
Waves and Particles

For many years, scientists debated the “true” nature of light: is it a wave, or a particle? Different experiments seem to give different answers. Physicists now recognize that light can act as a wave or a particle, depending on the situation. This is known as wave-particle duality, and it’s a reflection of trying to use classical ideas of waves and particles to describe something that’s really different from either (light). As strange as it may seem that light can change its characteristics, it was shown in the 20th century that electrons have the same kind of duality. In the right experiment, we see electrons (and other kinds of matter that we have so far taken for granted as being particles) behaving as waves, exhibiting diffraction and interference patterns. Possibly the best way to think about this is a mixture; light (and small bits of matter) move as waves (meaning they exhibit phenomena like interference and diffraction), but interact as particles (energy is absorbed in discrete lumps, particles collide like billiard balls).

Blackbody Radiation

Last semester we saw that the thermal radiation emitted by a hot object is largely independent of the composition of that object. The only important thing is the temperature. Because of this, we can draw blackbody curves which are the particular shapes of a graph of intensity vs. wavelength emitted for a particular temperature. As you can see below, the curves shown have noticeably different shapes at different temperatures.
If we wanted to do a little bit more math, we could determine the temperature of an object just by looking at its spectrum. If we know the intensity of 700 nm radiation compared to the intensity of 400 nm radiation, that ratio should give us a unique blackbody curve corresponding to a unique temperature (like the new fast-measuring ear thermometers).

In the early part of the 20th century, Max Planck suggested that the atoms which were oscillating to produce this radiation could only have certain particular values of energy (kind of like the difference between a 3-way light bulb and a bulb on a dimmer switch). Planck believed that the energy of the oscillator depended on the frequency of oscillation as

$$E = n hf$$

where \(n\) is a positive integer or zero and \(f\) is the frequency of oscillation. Possibly the most interesting part of the equation is the \(h\), now known as Planck’s constant. Its value is \(6.626 \times 10^{-34}\) J s, and its incredibly small size gives us a hint about why its existence took so long to be realized. This is similar to the way that we know (now) that a glass of water is really composed of many individual pieces of water (molecules). We don’t see water as being “lumpy” because the pieces are so ridiculously small on the scale of our senses. For most of our purposes, we could consider water to be a continuous substance; we could take a liter of water, divide it into two equal parts, then divide the half-liter parts in half again, etc. It would take some very specialized and expensive equipment for us to get to the point where we separate two water molecules from one another and can then go no further in dividing the water.

Just as the size of the atom tells us about the lumpiness of matter, Planck’s constant tells us about the lumpiness of energy.

**Photons**

Shortly after Planck made his findings known, Einstein realized that if the radiating oscillators can only have certain discrete values of energy, they must be radiating discrete values of energy (if you can only have $1 bills in your pockets, you can’t give someone a quarter!). Einstein called these discrete pieces of light energy photons.

One of the early experiments where light seemed to be behaving as a particle (or photon) is known as the photoelectric effect. The basic idea here is that when light shines on certain metals, the photons hit electrons in the metal hard enough to eject them from the rest of the metal entirely. Einstein said that this happened because the photons were carrying an energy equal to

$$E = hf$$

which should look familiar. We can calculate the energy of a red-light photon as an exercise. For red light, \(f\) is about \(4 \times 10^{14}\) Hz, so \(hf\) gives us \(2.65 \times 10^{-19}\) J. Obviously, the Joule is an inconveniently large unit of energy for this kind of situation, so we sometimes use the electron Volt (eV). We’ve seen this before when talking about the amount an electron’s energy changes
when moving through a potential difference of one volt, and it’s $1.6 \times 10^{-19}$ J. On this scale, our red photon has an energy of about 1.7 eV. Blue light, at about $7.9 \times 10^{14}$ Hz, would have an energy of around 3.3 eV.

The big idea here is that there is a connection between the frequency of light and its energy. Before this, people would have assumed that you could use something like a dimmer switch to turn the intensity of a light down as low as you’d care to, without limit. Running a blue light bulb off of this dimmer switch would allow you to produce, for example, a constant dim stream of blue light at a rate of 1 eV per second, for example. The connection between energy and frequency says you can’t do this – if you want blue light ($7.9 \times 10^{14}$ Hz, anyway), the smallest piece you can have is 3.27 eV. Anything smaller won’t be blue light. You won’t see a constant stream of light in this case, but instead you’ll see a discrete lump of blue light every few seconds, with nothing in between. This is very much like the situation with atoms: the smallest piece of Uranium you can possibly have has a mass of about $4 \times 10^{-25}$ kg. No one can give you $2 \times 10^{-25}$ kg of Uranium, because it just doesn’t exist. You can have $2 \times 10^{-25}$ kg of Hydrogen, because the basic pieces (atoms) are smaller. Similarly, you can have 1 eV of infrared radiation or microwave radiation, etc., but you can’t have a 1 eV visible-light photon, X-ray photon, etc.

Getting back to the photoelectric effect, it was observed that certain colors of light (near the blue end of the spectrum) could trigger this electron ejection, but other colors (near the red end) couldn’t. The intensity of the light (number of photons per second) wouldn’t allow red photons to eject electrons, no matter how bright the light bulb. Changing the intensity of a blue light bulb would change the current (number of electrons per second), however. A wave theory of light couldn’t explain this. A slowly vibrating wave should (over a longer time) be able to deliver the energy needed to free an electron just as well as a more quickly vibrating wave could do (in a shorter time). Also, no time lag was ever observed between turning on the light and catching the first ejected electrons. If waves were shaking them loose, it should take some time to get the first ones moving.

The photon theory is able to explain all of this. First, while free electrons in a metal may not be bound to particular atoms, they are certainly bound to the overall body of the metal. If they were totally free, you could pick up a piece of steel and shake electrons out of it! It takes a nonzero amount of energy to remove an electron from a metal. The exact amount depends on the metal, but it is called the work function. For potassium, it’s about 2.3 eV, while for platinum, it’s closer to 6.4 eV. What will happen if a red light photon hits a potassium atom? Since 1.7 eV < 2.3 eV, it won’t be able to eject it. The atom involved will either ignore the photon altogether or absorb and very quickly (nanoseconds) re-radiate it. Increasing the intensity of the red light just means more atoms are doing the same thing. What if two photons hit the atom at about the same time? Then you could have an ejection, but the problem is that no light source from the early 20th century was capable of this kind of intensity. The re-radiation happens so quickly that only powerful lasers can drop enough photons on the surface quickly enough. For our purposes, red light won’t give us any photoelectrons.

What about violet light? Since 3.3 eV > 2.3 eV, electrons will be ejected. In fact, there will be energy left over after the ejection. That extra energy goes into the kinetic energy of the electron, so it leaves with a greater speed. With greenish photons, having energies just slightly above 2.3 eV, we would get photoelectrons, but they’d be moving very slowly since there is very little energy remaining after freeing them. The formula is
\[ hf = KE_{\text{max}} + W_0 \]

where \( W_0 \) is the work function for the particular metal, \( hf \) is the energy of the incoming photon, and \( KE_{\text{max}} \) is the maximum kinetic energy an electron can leave with. More intense violet light gives more photoelectrons ejected per second which means a higher current. Digital cameras use this effect to give us pictures without film. In another popular application, we can collect these energetic electrons and use them to power an electric circuit (a battery, after all, is just a source of moving electrons) and we’ll have a solar cell.

**Photon Momentum**

We see another example of light acting like a particle when we look at X-ray photons striking electrons. The photon delivers some fraction of its energy and momentum to the electron, meaning the photon’s frequency drops to a lower value. This is known as the **Compton effect**. If we assume the electron is initially at rest (so it has zero KE) and is not bound to an atom (so we don’t have to do work to free the electron - this is usually a pretty good approximation since X-rays have energies in the thousands of electron volts and it typically takes only a few electron volts to free an outer electron from an atom anyway), conservation of energy requires

\[ hf_{\text{incident photon}} = hf'_{\text{scattered photon}} + KE_{\text{electron}} \]

The scattered photon has a lower frequency than the incident photon (to conserve energy), which means its wavelength is longer. We also will have conservation of momentum in this collision (no external forces), but we need to know the formula for the momentum of a photon. The photon’s momentum is \( p = E/c \) which we can also write \( p = h/\lambda \). For the collision, Compton found that the connection between the incident photon wavelength \( \lambda \), the scattered photon wavelength \( \lambda' \), and the angle between the directions of travel of the two photons \( \theta \) is

\[ \lambda' - \lambda = \frac{h}{mc} \left(1 - \cos \theta \right) \]

The mass in this formula is the electron’s mass, and \( h/mc \) is so commonly used that we call it the **Compton wavelength of the electron** and it is \( 2.43 \times 10^{-12} \) m. As the scattering angle gets closer to zero (meaning a smaller deflection in the photon’s direction), the shift in wavelength (proportional to the energy loss) gets closer to zero. Interestingly, there’s also a maximum shift in the wavelength which occurs when the angle is 180°. In this case, the shift in wavelength is twice the electron’s Compton wavelength, which would be \( 4.86 \times 10^{-12} \) m and corresponds to the photon being **backscattered** and retracing its path.
de Broglie Wavelengths

Louis de Broglie proposed the semi-crazy idea that particles have wavelengths(!) just like photons do, and that their wavelengths are related to their momenta in the same way:

\[ \lambda = \frac{h}{p} \]

We call this wavelength the de Broglie wavelength of the particle. This is a good time to emphasize the fact that this isn’t just a mathematical statement without physical significance. We can fire electrons through tiny slits and get the same interference effects that we see with light – it’s a real effect. If an electron has a kinetic energy of 500 eV, what is its wavelength? The electron’s rest energy is 511,000 eV, so this kinetic energy is so low in comparison to that that we can use Newtonian physics to relate KE and momentum (you can do it relativistically, but you’ll only see a tiny difference). Solving 0.5 m v^2 = 500 eV for the electron gives v = 1.33 x 10^7 m/s, meaning the momentum of the electron is 1.21 x 10^{-23} kg m/s. The wavelength is then

\[ \lambda = \frac{6.626 \times 10^{-34} \text{ J s}}{1.21 \times 10^{-23} \text{ kg m/s}} = 5.5 \times 10^{-11} \text{ m} \]

This is in the ballpark of the spacing between atoms in a crystal, so we can use crystals to see diffraction and interference in electron beams. More usefully, we can alter the energy of the electrons to give them different wavelengths. Watching the diffraction patterns will then tell us about the spacing and structure of the crystal.

Schrödinger’s Equation

If matter acts like a wave, its motion should be described by a wave equation. This wave will represent the probability amplitude for a particle to be found at a particular point in space at a particular time. The difference between probability and probability amplitude is that probability is always real and non-negative. It can be zero, but it can’t be less than zero. We could not get double-slit interference from just probabilities; with one slit open, there is a certain probability that a particle will be detected at position x on the screen. Opening another slit gives another way to get a particle to point x, so the probabilities would add and we would never see destructive interference.

Probability amplitudes become probabilities when we take their absolute square. Since these can be complex numbers, this means we multiply the amplitude by its complex conjugate (which is found by changing the sign on the imaginary part of the number; the complex conjugate of \( 4 + 7i \) is just \( 4 - 7i \)). We will always get a real and non-negative number from this process, so this will be our probability. Strictly speaking, it’s a probability density. It only makes sense when we integrate it over some volume and find the probability that a particle is inside that volume. The probability of finding a particle at the exact position \((x,y,z)\) at time \(t\) will be zero if we measure
the position to a high enough accuracy. In other words, if we can tell a difference between the
cosition \( x = 5 \) and the position \( x = 5.00\ldots001 \), we have a vanishingly small chance of ever seeing
a particle exactly at \( x = 5 \).

The variation in this probability amplitude in space and time is called the **wave function**. We
will only encounter wave functions which are **separable** in time and space; the general wave
function \( \Psi(x,y,z,t) \) will be expressible as \( \psi(x,y,z) e^{i\omega t} \) where \( \omega \) is the angular frequency of the
matter wave associated with the particle. Notice that this wave function is complex – when we
add two different wave functions, we can get destructive interference (the total after combining
and finding the absolute square is smaller than the absolute squares of the individual waves).

