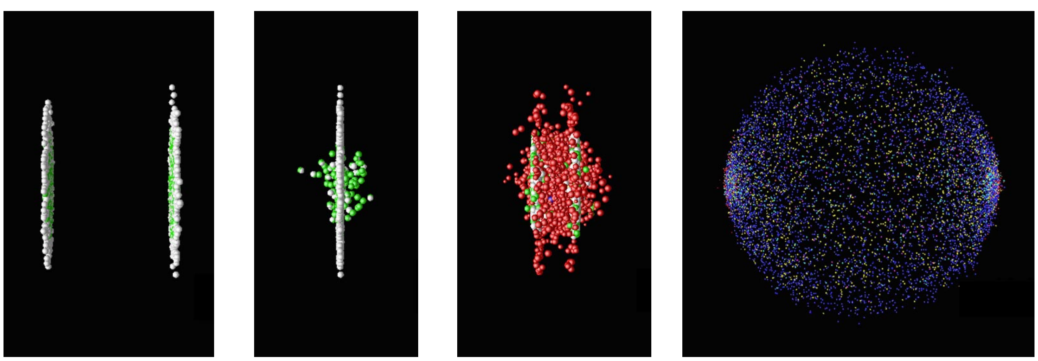
The Modern Revolution in Physics

Benjamin Crowell



Book 6 in the **Light and Matter** series of introductory physics textbooks
www.lightandmatter.com

The Modern Revolution in Physics

The **Light and Matter** series of introductory physics textbooks:

1 Newtonian Physics

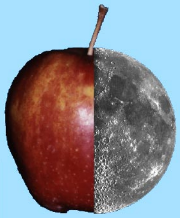
2 Conservation Laws

3 Vibrations and Waves

4 Electricity and Magnetism

5 Optics

6 The Modern Revolution in Physics



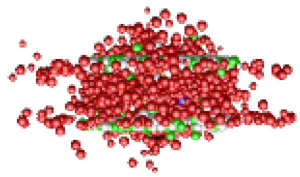
The Modern Revolution in Physics

Benjamin Crowell

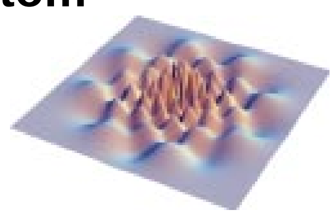
www.lightandmatter.com

Light and Matter
www.lightandmatter.com

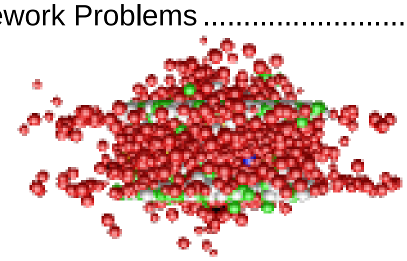
Edition 2.0 — 2000
rev. 2000-08-13

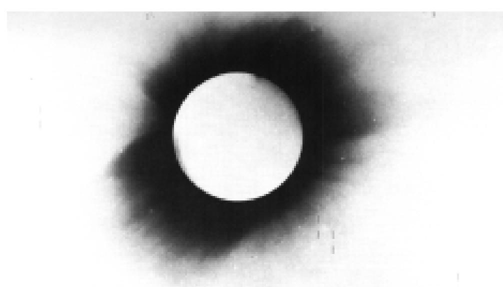


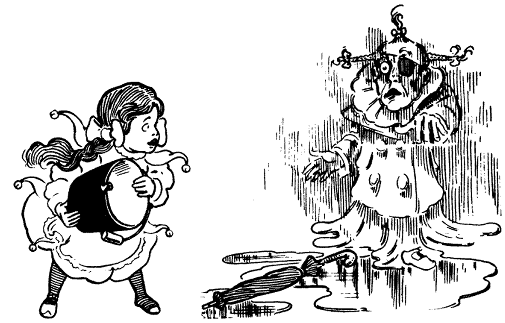
Contents

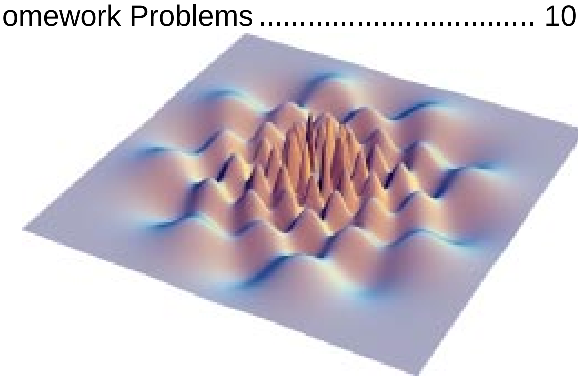


Contents









Albert Einstein in his days as a Swiss patent clerk, when he developed his theory of relativity.

1 Relativity, Part I

Complaining about the educational system is a national sport among professors in the U.S., and I, like my colleagues, am often tempted to imagine a golden age of education in our country's past, or to compare our system unfavorably with foreign ones. Reality intrudes, however, when my immigrant students recount the overemphasis on rote memorization in their native countries and the philosophy that what the teacher says is always right, even when it's wrong.

Albert Einstein's education in late-nineteenth-century Germany was neither modern nor liberal. He did well in the early grades (the myth that he failed his elementary-school classes comes from a misunderstanding based on a reversal of the German numerical grading scale), but in high school and college he began to get in trouble for what today's edspeak calls "critical thinking."

Indeed, there was much that deserved criticism in the state of physics at that time. There was a subtle contradiction between Maxwell's theory of electromagnetism and Galileo's principle that all motion is relative. Einstein began thinking about this on an intuitive basis as a teenager, trying to imagine what a light beam would look like if you could ride along beside it on a motorcycle at the speed of light. Today we remember him most of all for his radical and far-reaching solution to this contradiction, his theory of relativity, but in his student years his insights were greeted with derision from his professors. One called him a "lazy dog." Einstein's distaste for authority was typified by his decision as a teenager to renounce his German citizenship and become a stateless person, based purely on his distaste for the militarism and repressiveness of German society. He spent his most productive scientific years in Switzerland and Berlin, first as a patent clerk but later as a university professor. He was an outspoken pacifist and a stubborn opponent of World War I, shielded from retribution by his eventual acquisition of Swiss citizenship.

As the epochal nature of his work began to become evident, some liberal Germans began to point to him as a model of the “new German,” but with the Nazi coup d’etat, staged public meetings began to be held at which Nazi scientists criticized the work of this ethnically Jewish (but spiritually nonconformist) giant of science. Einstein had the good fortune to be on a stint as a visiting professor at CalTech when Hitler was appointed chancellor, and so escaped the Holocaust. World War II convinced Einstein to soften his strict pacifist stance, and he signed a secret letter to President Roosevelt urging research into the building of a nuclear bomb, a device that could not have been imagined without his theory of relativity. He later wrote, however, that when Hiroshima and Nagasaki were bombed, it made him wish he could burn off his own fingers for having signed the letter.

This chapter and the next are specifically about Einstein’s theory of relativity, but Einstein also began a second, parallel revolution in physics known as the quantum theory, which stated, among other things, that certain processes in nature are inescapably random. Ironically, Einstein was an outspoken doubter of the new quantum ideas, being convinced that “the Old One [God] does not play dice with the universe,” but quantum and relativistic concepts are now thoroughly intertwined in physics. The remainder of this book beyond the present pair of chapters is an introduction to the quantum theory, but we will continually be led back to relativistic ideas.

1.1 The Principle of Relativity

> Absolute, true, and mathematical time...flows at a constant rate without relation to anything external... Absolute space...without relation to anything external, remains always similar and immovable.
>
> Isaac Newton (tr. Andrew Motte)

Relativity according to Galileo and Einstein

Galileo’s most important physical discovery was that motion is relative. With modern hindsight, we restate this in a way that shows what made the teenage Einstein suspicious:

> **The Principle of Galilean Relativity**
> Matter obeys the same laws of physics in any inertial frame of reference, regardless of the frame’s orientation, position, or constant-velocity motion.

If this principle was violated, then experiments would have different results in a moving laboratory than in one at rest. The results would allow us to decide which lab was in a state of absolute rest, contradicting the idea that motion is relative. The new way of saying it thus appears equivalent to the old one, and therefore not particularly revolutionary, but note that it only refers to matter, not light.

Einstein’s professors taught that light waves obeyed an entirely different set of rules than material objects. They believed that light waves were a vibration of a mysterious medium called the ether, and that the speed of light should be interpreted as a speed relative to this ether. Even though Maxwell's treatment of electromagnetism made no reference to any ether, they could not conceive of a wave that was not a vibration of some medium. Thus although the cornerstone of the study of matter had for two centuries been the idea that motion is relative, the science of light seemed to contain a concept that certain frames of reference were in an absolute state of rest with respect to the ether, and were therefore to be preferred over moving frames.

Now let's think about Albert Einstein's daydream of riding a motorcycle alongside a beam of light. In cyclist Albert's frame of reference, the light wave appears to be standing still. He can stick measuring instruments into the wave to monitor the electric and magnetic fields, and they will be constant at any given point. This, however, violates Maxwell's theory of electromagnetism: an electric field can only be caused by charges or by time-varying magnetic fields. Neither is present in the cyclist's frame of reference, so why is there an electric field? Likewise, there are no currents or time-varying electric fields that could serve as sources of the magnetic field.

Einstein could not tolerate this disagreement between the treatment of relative and absolute motion in the theories of matter on the one hand and light on the other. He decided to rebuild physics with a single guiding principle:

> **Einstein's Principle of Relativity**
> Both light and matter obey the same laws of physics in any inertial frame of reference, regardless of the frame's orientation, position, or constant-velocity motion.

Maxwell's equations are the basic laws of physics governing light, and Maxwell's equations predict a specific value for the speed of light, $c=3.0\text{x}10^8$ m/s, so this new principle implies that *the speed of light must be the same in all frames of reference*.

1.2 Distortion of Time and Space

This is hard to swallow. If a dog is running away from me at 5 m/s relative to the sidewalk, and I run after it at 3 m/s, the dog's velocity in my frame of reference is 2 m/s. According to everything we have learned about motion, the dog *must* have different speeds in the two frames: 5 m/s in the sidewalk's frame and 2 m/s in mine. How, then, can a beam of light have the same speed as seen by someone who is chasing the beam?

In fact the strange constancy of the speed of light had shown up in the now-famous Michelson-Morley experiment of 1887. Michelson and Morley set up a clever apparatus to measure any difference in the speed of light beams traveling east-west and north-south. The motion of the earth around the sun at 110,000 km/hour (about 0.01% of the speed of light) is to our west during the day. Michelson and Morley believed in the ether hypothesis, so they expected that the speed of light would be a fixed value relative to the ether. As the earth moved through the ether, they thought they would observe an effect on the velocity of light along an east-west line. For instance, if they released a beam of light in a westward direction during the day, they expected that it would move away from them at less than the normal speed because the earth was chasing it through the ether. They were surprised when they found that the expected 0.01% change in the speed of light did not occur.

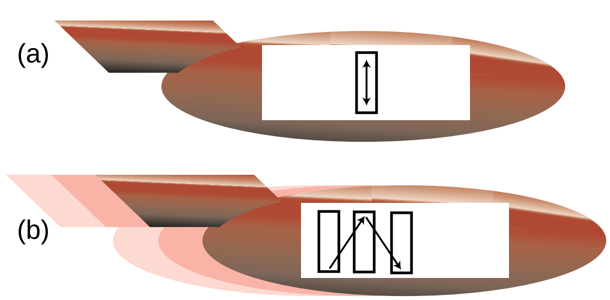
Although the Michelson-Morley experiment was nearly two decades in the past by the time Einstein published his first paper on relativity in 1905, he did not even know of the experiment until after submitting the paper. At this time he was still working at the Swiss patent office, and was isolated from the mainstream of physics.

How did Einstein explain this strange refusal of light waves to obey the usual rules of addition and subtraction of velocities due to relative motion? He had the originality and bravery to suggest a radical solution. He decided that space and time must be stretched and compressed as seen by observers in different frames of reference. Since velocity equals distance divided by time, an appropriate distortion of time and space could cause the speed of light to come out the same in a moving frame. This conclusion could have been reached by the physicists of two generations before, the day after Maxwell published his theory of light, but the attitudes about absolute space and time stated by Newton were so strongly ingrained that such a radical approach did not occur to anyone before Einstein.

An example of time distortion

Consider the situation shown in figures (a) and (b). Aboard a rocket ship we have a tube with mirrors at the ends. If we let off a flash of light at the bottom of the tube, it will be reflected back and forth between the top and bottom. It can be used as a clock: by counting the number of times the light goes back and forth we get an indication of how much time has passed. (This may not seem very practical, but a real atomic clock does work by essentially the same principle.) Now imagine that the rocket is cruising at a significant fraction of the speed of light relative to the earth. Motion is relative, so for a person inside the rocket, (a), there is no detectable change in the behavior of the clock, just as a person on a jet plane can toss a ball up and down without noticing anything unusual. But to an observer in the earth's frame of reference, the light appears to take a zigzag path through space, (b), increasing the distance the light has to travel.

