frac{\hbar^2}{2m}\left(\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2}\right) + U\psi \qquad (14)$$

Wavefunctions in a potential

Now that we have the general Schrödinger equation, we can begin to investigate the physics of electrons in atoms, molecules, wires, and other situations, as long as we can figure out the potential energy function (or a good approximation to it). The potential energy function is an input; it determines what physical system we are trying to describe. The other input is the initial wavefunction $\psi(x, t = 0)$. From these, we can determine the wavefunction $\psi(x,t)$ at any later time t.

For now, we can let a computer do the work and investigate the behavior of wavefunctions for particles in various kinds of potentials (e.g. using the PHET "Quantum Tunneling and Wave Packets" simulation where we can select various potentials and initial wavefunctions). There are many fascinating phenomena that result from studying this equation, but for now, let's just mention one. If we have a potential with a dip in it (i.e. a localized region with lower potential energy than everywhere else, as for the Coulomb potential near an atomic nucleus), a wavefunction that is initially localized near the dip can be "trapped". That is, the wavefunction in the region near the dip can stay nonzero forever. In detail, what we find is that these special trapped wavefunctions will oscillate up and down with particular frequencies (or be combinations of such oscillating functions). Since frequency is related to energy, this means that trapped wavefunctions can only exist at specific energies. The energies are "quantized"! It is this property that puts the "quantum" in "quantum mechanics".

One last thing: for the Coulomb potential, even through the potential goes to minus infinity at the location of the proton, there is a lowest possible energy for a quantum electron in this potential (e.g. in a Hydrogen atom). The reason is that if we try to localize an electron to the place where the potential is very negative, the momentum uncertainty goes up a lot, so the typical value of the kinetic energy for such a state is very large. If we want the lowest energy overall, the best we can do is have the electron in a wavefunction that is spread out a little bit: this is the wavefunction for the

s-wave orbital. Since there is no possible lower energy state, the electron in the s-wave orbital cannot radiate away any energy and fall into the nucleus. So the Heisenberg Uncertainty principle allows atoms to be stable and the universe as we know it to exist!

Aside: complex numbers

In the discussion on quantum superposition, we said that if $|a\rangle$ and $|b\rangle$ are are two states of a physical system (perhaps with definite values for some physical property such as position) then we can also have a state

$$\alpha|a\rangle + \beta|b\rangle \ .$$

Up until now, we have been assuming that α and β are real numbers. However, in order to describe the most general states, we need to allow α and β to be complex numbers.

The bottom line for the present discussion is that wavefunctions $\psi(x)$ that are real functions are just special cases that don't describe the most general electron states. Generally, the ψ at a location x will be a complex number. We can write this as

$$\psi(x) = \psi_R(x) + i\psi_I(x)$$

where the real functions $\psi_R(x)$ and $\psi_I(x)$ are known as the "real part" and the "imaginary part" of the wavefunction respectively.

To compute the probability density from a complex wavefunction, the rule is that we take the *magnitude squared* of the complex number. Thus, we have:

$$P(x) = |\psi(x)|^2 = (\psi_R(x))^2 + (\psi_I(x))^2 \ .$$

Back to wavefunctions for traveling electrons.

We said above that the wavefunction for a traveling electron with momentum p should look like a wave with wavelength h/p. The simplest guess would be

$$\psi(x) = cos(\frac{2\pi p}{h}x).$$

We said this should move with speed $v = p/m$. But you may have wondered: how do we know if this wave is moving to the left or to the right?

The answer is that the actual wavefunctions for traveling electrons are not real waves, but complex waves. You learned in math class that the cosine function is the real part of a function e^{itheta}. Generally:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta) \ .$$

So $e^{i\theta}$ is like a fancy complex version of $\cos(\theta)$

To describe a traveling electron with momentum p, it turns out that the actual wavefunction at some time is

$$\psi_p(x) = e^{i\frac{2\pi p}{h}x} = \cos(\frac{2\pi p}{h}x) + i\sin(\frac{2\pi p}{h}x) \tag{15}$$

instead of just $cos(\frac{2\pi p}{h}x)$.

Using the complex version resolves the puzzle about which direction the wave will move. If we change p to $-p$ in (15), we get a different wavefunction (the imaginary part switches sign). So we can say that for positive p the wave will move to the right, while for negative p, the wave will move to the left. This wouldn't have worked if the wavefunction was just $cos(\frac{2\pi p}{h}x)$, since in that case switching from p to $-p$ does nothing.

The fact that wavefunctions can be complex explains the mysterious i that appeared in the Schrodinger equation. You can check that

$$\psi_p(x,t) = e^{i\frac{2\pi p}{h}(x-(p/(2m))t)} \tag{16}$$

is a solution to the Schrodinger equation, while the same thing with just the cosine part doesn't quite work.