The equation that explains how this wave function evolves in time is known as **Schrödinger’s
equation** and is written in one dimension as

\[
\frac{d^2\psi}{dx^2} + \frac{8\pi^2 m}{h^2} [E - U(x)] \psi = 0
\]

where \( E \) is the particle’s total (nonrelativistic) energy and \( U(x) \) is the potential energy in the
neighborhood of the particle. It’s not obvious from the way it’s written, but this equation is
basically just a restatement of kinetic energy + potential energy = total energy. Remember that
the kinetic energy can be written in classical physics as \( p^2/2m \) and that de Broglie showed that \( p = h/\lambda \) for matter. From last semester’s study of waves, you may remember that, just as period and
frequency are inversely related, so are wavelength and **wave number**. Wave number is just the
number of waves per unit length (except for a factor of \( 2\pi \), so \( k = 2\pi/\lambda \)), while wavelength is
the number of unit lengths per wave. We could write the above equation (when \( U = 0 \) and we
have a free particle, so \( E = \frac{1}{2}m v^2 \)) as

\[
\frac{d^2\psi}{dx^2} + k^2 \psi = 0
\]

This is clearly solved by both sine and cosine, as well as both \( k \) and \(-k\). A convenient way to
write this solution involves the base of the natural logarithm, \( e \). It relies on **Euler’s formula**,
which is just

\[
e^{i\theta} = \cos \theta + i \sin \theta
\]

This can be proven using calculus, among other techniques. We can solve the differential
equation above in the most general way using

\[
\psi(x) = A e^{ikx} + B e^{-ikx}
\]
$A$ and $B$ are constants of integration, and are therefore determined by the initial conditions of a particular problem. If we further specialize the solution and restrict it to the case of a particle traveling towards positive values of $x$, the value of $B$ must be zero. What is the probability density associated with this wave function? It should just be $\psi$ times its complex conjugate (written as $\psi^*$), which would be $A^*e^{-i\lambda x}$. The two exponential functions cancel each other, and we’re left with a probability density equal to $|A|^2$. Notice that $A$ is a constant, so this probability density does not change with position. The probability of finding the particle is uniform everywhere in space; it is exactly as likely to be in one region of space as in every other similarly-sized region of space. We’ll see in the next section that the reason is tied to the exact value of $k$.

**Heisenberg’s Uncertainty Principle**

At the heart of the weird behavior we’ve seen so far while looking at interference and diffraction lies the Heisenberg Uncertainty Principle. We can boil this down to a relatively simple statement: We can’t measure both the position of a particle and its momentum to arbitrary accuracy at the same time. There is nothing like this in Newtonian physics – all of the uncertainty of measurement there comes about because of imperfections in our equipment. The uncertainty principle goes beyond saying that we don’t have perfect equipment, it says instead that we can’t have perfect equipment. To examine this more closely, look at what happens if we want to measure the position of an electron. How do we do that? At the most fundamental level, we look at it, which means we have to bounce photons off of it. In doing that, though, we transfer some momentum to the electron. If we want to measure the electron’s position to within one nanometer, for example, we need to use light with a wavelength of less than one nanometer. What momentum will the light have? We can use our previous results to see that

$$\text{if } \lambda < 10^{-9} \text{ m, } p = \frac{h}{\lambda} \text{ so } p > \frac{h}{10^{-9} \text{ m}}$$

Getting a better measurement of position means using light with a smaller wavelength, but that means each photon has a high momentum. Some (unknown) fraction of that momentum will be transferred to the electron, so measuring positional information causes us to alter, and therefore lose, momentum information. Why have we never noticed this? Because $h$ is so incredibly small. For macroscopic things (and even for many microscopic things), the momentum transferred by a visible-light photon is a ridiculously small fraction of the object’s total momentum. Also, we don’t often try to measure things to within a wavelength of visible light. The limits imposed by the uncertainty principle are not likely to affect the design of rulers anytime soon! We can write the formal expression for the uncertainties in momentum and position (written with deltas) as

$$\Delta p_x \geq \frac{h}{4\pi}$$
This tells us that the uncertainty is at least Planck’s constant divided by $4\pi$. No measurement can do better than this. We have a tradeoff between position measurements and momentum measurements – we can in principle find the position of a particle down to any accuracy, but the photon we hit it with will give it a correspondingly large increase in momentum.

There is also an uncertainty relationship between energy and time. It basically says that we are allowed to violate conservation of energy (!) very briefly and “borrow” some energy from the vacuum, as long as we return it very quickly. The product of the energy uncertainty and the time uncertainty satisfies the same inequality as above:

$$\Delta E \Delta t \geq \frac{h}{4\pi}$$

What this says is that the vacuum is to energy like a bank is to money. If we want to grab a large amount of money from the bank, we have to put it back very quickly if we’re going to do it before it is discovered. On the other hand, if we want to get a penny out of the bank, we can probably keep it for quite some time before it is noticed.

It’s strange and probably a little disturbing to see that one of the things we’ve based so much physics on (conservation of energy) seems to be disappearing. Before you get too worried, do a sample calculation: if you only want to borrow some energy for a picosecond ($10^{-12}$ seconds), how much can you have? Plugging $10^{-12}$ seconds in for $\Delta t$ in the formula above gives $5.3 \times 10^{-23}$ J, or about $3.3 \times 10^{-4}$ eV. Not exactly a huge amount of energy!

Knowing about the uncertainty principle lets us explain things like diffraction. If we force a photon to go through a tiny slit of width $d$, we obviously know its position to an uncertainty of $\Delta x = d$. That means we are unsure about its momentum (in the same direction that the slit runs) to about $\Delta p = h/4\pi d$.

In the picture above, the photon has some momentum $p_x$ as it passes through the slit. The angle of its final position on the screen is quickly seen to be the arctangent of $p_y/p_x$, so we can write
\[ \theta = \tan^{-1}\left( \frac{p_y}{p_x} \right) \]

We can know \( p_x \) as well as we like without affecting the y position or the y momentum – it’s only when we want position and momentum **along the same axis** that we run into the uncertainty principle. From earlier work, we know \( p_x = h / \lambda \). Also, we should replace \( p_y \) by \( \Delta p_y \), since we don’t really know what value we will get for it. Substituting in the value for \( \Delta p_y \) that results from sending the photon through a region where \( \Delta x = d \) gives

\[ \theta = \tan^{-1}\left( \frac{\Delta p_y}{p_x} \right) = \tan^{-1}\left( \frac{h}{4\pi d} \cdot \frac{1}{p_x} \right) = \tan^{-1}\left( \frac{h \lambda}{4\pi h d} \right) \approx \frac{\lambda}{d} \]

In other words, the amount of diffraction depends on the relationship between wavelength and slit width – just what we had before (we’re making a small angle approximation to drop the arctangent).

A more common use for the energy-time uncertainty relation tells us that if we only have a certain amount of time to measure the energy of a situation, our answer can’t be better than what is specified by the formula above. If an atom is only in an excited state for a brief period of time, we can’t be exactly sure about the energy of a photon it radiates to get back down to the ground state. More about this later.

**Barrier Tunneling**

One of the consequences of the uncertainty principle is that certain features of classical mechanics are radically altered. For example, if we imagine a ball bouncing inside a drinking glass, it is very straightforward to find out whether it will escape or not. If the kinetic + potential energy of the ball is less than the potential energy the ball would have at the rim of the glass, it absolutely won’t get out. The uncertainty principle allows the ball to “borrow” a small amount of energy from the empty space around it and return that energy once the ball has gone over the side. Since the time that energy can be borrowed is inversely proportional to the amount borrowed, and Planck’s constant is so small, we won’t see this happen with a ball & drinking glass.

Microscopic systems can show this effect, however. An electron with energy \( E \) approaching an energy barrier of height \( U \) (which is higher than \( E \)) has a certain probability of penetrating that barrier. The **transmission coefficient** \( T \) will be approximately equal to

\[ T \approx e^{-2bL} \quad b = \sqrt{8\pi^2 \frac{m(U - E)}{\hbar}} \]
where \( L \) is the thickness of the barrier. As you can see, the probability of penetrating a barrier drops very quickly as the barrier’s thickness increases. Also, if the barrier is very large compared to the particle’s energy, the probability of penetration drops rapidly. A common example of this phenomenon is found in your smoke detector. It contains Americium-241 which ejects an \( \alpha \) particle (2 protons + 2 neutrons, or a helium nucleus) when it decays. The process can be modeled by assuming the \( \alpha \) particle is trapped inside the nucleus like a particle in a well. The \( \alpha \) particle bounces back and forth within the nucleus until (by random chance) it grabs a small amount of energy from the vacuum, climbs the side of the well, and appears outside the confines of the nucleus. Once it’s on the outside, the repulsive force between the two protons of the \( \alpha \) particle and 93 remaining ones in the nucleus cause the \( \alpha \) particle to rocket away and be detected by circuitry in the smoke detector. This flow of \( \alpha \) particles is a current, which can be interrupted by something as unsubstantial as smoke. The interruption of current sets off the alarm.

Something interesting to notice is that the energy borrowed \textbf{really is} returned to the vacuum. We can calculate the energy an alpha particle needs to have to escape its confinement by the nucleus, and we find that the emitted alpha particle has a lower energy than this. If the borrowed energy didn’t have to be returned (and therefore conservation of energy could be violated on the large scale), the emitted alphas would be ejected much more quickly than they really are.

\textbf{The Confinement Principle and Energy}

Whenever a particle is restricted to a certain region of space (confined), it is also restricted to have only certain discrete values of energy. In other words, energies of bound particles are \textbf{quantized} (also known as \textbf{discrete}, like whole numbers), while energies of free particles are unrestricted (also known as \textbf{continuous} – like real numbers). To illustrate this, we will examine the problem of a particle trapped in an infinitely deep well. To keep things simple, we’ll work in one dimension only. A plot of potential energy vs. position for a well which sits on the \( x \)-axis between 0 and \( L \) is shown below:

Since the walls of this well are infinitely high, the electron (or whatever particle) has absolutely no chance of tunneling through. This means that its wave function must be zero at the edges (\( x=0 \) and \( x=L \)).
and \( x = L \). Just as is the case for a guitar string (tied down at both ends and plucked), there must be a half-integral number of wavelengths between the ends. We can then write

\[
L = \frac{n\lambda}{2} \quad \text{for} \quad n = 1, 2, 3, \ldots
\]

Using our earlier connections between wavelength and momentum as well as last semester’s energy-momentum relationship, we find

\[
p = \frac{n\hbar}{2L} = \sqrt{2mE} \quad \text{so} \quad E = \frac{\hbar^2 n^2}{8mL^2} \quad \text{for} \quad n = 1, 2, 3, \ldots
\]

This is a reasonable result in light of the uncertainty principle; the uncertainty in momentum is at least equal to \( \hbar / \text{size of uncertainty in position} \). As you can see, the energy spacing is not constant. The lowest energy levels (\( n = 2 \) and \( n = 1 \)) would be separated by \( \Delta E = E_2 - E_1 = 3\hbar^2/8mL^2 \) while the \( n = 10 \) and \( n = 9 \) levels would have a \( \Delta E = 19\hbar^2/8mL^2 \). If the electron is going to move from a lower level to a higher level (higher value of \( n \)), it can only happen by the absorption of energy, as would happen if a photon were absorbed. The electron can similarly only drop down to a lower level by emitting a photon of the required energy.