If we didn't believe in the principle of relativity, we could say that the



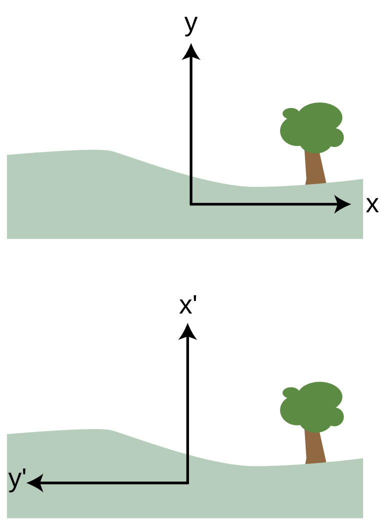
light just goes faster according to the earthbound observer. Indeed, this would be correct if the speeds were not close to the speed of light, and if the thing traveling back and forth was, say, a ping-pong ball. But according to the principle of relativity, the speed of light must be the same in both frames of reference. We are forced to conclude that time is distorted, and the light-clock appears to run more slowly than normal as seen by the earthbound observer. In general, a clock appears to run most quickly for observers who are in the same state of motion as the clock, and runs more slowly as perceived by observers who are moving relative to the clock.

Coordinate transformations

Speed relates to distance and time, so if the speed of light is the same in all frames of reference and time is distorted for different observers, presumably distance is distorted as well: otherwise the ratio of distance to time could not stay the same. Handling the two effects at the same time requires delicacy. Let's start with a couple of examples that are easier to visualize.

Rotation

For guidance, let's look at the mathematical treatment of a different part of the principle of relativity, the statement that the laws of physics are the same regardless of the orientation of the coordinate system. Suppose that two observers are in frames of reference that are at rest relative to each other, and they set up coordinate systems with their origins at the same point, but rotated by 90 degrees, as in figure (c). To go back and forth between the two systems, we can use the equations



(c) Two observers describe the same landscape with different coordinate systems.

$$\begin{aligned} x' &= y \\ y' &= -x \end{aligned}$$

A set of equations such as this one for changing from one system of coordinates to another is called a coordinate transformation, or just a transformation for short.

Similarly, if the coordinate systems differed by an angle of 5 degrees, we would have

$$\begin{aligned} x' &= (\cos 5°)\, x + (\sin 5°)\, y \\ y' &= (-\sin 5°)\, x + (\cos 5°)\, y \end{aligned}$$

Since cos 5°=0.997 is very close to one, and sin 5°=0.087 is close to zero, the rotation through a small angle has only a small effect, which makes sense. The equations for rotation are always of the form

$$\begin{aligned} x' &= (\text{constant \#1})\, x + (\text{constant \#2})\, y \\ y' &= (\text{constant \#3})\, x + (\text{constant \#4})\, y \quad . \end{aligned}$$

Galilean transformation for frames moving relative to each other

Einstein wanted to see if he could find a rule for changing between coordinate systems that were moving relative to each other. As a second warming-up example, let's look at the transformation between frames of reference in relative motion according to *Galilean* relativity, i.e. without any distortion of space and time. Suppose the x' axis is moving to the right at a speed v relative to the x axis. The transformation is simple:

$$\begin{aligned} x' &= x - vt \\ t' &= t \end{aligned}$$

Again we have an equation with constants multiplying the variables, but now the variables are distance and time. The interpretation of the $-vt$ term is the observer moving with the origin x' system sees a steady reduction in distance to an object on the right and at rest in the x system. In other words, the object appears to be moving according to the x' observer, but at rest according to x. The fact that the constant in front of x in the first equation equals one tells us that there is no distortion of space according to Galilean relativity, and similarly the second equation tells us there is no distortion of time.

Einstein's transformations for frames in relative motion

Guided by analogy, Einstein decided to look for a transformation between frames in relative motion that would have the form

$$\begin{aligned} x' &= Ax + Bt \\ t' &= Cx + Dt \quad . \end{aligned}$$

(Any form more complicated than this, for example equations including x^2 or t^2 terms, would violate the part of the principle of relativity that says the laws of physics are the same in all locations.) The constants A, B, C, and D would depend only on the relative velocity, v, of the two frames. Galilean relativity had been amply verified by experiment for values of v much less than the speed of light, so at low speeds we must have $A \approx 1$, $B \approx v$, $C \approx 0$, and $D \approx 1$. For high speeds, however, the constants A and D would start to become measurably different from 1, providing the distortions of time and space needed so that the speed of light would be the same in all frames of reference.

Self-Check

What units would the constants A, B, C, and D need to have?

Natural units

Despite the reputation for difficulty of Einstein's theories, the derivation of Einstein's transformations is fairly straightforward. The algebra, however, can appear more cumbersome than necessary unless we adopt a choice of units that is better adapted to relativity than the metric units of meters and seconds. The form of the transformation equations shows that time and

A relates distance to distance, so it is unitless, and similarly for *D*. Multiplying *B* by a time has to give a distance, so *B* has units of m/s. Multiplying *C* by distance has to give a time, so *C* has units of s/m.

space are not entirely separate entities. Life is easier if we adopt a new set of units:

Time is measured in *seconds*.

Distance is also measured in units of *seconds*. A distance of one second is how far light travels in one second of time.

In these units, the speed of light equals one by definition:

$$c = \frac{\text{1 second of distance}}{\text{1 second of time}} = 1$$

All velocities are represented by unitless numbers in this system, so for example v=0.5 would describe an object moving at half the speed of light.

Derivation of the transformations

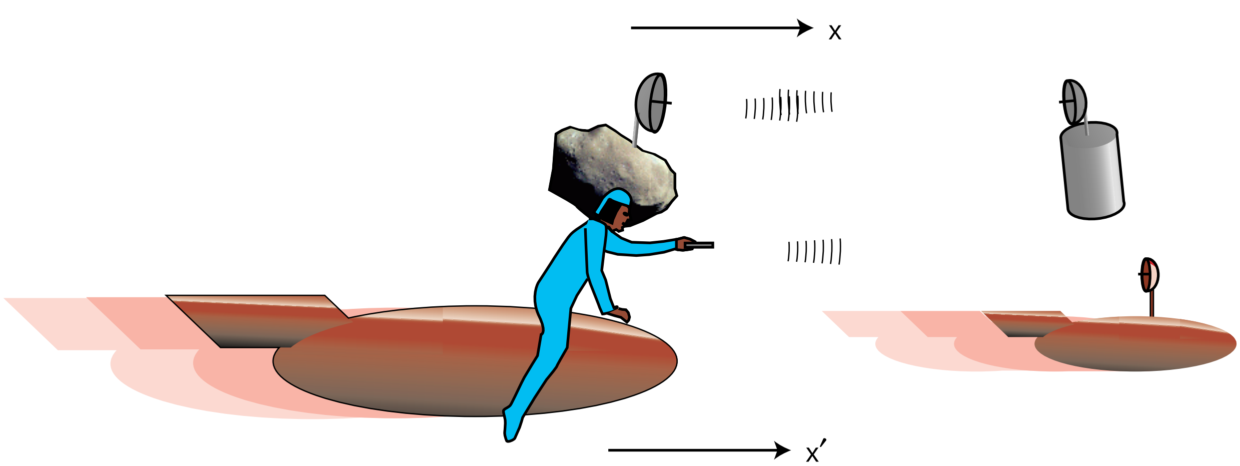
To find how the constants A, B, C, and D in the transformation equations

$$x' = Ax + Bt \qquad (1a)$$
$$t' = Cx + Dt \qquad (1b)$$

depend on velocity, we follow a strategy of relating the constants to one another by requiring that the transformation produce the right results in several different situations. By analogy, the rotation transformation for x and y coordinates has the same constants on the upper left and lower right, and the upper right and lower left constants are equal in absolute value but opposite in sign. We will look for similar rules for the frames-in-relative-motion transformations.

For vividness, we imagine that the x,t frame is defined by an asteroid at x=0, and the x',t' frame by a rocket ship at x'=0. The rocket ship is coasting at a constant speed v relative to the asteroid, and as it passes the asteroid they synchronize their clocks to read t=0 and t'=0.

We need to compare the perception of space and time by observers on the rocket and the asteroid, but this can be a bit tricky because our usual ideas about measurement contain hidden assumptions. If, for instance, we want to measure the length of a box, we imagine we can lay a ruler down on it, take in the scene visually, and take the measurement using the ruler's scale on the right side of the box while the left side of the box is simultaneously lined up with the butt of the ruler. The assumption that we can take in the whole scene at once with our eyes is, however, based on the



assumption that light travels with infinite speed to our eyes. Since we will be dealing with relative motion at speeds comparable to the speed of light, we have to spell out our methods of measuring distance.

We will therefore imagine an explicit procedure for the asteroid and the rocket pilot to make their distance measurements: they send electromagnetic signals (light or radio waves) back and forth to their own remote stations. For instance the asteroid's station will send it a message to tell it the time at which the rocket went by. The asteroid's station is at rest with respect to the asteroid, and the rocket's is at rest with respect to the rocket (and therefore in motion with respect to the asteroid).

The measurement of time is likewise fraught with danger if we are careless, which is why we have had to spell out procedures for the synchronization of clocks between the asteroid and the rocket. The asteroid must also synchronize its clock with its remote stations's clock by adjusting them until flashes of light released by both the asteroid and its station at equal clock readings are received on the opposite sides at equal clock readings. The rocket pilot must go through the same kind of synchronization procedure with her remote station.

Rocket's motion as seen by the asteroid

The origin of the rocket's x',t' frame is defined by the rocket itself, so the rocket always has $x'=0$. Let the asteroid's remote station be at position x in the asteroid's frame. The asteroid sees the rocket travel at speed v, so the asteroid's remote station sees the rocket pass it when x equals vt. Equation (1a) becomes $0=Avt+Bt$, which implies a relationship between A and B: $B/A=-v$. (In the Galilean version, we had $B=-v$ and $A=1$.) This restricts the transformation to the form

$$x' = Ax - Avt \qquad (2a)$$
$$t' = Cx + Dt \qquad (2b)$$

Asteroid's motion as seen by the rocket

Straightforward algebra can be used to reverse the transformation equations so that they give x and t in terms of x' and t'. The result for x is $x=(Dx'-Bt')/(AD-BC)$. The asteroid's frame of reference has its origin defined by the asteroid itself, so the asteroid is always at $x=0$. In the rocket's frame, the asteroid falls behind according to the equation $x'=-vt'$, and substituting this into the equation for x gives $0=(Dvt'-Bt')/(AD-BC)$. This requires us to have $B/D=-v$, i.e. D must be the same as A:

$$x' = Ax - Avt \qquad (3a)$$
$$t' = Cx + At \qquad (3b)$$

Agreement on the speed of light

Suppose the rocket pilot releases a flash of light in the forward direction as she passes the asteroid at $t=t'=0$. As seen in the asteroid's frame, we might expect this pulse to travel forward faster than normal because it was emitted by the moving rocket, but the principle of relativity tells us this is not so. The flash reaches the asteroid's remote station when x equals ct, and since we are working in natural units, this is equivalent to $x=t$. The speed of light must be the same in the rocket's frame, so we must also have $x'=t'$ when the flash gets there. Setting equations (3a) and (3b) equal to each other and substituting in $x=t$, we find $A-Av=C+A$, so we must have $C=-Av$:

$$x' = Ax - Avt \quad (4a)$$
$$t' = -Avx + At \quad (4b)$$

We have now determined the whole form of the transformation except for an overall multiplicative constant *A*.

Reversal of velocity

We can tie down this last unknown by considering what would have happened if the velocity of the rocket had been reversed. This would be entirely equivalent to reversing the direction of time, like playing a movie backwards, and it would also be equivalent to interchanging the roles of the rocket and the asteroid, since the rocket pilot sees the asteroid moving away from her to the left. The reversed transformation from the x',t' system to the x,t system must therefore be the one obtained by reversing the signs of t and t':

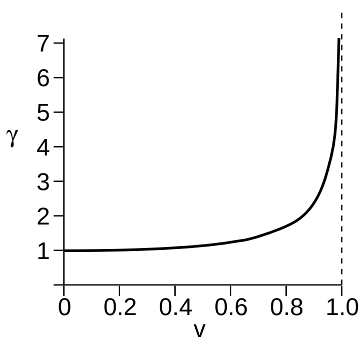
$$x = Ax' + Avt' \quad (5a)$$
$$-t = -Avx' - At' \quad (5b)$$

We now substitute equations 4a and 4b into equation 5a to eliminate x' and t', leaving only x and t:

$$x = A(Ax-Avt) + Av(-Avx+At)$$

The t terms cancel out, and collecting the x terms we find

$$x = A^2(1-v^2)x \quad ,$$



which requires $A^2(1-v^2)=1$, or $A=1/\sqrt{1-v^2}$. Since this factor occurs so often, we give it a special symbol, γ, the Greek letter gamma,

$$\gamma = \frac{1}{\sqrt{1-v^2}} \qquad \text{[definition of the } \gamma \text{ factor]}$$

Its behavior is shown in the graph on the left.

We have now arrived at the correct relativistic equation for transforming between frames in relative motion. For completeness, I will include, without proof, the trivial transformations of the y and z coordinates.

$$\begin{aligned} x' &= \gamma x - \gamma v t \\ t' &= -\gamma v x + \gamma t \\ y' &= y \\ z' &= z \end{aligned}$$

[transformation between frames in relative motion; v is the velocity of the x' frame relative to the x frame; the origins of the frames are assumed to have coincided at $x=x'=0$ and $t=t'=0$]

Self-Check

What is γ when v=0? Interpret the transformation equations in the case of v=0.

Discussion Question

A. If you were in a spaceship traveling at the speed of light (or extremely close to the speed of light), would you be able to see yourself in a mirror?

B. A person in a spaceship moving at 99.99999999% of the speed of light relative to Earth shines a flashlight forward through dusty air, so the beam is visible. What does she see? What would it look like to an observer on Earth?

Looking at the definition of γ, we see that γ=1 when v=0. The transformation equations then reduce to $x'=x$ and $t'=t$, which makes sense.

1.3 Applications

We now turn to the subversive interpretations of these equations.

Nothing can go faster than the speed of light.

Remember that these equations are expressed in natural units, so v=0.1 means motion at 10% of the speed of light, and so on. What happens if we want to send a rocket ship off at, say, twice the speed of light, v=2? Then γ will be $1/\sqrt{-3}$. But your math teacher has always cautioned you about the severe penalties for taking the square root of a negative number. The result would be physically meaningless, so we conclude that no object can travel faster than the speed of light. Even travel exactly at the speed of light appears to be ruled out for material objects, since then γ would be infinite.

Einstein had therefore found a solution to his original paradox about riding on a motorcycle alongside a beam of light, resulting in a violation of Maxwell's theory of electromagnetism. The paradox is resolved because it is impossible for the motorcycle to travel at the speed of light.

Most people, when told that nothing can go faster than the speed of light, immediately begin to imagine methods of violating the rule. For instance, it would seem that by applying a constant force to an object for a long time, we would give it a constant acceleration which would eventually result in its traveling faster than the speed of light. We will take up these issues in section 2.2.

No absolute time

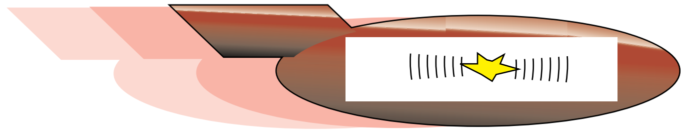
The fact that the equation for time is not just $t'=t$ tells us we're not in Kansas anymore — Newton's concept of absolute time is dead. One way of understanding this is to think about the steps described for synchronizing the four clocks:

(1) The asteroid's clock — call it A1 — was synchronized with the clock on its remote station, A2.

(2) The rocket pilot synchronized her clock, R1, with A1, at the moment when she passed the asteroid.

(3) The clock on the rocket's remote station, R2, was synchronized with R1.

Now if A2 matches A1, A1 matches R1, and R1 matches R2, we would expect A2 to match R2. This cannot be so, however. The rocket pilot released a flash of light as she passed the asteroid. In the asteroid's frame of reference, that light had to travel the full distance to the asteroid's remote station before it could be picked up there. In the rocket pilot's frame of reference, however, the asteroid's remote station is rushing at her, perhaps at a sizeable fraction of the speed of light, so the flash has less distance to travel before the asteroid's station meets it. Suppose the rocket pilot sets things up so that R2 has just enough of a head start on the light flash to reach A2 at the same time the flash of light gets there. Clocks A2 and R2 cannot agree, because the time required for the light flash to get there was different in the two frames. Thus, two clocks that were initially in agreement will disagree later on.

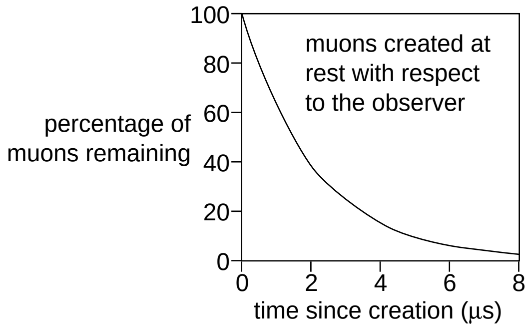


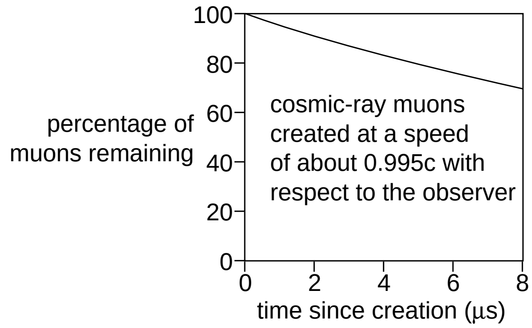
No simultaneity

Part of the concept of absolute time was the assumption that it was valid to say things like, "I wonder what my uncle in Beijing is doing right now." In the nonrelativistic world-view, clocks in Los Angeles and Beijing could be synchronized and stay synchronized, so we could unambiguously define the concept of things happening simultaneously in different places. It is easy to find examples, however, where events that seem to be simultaneous in one frame of reference are not simultaneous in another frame. In the figure above, a flash of light is set off in the center of the rocket's cargo hold. According to a passenger on the rocket, the flashes have equal distances to travel to reach the front and back walls, so they get there simultaneously. But an outside observer who sees the rocket cruising by at high speed will see the flash hit the back wall first, because the wall is rushing up to meet it, and the forward-going part of the flash hit the front wall later, because the wall was running away from it. Only when the relative velocity of two frames is small compared to the speed of light will observers in those frames agree on the simultaneity of events.

Time dilation

Let's compare the rate at which time passes in two frames. A clock that stays on the asteroid will always have $x=0$, so the time transformation equation $t'=-v\gamma x+\gamma t$ becomes simply $t'=\gamma t$. If the rocket pilot monitors the ticking of a clock on the asteroid via radio (and corrects for the increasingly long delay for the radio signals to reach her as she gets farther away from it), she will find that the rate of increase of the time t' on her wristwatch is always greater than the rate at which the time t measured by the asteroid's clock increases. It will seem to her that the asteroid's clock is running too slowly by a factor of γ. This is known as the time dilation effect: clocks seem to run fastest when they are at rest relative to the observer, and more slowly when they are in motion. The situation is entirely symmetric: to people on the asteroid, it will appear that the rocket pilot's clock is the one that is running too slowly.





Example: Cosmic-ray muons

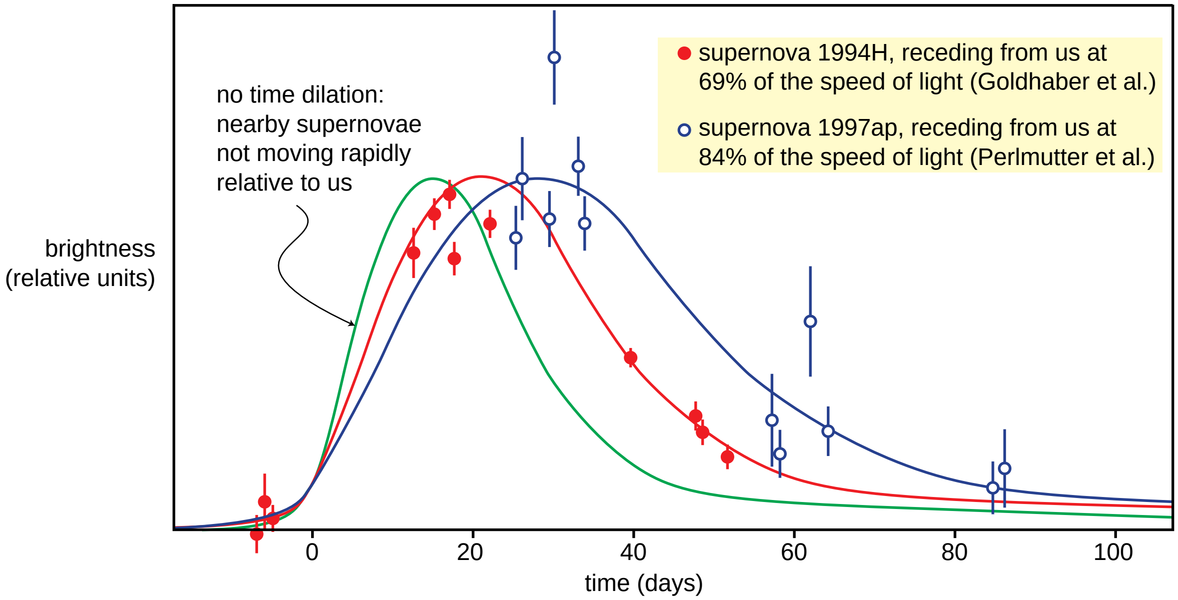
Cosmic rays are protons and other atomic nuclei from outer space. When a cosmic ray happens to come the way of our planet, the first earth-matter it encounters is an air molecule in the upper atmosphere. This collision then creates a shower of particles that cascade downward and can often be detected at the earth's surface. One of the more exotic particles created in these cosmic ray showers is the muon (named after the Greek letter mu, μ). The reason muons are not a normal part of our environment is that a muon is radioactive, lasting only 2.2 microseconds on the average before changing itself into an electron and two neutrinos. A muon can therefore be used as a sort of clock, albeit a self-destructing and somewhat random one! The graphs above show the average rate at which a sample of muons decays, first for muons created at rest and then for high-velocity muons created in cosmic-ray showers. The second graph is found experimentally to be stretched out by a factor of about ten, which matches well with the prediction of relativity theory:

$$\begin{aligned}\gamma &= 1 / \sqrt{1-v^2} \\ &= 1 / \sqrt{1-0.995^2} \\ &\approx 10\end{aligned}$$

Since a muon takes many microseconds to pass through the atmosphere, the result is a marked increase in the number of muons that reach the surface.

Example: Time dilation for objects larger than the atomic scale

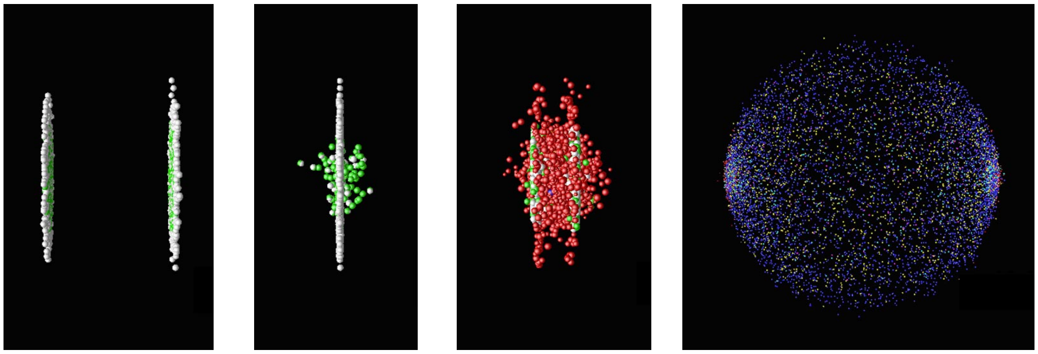
Our world is (fortunately) not full of human-scale objects moving at significant speeds compared to the speed of light. For this reason, it took over 80 years after Einstein's theory was published before anyone could come up with a conclusive example of drastic time dilation that wasn't confined to cosmic rays or particle accelerators. Recently, however, astronomers have found definitive proof that entire stars undergo time dilation. The universe is expanding in the aftermath of the Big Bang, so in general everything in the universe is getting farther away from everything else. One need only find an astronomical process that takes a standard amount of time, and then observe how long it appears to take when it occurs in a part of the universe that is receding from us rapidly. A type of exploding star called a type Ia supernova fills the bill, and technology is now sufficiently advanced to allow them to be detected across vast distances. The graph on the following page shows convincing evidence for time dilation in the brightening and dimming of two distant supernovae.



The twin paradox

A natural source of confusion in understanding the time-dilation effect is summed up in the so-called twin paradox, which is not really a paradox. Suppose there are two teenaged twins, and one stays at home on earth while the other goes on a round trip in a spaceship at relativistic speeds (i.e. speeds comparable to the speed of light, for which the effects predicted by the theory of relativity are important). When the traveling twin gets home, he has aged only a few years, while his brother is now old and gray. (Robert Heinlein even wrote a science fiction novel on this topic, although it is not one of his better stories.)

The paradox arises from an incorrect application of the theory of relativity to a description of the story from the traveling twin's point of view. From his point of view, the argument goes, his homebody brother is the one who travels backward on the receding earth, and then returns as the earth approaches the spaceship again, while in the frame of reference fixed to the spaceship, the astronaut twin is not moving at all. It would then seem that the twin on earth is the one whose biological clock should tick more slowly, not the one on the spaceship. The flaw in the reasoning is that the principle of relativity only applies to frames that are in motion at constant velocity relative to one another, i.e. inertial frames of reference. The astronaut twin's frame of reference, however, is noninertial, because his spaceship must accelerate when it leaves, decelerate when it reaches its destination, and then repeat the whole process again on the way home. What we have been studying is Einstein's special theory of relativity, which describes motion at constant velocity. To understand accelerated motion we would need the general theory of relativity (which is also a theory of gravity). A correct treatment using the general theory shows that it is indeed the traveling twin who is younger when they are reunited.



Length contraction

The treatment of space and time in the transformation between frames is entirely symmetric, so distance intervals as well as time intervals must be reduced by a factor of γ for an object in a moving frame. The figure above shows a an artist's rendering of this effect for the collision of two gold nuclei at relativistic speeds in the RHIC accelerator in Long Island, New York, scheduled to come on line in 2000. The gold nuclei would appear nearly spherical (or just slightly lengthened like an American football) in frames moving along with them, but in the laboratory's frame, they both appear drastically foreshortened as they approach the point of collision. The later pictures show the nuclei merging to form a hot soup, in which experimenters hope to observe a new form of matter.

Perhaps the most famous of all the so-called relativity paradoxes involves the length contraction. The idea is that one could take a schoolbus and drive it at relativistic speeds into a garage of ordinary size, in which it normally would not fit. Because of the length contraction, the bus would supposedly fit in the garage. The paradox arises when we shut the door and then quickly slam on the brakes of the bus. An observer in the garage's frame of reference will claim that the bus fit in the garage because of its contracted length. The driver, however, will perceive the garage as being contracted and thus even less able to contain the bus than it would normally be. The paradox is resolved when we recognize that the concept of fitting the bus in the garage "all at once" contains a hidden assumption, the assumption that it makes sense to ask whether the front and back of the bus can simultaneously be in the garage. Observers in different frames of reference moving at high relative speeds do not necessarily agree on whether things happen simultaneously. The person in the garage's frame can shut the door at an instant he perceives to be simultaneous with the front bumper's arrival at the opposite wall of the garage, but the driver would not agree about the simultaneity of these two events, and would perceive the door as having shut long after she plowed through the back wall.