**Wave Functions**

If the electron in a well acts like a wave, we should be able to find its wave function which will more completely describe it. Since the potential is constant (and therefore usually set equal to zero) inside the well, the particle is free until it hits the edges of the well. This means we need to solve the Schrödinger equation for zero potential, subject to the condition that our answer must equal zero when \( x = 0 \) or \( L \). You can quickly verify that the solution below will work:

\[
\psi_n(x) = A \sin \left( \frac{n\pi x}{L} \right) \quad \text{for} \quad n = 1, 2, 3, \ldots
\]

One of the things this function can do for us is to help us find the electron’s position. Remember that we’re not using Newtonian physics, so we will be looking for probabilities for the electron to be in various locations. The probability will just be this wave function (which is a probability amplitude) multiplied by its complex conjugate. Since the wave function is real in this case, we can just square it and get the same result. Therefore, we can say that the probability function \( p(x) \) is

\[
p(x) = \psi_n^2(x) \, dx = A^2 \sin^2 \left( \frac{n\pi x}{L} \right) \, dx
\]

What’s \( A \)? It’s the normalization constant which sets the scale for the overall problem. We can find it through a very simple
argument: the probability of finding the electron somewhere on the line from $-\infty$ to $+\infty$ must be 1. The wave function will be zero outside of the well, so the probability function will also be zero there. We’re left with the logical argument that the integral of $p(x)$ from 0 to L must be 1. We can write that as

$$\int_{-\infty}^{\infty} p(x) \, dx = \int_{0}^{L} A^2 \sin^2 \left( \frac{n\pi x}{L} \right) \, dx = 1$$

After some substitutions and integration, we are left with $A = \sqrt{\frac{2}{L}}$. We can now put this back into our expression for $\psi$ and the only variables remaining are $L$, $n$, and $x$. Imagine what our probability function inside the well will look like for various values of $n$. If $n = 1$, we’ll have a single bump in the center trailing off to zero at the edges, meaning that if an electron is in its lowest energy state (ground state), we’re more likely to find it at the center of the well than near the edges. As $n$ increases, we fit more cycles of $\sin^2$ in the well. At very large values of $n$, the electron is essentially equally likely to be anywhere in the well, which is what classical physics predicts. This is an example of Bohr’s correspondence principle, which just says that we will get classical physics back from quantum physics at large enough values of $n$.

If the well’s walls are not infinitely high, we have a finite square well, which can also be solved using Schrödinger’s equation. The only problem with this is that the electron does have a chance of tunneling out of the well, and that means that its wavefunction does not drop to zero exactly at $x=0$ and $L$. Because of this, finding the solutions to this equation becomes significantly harder; we have to match the solution outside the well with the one inside the well. This was much easier in the infinite potential case, since we just needed to make the wave function go to zero at the wall. Here it will have to go to some number, the particular value of which depends on the size and depth of the well and the particle’s total energy. It’s too mathematically ugly for us to do, but we can notice the general results that 1) there is penetration of the wave function into the walls of the well and 2) the energy of a given level in a finite square well will always be less than the energy of the same state (same $n$) in an infinite square well of the same width.

**The Bohr Model**

Bohr’s explanation of the energy levels in the hydrogen atom fits observed spectral lines very well. Bohr, like Planck, also used the idea that atomic energy levels were quantized (i.e., they could only assume certain fixed values). Bohr also assumed (here’s the crazy idea) that the electrons in these energy levels don’t radiate energy. In other words, the classical-physics idea that the electrons are accelerated and therefore radiate doesn’t work. Bohr predicted that radiation only happened when an electron moved from one energy level to another one. If a photon strikes an atom, the electron can be promoted to a higher energy level (kind of like moving a mass up to a higher shelf in a gravitational field). Eventually, the electron will fall back down to a lower level (mass falls off of shelf) and emit a photon. If we want to preserve conservation of energy, the two energy levels must be related to the photon’s energy by

$$E_i - E_f = hf$$

12
where $E_i$ and $E_f$ are the initial and final energy levels, and $hf$ is the photon’s energy. How do we find the allowed energy levels?

The energy of an electron in a hydrogen atom has two parts: potential energy, due to the attraction between the electron and a proton, and kinetic energy due to the motion of the electron. The total energy is the sum of these, and it must be negative. The reason is that we know we have to invest energy into the atom to separate the electron and proton. When they are completely separated and at rest, the electron’s KE is zero and the PE between the electron and proton is zero (if they are separated by an infinite distance). By conservation of energy, if we have to add energy to the system to get to the zero energy point, it must be negative to start with. We get:

$$E = \frac{1}{2} m v^2 - \frac{k Z e^2}{r}$$

where $Z$ is the charge on the nucleus (+1 for hydrogen, +2 for helium, +3 for lithium, ...). What’s $v$? In the planetary model, the centripetal force is supplied by the electrostatic attraction between the nucleus and the electron, so we get

$$\frac{m v^2}{r} = \frac{k Z e^2}{r^2}$$

There are two unknowns here ($v$ and $r$) and only one equation. We need another equation to fully specify everything. Bohr assumed that the angular momentum of the electron was quantized. This is a big leap – in Newtonian physics, there’s no lower limit to the amount of angular momentum that can be added to a system, or that the system can have (besides zero). He assumed that the electron’s angular momentum satisfied

$$L = m v r = \frac{n h}{2\pi}$$

where $n$ has to be an integer. This weird idea was explained later by de Broglie, who said that the wave representing the electron (remember particle-wave duality?) was like a standing wave in a pipe. He believed that the circumference of the electron’s orbit had to hold an integral number of wavelengths. In other words,

$$2\pi r = n \lambda$$

However, de Broglie had already connected the electron’s wavelength to its momentum. We can change the formula above into
2\pi r = n\frac{h}{mv} \implies mvr = \frac{nh}{2\pi}

Anyway, we can take this and plug it into the centripetal force/electrostatic force equation above and get

\[ m v^2 r = kZ e^2 \implies \frac{L^2}{mr} = k Z e^2 \implies \frac{n^2 h^2}{4\pi^2 mr} = k Z e^2 \]

so solving for \( r \), we get

\[ r = \frac{n^2 h^2}{4\pi^2 m k Z e^2} \]

For hydrogen, this predicts a radius (known as the \textbf{Bohr radius}) of 5.29 x 10^{-11} m. Plugging this back into the expression for the energy of the atom gives

\[ E = -kZ e^2 \frac{2\pi^2 m k^2 e^4}{h^2} \frac{Z^2}{n^2} \]

The quantity in parentheses is made up of constants, and its value is 13.6 eV. This tells us that the ground state \((n = 1)\) the lowest energy state the atom can have) of a hydrogen atom \((Z = 1)\) has an energy of \(-13.6\) eV, and that’s the amount of energy we have to supply to ionize the atom.

This formula explains that energy levels aren’t evenly spaced. The factor of \(1/n^2\) tells us that the levels get closer together as the electron gets closer to being freed. For example, to jump from the ground state to the 1\textsuperscript{st} excited state \((n = 2)\), we need to supply

\[ E_f - E_i = -13.6 eV \left( \frac{1}{2^2} - \frac{1}{1^2} \right) = 10.2 eV \]

but if we want to move from the 48\textsuperscript{th} excited state \((n = 49)\) to the 49\textsuperscript{th} \((n = 50)\), we only need

\[ E_f - E_i = -13.6 eV \left( \frac{1}{50^2} - \frac{1}{49^2} \right) = 2.24 \times 10^{-4} eV \]
It’s occasionally helpful to think of the electron moving between energy levels as being like a ball on a staircase (though the stairs are spaced strangely). Moving from a lower step to a higher one takes an investment of energy; the move back down the steps releases energy. Once the ball reaches the bottom of the steps (ground state), no lower drops are possible.

If we want to know how much energy it will take to remove the electron entirely from the atom, that also depends on the starting state of the electron. The final state can be considered to be \( n = \infty \), so our formula is just

\[
E_f - E_i = -13.6eV \left(0 - \frac{1}{n^2}\right) = \frac{13.6 eV}{n^2}
\]

This tells us about the line spacing in the hydrogen spectrum. To put this in the same form as the Lyman series, Balmer series, etc., we recognize that \( \frac{1}{\lambda} = \frac{f}{c} \). We can then write

\[
\frac{1}{\lambda} = 2\pi^2 \frac{m^2 k^2 e^4}{\hbar^3 c^2} Z^2 \left( \frac{1}{n_f^2} - \frac{1}{n_i^2} \right) \quad n_i > n_f
\]

Plugging the values for these constants into the formula (including \( Z = 1 \)) gives us (using \( R \) for the Rydberg constant as before)

\[
\frac{1}{\lambda} = R \left( \frac{1}{n_f^2} - \frac{1}{n_i^2} \right)
\]

If the final state of the atom is the first excited state \( (n_f = 2) \), we’ll get the Balmer series. If the final state is the second excited state \( (n_f = 3) \), we’ll get the Paschen series. Once this discovery was made, the empirical formulae discovered by Balmer, Paschen, Lyman, etc. were now understood and explained. The lines produced by the excited gas atoms form an emission spectrum for the gas. Similarly, if the full rainbow of colors (white light) is passed through a cool, low-pressure gas, certain wavelengths will be absorbed by the gas (they will very quickly be re-radiated, but in random directions) forming an absorption spectrum. The connection between emission, absorption, and blackbody spectra can be seen below:
The Sun (a dense ball of hot gas) produces a blackbody spectrum – all colors within a range. Someone looking through a prism directly at the Sun (point A) will see the full rainbow (and wavelengths outside the visible, but that’s not important right now). Someone looking at the Sun through a thin, cool gas will notice certain lines missing. This is because light matching one of the energy-level transitions of the atoms in the gas gets absorbed by the gas and used to move the electrons to a higher state. The electrons will very quickly drop back down, but they’ll emit the photons (matching the missing parts of the rainbow) in random directions. Only a tiny percentage of them will be emitted in the original direction, so the person at point B will see very dark lines when comparing them to the full spectrum nearby. This is the absorption spectrum, and it’s really all we can see of the Sun (since it has its own atmosphere around it). Someone at point C, who wouldn’t be seeing anything if not for the gas, will see some of the scattered photons emitted by the gas. Since they are comparing those photons to the darkness beside them, the lines look bright in comparison and they get an emission spectrum.

**Schrödinger’s Equation for Hydrogen Atoms**

In a hydrogen atom, the electron’s potential energy varies with radial distance to the proton at the center \( r \) as

\[
U(r) = -\frac{e^2}{4\pi \varepsilon_0 \, r}
\]

While one number \( n \) is enough to specify the electron’s state in the one-dimensional square well, for an atom in three dimensions of space, we need three numbers. The first one is also called \( n \), which is the **principal quantum number**. The next two are \( l \) and \( m \), known as the **orbital quantum number** and **orbital magnetic quantum number**, respectively. Just as \( n \) must
be a positive whole number greater than zero, \( l \) and \( m \) are also restricted to certain values. \( l \) can range from 0 to \( n-1 \) and \( m \) goes from \(-l\) to \(+l\) (both \( l \) and \( m \) must be whole numbers). The wave function for an electron in its ground state in a hydrogen atom (found by solving the Schrödinger equation in three dimensions in spherical coordinates) is

\[
\psi(r) = \frac{e^{-r/a}}{\sqrt{\pi} \ a^{3/2}}
\]

where \( a \) is the Bohr radius found earlier. To find the probability density function, we just square this as before (again, since \( \psi \) is real, squaring is the same as multiplying by the complex conjugate) and get ready to integrate. We have to pay a little attention to how we do the integration in 3-D. This is a spherically symmetric problem since there is no top, left, right, etc. of a proton, so the only important thing is the distance from the proton to the electron (radial distance, or \( r \)). Also, in 3-D, we need to integrate the probability density over some volume \( dV \) rather than some part of the one-dimensional well \( dx \). We need an expression for \( dV \). We could use \( dV = dx \ dy \ dz \), but that’s not spherically symmetric, which means our integration will be painful. We can instead use \( dV = 4 \pi r^2 \ dr \) (notice that if you integrate this, you get the volume of a sphere, or \( 4/3 \pi r^3 \)). We’ll absorb the \( 4 \pi r^2 \) into our probability function so that the integral will look almost the same as it did in the square well case. This means that we get

\[
P(r) = \psi^2(r) \ 4 \pi \ r^2 \ \Rightarrow \ \frac{4r^2}{a^3} e^{-2r/a}
\]

and our probability integral is

\[
\int_0^\infty P(r) \ dr = 1
\]

Notice that the integral ranges from 0 to \( \infty \) since we are in radial coordinates; there’s no such thing as a negative distance between the proton and electron. This result tells us that the idea of electrons in planet-like orbits must be wrong. The probability for finding the electron at various distances looks like this.
Notice the peak at $r = 1$ Bohr radius. This is the most likely place to find the electron, but it is by no means guaranteed to be there. There’s still a substantial chance to find it anywhere from practically zero radius to 4 Bohr radii or more.