Discussion Questions

A. A question that students often struggle with is whether time and space can really be distorted, or whether it just seems that way. Compare with optical illusions are magic tricks. How do we know that these illusions are not real? Are relativistic effects the same or not?

B. On a spaceship moving at relativistic speeds, would a lecture seem even longer and more boring than normal?

C. Mechanical clocks can be affected by motion. For example, it was a significant technological achievement to build a clock that could sail aboard a ship and still keep accurate time, allowing longitude to be determined. How is this similar to or different from relativistic time dilation?

D. What would the shapes of the two nuclei in the RHIC experiment look like to a microscopic observer riding on the left-hand nucleus? To an observer riding on the right-hand one? Can they agree on what is happening? If not, why not — after all, shouldn't they see the same thing if they both compare the two nuclei side-by-side at the same instant in time?

E. If you stick a piece of foam rubber out the window of your car while driving down the freeway, the wind may compress it a little. Does it make sense to interpret the relativistic length contraction as a type of strain that pushes an object's atoms together like this? How does this relate to the previous discussion question?

Summary

Selected Vocabulary

transformation .... the mathematical relationship between the variables such as x and t, as observed in different frames of reference

Terminology Used in Some Other Books

Lorentz transformation ...... the transformation between frames in relative motion

Notation

γ .... an abbreviation for $1/\sqrt{1-v^2}$

Summary

Einstein's principle of relativity states that both light and matter obey the same laws of physics in any inertial frame of reference, regardless of the frame's orientation, position, or constant-velocity motion. Maxwell's equations are the basic laws of physics governing light, and Maxwell's equations predict a specific value for the speed of light, c=3.0x10^8 m/s, so this new principle implies that *the speed of light must be the same in all frames of reference*, even when it seems intuitively that this is impossible because the frames are in relative motion. This strange constancy of the speed of light was experimentally supported by the 1887 Michelson-Morley experiment. Based only on this principle, Einstein showed that time and space as seen by one observer would be distorted compared to another observer's perceptions if they were moving relative to each other. This distortion is spelled out in the transformation equations:

$$\begin{aligned} x' &= \gamma x - \gamma v t \\ t' &= -\gamma v x + \gamma t \end{aligned} ,$$

where v is the velocity of the x',t' frame with respect to the x,t frame, and γ is an abbreviation for $1/\sqrt{1-v^2}$. Here, as throughout the chapter, we use the natural system of units in which the speed of light equals 1 by definition, and both times and distances are measured in units of seconds. One second of distance is how far light travels in one second. To change natural-unit equations back to metric units, we must multiply terms by factors of c as necessary in order to make the units of all the terms on both sides of the equation come out right.

Some of the main implications of these equations are:

(1) Nothing can move faster than the speed of light.

(2) The size of a moving object is shrunk. An object appears longest to an observer in a frame moving along with it (a frame in which the object appears is at rest).

(3) Moving clocks run more slowly. A clock appears to run fastest to an observer in a frame moving along with it (a frame in which the object appears is at rest).

(4) There is no well-defined concept of simultaneity for events occurring at different points in space.

Homework Problems

1.(a) Reexpress the transformation equations for frames in relative motion using ordinary units where $c \neq 1$. (b) Show that for speeds that are small compared to the speed of light, they are identical to the Galilean equations.

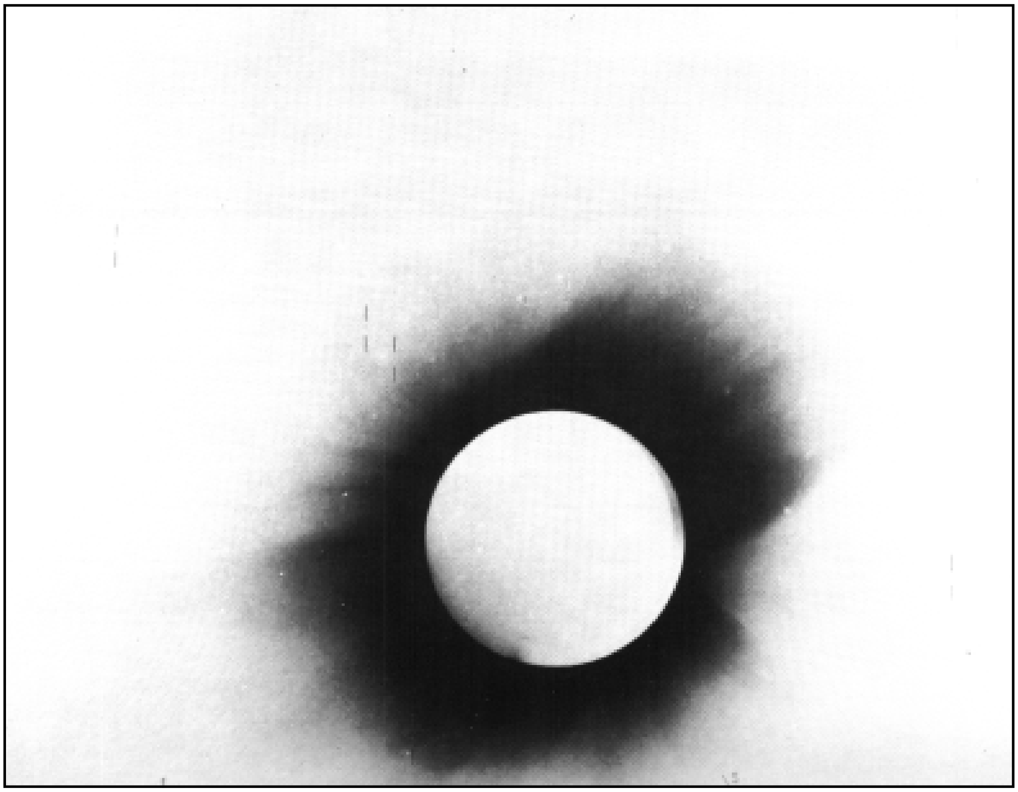
2. Atomic clocks can have accuracies of better than one part in 10^{13}. How does this compare with the time dilation effect produced if the clock takes a trip aboard a jet moving at 300 m/s? Would the effect be measurable? [Hint: Your calculator will round γ off to one. Use the low-velocity approximation $\gamma=1+v^2/2c^2$, which will be derived in chapter 2.]

3. (a) Find an expression for v in terms of γ in natural units. (b) Show that for very large values of γ, v gets close to the speed of light.

4 ★. Of the systems we ordinarily use to transmit information, the fastest ones — radio, television, phone conversations carried over fiber-optic cables — use light. Nevertheless, we might wonder whether it is possible to transmit information at speeds greater than c. The purpose of this problem is to show that if this was possible, then special relativity would have problems with *causality*, the principle that the cause should come earlier in time than the effect. Suppose an event happens at position and time x_1 and t_1 which causes some result at x_2 and t_2. Show that if the distance between x_1 and x_2 is greater than the distance light could cover in the time between t_1 and t_2, then there exists a frame of reference in which the event at x_2 and t_2 occurs *before* the one at x_1 and t_1.

5 ★. Suppose one event occurs at x_1 and t_1 and another at x_2 and t_2. These events are said to have a *spacelike* relationship to each other if the distance between x_1 and x_2 is greater than the distance light could cover in the time between t_1 and t_2, *timelike* if the time between t_1 and t_2 is greater than the time light would need to cover the distance between x_1 and x_2, and *lightlike* if the distance between x_1 and x_2 is the distance light could travel between t_1 and t_2. Show that spacelike relationships between events remain spacelike regardless of what coordinate system we transform to, and likewise for the other two categories. [It may be most elegant to do problem 9 from ch. 2 first and then use that result to solve this problem.]

S A solution is given in the back of the book.
✓ A computerized answer check is available.
★ A difficult problem.
∫ A problem that requires calculus.



Einstein's famous equation $E=mc^2$ states that mass and energy are equivalent. The energy of a beam of light is equivalent to a certain amount of mass, and the beam is therefore deflected by a gravitational field. Einstein's prediction of this effect was verified in 1919 by astronomers who photographed stars in the dark sky surrounding the sun during an eclipse. (This is a photographic negative, so the circle that appears bright is actually the dark face of the moon, and the dark area is really the bright corona of the sun.) The stars, marked by lines above and below them, appeared at positions slightly different than their normal ones, indicating that their light had been bent by the sun's gravity on its way to our planet.

2 Relativity, Part II

So far we have said nothing about how to predict motion in relativity. Do Newton's laws still work? Do conservation laws still apply? The answer is yes, but many of the definitions need to be modified, and certain entirely new phenomena occur, such as the conversion of mass to energy and energy to mass, as described by the famous equation $E=mc^2$. To cut down on the level of mathematical detail, I have relegated most of the derivations to optional section 2.6, presenting mainly the results and their physical explanations in this section.

2.1 Invariants

The discussion has the potential to become very confusing very quickly because some quantities, force for example, are perceived differently by observers in different frames, whereas in Galilean relativity they were the same in all frames of reference. To clear the smoke it will be helpful to start by identifying quantities that we can depend on *not* to be different in different frames. We have already seen how the principle of relativity requires that the speed of light is the same in all frames of reference. We say that c is invariant.

Another important invariant is mass. This makes sense, because the principle of relativity states that physics works the same in all reference frames. The mass of an electron, for instance, is the same everywhere in the universe, so its numerical value is one of the basic laws of physics. We should therefore expect it to be the same in all frames of reference as well. (Just to make things more confusing, about 50% of all books say mass is invariant, while 50% describe it as changing. It is possible to construct a self-consistent framework of physics according to either description. Neither way is right or wrong, the two philosophies just require different sets of definitions of quantities like momentum and so on. For what it's worth, Einstein eventually weighed in on the mass-as-an-invariant side of the argument. The main thing is just to be consistent.)

A second invariant is electrical charge. This has been verified to high precision because experiments show that an electric field does not produce any measurable force on a hydrogen atom. If charge varied with speed, then the electron, typically orbiting at about 1% of the speed of light, would not exactly cancel the charge of the proton, and the hydrogen atom would have a net charge.

2.2 Combination of Velocities

The impossibility of motion faster than light is the single most radical difference between relativistic and nonrelativistic physics, and we can get at most of the issues in this chapter by considering the flaws in various plans for going faster than light. The simplest argument of this kind is as follows. Suppose Janet takes a trip in a spaceship, and accelerates until she is moving at v=0.9 (90% of the speed of light in natural units) relative to the earth. She then launches a space probe in the forward direction at a speed u=0.2 relative to her ship. Isn't the probe then moving at a velocity of 1.1 times the speed of light relative to the earth?

The problem with this line of reasoning is that the distance covered by the probe in a certain amount of time is shorter as seen by an observer in the earthbound frame of reference, due to length contraction. Velocities are therefore combined not by simple addition but by a more complex method, which we derive in section 2.6 by performing two transformations in a row. In our example, the first transformation would be from the earth's frame to Janet's, the second from Janet's to the probe's. The result is

$$v_{\text{combined}} = \frac{u+v}{1+uv} \quad . \quad \text{[relativistic combination of velocities]}$$

Example: Janet's probe

Applying the equation to Janet's probe, we find

$$v_{\text{combined}} = \frac{0.9+0.2}{1+(0.9)(0.2)} = 0.93 \quad ,$$

so it is still going quite a bit slower than the speed of light

Example: Combination of velocities in unnatural units

In a system of units, like the metric system, with $c \neq 1$, all our symbols for velocity should be replaced with velocities divided by c, so we have

$$\frac{v_{\text{combined}}}{c} = \frac{\frac{u}{c}+\frac{v}{c}}{1+\left(\frac{u}{c}\right)\left(\frac{v}{c}\right)} \quad ,$$

or

$$v_{\text{combined}} = \frac{u+v}{1+uv/c^2} \quad .$$

When u and v are both much less than the speed of light, the quantity uv/c^2 is very close to zero, and we recover the nonrelativistic approximation, $v_{\text{combined}}=u+v$.

The line of reasoning given in the second example shows the correspondence principle at work: when a new scientific theory replaces an old one, the two theories must agree within their common realm of applicability.

2.3 Momentum and Force

Momentum

We begin our discussion of relativistic momentum with another scheme for going faster than light. Imagine that a freight train moving at a velocity of 0.6 (v=0.6c in unnatural units) strikes a ping-pong ball that is initially at rest, and suppose that in this collision no kinetic energy is converted into other forms such as heat and sound. We can easily prove based on conservation of momentum that in a very unequal collision of this kind, the smaller object flies off with double the velocity with which it was hit. (This is because the center of mass frame of reference is essentially the same as the frame tied to the freight train, and in the center of mass frame both objects must reverse their initial momenta.) So doesn't the ping-pong ball fly off with a velocity of 1.2, i.e. 20% faster than the speed of light?

The answer is that since $p=mv$ led to this contradiction with the structure of relativity, $p=mv$ must not be the correct equation for relativistic momentum. Apparently $p=mv$ is only a low-velocity approximation to the correct relativistic result. We need to find a new expression for momentum that agrees approximately with $p=mv$ at low velocities, and that also agrees with the principle of relativity, so that if the law of conservation of momentum holds in one frame of reference, it also is obeyed in every other frame. A proof is given in section 2.6 that such an equation is

$$p = m\gamma v \quad , \qquad \text{[relativistic equation for momentum]}$$

which differs from the nonrelativistic version only by the factor of γ. At low velocities γ is very close to 1, so $p=mv$ is approximately true, in agreement with the correspondence principle. At velocities close to the speed of light, γ approaches infinity, and so an object would need infinite momentum to reach the speed of light.

Force

What happens if you keep applying a constant force to an object, causing it to accelerate at a constant rate until it exceeds the speed of light? The hidden assumption here is that Newton's second law, $a=F/m$, is still true. It isn't. Experiments show that at speeds comparable to the speed of light, $a=F/m$ is wrong. The equation that still *is* true is

$$F = \frac{\Delta p}{\Delta t} \quad .$$

You could apply a constant force to an object forever, increasing its momentum at a steady rate, but as the momentum approached infinity, the velocity would approach the speed of light. In general, a force produces an acceleration significantly *less* than F/m at relativistic speeds.

Would passengers on a spaceship moving close to the speed of light perceive every object as being more difficult to accelerate, as if it was more massive? No, because then they would be able to detect a change in the laws of physics because of their state of motion, which would violate the principle of relativity. The way out of this difficulty is to realize that force is not an invariant. What the passengers perceive as a small force causing a small change in momentum would look to a person in the earth's frame of reference like a large force causing a large change in momentum. As a practical matter, conservation laws are usually more convenient tools for relativistic problem solving than procedures based on the force concept.

2.4 Kinetic Energy

Since kinetic energy equals $\frac{1}{2}mv^2$, wouldn't a sufficient amount of energy cause v to exceed the speed of light? You're on to my methods by now, so you know this is motivation for a redefinition of kinetic energy. Section 2.6 derives the work-kinetic energy theorem using the correct relativistic treatment of force. The result is

$$KE = m(\gamma-1) \quad . \quad \text{[relativistic kinetic energy]}$$

Since γ approaches infinity as velocity approaches the speed of light, an infinite amount of energy would be required in order to make an object move at the speed of light.

Example: Kinetic energy in unnatural units

How can this equation be converted back into units in which the speed of light does not equal one? One approach would be to redo the derivation in section 2.6 in unnatural units. A far simpler approach is simply to add factors of c where necessary to make the metric units look consistent. Suppose we decide to modify the right side in order to make its units consistent with the energy units on the left. The ordinary nonrelativistic definition of kinetic energy as $\frac{1}{2}mv^2$ shows that the units on the left are

$$\text{kg}\cdot\frac{\text{m}^2}{\text{s}^2} \quad .$$

The factor of γ–1 is unitless, so the mass units on the right need to be multiplied by m^2/s^2 to agree with the left. This means that we need to multiply the right side by c^2:

$$KE = mc^2(\gamma-1)$$

This is beginning to resemble the famous $E=mc^2$ equation, which we will soon attack head-on.

Example: The correspondence principle for kinetic energy

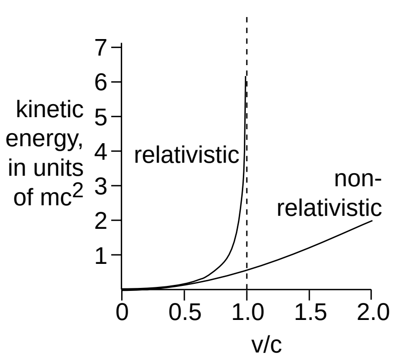
It is far from obvious that this result, even in its metric-unit form, reduces to the familiar $\frac{1}{2}mv^2$ at low speeds, as required by the correspondence principle. To show this, we need to find a low-velocity approximation for γ. In metric units, the equation for γ reads as

$$\gamma = \frac{1}{\sqrt{1-v^2/c^2}} \quad .$$

Reexpressing this as $\left(1-v^2/c^2\right)^{-1/2}$, and making use of the approximation $\left(1+\varepsilon\right)^p \approx 1+p\varepsilon$ for small ε, the equation for gamma becomes

$$\gamma \approx 1+\frac{v^2}{2c^2} \quad ,$$

which can readily be used to show $mc^2(\gamma-1)\approx\frac{1}{2}mv^2$.



The Large Hadron Collider. The red circle shows the location of the underground tunnel which the LHC will share with a preexisting accelerator.

Example: the large hadron collider

Question: The Large Hadron Collider (LHC), being built in Switzerland, is a ring with a radius of 4.3 km, designed to accelerate two counterrotating beams of protons to energies of 7 TeV per proton. (The word "hadron" refers to any particle that participates in strong nuclear forces.) The TeV is a unit of energy equal to 10^{12} eV, where 1 eV=$1.60\text{x}10^{-19}$ J is the energy a particle with unit charge acquires by moving through a voltage difference of 1 V.) The ring has to be so big because the inward force from the accelerator's magnets would not be great enough to make the protons curve more tightly at top speed.

(a) What inward force must be exerted on each proton?

(b) In a purely Newtonian world where there were no relativistic effects, how much smaller could the LHC be if it was to produce proton beams moving at speeds close to the speed of light?

Solution:

(a) Since the protons have velocity vectors with constant magnitudes, γ is constant, so let's start by computing it. We'll work the whole problem in mks, since none of the data are given in natural units. The kinetic energy of each proton is

$$\begin{aligned} KE &= 7\ \text{TeV} \\ &= (7\ \text{TeV})(10^{12}\ \text{eV/TeV})(1.60\text{x}10^{-19}\ \text{J/eV}) \\ &= 1.1\text{x}10^{-6}\ \text{J} \quad . \end{aligned}$$

A microjoule is quite a healthy energy for a subatomic particle! Looking up the mass of a proton, we have

$$\begin{aligned} mc^2 &= (1.7\text{x}10^{-27}\ \text{kg})(3.0\text{x}10^{8}\ \text{m/s})^2 \\ &= 1.5\text{x}10^{-10}\ \text{J} \quad . \end{aligned}$$

The kinetic energy is thousands of times greater than mc^2, so the protons go very close to the speed of light. Under these conditions there is no significant difference between γ and γ–1, so

$$\begin{aligned} \gamma &\approx KE\ /\ mc^2 \\ &= 7.3\text{x}10^3 \end{aligned}$$

We analyze the circular motion in the laboratory frame of reference, since that is the frame of reference in which the LHC's magnets sit, and their fields were calibrated by instruments at rest with respect to them. The inward force required is

$$\begin{aligned} \mathbf{F} &= \Delta\mathbf{p}/\Delta t \\ &= \Delta(m\gamma\mathbf{v})/\Delta t \\ &= m\,\gamma\,\Delta\mathbf{v}/\Delta t \\ &= m\,\gamma\,\mathbf{a} \quad . \end{aligned}$$

Except for the factor of γ, this is the same result we would have had in Newtonian physics, where we already know the equation $a=v^2/r$ for the inward acceleration in uniform circular motion. Since the velocity is essentially the speed of light, we have $a=c^2/r$. The force required is

$$\begin{aligned} \text{F} &= m\,\gamma\,c^2/r \\ &= KE\ /\ r \quad . \qquad [\text{since } \gamma \approx \gamma\text{–}1] \end{aligned}$$

This looks a little funny, but the units check out, since a joule is the same as a newton-meter. The result is

$$\text{F} = 2.6\text{x}10^{-10}\ \text{N}$$

(b)

$$\begin{aligned} \text{F} &= mv^2/r \qquad [\text{nonrelativistic equation}] \\ &= mc^2/r \\ r &= mc^2/F \\ &= 0.59\ \text{m} \end{aligned}$$

In a nonrelativistic world, it would be a table-top accelerator! The energies and momenta, however, would be smaller.

2.5 Equivalence of Mass and Energy

The treatment of relativity so far has been purely mechanical, so the only form of energy we have discussed is kinetic. For example, the storyline for the introduction of relativistic momentum was based on collisions in which no kinetic energy was converted to other forms. We know, however, that collisions can result in the production of heat, which is a form of kinetic energy at the molecular level, or the conversion of kinetic energy into entirely different forms of energy, such as light or potential energy.

Let's consider what happens if a blob of putty moving at velocity v hits another blob that is initially at rest, sticking to it, and that as much kinetic energy as possible is converted into heat. (It is not possible for all the KE to be converted to heat, because then conservation of momentum would be violated.) The nonrelativistic result is that to obey conservation of momentum the two blobs must fly off together at $v/2$.

Relativistically, however, an interesting thing happens. A hot object has more momentum than a cold object! This is because the relativistically correct expression for momentum is $p=m\gamma v$, and the more rapidly moving molecules in the hot object have higher values of γ. There is no such effect in nonrelativistic physics, because the velocities of the moving molecules are all in random directions, so the random motion's contribution to momentum cancels out.

In our collision, the final combined blob must therefore be moving a little more slowly than the expected $v/2$, since otherwise the final momentum would have been a little greater than the initial momentum. To an observer who believes in conservation of momentum and knows only about the overall motion of the objects and not about their heat content, the low velocity after the collision would have to seem to require a magical change in the mass, as if the mass of two combined, hot blobs of putty was more than the sum of their individual masses.

Heat energy is equivalent to mass.

Now we know that mass is invariant, and no molecules were created or destroyed, so the masses of all the molecules must be the same as they always were. The change is due to the change in γ with heating, not to a change in m. But how much does the mass *appear* to change? In section 2.6 we prove that the perceived change in mass exactly equals the change in heat energy between two temperatures, i.e. changing the heat energy by an amount E changes the effective mass of an object by E as well. This looks a bit odd because the natural units of energy and mass are the same. Converting back to ordinary units by our usual shortcut of introducing factors of c, we find that changing the heat energy by an amount E causes the apparent mass to change by $m=E/c^2$. Rearranging, we have the famous $E=mc^2$.

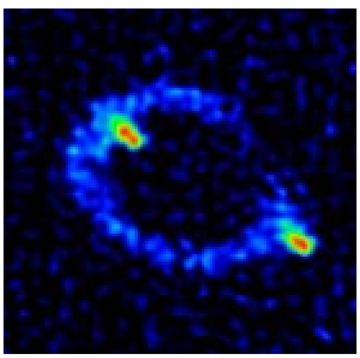
All energy is equivalent to mass.

But this whole argument was based on the fact that heat is a form of kinetic energy at the molecular level. Would $E=mc^2$ apply to other forms of energy as well? Suppose a rocket ship contains some electrical potential energy stored in a battery. If we believed that $E=mc^2$ applied to forms of kinetic energy but not to electrical potential energy, then we would have to expect that the pilot of the rocket could slow the ship down by using the battery to run a heater! This would not only be strange, but it would violate the principle of relativity, because the result of the experiment would be different depending on whether the ship was at rest or not. The only logical conclusion is that *all forms of energy are equivalent to mass*. Running the heater then has no effect on the motion of the ship, because the total energy in the ship was unchanged; one form of energy was simply converted to another.

Example: A rusting nail
Question: A 50-gram iron nail is left in a cup of water until it turns entirely to rust. The energy released is about 0.5 MJ (mega-joules). In theory, would a sufficiently precise scale register a change in mass? If so, how much?
Solution: The energy will appear as heat, which will be lost to the environment. So the total mass-energy of the cup, water, and iron will indeed be lessened by 0.5 MJ. (If it had been perfectly insulated, there would have been no change, since the heat energy would have been trapped in the cup.) Converting to mass units, we have

$$
\begin{aligned}
m &= E/c^2 \\
&= (0.5\text{x}10^6 \text{ J}) / (3.0\text{x}10^8 \text{ m/s})^2 \\
&= 6\text{x}10^{-12} \text{ J/(m}^2\text{/s}^2) \\
&= 6\text{x}10^{-12} \text{ (kg}\cdot\text{m}^2\text{/s}^2)/(\text{m}^2\text{/s}^2) \\
&= 6\text{x}10^{-12} \text{ kg} \quad ,
\end{aligned}
$$

so the change in mass is too small to measure with any practical technique. This is because the square of the speed of light is such a large number in metric units.



This telescope picture shows two images of the same distant object, an exotic, very luminous object called a quasar. This is interpreted as evidence that a massive, dark object, possibly a black hole, happens to be between us and it. Light rays that would otherwise have missed the earth on either side have been bent by the dark object's gravity so that they reach us. The actual direction to the quasar is presumably in the center of the image, but the light along that central line don't get to us because they are absorbed by the dark object. The quasar is known by its catalog number, MG1131+0456, or more informally as Einstein's Ring.

Energy participates in gravitational forces.

In the example we tacitly assumed that the increase in mass would show up on a scale, i.e. that its gravitational attraction with the earth would increase. Strictly speaking, however, we have only proven that energy relates to *inertial* mass, i.e. to phenomena like momentum and the resistance of an object to a change in its state of motion. Even before Einstein, however, experiments had shown to a high degree of precision that any two objects with the same inertial mass will also exhibit the same gravitational attractions, i.e. have the same *gravitational* mass. For example, the only reason that all objects fall with the same acceleration is that a more massive object's inertia is exactly in proportion to the greater gravitational forces in which it participates. We therefore conclude that energy participates in gravitational forces in the same way mass does. The total gravitational attraction between two objects is proportional not just to the product of their masses, m_1m_2, as in Newton's law of gravity, but to the quantity $(m_1+E_1)(m_2+E_2)$. (Even this modification does not give a complete, self-consistent theory of gravity, which is only accomplished through the general theory of relativity.)

Example: Gravity bending light
The first important experimental confirmation of relativity came when stars next to the sun during a solar eclipse were observed to have shifted a little from their ordinary position. (If there was no eclipse, the glare of the sun would prevent the stars from being observed.) Starlight had been deflected by gravity.