We’ve said earlier that 1) the potential is spherically symmetric and 2) we need 3 quantum numbers for 3 dimensions of space. For this reason (and our choice of variables), $n$ is the only number that determines the energy of the electron in an isolated hydrogen atom. The other numbers only have an effect on the energy of an electron if there is something present to single out one particular direction in space, like a magnetic field or electric field. In the absence of those, all $l$ and $m$ for a given $n$ have the same energies. If a magnetic field is present, the different values of $l$ and $m$ will shift the electron’s probability function around and make certain locations more or less likely. For example, large values of $l$ (= high angular momentum) make it unlikely that the electron will be found in the center of the atom. States with high values of $l$ (and therefore high values of $n$) look more like the classical idea of planets orbiting the Sun in circles – the correspondence principle again.

In chemistry and atomic physics, electrons with the same value of $n$ in an atom are said to be in the same shell and those with the same value of $l$ are in the same subshell.

**Electron Spin**

In addition to the three quantum numbers $n$, $l$, and $m$, an electron in an atom also has spin angular momentum, which is also quantized. This is the same spin we used to explain the various kinds of magnetism in earlier chapters. We can’t think of it as being exactly like the electron is spinning on its axis like a planet, since the electron is apparently a point particle with no physical size. However, the reason for calling this quantity spin angular momentum is that it works in essentially every respect like a classical spin would. One of the few differences is that all particles of the same type have the same total spin. All electrons have spin $= \frac{1}{2}$ (as do all protons, neutrons, and quarks – more about them later). When we look at the projection of spin along any particular axis, we could measure either $+\frac{1}{2}$ or $-\frac{1}{2}$. This provides us with two new quantum numbers; $s$ is the value of the intrinsic spin of the particle (again, $\frac{1}{2}$ for the electron).
and \( m_s \) is the projection of that spin along a particular axis, ranging from \(-s\) to \(+s\) in steps of one. For an electron, that means the only two possible values of \( m_s \) are \(+\frac{1}{2}\) or \(-\frac{1}{2}\). These five numbers are enough to completely specify any particular electron in any atom. **No electron can have the same quantum numbers as another electron in the same atom.** This is known as the Pauli exclusion principle, and it’s what makes atoms stable. The addition of spin means we can fit twice as many electrons in a given shell or subshell. For example, look at the periodic table below (from http://www.physics.ohio-state.edu/~lvw/physics/bigpertable.jpg)

Notice that the top row contains only hydrogen and helium. In hydrogen’s lowest energy state, it will have one electron in the \( n = 1 \) state. For this electron, if \( n = 1 \), \( l = 0 \) and \( m_l = 0 \). On the opposite side of the table we have helium, which has two electrons. They can both have \( n = 1 \), \( l = 0 \) and \( m_l = 0 \) (meaning they are both in the ground state) if they have different values of \( m_s \) (they can’t have different values of \( s \) since they’re both electrons and all electrons have \( s = \frac{1}{2} \)). The electrons in helium’s ground state will then have the same quantum numbers for everything except \( m_s \), and one will have \( m_s = +\frac{1}{2} \) while the other has \( m_s = -\frac{1}{2} \). This means helium has no more room for electrons in the \( n = 1 \) level and we can then say that its outer shell is filled. This makes it extremely unlikely that helium will try to combine with any other element since having a filled shell is the “goal” of any atom.
The next row starts with lithium which has the filled $n=1$ shell and one electron in the $n=2$ shell. This shell has subshells with $l=0$ and $l=1$. In those subshells, $m_l$ can be either 0 ($l=0$) or –1, 0, or +1 ($l=1$). That gives 4 possible combinations of $l$ and $m_l$, which would mean that the second shell would be filled by the time we get to element number 6. However, the fact that we now have the spin magnetic quantum number $m_s$, means we can double the available space for electrons, and the second shell doesn’t fill up until we reach element number 10 (neon, another element that generally refuses to combine with anything else because of its filled shell).

**More About Angular Momentum**

Now it’s time to combine a few of the things we’ve seen before. We know that electrons have both spin and orbital angular momentum, and we know that these things are quantized (remember that this means they are restricted to some limited set of values rather than infinitely variable). How are the magnitudes of spin and orbital angular momentum related to the components of these things along a particular axis (generally taken to be the $z$ axis)?

It might seem that, in the cases when $m_l = l$ or $m_s = \frac{1}{2}$, since these are maximal values, the vectors must be aligned with the $z$ axis. If that were true, though, we would also know the components of angular momentum along the $x$ and $y$ axes as well: they’d have to be zero. Because of the uncertainty principle, we are not allowed to know both the magnitude of angular momentum and more than one of its components. This means that the magnitude will be a little different than we might expect. For the orbital angular momentum, we will get

$$L = \sqrt{l(l+1)} \hbar$$

You can verify for yourself that even if $m_l = l$, which is as large as it can get, it will still be smaller than $L$. This means there will be $x$ and $y$ components (which we can’t know). The $z$ component is then

$$L_z = m_l \hbar$$

A similar result holds for the spin angular momentum, where we find that

$$S = \sqrt{s(s+1)} \hbar = \sqrt{\frac{1}{2}(\frac{3}{2})} \hbar$$

and

$$S_z = m_s \hbar$$

The spin and angular momentum vectors will combine to form a **total angular momentum vector** $\mathbf{J} = \mathbf{L} + \mathbf{S}$ (all adding vectorially, of course). If there are multiple electrons in an atom, we will add the $L$’s and $S$’s separately, and then put the totals together in $\mathbf{J}$. This total angular momentum will determine the magnetic properties of the atom.
Where’s the proof that electrons have spin, and that this has anything to do with the world? What if we take a beam of atoms with nonzero total angular momentum (and therefore some magnetic properties) and put them in a nonuniform magnetic field? We know that a uniform magnetic field will only align dipoles, not move them around. If the field is not uniform, though, there is a stronger push one way than the other, and the dipoles (atoms) will move in one direction or the other, depending on the strength of their magnetic dipole moment. In the 1920’s, an experiment (done by Stern & Gerlach) involving silver atoms moving through a magnetic field produced an interesting result. If there was no magnetic field present (or in the presence of a uniform field), a beam of silver atoms would just fly off in a straight line and hit a target. All the atoms were the same, so there would only be one place they could land.

When the nonuniform magnetic field was activated, different strengths of magnetic dipoles would feel different forces. In classical physics, we could expect any value for this strength since the component of angular momentum along the field direction could take any value from a negative maximum to a positive maximum or anything in between. We would therefore expect the spot to become a line, where the atoms landing at the ends of the line would have had the largest deflections (negative or positive) corresponding to the largest dipole strengths. In fact, the beam split into two spots instead of a line. This meant there were only two possible values for the magnetic dipole strength and, from what we already know, two possible values for the projection of the electron’s spin. Although silver has 43 electrons, 42 of them are paired up and cancel each other’s contribution to the total angular momentum. The 43rd electron is in the 5s state, meaning it has zero orbital angular momentum (that’s what s means) and a spin angular momentum of either + or −½. This is solid proof that electrons are allowed only two values for $m_s$. (You can see the arrangements of silver’s electrons at [http://www.webelements.com/webelements/elements/text/Ag/econ.html](http://www.webelements.com/webelements/elements/text/Ag/econ.html))

In contrast, look at iron ([http://www.webelements.com/webelements/elements/text/Fe/econ.html](http://www.webelements.com/webelements/elements/text/Fe/econ.html)). It has 4 unpaired electrons in the 3d shell (meaning $l = 2$). They combine to give a large value of total angular momentum, which is why iron is so magnetic.

Magnetic Resonance

If an isolated electron can have a magnetic moment because of its intrinsic spin, the same must be true of an isolated proton (or a proton in a hydrogen atom, which is alone in the nucleus). However, the size of the magnetic moment is inversely proportional to the particle’s mass, so the proton’s magnetic moment will be much smaller than the electron’s (it won’t be quite as simple as taking $1/1836$ times the electron’s magnetic moment $\mu_B$, though, because other factors involving the proton’s internal structure are also involved). The value for the proton is $1.41 \times 10^{-26}$ J/T. If the proton is placed in a strong magnetic field, it will try to align its spin with the field. If the proton is pointing in the wrong direction (since it has spin $1/2$, the only choices are aligning as close as possible to the direction of the field or as far opposite to it as possible; there’s no in between), it will have to get rid of some energy to line up. That energy is equal to
Inside a large MRI machine (which used to be called an NMR machine, for Nuclear Magnetic Resonance, until people decided that the word “nuclear” tended to scare patients), the magnetic field is typically in the 1 to 3 Tesla range. That means the proton would have to get rid of $2 \mu B = 2 \times 10^{-7}$ eV of energy. This tiny amount of energy corresponds to a photon in the 80 MHz range. The absorption of photons of this frequency will push aligned protons into the unaligned position. The back and forth absorption and emission are not balanced since the protons are more likely to be in the lower energy state than in the upper energy state. Since this energy difference is so tiny compared to the thermal energies at room temperature (somewhere around 0.025 eV), the population difference between aligned and unaligned states is small (but detectable).

This method can be used to identify the location of protons (= hydrogen nuclei = part of water molecules) in the body or in other materials. Many different isotopes of nuclei have nonzero spins, and that’s all that is necessary to use this method. Additionally, the magnetic field of nearby nuclei and atoms can cause measurable and specific changes in the energy of the emitted or absorbed photon; this method is used to identify particular materials in a process known as NMR spectroscopy.

### The Pauli Exclusion Principle and the Periodic Table

When an atom has multiple electrons, finding the energy of a given electron is much harder to do, since the electrons interact with (repel) one another, and also tend to screen out some of the positive charges in the nucleus. For example, if an electron is some small distance away from a neutral hydrogen atom, it shouldn’t see much of the charge of the nucleus at all – there’s a positive charge in the center, and a negative charge surrounding it.

For these more complicated atoms, the energy of a particular electron is still most critically dependent on the value of $n$, but the other quantum numbers become important as well. For a given value of $n$, larger values of $l$ mean larger energies. The values of $m_l$ and $m_s$ don’t affect the energy (in our approximation) except in the presence of a magnetic field. For this reason, we can group electrons in an atom which share the same value of $n$ into energy-based classes called shells. If $n=1$, $l$ can only equal zero, so $m_l$ can also only equal zero, and our final quantum number $m_s$ can be $+1/2$ or $-1/2$. That gives us room for two electrons in the lowest shell. This is because of the Pauli exclusion principle, which says that each electron in an atom has to have a unique value of $(n, l, m_l, m_s)$.

So, the first shell holds 2 electrons. In hydrogen, we’ve only got one electron, so it can have an $m_s$ of $+1/2$ or $-1/2$ (commonly called up or down). In helium, we’ve got two electrons. Both would like to be in the ground state, and both can be if their spins are opposite (different values of $m_s$). This fills the first shell completely, and there is no particular incentive for helium to take another electron, and it definitely doesn’t want to give up one of its two. Helium is therefore essentially chemically inert. When we go to three electrons in a lithium atom, one of them has to go to the 1st excited state ($n = 2$). It is further from the nucleus than the other two, and the positive charge of the nucleus is screened reasonably well by the two inner electrons. This is
why lithium doesn’t hold that 3rd electron too tightly. An atom that is willing to grab other electrons or release one of its own is usually very chemically active.

The \( n=2 \) shell allows electrons to have \( l \) values of 0 or 1. For \( l = 0 \), the choice for \( m_l \) is just 0, as in the ground state. That means two electrons (differing values of \( m_s \)) can fit in the \( l = 0 \) subshell. For the \( l = 1 \) subshell, \( m_l \) can be \(-1, 0, +1\). The three possibilities multiplied by the two choices for \( m_s \) give a total of 6 more places to put electrons. It took 2 electrons to fill the first shell; it takes 2 more to fill the \( n = 2, l = 0 \) subshell and 6 more to fill the \( n = 2, l = 1 \) subshell. That means an atom with 10 electrons should have both of its first two shells filled, and should also be very stable. That atom is neon, and it’s another noble gas (not interested in trading electrons with anything else). The electrons with larger \( l \) values in the \( n = 2 \) shell have higher energies, but all \( m_l \) values (and the \( m_s \) values) represent equal energies (in our approximation) unless there is a magnetic field around.