Example: Black holes
A star with sufficiently strong gravity can prevent light from leaving. Quite a few black holes have been detected via their gravitational forces on neighboring stars or clouds of dust.

Creation and destruction of particles

Since mass and energy are beginning to look like two sides of the same coin, it may not be so surprising that nature displays processes in which particles are actually destroyed or created; energy and mass are then converted back and forth on a wholesale basis. This means that in relativity there are no separate laws of conservation of energy and conservation of mass. There is only a law of conservation of mass plus energy (referred to as mass-energy). In natural units, $E+m$ is conserved, while in ordinary units the conserved quantity is $E+mc^2$.

Example: Electron-positron annihilation

Natural radioactivity in the earth produces positrons, which are like electrons but have the opposite charge. A form of antimatter, positrons annihilate with electrons to produce gamma rays, a form of high-frequency light. Such a process would have been considered impossible before Einstein, because conservation of mass and energy were believed to be separate principles, and the process eliminates 100% of the original mass. In metric units, the amount of energy produced by annihilating 1 kg of matter with 1 kg of antimatter is

$$\begin{aligned} E &= mc^2 \\ &= (2\text{ kg})(3.0\text{x}10^8\text{ m/s})^2 \\ &= 2\text{x}10^{17}\text{ J} \quad , \end{aligned}$$

which is on the same order of magnitude as a day's energy consumption for the entire world!

Positron annihilation forms the basis for the medical imaging procedure called a PET (positron emission tomography) scan, in which a positron-emitting chemical is injected into the patient and mapped by the emission of gamma rays from the parts of the body where it accumulates.

Note that the idea of mass as an invariant is separate from the idea that mass is not separately conserved. Invariance is the statement that all observers agree on a particle's mass regardless of their motion relative to the particle. Mass may be created or destroyed if particles are created or destroyed, and in such a situation mass invariance simply says that all observers will agree on how much mass was created or destroyed.

2.6* Proofs

Combination of velocities

We proceed by transforming from the x,t frame to the x',t' frame moving relative to it at a velocity v_1, and then from that from to a third frame, x'',t'', moving with respect to the second at v_2. The result must be equivalent to a single transformation from x,t to x'',t'' using the combined velocity. Transforming from x,t to x',t' gives

$$\begin{aligned} x' &= \gamma x - v_1\gamma_1 t \\ t' &= -v_1\gamma_1 x + \gamma_1 t \quad , \end{aligned}$$

and plugging this into the second transformation results in

$$\begin{aligned} x'' &= \gamma_2(\gamma x - v_1\gamma_1 t) - v_2\gamma_2(-v_1\gamma_1 x + \gamma_1 t) \\ t'' &= \ldots + \ldots \quad , \end{aligned}$$

where "…" indicates terms that we don't need to complete the derivation. Collecting terms gives

$$x'' = (\ldots)x - (v_1+v_2)\gamma_1\gamma_2 t \quad ,$$

where the coefficient of t, $-(v_1+v_2)\gamma_1\gamma_2$, must be the same as it would have been in a direct transformation from x,t to x'',t'':

$$-v_{\text{combined}}\gamma_{\text{combined}} = -(v_1+v_2)\gamma_1\gamma_2$$

Straightforward algebra then produces the equation in section 2.2.

Relativistic momentum

We want to show that if $p=m\gamma v$, then any collision that conserves momentum in the center of mass frame will also conserve momentum in any other frame. The whole thing is restricted to two-body collisions in one dimension in which no kinetic energy is changed to any other form, so it is not a general proof that $p=m\gamma v$ forms a consistent part of the theory of relativity. This is just the minimum test we want the equation to pass.

Let the new frame be moving at a velocity u with respect to the center of mass and let Γ (capital gamma) be $1/\sqrt{1-u^2}$. Then the total momentum in the new frame (at any moment before or after the collision) is

$$p' = m_1\gamma_1'v_1' + m_2\gamma_2'v_2' \quad .$$

The velocities v_1' and v_2' result from combining v_1 and v_2 with u, so making use of the result from the previous proof,

$$\begin{aligned} p' &= m_1(v_1+u)\Gamma\gamma_1 + m_2(v_2+u)\Gamma\gamma_2 \\ &= (m_1\gamma_1 v_1+m_2\gamma_2 v_2)\Gamma + (m_1\gamma_1+m_2\gamma_2)\Gamma u \\ &= p\Gamma + (KE_1+m_1+KE_2+m_2)\Gamma u \quad . \end{aligned}$$

If momentum is conserved in the center of mass frame, then there is no change in p, the momentum in the center of mass frame, after the collision. The first term is therefore the same before and after, and the second term is also the same before and after because mass is invariant, and we have assumed no KE was converted to other forms of energy. (We shouldn't expect the proof to work if KE is changed to other forms, because we have not taken into account the effects of any other forms of mass-energy.)

Relativistic work-kinetic energy theorem

This is a straightforward application of calculus, albeit with a couple of tricks to make it easier to do without recourse to a table of integrals. The kinetic energy of an object of mass m moving with velocity v equals the work done in accelerating it to that speed from rest:

$$
\begin{aligned}
KE &= \int_{v=0}^{v} F\,\mathrm{d}x \\
&= \int_{v=0}^{v} \frac{\mathrm{d}p}{\mathrm{d}t}\,\mathrm{d}x \\
&= \int_{v=0}^{v} \frac{\mathrm{d}(m\gamma v)}{\mathrm{d}t}\,\mathrm{d}x \\
&= m\int_{v=0}^{v} v\,\mathrm{d}(\gamma v) \\
&= m\int_{v=0}^{v} v^2\,\mathrm{d}\gamma + m\int_{v=0}^{v} v\gamma\,\mathrm{d}v \\
&= m\int_{v=0}^{v} \left(1-\gamma^{-2}\right)\mathrm{d}\gamma + m\int_{v=0}^{v} \frac{v\,\mathrm{d}v}{\sqrt{1-v^2}} \\
&= m\left(\gamma+\frac{1}{\gamma}\right)\Bigg]_{v=0}^{v} - m\sqrt{1-v^2}\,\Bigg]_{v=0}^{v} \\
&= m\left(\gamma+\frac{1}{\gamma}-\sqrt{1-v^2}\right)\Bigg]_{v=0}^{v} \\
&= m\gamma\Big]_{v=0}^{v} \quad = \quad m(\gamma-1)
\end{aligned}
$$

Change in inertia with heating

Technical note
This proof really only applies to an ideal gas, which expresses all of its heat energy as kinetic energy. In general heat energy is expressed partly as kinetic energy and partly as electrical potential energy.

We prove here that the inertia of a heated object (its apparent mass) increases by an amount equal to the heat. Suppose an object moving with velocity v_{cm} consists of molecules with masses m_1, m_2, ..., which are moving relative to the origin at velocities v_{o1}, v_{o2}, ... and relative to the object's center of mass at velocities v_1, v_2, ... The total momentum is

$$
\begin{aligned}
p_{total} &= m_1 v_{o1} \gamma_{o1} + \ldots \\
&= m_1(v_{cm}+v_1)\gamma_{cm}\gamma_1 + \ldots
\end{aligned}
$$

where we have used the result from the first subsection. Rearranging,

$$p_{total} = \gamma_{cm}\,[(m_1\gamma_1 v_{cm}+\ldots) + (m_1\gamma_1 v_1+\ldots)]$$

The second term, which is the total momentum in the c.m. frame, vanishes.

$$p_{total} = (m_1\gamma_1+\ldots)\gamma_{cm} v_{cm}$$

The quantity in parentheses is the total mass plus the total thermal energy.

Summary

Selected Vocabulary

invariant .... a quantity that does not change when transformed

Terminology Used in Some Other Books

rest mass .... referred to as mass in this book; written as m_0 in some books

mass .... What some books mean by "mass" is our $m\gamma$.

Summary

Other quantities besides space and time, including momentum, force, and energy, are distorted when transformed from one frame to another, just as time and space are. But some quantities, notably mass, electric charge, and the speed of light, are invariant: they are the same in all frames.

If object A moves at velocity u relative to object B, and B moves at velocity v relative to object C, the combination of the velocities, i.e. A's velocity relative to C, is not given by $u+v$ but rather by

$$v_{combined} = \frac{u+v}{1+uv} \quad \text{[natural units]} \qquad = \frac{u+v}{1+uv/c^2} \quad \text{[ordinary units]} \quad .$$

Relativistic momentum is the same in either system of units,

$$p = m\gamma v \quad \text{[natural units]} \qquad = m\gamma v \quad \text{[ordinary units]} \quad ,$$

and kinetic energy is

$$KE = m(\gamma - 1) \quad \text{[natural units]} \qquad = mc^2(\gamma - 1) \quad \text{[ordinary units]} \quad .$$

A consequence of the theory of relativity is that mass and energy do not obey separate conservation laws. Instead, the conserved quantity is the mass-energy. Mass and energy may be converted into each other according to the famous equation

$$E = m \quad \text{[natural units]} \qquad = mc^2 \quad \text{[ordinary units]} \quad .$$

Homework Problems

1✓. (a) A spacecraft traveling at 1.0000×10^7 m/s relative to the earth releases a probe in the forward direction at a relative speed of 2.0000×10^7 m/s. How fast is the probe moving relative to the earth? How does this compare with the nonrelativistic result? (b) Repeat the calculation, but with both velocities equal to $c/2$. How does this compare with the nonrelativistic result?

2. (a) Show that when two velocities are combined relativistically, and one of them equals the speed of light, the result also equals the speed of light. (b) Explain why it has to be this way based on the principle of relativity. [Note that it doesn't work to say that it has to be this way because motion faster than c is impossible. That isn't what the principle of relativity says, and it also doesn't handle the case where the velocities are in opposite direction.]

3✓. Cosmic-ray particles with relativistic velocities are continually bombarding the earth's atmosphere. They are protons and other atomic nuclei. Suppose a carbon nucleus (containing six protons and six neutrons) arrives with an energy of 10^{-7} J, which is unusually high, but not unheard of. By what factor is its length shortened as seen by an observer in the earth's frame of reference? [Hint: You can just find γ, and avoid finding v.]

4✓. (a) A free neutron (as opposed to a neutron bound into an atomic nucleus) is unstable, and decays radioactively into a proton, an electron, and a particle called a neutrino. (This process can also occur for a neutron in a nucleus, but then other forms of mass-energy are involved as well.) The masses are as follows:

neutron	1.67495×10^{-27} kg
proton	1.67265×10^{-27} kg
electron	0.00091×10^{-27} kg
neutrino	negligible

Find the energy released in the decay of a free neutron.

(b) We might imagine that a proton could decay into a neutron, a positron, and a neutrino. Although such a process can occur within a nucleus, explain why it cannot happen to a free proton. (If it could, hydrogen would be radioactive!)

5. (a) Find a relativistic equation for the velocity of an object in terms of its mass and momentum (eliminating γ). Work in natural units. (b) Show that your result is approximately the same as the classical value, p/m, at low velocities. (c) Show that very large momenta result in speeds close to the speed of light.

S A solution is given in the back of the book.
✓ A computerized answer check is available.
★ A difficult problem.
∫ A problem that requires calculus.

6. (a) Prove the equation $E^2-p^2=m^2$ for a material object, where $E=m\gamma$ is the total mass-energy. (b) Using this result, show that an object with zero mass must move at the speed of light. (c) This equation can be applied more generally, to light for instance. Use it to find the momentum of a beam of light having energy E. (d) Convert your answer from the previous part into ordinary units.

7. Starting from the equation $v_{combined}\gamma_{combined} = (v_1+v_2)\gamma_1\gamma_2$ derived in section 2.6, complete the proof of $v_{combined} = (v_1+v_2)/(1+v_1v_2)$.

8★. A source of light with frequency f is moving toward an observer at velocity v (or away from the observer if v is negative). Find the relativistically correct equation for the Doppler shift of the light. [Hint: Write down an equation for the motion of one wavefront in the source's frame, and then a second equation for the motion of the next wavefront in the source's frame. Then transform to the observer's frame and find the separation in time between the arrival of the first and second wavefronts at the same point in the observer's frame.]

9 ★. Suppose one event occurs at x_1 and t_1 and another at x_2 and t_2. Prove that the quantity $(t_2-t_1)^2-(x_2-x_1)^2$ is the same even when we transform into another coordinate system. This quantity is therefore a kind of invariant, albeit an invariant of a more abstract kind than the ones discussed until now. [When the relationship between the events is timelike in the sense of problem 5 in ch. 1, the square root of $(t_2-t_1)^2-(x_2-x_1)^2$ can be interpreted as the amount of time that would be measured by a clock that moved from one event to the other at constant velocity. It is therefore known as the *proper time* between events 1 and 2. The way the proper time relates to space and time is very much like the way Pythagorean theorem relates distance to two space dimensions, the difference being the negative sign that occurs in the former. Proper time is unaffected by Einstein-style transformations, whereas distance is unaffected by rotations.]

10. An antielectron collides with an electron that is at rest. (An antielectron is a form of antimatter that is just like an electron, but with the opposite charge.) The antielectron and electron annihilate each other and produce two gamma rays. (A gamma ray is a form of light. It has zero mass.) Gamma ray 1 is moving in the same direction as the antielectron was initially going, and gamma ray 2 is going in the opposite direction. Throughout this problem, you should work in natural units and use the notation E to mean the total mass-energy of a particle, i.e. its mass plus its kinetic energy. Find the energies of the two gamma-rays, E_1 and E_2, in terms of m, the mass of an electron or antielectron, and E_o, the initial mass-energy of the antielectron. [Hint: See problem 6a.]

Left: In 1980, the continental U.S. got its first taste of active volcanism in recent memory with the eruption of Mount St. Helens.