By the time we get to \( n = 3 \) and 4, the extra energy of a high-\( l \) state becomes significant. In fact, the \( n = 3, l = 2 \) state has a higher energy than \( n = 4, l = 0 \). That means we’ll fill up \( n = 3, l = 0 \), then \( n = 3, l = 1 \), and then skip \( n = 3, l = 2 \) to first fill \( n = 4, l = 0 \). This kind of thing happens for other values of \( n \) and \( l \). Since the periodic table groups atoms which have similar chemical behaviors in columns, and since that behavior is largely dependent on the outer electrons, we can see that we’ll get a string of atoms with similar chemical properties. They’ll have the same outer electron arrangement, but the inner electron levels will be filling up as we move to higher atomic numbers. That’s why we have those weird groupings of elements off to themselves like the Lanthanides (all have the \( n=6, l=0 \) level filled but have various vacancies in lower \( n \) subshells) and the Actinides (all have the \( n=7, l=0 \) filled, etc.)

**X-rays**

When fast-moving electrons hit a solid substance, they interact with the atoms in the material (both the electrons in the atoms and the nuclei) and rapidly lose energy. This comes out in the form of photons, and we can plot the spectrum of the radiation (intensity vs. frequency or wavelength). We’ll see a broad, relatively flat curve representing the continuous radiation known as Bremsstrahlung, which means “braking radiation”. It comes from the interaction between nuclei in the material and the speeding electron. The electron is decelerated by the interactions, and therefore emits photons. We’ll also see a few very sharp peaks in the curve which represent what are known as characteristic X-rays. These are nothing more than the spectral lines we’ve already seen, except they belong to inner-shell electrons. These electrons are very tightly bound, so the ionization energies can be in the thousands of electron volts instead of just a few eV.

There is also a sharp cutoff to the spectrum, because if the electron has moved through a potential difference of 70,000 volts, for example, the highest-energy photon it can possibly generate is 70,000 eV. There’s no way to make one larger than that, so the short-wavelength cutoff is sharp, but the longer wavelengths trail off to all lower energies.

This is the process used to generate X-rays for hospitals – electrons are accelerated through a potential difference and then allowed to slam into a dense material. It turns out that materials with a high atomic number (large \( Z \)) are best suited for making X-rays. Not only are their inner-
shell electrons bound more tightly (remember the factor of $Z^2$ in the expression for energy levels in an atom), giving characteristic X-rays in the range needed for making images (20,000 eV and up), but they are more efficient at producing Bremsstrahlung (production depends on $Z^2$ here as well).

**Lasers**

One of the most interesting and important discoveries of the 20th century was made by Einstein when he realized that atoms could emit light in two ways. The “normal” way for an excited atom to get to the ground state is by spontaneous emission of a photon. This is called spontaneous because it will happen in the (apparent) absence of any external influence. Einstein noticed that, for thermodynamic equilibrium to be maintained, another kind of emission must also exist. This is known as stimulated emission, and it is caused by a photon striking the (already excited) atom. As we know, the atom can move from a lower state to a higher one if a photon carrying an energy equal to the energy difference between levels hits it. It turns out, though, that stimulated emission can occur if the photon is carrying the amount of energy that would be released if the atom made a downward transition. In other words, if we call the three lowest energy levels $E_1$, $E_2$, and $E_3$, an atom in $E_2$ can go in either direction: it can absorb a photon of frequency $(E_3 - E_2)/h$ and move up to the third level, or it can interact with a photon of frequency $(E_2 - E_1)/h$ and drop down to the first level, emitting a photon of its own in the process. The special part about stimulated emission is that the second photon (emitted by the formerly-excited atom) is exactly the same as the first photon (the one that caused the emission). We expect the energies to be the same, because of the condition that the first photon’s energy matches the difference in energy levels. We also see that the phase and direction of the two photons are the same.

This means laser light is monochromatic (single frequency), coherent (in phase), low divergence (not much spread), and able to be focused to a very small region for high intensities.

We don’t see this behavior from other light sources because a typical excited atom will undergo spontaneous emission in a few nanoseconds. There isn’t time for another photon to stimulate it. Additionally, if we imagine our first photon moving through a gas of these atoms, it could just as easily be absorbed by an atom in level 1 as it could trigger the de-excitation of an atom from level 2 to level 1. In fact, the first process is far more likely: atoms in lower states are much more common than those in excited states. The number of atoms $N_x$ in a higher energy state $E_x$ compared to the number $N_0$ in the ground state of energy $E_0$ can be found by

$$N_x = N_0 e^{- (E_x - E_0)/kT}$$

For small $kT$ (room temperature would be $kT = 0.025$ eV or so) and large differences in the energies of the states (a few eV, like visible light photons), very few atoms will be found in the excited state, so free photons are probably going to be moving back and forth between level 1 atoms and promoting them to level 2. Only occasionally will one hit a level 2 atom and cause stimulated emission. When it does, the two identical photons will probably be absorbed by level 1 atoms anyway. What we need to do to fix this is to put most of the atoms into a higher level.
That way, a photon has a better chance of hitting a level 2 atom and stimulating a copy of itself than it does of being used to move a level 1 atom to level 2. This situation, if we can cause it, is called a population inversion.

We can make this by understanding that some atoms have metastable states. These are excited states that will last much longer than the typical excited state (possibly ms as opposed to ns!). If we can put atoms in this relatively-long-term storage, we can move many of them there and then trigger their de-excitation by a photon when we want. Each photon that hits an excited atom will cause the release of a copy of it, as well as continuing on unharmed. Very quickly, we’ll build up many photons which are duplicates. We can let the photons bounce back and forth along the gas by confining them with mirrors. If one mirror is partially transmitting, we’ll get a laser beam out of the end. Lasers are an incredibly efficient way to generate light. A laser pointer (bright enough to damage your eye) typically uses around a few milliwatts of power! Lasers which use hundreds of watts (less than a toaster) could cut through you (or steel, etc.). The wavelength emitted by the laser depends on the energy of the metastable state of the atoms which fuel it.

Conductors, Insulators, and Semiconductors

The electrical properties of materials are very important in a world filled with computers. We have so far talked in general terms about whether a given material is a conductor or insulator based on its resistivity. This is a major factor, but there are two other physical properties that we should examine. First, there is the temperature coefficient of resistivity, which describes how resistivity varies with temperature, and is symbolized by $\alpha$. We can define $\alpha$ using

$$\alpha = \frac{1}{\rho} \frac{d\rho}{dt}$$

Finally, the number density of charge carriers is also important. We usually use $n$ to represent this quantity, which we can determine using the Hall effect. The units of this will just be $\text{m}^{-3}$.

We can organize these materials by resistivity, and we’ll find $\rho_{\text{insulator}} > \rho_{\text{semiconductor}} > \rho_{\text{metal}}$. The change of resistivity with temperature is $\alpha_{\text{metal}} > 0$ and $\alpha_{\text{semiconductor}} << 0$ (large and negative, in other words). Also, $n_{\text{semiconductor}} < n_{\text{metal}}$. We need a theory that explains these differences.

The key to this is the realization that our earlier work with energy levels (shells) relied on atoms being isolated from each other (far away on the scale of the outer electron’s orbit). As we bring atoms closer together, the outer electrons start to feel forces from the neighboring atom, and the energy levels split. As we bring more atoms together, the available energy levels continue to multiply. Eventually, the splitting is so extreme that bands develop, which are energy ranges where there are so many available states that the levels are tiny fractions of an eV apart (we can find the number by dividing the band size, which will be a few eV, by the number of atoms present, typically $10^{24}$ or so. This gives a spacing so small as to almost be a continuum.)

The electrons have so many available states spaced so closely that, in some ways, it’s like they are free electrons (remember that free electrons have no restrictions on their energies). There are
still energies that are not allowed, though, just as in the single-atom case. These energy ranges are called **gaps**.

![Energy level spacing – single atom](image1)

![Energy level spacing – macroscopic body](image2)

The size of the gap is comparable to the size of the bands around it, in general, meaning a few eV or so. Although the energy levels in the band are very closely spaced, there are still only a finite number of levels in any band, and we still have the Pauli exclusion principle to consider. If one band is completely filled, any electron that wanted to have a higher energy would have to jump the gap into the next band. For a current to flow, the electrons have to gain energy. What this means is that, if a band is completely filled and any electrons would have to make a very large jump to get to an unoccupied band, this material will be an insulator. Putting a “normal” voltage across an insulator doesn’t give the electrons in the filled band enough energy to jump up to the next band, so no current flows. What if we keep increasing the voltage to higher and higher values? Eventually, the insulator will break down and conduct, meaning we’re now giving the electrons enough energy to jump the gap and get into the free band.

Metals obey the same rules, but the top band in a metal is only partially filled. When you think about it, a half-full band is ideal for conduction; a band that is almost empty will have only a few electrons that can participate in conduction, and if the band is almost full, only a few electrons will be able to gain energy and move to the higher regions of the band (thereby carrying current).

Where will these electrons sit as we lower the temperature? The energy bands will fill in from lowest to highest until all the electrons have found an energy level. At $T = 0$ K (absolute zero), all electrons will be in the lowest possible state for each. Going to the highest occupied band, we set the minimum energy in that band to be zero (this just means electrons in lower bands will have negative energies, but that’s not a problem). The energy of the highest occupied level in this band (and therefore the highest level in the whole material) is called the **Fermi energy** $E_F$. This energy is about 7 eV for copper and about 11.7 eV for aluminum. These values are so large compared to the thermal energies ($k_B T = $ about 0.025 eV at room temperature) that the energy level occupation looks almost the same at room temperature as it would at absolute zero. In other words, almost all of the electrons are in the lowest state available, and only a small fraction are “excited” above the lowest free state. The vast majority of electrons in a metal **don’t** participate in the conduction of electricity. Only the **valence** (outer) electrons of a metal are involved. Each
atom in the metal will therefore contribute its outer electrons (typically one or two) to conduction.

The probability that a given energy level will be occupied is given by the formula below

\[
P(E) = \frac{1}{e^{(E - E_F) / k_B T} + 1}
\]

At 0 K, this function tells us we will get a probability of 1 for every state below \(E_F\) to be occupied and zero chance of any higher states being filled. This will look like a **step function**, named for its obvious similarity to a step:

![Step Function](image)

When \(k_B T\) is small as it is at room temperature, this graph doesn’t change much. The corner at the Fermi energy is rounded and the function has a small value for \(E > E_F\):

![Step Function](image)

As small as this change is, it’s actually exaggerated for clarity.

**Semiconductors**

Semiconductors at 0 K look very much like insulators in terms of energy bands; the topmost occupied band (valence band) is filled and the next band above it (conduction band) is empty, and a gap separates the two. The gap is smaller than in a typical insulator, which means that at room temperature, the electrons actually have a reasonable chance of making the jump from the
valence band to the conduction band. When they do this, the electrons are now participating in the current carrying process. They have also left a space behind in the valence band, known as a hole. This is just an opening for another electron (lower in the same valence band) to now take its place. That means this electron is also participating in the conduction process. It’s common to simplify the description of this by considering the holes to be positive charge carriers which move in the same direction as the conventional current, just as the electrons are negative charge carriers moving opposite to the conventional current. This is the reason why resistivity drops with temperature (so $\alpha$ is negative) in semiconductors.

We can enhance this effect by replacing a tiny fraction (about 1 in 10 million) Si atoms in a semiconductor with something like a phosphorus atom or an aluminum atom in a process called doping. Phosphorus has one extra electron in its outer shell compared to silicon, so that electron is sort of unnecessary when it comes to binding the atoms into a solid lattice. The electron is only loosely bound, so it takes only a small amount of energy to free it and let it carry current. Because phosphorus has one extra electron, it’s known as a donor atom. Adding more free electrons (negative) particles to the semiconductor makes it an n-type semiconductor. We could substitute aluminum for phosphorus also. Aluminum only has 3 electrons in its outer shell (1 less than silicon), so there’s a hole there that electrons on neighboring atoms could fall into without much encouragement. The holes are sort of like positive charge carriers, so a piece of silicon doped with aluminum is a p-type semiconductor.