Top: An eruption of the Hawaiian volcano Pu'u O'o.

3 Rules of Randomness

Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the things which compose it...nothing would be uncertain, and the future as the past would be laid out before its eyes.

Pierre Simon de Laplace, 1776

The energy produced by the atom is a very poor kind of thing. Anyone who expects a source of power from the transformation of these atoms is talking moonshine.

Ernest Rutherford, 1933

The Quantum Mechanics is very imposing. But an inner voice tells me that it is still not the final truth. The theory yields much, but it hardly brings us nearer to the secret of the Old One. In any case, I am convinced that He does not play dice.

Albert Einstein

However radical Newton's clockwork universe seemed to his contemporaries, by the early twentieth century it had become a sort of smugly accepted dogma. Luckily for us, this deterministic picture of the universe breaks down at the atomic level. The clearest demonstration that the laws of physics contain elements of randomness is in the behavior of radioactive atoms. Pick two identical atoms of a radioactive isotope, say the naturally occurring uranium 238, and watch them carefully. They will undergo fission at different times, even though there was no difference in their initial behavior.

We would be in big trouble if these atoms' behavior was as predictable as expected in the Newtonian world-view, because radioactivity is an important source of heat for our planet. In reality, each atom chooses a random moment at which to release its energy, resulting in a nice steady heating effect. The earth would be a much colder planet if only sunlight heated it and not radioactivity. Probably there would be no volcanoes, and the oceans would never have been liquid. The deep-sea geothermal vents in which life first evolved would never have existed. But there would be an even worse consequence if radioactivity was deterministic: after a few billion years of peace, all the uranium 238 atoms in our planet would presumably pick the same moment to decay. The huge amount of stored nuclear energy, instead of being spread out over eons, would all be released at one instant, blowing our whole planet to Kingdom Come.

The new version of physics, incorporating certain kinds of randomness, is called quantum physics (for reasons that will become clear later). It represented such a dramatic break with the previous, deterministic tradition that everything that came before is considered "classical," even the theory of relativity. The remainder of this book is a basic introduction to quantum physics.

3.1 Randomness Isn't Random

Einstein's distaste for randomness, and his association of determinism with divinity, goes back to the Enlightenment conception of the universe as a gigantic piece of clockwork that only had to be set in motion initially by the Builder. Many of the founders of quantum mechanics were interested in possible links between physics and Eastern and Western religious and philosophical thought, but every educated person has a different concept of religion and philosophy. Bertrand Russell remarked, "Sir Arthur Eddington deduces religion from the fact that atoms do not obey the laws of mathematics. Sir James Jeans deduces it from the fact that they do."

Russell's witticism, which implies incorrectly that mathematics cannot describe randomness, reminds us how important it is not to oversimplify this question of randomness. You should not simply surmise, "Well, it's all random, anything can happen." For one thing, certain things simply cannot happen, either in classical physics or quantum physics. The conservation laws of mass, energy, momentum, and angular momentum are still valid, so for instance processes that create energy out of nothing are not just unlikely according to quantum physics, they are impossible.

A useful analogy can be made with the role of randomness in evolution. Darwin was not the first biologist to suggest that species changed over long periods of time. His two new fundamental ideas were that (1) the changes arose through random genetic variation, and (2) changes that enhanced the organism's ability to survive and reproduce would be preserved, while maladaptive changes would be eliminated by natural selection. Doubters of evolution often consider only the first point, about the randomness of natural variation, but not the second point, about the systematic action of natural selection. They make statements such as, "the development of a complex organism like Homo sapiens via random chance would be like a whirlwind blowing through a junkyard and spontaneously assembling a jumbo jet out of the scrap metal." The flaw in this type of reasoning is that it ignores the deterministic constraints on the results of random processes. For an atom to violate conservation of energy is no more likely than the conquest of the world by chimpanzees next year.

Discussion Question

Economists often behave like wannabe physicists, probably because it seems prestigious to make numerical calculations instead of talking about human relationships and organizations like other social scientists. Their striving to make economics work like Newtonian physics extends to a parallel use of mechanical metaphors, as in the concept of a market's supply and demand acting like a self-adjusting machine, and the idealization of people as economic automatons who consistently strive to maximize their own wealth. What evidence is there for randomness rather than mechanical determinism in economics?

3.2 Calculating Randomness

You should also realize that even if something is random, we can still understand it, and we can still calculate probabilities numerically. In other words, physicists are good bookmakers. A good bookmaker can calculate the odds that a horse will win a race much more accurately that an inexperienced one, but nevertheless cannot predict what will happen in any particular race.

Statistical independence

As an illustration of a general technique for calculating odds, suppose you are playing a 25-cent slot machine. Each of the three wheels has one chance in ten of coming up with a cherry. If all three wheels come up cherries, you win $100. Even though the results of any particular trial are random, you can make certain quantitative predictions. First, you can calculate that your odds of winning on any given trial are 1/10x1/10x1/10=1/1000=0.001. Here, I am representing the probabilities as numbers from 0 to 1, which is clearer than statements like "The odds are 999 to 1," and makes the calculations easier. A probability of 0 represents something impossible, and a probability of 1 represents something that will definitely happen.

Also, you can say that any given trial is equally likely to result in a win, and it doesn't matter whether you have won or lost in prior games. Mathematically, we say that each trial is statistically independent, or that separate games are uncorrelated. Most gamblers are mistakenly convinced that, to the contrary, games of chance are correlated. If they have been playing a slot machine all day, they are convinced that it is "getting ready to pay," and they do not want anyone else playing the machine and "using up" the jackpot that they "have coming." In other words, they are claiming that a series of trials at the slot machine is negatively correlated, that losing now makes you more likely to win later. Craps players claim that you should go to a table where the person rolling the dice is "hot," because she is likely to keep on rolling good numbers. Craps players, then, believe that rolls of the dice are positively correlated, that winning now makes you more likely to win later.

My method of calculating the probability of winning on the slot machine was an example of the following important rule for calculations based on independent probabilities:

The Law of Independent Probabilities
If the probability of one event happening is P_A, and the probability of a second statistically independent event happening is P_B, then the probability that they will both occur is the product of the probabilities, $P_A P_B$. If there are more than two events involved, you simply keep on multiplying.

Note that this only applies to independent probabilities. For instance, if you have a nickel and a dime in your pocket, and you randomly pull one out, there is a probability of 0.5 that it will be the nickel. If you then replace the coin and again pull one out randomly, there is again a probability of 0.5 of coming up with the nickel, because the probabilities are independent. Thus, there is a probability of 0.25 that you will get the nickel both times.

Suppose instead that you do not replace the first coin before pulling out the second one. Then you are bound to pull out the other coin the second time, and there is no way you could pull the nickel out twice. In this situation, the two trials are not independent, because the result of the first trial has an effect on the second trial. The law of independent probabilities does not apply, and the probability of getting the nickel twice is zero, not 0.25.

Experiments have shown that in the case of radioactive decay, the probability that any nucleus will decay during a given time interval is unaffected by what is happening to the other nuclei, and is also unrelated to how long it has gone without decaying. The first observation makes sense, because nuclei are isolated from each other at the centers of their respective atoms, and therefore have no physical way of influencing each other. The second fact is also reasonable, since all atoms are identical. Suppose we wanted to believe that certain atoms were “extra tough,” as demonstrated by their history of going an unusually long time without decaying. Those atoms would have to be different in some physical way, but nobody has ever succeeded in detecting differences among atoms. There is no way for an atom to be changed by the experiences it has in its lifetime.

Addition of probabilities

The law of independent probabilities tells us to use multiplication to calculate the probability that both A and B will happen, assuming the probabilities are independent. What about the probability of an “or” rather than an “and”? If two events A and B are mutually exclusive, then the probability of one or the other occurring is the sum P_A+P_B. For instance, a bowler might have a 30% chance of getting a strike (knocking down all ten pins) and a 20% chance of knocking down nine of them. The bowler's chance of knocking down either nine pins or ten pins is therefore 50%.

It does not make sense to add probabilities of things that are not mutually exclusive, i.e. that could both happen. Say I have a 90% chance of eating lunch on any given day, and a 90% chance of eating dinner. The probability that I will eat either lunch or dinner is not 180%.

Normalization

If I spin a globe and randomly pick a point on it, I have about a 70% chance of picking a point that's in an ocean and a 30% chance of picking a point on land. The probability of picking either water or land is 70%+30%=100%. Water and land are mutually exclusive, and there are no other possibilities, so the probabilities had to add up to 100%. It works the same if there are more than two possibilities — if you can classify all possible outcomes into a list of mutually exclusive results, then all the probabilities have to add up to 1, or 100%. This property of probabilities is known as normalization.

Averages

Another way of dealing with randomness is to take averages. The casino knows that in the long run, the number of times you win will approximately equal the number of times you play multiplied by the probability of winning. In the game mentioned above, where the probability of winning is 0.001, if you spend a week playing, and pay $2500 to play 10,000 times, you are likely to win about 10 times (10,000x0.001=10), and collect $1000. On the average, the casino will make a profit of $1500 from you. This is an example of the following rule.

> **Rule for Calculating Averages**
> If you conduct N identical, statistically independent trials, and the probability of success in each trial is P, then on the average, the total number of successful trials will be NP. If N is large enough, the relative error in this estimate will become small.

The statement that the rule for calculating averages gets more and more accurate for larger and larger N (known popularly as the “law of averages”) often provides a correspondence principle that connects classical and quantum physics. For instance, the amount of power produced by a nuclear power plant is not random at any detectable level, because the number of atoms in the reactor is so large. In general, random behavior at the atomic level tends to average out when we consider large numbers of atoms, which is why physics seemed deterministic before physicists learned techniques for studying atoms individually.

We can achieve great precision with averages in quantum physics because we can use identical atoms to reproduce exactly the same situation many times. If we were betting on horses or dice, we would be much more limited in our precision. After a thousand races, the horse would be ready to retire. After a million rolls, the dice would be worn out.

Self-Check

Which of the following things *must* have independent, which *could* be independent, and which definitely are *not* independent?
(1) the probability of successfully making two free-throws in a row in basketball
(2) the probability that it will rain in London tomorrow and the probability that it will rain on the same day in a certain city in a distant galaxy
(3) your probability of dying today and of dying tomorrow

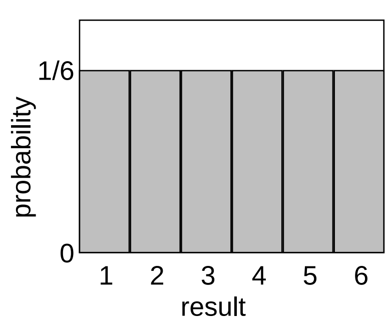
Discussion questions

A. Newtonian physics is an essentially perfect approximation for describing the motion of a pair of dice. If Newtonian physics is deterministic, why do we consider the result of rolling dice to be random?

B. Why isn't it valid to define randomness by saying that randomness is when all the outcomes are equally likely?

C. The sequence of digits 121212121212121212 seems clearly nonrandom, and 41592653589793 seems random. The latter sequence, however, is the decimal form of pi, starting with the third digit. There is a story about the Indian mathematician Ramanujan, a self-taught prodigy, that he was riding in a cab in New York with another mathematician who pointed out the license plate of a truck, saying that the number seemed completely uninteresting. Ramanujan replied that on the contrary, it was a very interesting number because it was the smallest integer that could not be represented as the sum of the squares of three prime numbers. The Argentine author Jorge Luis Borges wrote a short story called "The Library of Babel," in which he imagined a library containing every book that could possibly be written using the letters of the alphabet. It would include a book containing only the repeated letter "a"; all the ancient Greek tragedies known today, all the lost Greek tragedies, and millions of Greek tragedies that were never actually written; your own life story, and various incorrect versions of your own life story; and countless anthologies containing a short story called "The Library of Babel." Of course, if you picked a book from the shelves of the library, it would almost certainly look like a nonsensical sequence of letters and punctuation, but it's always possible that the seemingly meaningless book would be a science-fiction screenplay written in the language of a Neanderthal tribe, or a set of incomparably beautiful love poems written in a language that never existed. In view of these examples, what does it really mean to say that something is random?

3.3 Probability Distributions

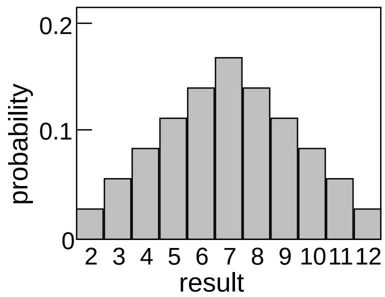


(a) Probability distribution for the result of rolling a single die.

So far we've discussed random processes having only two possible outcomes: yes or no, win or lose, on or off. More generally, a random process could have a result that is a number. Some processes yield integers, as when you roll a die and get a result from one to six, but some are not restricted to whole numbers, for example the number of seconds that a uranium-238 atom will exist before undergoing radioactive decay.

Consider a throw of a die. If the die is "honest," then we expect all six values to be equally likely. Since all six probabilities must add up to 1, then probability of any particular value coming up must be 1/6. We can summarize this in a graph, (a). Areas under the curve can be interpreted as total probabilities. For instance, the area under the curve from 1 to 3 is 1/6+1/6+1/6=1/2, so the probability of getting a result from 1 to 3 is 1/2. The function shown on the graph is called the probability distribution.

(1) Most people would think they were positively correlated, but it's possible that they're independent. (2) These must be independent, since there is no possible physical mechanism that could make one have any effect on the other. (3) These cannot be independent, since dying today guarantees that you won't die tomorrow.