The p-n Junction

If we combine these two materials, one with an excess of free electrons and one with an excess of mobile holes, the place where they touch is called a p-n junction. The excess electrons will cross the junction to try to fill the excess of holes. This motion of charged particles (current) is called the diffusion current. We could just as easily think of this as positively charged holes moving across the junction in the other direction. The conventional current (what a positive charge would do) goes from the p to the n side. However, each side of the junction was originally electrically neutral; there might have been “extra” electrons in the n-type semiconductor, but there were just as many “extra” protons in the phosphorus nuclei. The same is true on the p side. Now that electrons have migrated, the n side is actually positively charged, and the p side is negatively charged. This creates an electric field which tends to stop more electrons from crossing over. The region around the junction where electrons and holes have found each other (so there are few mobile charge carriers present) is called the depletion zone.

Since there will be some mobile holes in an n-type semiconductor, and some mobile electrons in a p-type, they will see the electric field in the depletion zone as an energetically favorable move. Their motion forms the drift current. This will exactly balance the diffusion current in a p-n junction which is not connected to any circuit. There will be a potential difference between the two sides of the junction which acts as a barrier, preferentially keeping holes on the p side and electrons on the n side.

Diodes

The balance between the drift and diffusion currents can be tilted by connecting the p side of the junction to the positive terminal of a battery and the n side to the negative terminal of the same
battery. This is called applying a **forward bias** to the diode, and it tends to remove the potential barrier discussed above and shrink the size of the depletion zone. In this state, the diffusion current climbs while the drift current is unaffected. This gives a net current from the p side to the n side. If we connect the p-n junction to the opposite battery terminals (known as a **reverse bias**), the depletion zone gets larger, the potential barrier gets higher, and the diffusion current will be smaller than the drift current. This gives a (very small) net current through the diode, but it is in many cases so small that we can consider it to be zero.

When the electrons and holes combine, we can either get thermal energy, as in a “normal” diode, or we can get the emission of light, in a **light-emitting diode (LED)**. Here, the junction is doped heavily and the energy gap is chosen to be somewhere around 2-4 eV. This means the combination of an electron with a hole (which must release a photon) will produce radiation in the visible part of the spectrum. The frequency of light emitted is related to the gap energy as you might expect: \( f = \frac{E_g}{h} \)

Reversing this process by shining light on the p-n junction can produce an electric current. This is the basic design of the **photodiode**.

When properly designed, these junctions can be combined (either as p-n-p or n-p-n) and used to make a **transistor**. The basic idea behind a transistor is to allow a small current to control a large current. The small current signal is used to operate a **gate** which allows a large current to flow through the transistor. For example, we know that a microphone (which can be simply described as a paper cone attached to a coil of wire near a magnet) produces a current which varies as the coil in the microphone moves back and forth in response to sound waves. The current from the microphone is tiny, but it can be used to control the gate in a transistor and let a much larger current move through it. As the signal from the microphone gets larger or smaller, the large current flow gets larger or smaller. If that current flow is attached to a speaker, we have a public address system.

**Nuclear Physics**

Inside the cloud of (negatively charged) electrons which make up the exterior of an atom, we find a positively charged nucleus. The nucleus is composed of protons and neutrons. The protons are positively charged, and the neutrons are neutral. Their masses are almost the same, and both are much larger than the electron. If we measure in electron masses, the proton’s mass is about 1836 \( m_e \) and the neutron has a mass of about 1839 \( m_e \). For an atom to be electrically neutral, the number of protons and number of electrons must be equal. The element’s identity, and therefore its place on the periodic table, is determined by the number of protons in its nucleus. Hydrogen has one, helium two, lithium three, etc. on up to 116 protons in ununhexium. For a given number of protons, a nucleus can contain a (small) range of numbers of neutrons. Atoms with the same number of protons but different numbers of neutrons are called different **isotopes** of the element. Ordinary hydrogen has zero neutrons, deuterium is an isotope of hydrogen with one neutron (and one proton, of course), and tritium has 2 neutrons and a proton. The number of protons + the number of neutrons is sometimes called the number of **nucleons** and is usually denoted by \( A \). The number of protons in the nucleus is usually represented by \( Z \). Since we can get the same information from the chemical symbol for the element, it is sometimes omitted, but we’ll generally give the element’s symbol a superscript which tells us the total nucleon number
Physical Properties & Mass Defect

Although electrons are considered to be point particles, with no measurable physical size, nucleons (protons & neutrons) have reasonably well-determined radii, which are around $10^{-15}$ m. This distance is so common in nuclear physics that, in addition to being known as a femtometer, it’s also called a Fermi. Since nuclei can be thought of as roughly spherical, and the nucleons are basically touching each other, the radius $r$ of a nucleus can be approximated by

$$r = r_0 A^{1/3} \quad \text{where} \quad r_0 = 1.2 \times 10^{-15} \text{ m}$$

As you might guess, kilograms are inconveniently large units for nuclei. Atomic mass units, or amu, are more commonly used. These units (represented symbolically by $u$) are $1.6605 \times 10^{-27}$ kg. In energy terms, multiplying this by $c^2$ and converting to MeV, we get $1 \text{ u} \times c^2 = 931.49$ MeV.

If the particles in a nucleus are held to one another very tightly, we might ask how much energy it would take to remove one of them. This is kind of like our earlier effort to find out how much energy we have to add to an atom to remove an electron. For an outer-shell electron, the energies we’re talking about are typically a few eV. The energy equivalent of an electron’s mass is $511,000$ eV, so this is only a small change in the total energy of the electron. In the nucleus, while masses of nucleons are much larger, the energy in the attraction between nucleons is larger still. If we talk about the energy needed to separate a nucleus into individual protons and neutrons, that is called the binding energy. This is easily determined by looking at the mass of the nucleus compared to the sum of the individual masses of the nucleons. In our everyday experience, the two should be exactly equal: a dozen identical eggs will weigh exactly 12 times what one egg weighs.

It’s the binding energy that makes things different. We know that we have to invest energy in the nucleus to get it to split into many nucleons. When it has split, we’ll have a mass-energy equal to $Z$ protons and $(A-Z)$ neutrons. If we had to add energy to get to that point, we must have had a smaller mass-energy when the nucleons were all bound together. The difference in mass between the nucleus and its individual parts is called the mass defect, and it’s related to the binding energy by

$$\text{Binding Energy} = (\text{mass defect}) c^2$$

For example, a helium nucleus has a mass of $6.6447 \times 10^{-27}$ kg, and is composed of two protons (mass = $1.6726 \times 10^{-27}$ kg) and two neutrons ($1.6749 \times 10^{-27}$ kg), giving a difference (mass defect) of $5.03 \times 10^{-29}$ kg. Multiplying that by $c^2$ and converting to electron volts, we get about 28 MeV. Later, when we look at other atoms, we’ll see that the binding energy goes up as the size of the nucleus goes up. What will be a more interesting number for us is the binding energy per
nucleon. To find it for helium, divide the 28 MeV we found by the 4 nucleons in the nucleus, and we get about 7 MeV per nucleon. As it turns out, this will be a very useful measure of the stability of a given nucleus. A helium nucleus is very stable.

Notice that the energies needed to change things in the nucleus are in the neighborhood of a million times as large as those needed to move electrons around. This will explain (later) the incredible power of nuclear bombs & reactors, as well as the inability of alchemists to successfully turn lead into gold.

**Radioactivity**

When a nucleus has gotten too large (or when the number of protons and the number of neutrons are out of balance, or when the nucleus has been excited somehow), it will emit a radioactive particle of some kind. The three major kinds we’re interested in are called $\alpha$, $\beta$, and $\gamma$ particles. If these had been discovered today, we would probably have given them different names. For example, the $\alpha$ particle is just a helium nucleus: 2 protons and 2 neutrons stuck together. When a nucleus (usually a very large one) emits an $\alpha$ particle, it will change the chemical identity of the element. For example, if U$^{238}$ undergoes $\alpha$ decay, it will turn into Th$^{234}$, changing its chemical identity. In this case, the U$^{238}$ would be called the parent nucleus and Th$^{234}$ the daughter nucleus. In general, then, the decay will look like this:

$$A_z \ P = \ A_{z-4} \ D + \alpha$$

Notice that the total number of nucleons is conserved during the decay. Four of them left in the form of an $\alpha$ particle, but the total number on the right is equal to the total number on the left. Also, charge has been conserved. The daughter nucleus has lost two units of positive charge, but they are part of the $\alpha$ particle.

The $\alpha$ particle will be emitted with a large amount of kinetic energy. This makes sense if you consider that the positively-charged $\alpha$ particle will find itself sitting very close to the positively-charged daughter nucleus. The Coulomb repulsion will be incredibly strong, and the nucleus will be ejected with a relatively large kinetic energy. The $\alpha$ particle won’t travel very far in matter, though: it’s very heavy, meaning its speed is not huge for a given KE. Also, it carries a net charge of +2, so it will be interacting with the electron clouds of every atom it passes. $\alpha$ particles typically travel distances of a few mm or less, depending on the density and atomic number of the material they’re passing through. The dead layer of skin on the outside of your body is enough to stop these particles. Smoke detectors rely on the ability of smoke to stop $\alpha$ particles. A very small radioactive source provides a current made up of $\alpha$ particles to a sensor. If smoke gets between the source and the detector, the current drops and the alarm is triggered.

$\beta$ decay changes the relative number of protons and neutrons in the nucleus. If a nucleus has too many protons for the number of neutrons in it, or vice versa, it can emit an electron (written as $\beta^-$) or a positron, which is the antimatter partner of an electron (written $\beta^+$). For electron emission, a neutron in the nucleus will become a proton, and we can write
If a proton becomes a neutron instead, we’ll get

\[ {}^A_Z P = {}^{A+1}_{Z+1} D + \beta^- \]

A third possibility is that a proton in the nucleus of a heavy atom can grab one of its innermost electrons and the two will be converted into a neutron. This is called electron capture. When an electron leaves the nucleus, it will wander around until it finds a new home. If a positron is ejected, though, it will move around only until it finds an electron. When it does, the two will annihilate in a process which converts mass to energy with 100% efficiency. The two particles will collide and produce two gamma rays (each with an energy of about 511 keV), conserving charge (1 + -1 before = 0 + 0 after) as well as energy. This is the basis for PET imaging, where the near-simultaneous arrival of the two gamma rays provides information about the location of the radioactive source. Electrons from β decay are more penetrating than α particles because of their smaller mass (and therefore higher speeds for a given kinetic energy) and lower charge. A thin sheet of metal will stop most of them, though.

γ decay is the final type we’ll examine. Sometimes, a nucleus will end up in an excited state (frequently after some other kind of decay has already happened) and will need to rid itself of some extra energy. In much the same way as an electron dropping from a higher state to the ground state, the nucleus will de-excite and emit a gamma ray (energies here are typically in the MeV range, although they can drop down into the keV range as well). The identity of the element doesn’t change, since a photon can’t carry away any charge, and the nucleon number won’t change since photons aren’t nucleons. We write the decay as

\[ {}^A_Z P^* = {}^A_Z P + \gamma \]

where the asterisk indicates an excited state. Since γ rays are just high energy photons, they carry no charge and move at the speed of light. They are significantly more penetrating than electrons from β decay, and may go through several cm (or more) of lead.

**Activity & Decay**

The decay of atoms is very different from the gradual aging and death of plants and animals. For radioactive atoms, we define a half-life (written as \( T_{1/2} \)) as the amount of time necessary for half of a given sample of atoms to decay. This is really a statistical measure of the chance for any given atom to decay in a certain period of time, and since it is statistical, we will only be able to apply it to large numbers of atoms (not usually a problem, since a billion billion atoms would only be micrograms of most elements). In the same way, if we flip a coin twice and get heads
both times, that doesn’t prove it’s not a regular coin. If we do it many, many times and get a significant deviation from 50% heads, we may have a trick coin.

One of the big differences between atoms and living things is that the chance of decay is independent of the age of the atom. For example, $^{14}\text{C}$ has a $T_{1/2}$ of about 5700 years. The independence of age means that a 20,000 year-old $^{14}\text{C}$ atom has as much chance of decaying in a given time period as a brand new $^{14}\text{C}$ atom. This also means that the number of atoms decaying at any point in time depends linearly on the number of atoms which have not yet decayed. If $N$ is the number of atoms at any given time, the rate of decay (measured in decays per second and known as the activity of the sample) is given by

$$\frac{dN}{dt} = -\lambda N$$

where $\lambda$ is a constant which depends on the particular type of atom, and is called the decay constant. Notice the negative sign – that’s what tells us the sample is getting smaller over time. In the SI system, we measure activity in becquerels (Bq) where 1 Bq = 1 decay per second (a very low activity). An older unit is the Curie, which is equal to 37 billion decays per second.

Anyway, to find the amount of an element remaining after time $t$, we need to solve the equation above. When we do that, we get that

$$N(t) = N_0 e^{-\lambda t}$$

where $N_0$ is the amount of the element you had when you started your watch ($t = 0$). Since we know that when $t = T_{1/2}$, $N$ will equal $N_0/2$, we can use that to relate the decay constant to the half life and we get

$$T_{1/2} = \frac{\ln 2}{\lambda} \approx 0.693 \frac{\lambda}{\lambda}$$

Scientists use this idea to find the dates when a given sample of material was formed. This process is used to find the approximate year of death of animals and plants, which take in $^{14}\text{C}$ along with ordinary $^{12}\text{C}$ while they are alive. Once dead, the amount of $^{14}\text{C}$ starts to decay. If measurements show that only 3.5% of the original amount of $^{14}\text{C}$ in an organism remains today, we can solve the math above to find the time of death:
Sometimes, a radioactive element decays into a daughter atom which is itself still radioactive. In this case, we have a radioactive decay series. The decay of $^{238}\text{U}$ continues until it becomes $^{206}\text{Pb}$, meaning there are several $\alpha$ decays (along with other kinds of decay). This kind of decay is responsible for Earth’s supply of helium, which does not combine chemically with anything and has such a high velocity at normal temperatures that it quickly escapes Earth’s gravity when released from its underground deposits.

**Neutrinos**

Scientists in the early part of the 20\textsuperscript{th} century noticed, after careful observation of the end products of $\beta$ decay, that the electron was usually ejected with a range of kinetic energies, up to a maximum of the total energy available in the decay. When the electron left with less than the maximum KE, where was the rest of the energy going? Either conservation of energy didn’t work at the nuclear level, or something else was going on. Pauli (of the Pauli exclusion principle) postulated the existence of another particle that was not being detected. It must have no charge (or it would be easily detected, as well as indicating a violation of the conservation of charge, which would be bad news) and an incredibly tiny rest mass (otherwise, the electron could not take such a high fraction of the decay energy since some of it would have to be used to produce this particle). The new particle was called the neutrino, and its discovery gives us another conservation law. This one is called conservation of lepton number. Lepton comes from the Greek word for “light”, since the primary leptons (neutrinos and electrons) are among the lightest of matter particles. Conservation of lepton number means that we have to have the same number of neutrinos + electrons before and after a decay. Each neutrino or electron will have a lepton number of +1, while positrons and antineutrinos will have lepton numbers of −1.

In positron decay, we have an electron (lepton number +1) hitting a positron (lepton number −1) and leaving us with two photons (each lepton number 0). That means, though, that our previous decay diagrams must have been incomplete. For electron emission, we should write

$$A \quad Z 
\begin{array}{c}
\text{P} = Z + 1 \\
\text{D + } \beta^- + \overline{\nu}
\end{array}$$
where $\nu$ is a neutrino and $\bar{\nu}$ with a bar over it means an antineutrino. Positron emission becomes

$$^Z_A P = ^{Z-1}_A D + \beta^+ + \nu$$

Why were these not detected for so long? Neutrinos (and their antimatter counterparts) have no charge, so they don’t interact via the electromagnetic force. Leptons in general (electrons & positrons included) don’t feel the strong force. Gravity is the weakest of all forces, and while neutrinos feel it, gravity is just too weak to play a role in nuclear processes. How do neutrinos interact at all? They feel the weak force. This is the last of the four fundamental forces of nature, and it governs radioactive decay. Leptons do feel this force, but it is so weak (though still stronger than gravity) that it is incredibly unlikely for a neutrino to interact with any of the matter in a particle detector of any kind. A neutrino has the ability to pass through trillions of km of lead without interacting!! The only way these are ever detected is when the flux of neutrinos is ridiculously high and the detector is the size of a building. When those conditions are met, you might detect a neutrino a day or so. Years after the discovery of the neutrino, it was realized that the electromagnetic and weak forces are really part of the same force, just as the electric and magnetic forces are really part of the electromagnetic force. The new force is called the **electroweak force**, and it drops the number of fundamental forces in the universe to three, with the addition of gravity and the strong force.

**Radiation Detectors**

There are many kinds of devices which detect radiation. Among the most common are **Geiger counters** which are just gas-filled cylinders with a wire running down their center. The wire and the wall of the cylinder are connected to opposite sides of a high-voltage source (kind of like 2 capacitor plates), and there is a thin window in the cylinder to allow $\alpha$, $\beta$, and $\gamma$ rays to enter. When they do, they ionize the gas between the wire and the cylinder wall, allowing a current to flow between the two. This brief current is measured and frequently fed to a speaker, providing a “click”.

**Scintillation detectors** work by using a special compound in the form of a crystal which emits (many) visible-light photons when struck by radiation. A **photomultiplier tube** (PMT) is placed against the crystal’s face and the photons cause the release of electrons in the tube (photoelectric effect). A potential difference in the tube accelerates the now-freed electrons on to a different part of the tube, where they strike a positive electrode and release more electrons. These are accelerated further to the next (even more) positive electrode, and the process continues. In this way, a few electrons at one end are amplified into many electrons at the output of the PMT. This electrical pulse can be recorded and analyzed later.

Other kinds of detectors include **cloud chambers** and **bubble chambers**, where a gas is on the verge of condensing (cloud) or a liquid is on the verge of boiling (bubble) and a passing $\alpha$, $\beta$, or $\gamma$ causes droplets to form along its path (cloud) or bubbles to form along the path.
**Biological Effects of Radiation**

How can we quantify the danger posed by radioactive materials around us? Various things are important: first, if the radiation is not ionizing, we don’t generally run into intensities of it which will cause problems. Ionizing radiation means radiation which can knock electrons free from atoms, and photons with energies in the UV and above can generally do this. Ions that don’t belong in a cell can cause trouble in there, so ionizing radiation is the kind we worry about.

One way to define the intensity of radiation is through exposure, which measures the number of ion pairs formed by radiation as it passes through a given amount of air (at standard temperature and pressure, known as STP). This would be a number of coulombs (of ions) per kilogram (of air), but an older unit for this is called the roentgen. We can convert using

\[
Roentgens = \frac{1}{2.58 \times 10^{-4}} \left( \frac{q}{m} \right)
\]

For living things, we don’t care about what’s happening in air as much as we care about what happens in the organism. For this, we have the absorbed dose, measured in grays (Gy) or rads (\(= 0.01\) Gy). This is a measure of the energy absorbed per kilogram of material doing the absorbing. 1 Gy = 1 Joule/kg (or 100 rads – a very high dose for a person).

Finally, we can include a factor that differentiates between the types of radiation based on the damage they cause to us, and this factor is called the relative biological effectiveness (RBE). This factor is measured relative to a dose of 200 keV X-rays. In other words, if something has an RBE of 10, a given dose of it will do 10 times as much damage as the same dose of 200 keV X-rays. \(\gamma\) and \(\beta\) particles have RBE’s of 1, but \(\alpha\) particles have an RBE of 10 to 20. We use this factor and the absorbed dose to find the biologically equivalent dose, measured in rems:

\[
Bio.\,Equiv.\,Dose = RBE \times Absorbed\,\,Dose
\]

One rem is actually a fairly large dose of radiation. Typically, most people get something less than half a rem per year from exposure to natural sources of radiation, plane flights, etc. A few hundred rem will kill most people, and 1000 rem is almost certainly lethal to anyone. A CT scan will typically give you about one rem of exposure (not known to be enough to cause anything in particular).

**Induced Nuclear Reactions**

Instead of the random decays characteristic of naturally radioactive elements, we can cause decays if we add energy and/or particles to a nucleus. For example, we can bombard nuclei of \(^{32}\)S with neutrons. We’ll get \(^{32}\)P and protons out, which we can write as
\[ ^{32}S + n \rightarrow ^{32}P + p \]

Notice that the total number of nucleons is conserved (32 +1 before, 32+1 after) as is the total charge (+16 + 0 before, +15 + 1 after). A shorthand way to write this is in the form below:

\[ \text{parent} \{ \text{bullet, shrapnel} \} \text{daughter} \]

where parent is the target being bombarded by bullet, producing shrapnel and daughter. For example, we can write

\[ ^{10}_5 B (n, \alpha) ^{7}_3 Li \]

if we mean that firing neutrons at a \(^{10}B\) nucleus gives us \(\alpha\) particles and \(^{7}Li\).

**Nuclear Fission**

In a variant of the kind of induced reaction above, the addition of a neutron to an already-large nucleus can cause it to split into two approximately equal parts. This is known as nuclear fission. In some reactions, more neutrons are released in addition to the two nuclear fragments. Depending on the material, these extra neutrons can cause more fission events, releasing even more neutrons, causing more fissions, etc. This runaway process is known as an uncontrolled chain reaction and can release an incredible amount of energy in a very short time (nuclear bombs work on this principle). In a typical fission event, \(> 100 \text{ MeV}\) can be released. A thousand tons of TNT will release about \(4.2 \times 10^{12} \text{ J}\). It will only take the fissioning of about a hundred grams (1/4 pound) of uranium to release the same energy!

If the chain reaction above can be controlled, so that only one neutron from each fission is left free to cause another fission, useful work can be produced instead of a giant explosion. For \(^{235}U\), the ideal way to cause fission is to hit it with a thermal neutron, meaning one which has a kinetic energy of about 0.03 eV. These are called thermal because their energies are around \(kT\) where \(k\) is Boltzmann’s constant and \(T\) is room temperature in Kelvins (about 300K). Faster neutrons are less likely to cause fissions in \(^{235}U\), but fission reactions typically produce very fast neutrons. For these to be usable to cause more fissions, they have to be slowed through collisions with a moderator, sometimes water. The slowed neutrons can now cause more fissions. Each fission event releases about 200 MeV of energy which eventually becomes heat and is used to boil water. The steam is then used to spin a turbine, which turns a generator.

If each fission reaction produces (on average) less than 1 new neutron which causes another nucleus to fission, the reaction will die out eventually and is called subcritical. If each reaction causes exactly 1 new fission, the reaction is self-sustaining and critical. If more than 1 new fission-causing neutron per fission is produced, the reaction is supercritical and will end in a meltdown or explosion. The rate of neutron production is regulated by control rods which contain material that is good at absorbing neutrons (boron, for example) but which does not
fission and produce more neutrons. As the rods are pushed deeper into the reactor, they soak up more neutrons and keep the reaction below the critical point.

Nuclear Fusion

Another way to produce large amounts of energy is to join two small nuclei to make a larger nucleus. This is called fusion and is the source of the Sun’s energy output. Fusion is more efficient (energy output for a given mass input) than fission. For example, we can combine two isotopes of hydrogen (deuterium and tritium) to make helium:

\[
\begin{align*}
^2_1H + ^3_1H & \rightarrow ^4_2He + n
\end{align*}
\]

Notice that we still have to conserve nucleon number and charge. This reaction releases about 18 MeV. This is less than the uranium fission reaction above, but the energy released per nucleon is about 4 times greater! What does it take to cause fusion? We have to overcome the incredibly strong Coulomb repulsion between the two nuclei we want to fuse. They have to be moving so fast that the repulsive Coulomb force can’t slow them down enough to stop them before they hit. This requires very high temperatures and/or pressures, such as are found in the center of the Sun or the center of a fission bomb. That’s why there is a fission bomb at the heart of every thermonuclear (or hydrogen, or H) bomb. It’s the match that lights the larger explosive. In our sun, the pressures and temperatures work to combine 6 hydrogen nuclei (= protons) into a helium nucleus plus 2 of the original protons and some gamma rays (as well as the positrons and neutrinos needed to enforce conservation of charge and lepton number). About 25 MeV is released in this reaction. These very-high-energy \( \gamma \) rays are scattered over and over in the center of the Sun, losing a little energy each time and adding to the overall temperature of the Sun. After thousands of years, those few MeV photons have been scattered into millions of photons of a couple of eV each which finally escape the surface of the Sun.

Fusion would be a great energy source for us to use on the Earth – the oceans are full of hydrogen, and the by-product (helium) is harmless. The problem of nuclear waste would almost disappear (although the reactor materials themselves could get radioactively nasty). Unfortunately, we don’t yet have a way to hold the hydrogen nuclei together as we heat them to ridiculous temperatures. The Sun uses gravity, but we can’t. We also can’t use any kind of traditional bottle, since it would heat up and be destroyed while at the same time cooling the hydrogen it holds. There has been some progress in holding the fuel using a magnetic field (magnetic confinement) and in relying on inertia to hold the fuel motionless momentarily while large lasers blast away at it (inertial confinement). So far, this method looks to be a few decades away from running a commercial power plant.

If we got this to work, could we hook up a fusion reactor to one of our old fission reactors and just let them trade fuel and waste? Could we let the fusion reactor build big nuclei that the fission reactor could then split back into little nuclei? No. This would be equivalent to perpetual motion. The physical stumbling block here is in the binding energy per nucleon. A graph of this curve is shown in your book. The curve peaks at about 60 nucleons (around iron). This means that fusing large atoms or splitting small atoms (where large and small are on opposite sides of the 60 nucleon peak) releases some energy, but it costs more than it releases.
Elementary Particles

Particles can be put into two broad classes: matter particles and force-carrying particles. The matter particles we’ve seen so far are the electron (and its antimatter equivalent, the positron), the neutrino (and antineutrino), and the proton and neutron. In fact, protons and neutrons are made up of smaller particles known as quarks, which we believe are truly fundamental. There are six kinds of quarks, arranged into three families, written as below:

\[
\begin{pmatrix}
 u \\
 d \\
 c \\
 s \\
 t \\
 b
\end{pmatrix}
\]

These letters stand for up, down, charmed, strange, top, and bottom. These quarks are arranged so that the upper quark in each family has a charge of \(+2/3\) \(e\) and the lower one has a charge of \(-1/3\) \(e\). Antimatter copies of all of these quarks exist, and their charges are reversed relative to the matter versions. Things made of quarks are called hadrons and they are the only particles that feel the strong force. We never see free quarks; we always see them paired either in groups of three quarks or three anti-quarks, or in pairs of one quark and one anti-quark. The groups of 3 are known as baryons and the quark-anti-quark pairs are called mesons. A neutron is composed of a single up quark and two down quarks. Verify that its charge must be zero. A proton, on the other hand, is made up of two up quarks and a single down quark, giving a charge of \(+e\). The matter particles which don’t feel the strong force are known as leptons. There are also three families of 2 leptons each, and we can write them as

\[
\begin{pmatrix}
 e^- \\
 \nu_e \\
 \mu^- \\
 \nu_\mu \\
 \tau^- \\
 \nu_\tau
\end{pmatrix}
\]

We now see that there are really 3 different kinds of neutrino, each associated with a different family. The other leptons are the electron, the muon, and the tau particle. The associated neutrinos are specified as the electron-neutrino, muon-neutrino, and tau-neutrino. Notice that all “ordinary” matter is made from the first family of quarks and the first family of leptons. The second and third groups only appear in high-energy collisions.

These 12 basic particles, and their anti-particles, are the fundamental constituents of matter (or anti-matter). The force-carrying particles are: the photon, which carries the electromagnetic force and is therefore exchanged only by charged particles; the gluon, which carries the strong force, and is only exchanged by quarks; the \(W^+\), \(W^-\), and \(Z^0\) particles, which carry the weak force, and are exchanged by all matter particles, and the (as-yet undiscovered) graviton, which carries the gravitational force, and is felt by all particles, both force-carrying and matter.

In yet another division of particles, those with integral spin are called bosons and those with half-integral spin are called fermions. Force carrying particles like those mentioned above are all bosons. They do not obey the Pauli exclusion principle – in fact, they like to be in the same state.
This is part of the reason that lasers work as they do. All fermions obey the exclusion principle, and they are commonly considered to be matter particles.

The point of these divisions is to group particles by the rules governing their behavior. Baryon number is conserved, for example. Whenever a neutron decays, it will produce a proton. Both have baryon number = +1, and the electron produced has a baryon number of 0 (since it’s not a baryon), so it is conserved. Because lepton number is also conserved, the electron is not the only other particle to be produced in the decay. It is present so that charge is conserved, but something else has to be produced to conserve lepton number. The antineutrino does this without disturbing the balance already present for charge or baryon number.

In addition to the conservation of overall lepton number, the lepton numbers of the individual families are separately conserved. That means that the decay of a $\mu$ into an electron must look like

$$\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e$$

The $\nu_\mu$ neutrino is there to preserve $\mu$ lepton number, and the $\nu_e$ antineutrino is there to balance the presence of the electron to keep $e$ lepton number at zero.

**The Strong Force**

We know that the negative electrons are held near the nucleus by the Coulomb force between them and the protons, but how are the protons able to exist so close to one another? We can estimate the magnitude of the repulsive force between two protons in the nucleus by assuming they are $10^{-15}$ m apart. We get

$$F = \frac{k q_1 q_2}{r^2} = \frac{\left(9 \times 10^9 \text{ N m}^2 / \text{C}^2\right)\left(1.6 \times 10^{-19} \text{ C}\right)^2}{\left(10^{-15} \text{ m}\right)^2} = 230 \text{ N}$$

That’s about 45 pounds!! That’s an incredible amount of force on these two tiny masses. There must be an opposing force which is very strong and acts to hold the protons together. This force is called, simply, the strong force, and it’s the last basic kind of force we’ve seen (the others being gravity, electromagnetism, and the weak force).

The strong force is incredibly strong, and it operates between both kinds of nucleons, almost independent of their type (protons attract protons and neutrons almost equally well). The strong force is also only active over a very short range – around $10^{-15}$ m or so. This means that a nucleon is held in place mainly by its nearest neighbors. This fact sets a limit on the size of nucleus that can exist. Every proton we add to the nucleus is attracted by a few nearby nucleons, but it’s repelled by every other proton in the nucleus since the electromagnetic force has an infinite range. One way we can postpone the inevitable is by adding lots of neutrons – they attract the neighboring protons, and they serve as “spacers” to separate protons from each other, thereby reducing the repulsive force between them. As the nucleus grows, these neutrons become more important.
For the light elements, there are typically as many neutrons as protons in the nucleus (the most common carbon isotope has 6 of each, the most common nitrogen isotope has 7 of each, etc.). As we move up the periodic table and add more protons, it takes more neutrons to separate them to have a stable nucleus. By the time we get to Bismuth, we need about 3 neutrons for every 2 protons to keep things stable. Eventually, we have too many protons for the nucleus to remain stable, and we’ll have **radioactive decay**, which is controlled by the **weak force**.

**The Electroweak Force**

In the 1970’s, physicists succeeded in developing a theory that unified both the electromagnetic and weak forces into one **electroweak force**. The force-carrying particles we see are the $W^+$, $W^-$, and $Z^0$. For example, free neutrons are unstable and will decay into protons with a half-life of about 10 minutes. The decay process involves one of the down quarks (charge = -1/3) becoming an up quark (charge = +2/3). For this to happen, something must carry away a charge of $-e$. This is the $W^-$ particle. The $W^-$ will decay into the electron and electron anti-neutrino (to preserve electron lepton number) so that we get

$$d \rightarrow u + W^- \rightarrow u + e^- + \bar{\nu}_e$$

This will conserve charge, baryon number, and electron lepton number.

**The Universe**

We know from the Doppler shift that the spectra of moving objects will be shifted towards higher or lower frequencies when those objects are moving towards or away from us. In the early 20th century, astronomer Edwin Hubble noticed that almost every galaxy visible is redshifted, meaning they appear to be moving away from us. Using various techniques of distance measurement, he discovered that the rate of recession is proportional to the recessional velocity; in other words, things farther from us are moving away from us faster. This is summarized as **Hubble’s Law**, which says that

$$v = H r$$

where $v$ is the measured recessional velocity, $r$ is the distance to the receding galaxy, and $H$ is the **Hubble constant**, which is currently estimated at about 70 km/s/Mpc (Mpc stands for Megaparsec, a distance equal to 3,260,000 light years). In other words, if an object is measured to have a redshift corresponding to a recessional velocity of 7,000 km/s, it is about 100 Mpc away. Notice we could rewrite this so that our units would be s$^{-1}$.

This effect is only noticeable when objects are so far apart that all other forces of interaction are essentially negligible. For example, the nearest large galaxy to our Milky Way is the Andromeda galaxy, which is about 2 million light years away. Hubble’s law would predict it is moving away from us at around 50 km/s or so, but it’s actually moving towards us due to the mutual gravitational attraction between it and our galaxy. Also, just because distant things in all
directions seem to be moving away from us, that doesn’t make us the center of the universe. Imagine dots painted on a balloon that is then inflated. Each dot will be moving away from every other dot, and the velocity of recession will be largest between the most distant dots because there’s more balloon in between them to expand.

Since we have a measurement of how fast the universe is expanding, we can work backwards and find out when all of the matter in the universe would have been found at the same location. This time is just $1/H$, and it gives us the time since the explosion that started the expansion, known as the **Big Bang**. This time is around 14 billion years or so.

Possibly the most compelling piece of evidence for the big bang theory is the discovery of the **cosmic background radiation**. This is a microwave signal which permeates all of space, and it looks like the blackbody radiation curve for an object with a temperature of about 2.7 K. If we start running the universe’s clock backwards and examining the physics at each point in time, we find that the temperature of the explosion dropped steadily as the universe expanded (no different than any other explosion). For many years after the big bang, temperatures would have been too high for atoms to exist.

After about 380,000 years, the average temperature would have dropped to just under 3,000 K and electrons would be able to recombine with ions to form neutral atoms. This is the first time when light would be able to propagate freely, since a cloud of ions and electrons (or **plasma**) would not be transparent to light. The universe has expanded by a factor of about 1000 in the billions of years since this **decoupling** of light and matter happened. That expansion stretched the light waves from the decoupling so that, instead of looking like a 3,000 K blackbody, it now looks more like a $3,000 / 1,000 = 3$ K blackbody.

**Dark Matter, etc.**

One of the strangest predictions made by astrophysicists is that there is more matter in the universe than we have been able to detect. This **dark matter** is believed to exist because plots of the rotational speeds of stars in galaxies as a function of their distance from the center place very strict limits on the mass that must be present inside their orbits. Gravity is the only force available to hold these (relatively) fast-moving stars in their orbits around the centers of their galaxies, but the amount of matter astronomers can see in the galaxies is insufficient to provide that gravitational force. In other words, based on the masses predicted by the brightness of the stars, there is just not enough mass to keep these stars in orbit.

Early ideas to fix this included black holes and brown dwarfs (things which contain lots of mass in the form of hydrogen and helium, but not quite enough to ignite and be stars), but nuclear and particle physics provides information connecting the relative abundances of hydrogen and helium in today’s universe with the amount of baryonic matter produced in the big bang. According to these studies, there **cannot** be enough matter tied up in black holes & brown dwarfs to account for the dark matter, because there wasn’t enough “ordinary” matter produced in the big bang in the first place.

This suggests an exotic type of matter that has never been detected on Earth. The matter has to interact gravitationally, but other than that, there’s not much we can predict about it. In addition
to the galactic rotation curves, the existence of dark matter is also predicted by the “clumping” of matter in galaxies in the first place. Given the time since the big bang and the temperatures shortly after it, it seems that ordinary matter would not have been able to condense into structures like galaxies very quickly (we can place age limits on galaxies by noticing that they exist very far away from us, meaning the light has had to travel a very long time to get to us, meaning they must have already existed billions of years ago). Dark matter, which wouldn’t exchange energy (through the electromagnetic force = photons) with a hot, expanding early universe, would be able to clump together quickly and thereby serve as sites of high mass density which would attract the ordinary matter. This allows a hot early universe to form galaxies earlier than would otherwise be possible.

Even stranger than dark matter is the recent suggestion of the existence of dark energy. This is believed by some astronomers to exist because some calculations suggest that the expansion of the universe is actually speeding up instead of slowing down! This would be like throwing a baseball into the air and watching it gain speed. There must be some other source of energy present if this happens, and right now (since it is unknown, like dark matter), it’s just called dark energy.

So far, there has been no experimental evidence for either dark matter or dark energy on Earth, and it is still possible that some other theory will explain the observed phenomena without the need to involve these things.