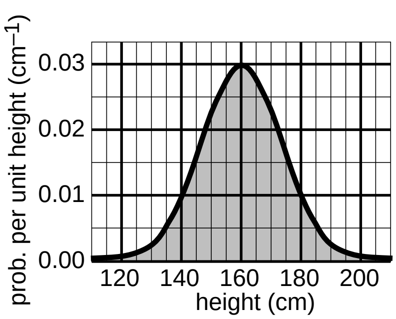


(b) Rolling two dice and adding them up.

Figure (b) shows the probabilities of various results obtained by rolling two dice and adding them together, as in the game of craps. The probabilities are not all the same. There is a small probability of getting a two, for example, because there is only one way to do it, by rolling a one and then another one. The probability of rolling a seven is high because there are six different ways to do it: 1+6, 2+5, etc.

If the number of possible outcomes is large but finite, for example the number of hairs on a dog, the graph would start to look like a smooth curve rather than a ziggurat.

What about probability distributions for random numbers that are not integers? We can no longer make a graph with probability on the *y* axis, because the probability of getting a given exact number is typically zero. For instance, there is zero probability that a radioactive atom will last for *exactly* 3 seconds, since there is are infinitely many possible results that are close to 3 but not exactly three: 2.999999999999999996876876587658465436, for example. It doesn't usually make sense, therefore, to talk about the probability of a single numerical result, but it does make sense to talk about the probability of a certain range of results. For instance, the probability that an atom will last more than 3 and less than 4 seconds is a perfectly reasonable thing to discuss. We can still summarize the probability information on a graph, and we can still interpret areas under the curve as probabilities.



(c) A probability distribution for height of human adults. (Not real data.)

But the y axis can no longer be a unitless probability scale. In radioactive decay, for example, we want the x axis to have units of time, and we want areas under the curve to be unitless probabilities. The area of a single square on the graph paper is then

(unitless area of a square)
= (width of square with time units) x (height of square) .

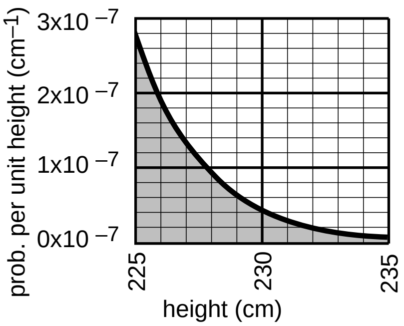
If the units are to cancel out, then the height of the square must evidently be a quantity with units of inverse time. In other words, the y axis of the graph is to be interpreted as probability per unit time, not probability.

Figure (c) shows another example, a probability distribution for people's height. This kind of bell-shaped curve is quite common.

Self-Check

Compare the number of people with heights in the range of 130-135 cm to the number in the range 135-140.

The area under the curve from 130 to 135 cm is about 3/4 of a rectangle. The area from 135 to 140 cm is about 1.5 rectangles. The number of people in the second range is about twice as much. We could have converted these to actual probabilities (1 rectangle = 5 cm x 0.005 cm-1 = 0.025), but that would have been pointless because we were just going to compare the two areas.



(d) A close-up of the right-hand tail of the distribution shown in the previous figure.

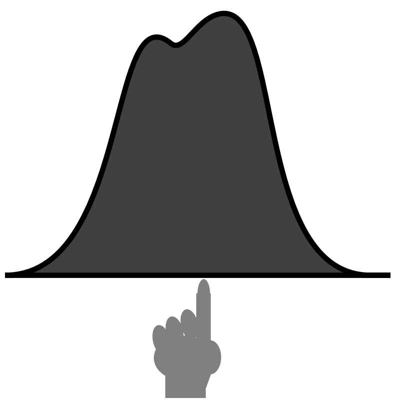
Example: Looking for tall basketball players

Question: A certain country with a large population wants to find very tall people to be on its Olympic basketball team and strike a blow against western imperialism. Out of a pool of 10^8 people who are the right age and gender, how many are they likely to find who are over 225 cm (7'4") in height? Figure (d) gives a close-up of the "tails" of the distribution shown previously.

Solution: The shaded area under the curve represents the probability that a given person is tall enough. Each rectangle represents a probability of 0.2×10^{-7} cm^{-1} x 1 cm = 2×10^{-9}. There are about 35 rectangles covered by the shaded area, so the probability of having a height greater than 230 cm is 7×10^{-8}, or just under one in ten million. Using the rule for calculating averages, the average, or expected number of people this tall is $(10^8)\times(7\times10^{-8})=7$.

Average and width of a probability distribution

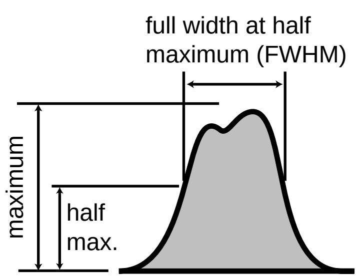
If the next Martian you meet asks you, "How tall is an adult human?," you will probably reply with a statement about the average human height, such as "Oh, about 5 feet 6 inches." If you wanted to explain a little more, you could say, "But that's only an average. Most people are somewhere between 5 feet and 6 feet tall." Without bothering to draw the relevant bell curve for your new extraterrestrial acquaintance, you've summarized the relevant information by giving an average and a typical range of variation.



(e) The average of a probability distribution.

The average of a probability distribution can be defined geometrically as the horizontal position at which it could be balanced if it was constructed out of cardboard. A convenient numerical measure of the amount of variation about the average, or amount of uncertainty, is the full width at half maximum, or FWHM, shown in the figure.

A great deal more could be said about this topic, and indeed an introductory statistics course could spend months on ways of defining the center and width of a distribution. Rather than force-feeding you on mathematical detail or techniques for calculating these things, it is perhaps more relevant to point out simply that there are various ways of defining them, and to inoculate you against the misuse of certain definitions.



(f) The full width at half maximum (FWHM) of a probability distribution.

The average is not the only possible way to say what is a typical value for a quantity that can vary randomly; another possible definition is the median, defined as the value that is exceeded with 50% probability. When discussing incomes of people living in a certain town, the average could be very misleading, since it can be affected massively if a single resident of the town is Bill Gates. Nor is the FWHM the only possible way of stating the amount of random variation; another possible way of measuring it is the standard deviation (defined as the square root of the average squared deviation from the average value).

3.4 Exponential Decay and Half-Life

Most people know that radioactivity "lasts a certain amount of time," but that simple statement leaves out a lot. As an example, consider the following medical procedure used to diagnose thyroid function. A very small quantity of the isotope ^{131}I, produced in a nuclear reactor, is fed to or injected into the patient. The body's biochemical systems treat this artificial, radioactive isotope exactly the same as ^{127}I, which is the only naturally occurring type. (Nutritionally, iodine is a necessary trace element. Iodine taken into the body is partly excreted, but the rest becomes concentrated in the thyroid gland. Iodized salt has had iodine added to it to prevent the nutritional deficiency known as goiters, in which the iodine-starved thyroid becomes swollen.) As the ^{131}I undergoes beta decay, it emits electrons, neutrinos, and gamma rays. The gamma rays can be measured by a detector passed over the patient's body. As the radioactive iodine becomes concentrated in the thyroid, the amount of gamma radiation coming from the thyroid becomes greater, and that emitted by the rest of the body is reduced. The rate at which the iodine concentrates in the thyroid tells the doctor about the health of the thyroid.

If you ever undergo this procedure, someone will presumably explain a little about radioactivity to you, to allay your fears that you will turn into the Incredible Hulk, or that your next child will have an unusual number of limbs. Since iodine stays in your thyroid for a long time once it gets there, one thing you'll want to know is whether your thyroid is going to become radioactive forever. They may just tell you that the radioactivity "only lasts a certain amount of time," but we can now carry out a quantitative derivation of how the radioactivity really will die out.

Let $P_{surv}(t)$ be the probability that an iodine atom will survive without decaying for a period of at least t. It has been experimentally measured that half all ^{131}I atoms decay in 8 hours, so we have

$$P_{surv}(8\text{ hr}) = 0.5 \quad .$$

Now using the law of independent probabilities, the probability of surviving for 16 hours equals the probability of surviving for the first 8 hours multiplied by the probability of surviving for the second 8 hours,

$$\begin{aligned} P_{surv}(16\text{ hr}) &= 0.5 \times 0.5 \\ &= 0.25 \quad . \end{aligned}$$

Similarly we have

$$\begin{aligned} P_{surv}(24\text{ hr}) &= 0.5 \times 0.5 \times 0.5 \\ &= 0.125 \quad . \end{aligned}$$

Generalizing from this pattern, the probability of surviving for any time t that is a multiple of 8 hours is

$$P_{surv}(t) = 0.5^{t/(8\text{ hr})}$$

We now know how to find the probability of survival at intervals of 8 hours, but what about the points in time in between? What would be the probability of surviving for 4 hours? Well, using the law of independent probabilities again, we have

$$P_{surv}(8\text{ hr}) = P_{surv}(4\text{ hr}) \times P_{surv}(4\text{ hr}) \quad ,$$

which can be rearranged to give

$$P_{surv}(4\text{ hr}) = \sqrt{P_{surv}(8\text{ hr})} = \sqrt{0.5} = 0.707 \quad .$$

This is exactly what we would have found simply by plugging in $P_{surv}(t) = 0.5^{t/(8\text{ hr})} = 0.5^{1/2}$ and ignoring the restriction to multiples of 8 hours. Since 8 hours is the amount of time required for half of the atoms to decay, it is known as the half-life, written $t_{1/2}$. The general rule is as follows:

Exponential Decay Formula

$$P_{surv}(t) = 0.5^{t/t_{1/2}}$$

Using the rule for calculating averages, we can also find the number of atoms, $N(t)$, remaining in a sample at time t:

$$N(t) = N(0) \times 0.5^{t/t_{1/2}}$$

Both of these equations have graphs that look like dying-out exponentials, as in the example below.

Example: Radioactive contamination at Chernobyl

Question: One of the most dangerous radioactive isotopes released by the Chernobyl disaster in 1986 was ^{90}Sr, whose half-life is 28 years. (a) How long will it be before the contamination is reduced to one tenth of its original level? (b) If a total of 10^{27} atoms was released, about how long would it be before not a single atom was left?

Solution: (a) We want to know the amount of time that a ^{90}Sr nucleus has a probability of 0.1 of surviving. Starting with the exponential decay formula,

$$P_{surv} = 0.5^{t/t_{1/2}} \quad ,$$

we want to solve for t. Taking natural logarithms of both sides,

$$\ln P = \frac{t}{t_{1/2}} \ln 0.5 \quad ,$$

so

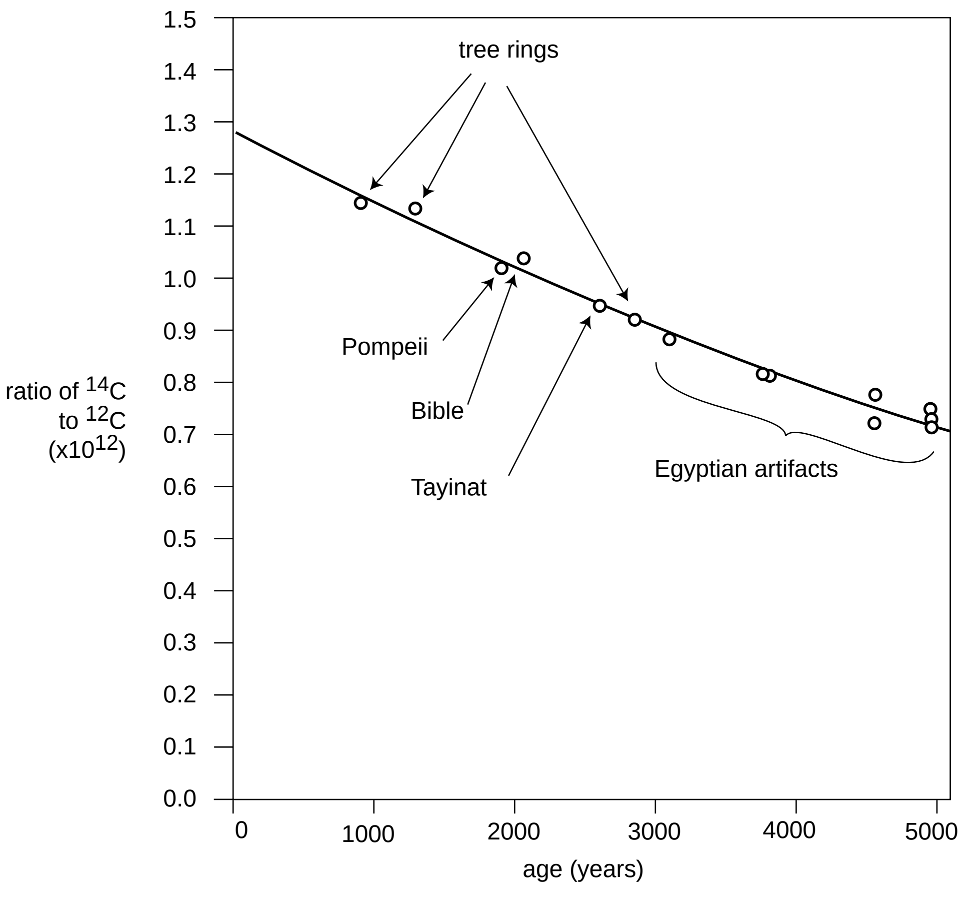
$$t = \frac{t_{1/2} \ln P}{\ln 0.5}$$

Plugging in P=0.1 and $t_{1/2}$=28 years, we get t=93 years.

(b) This is just like the first part, but $P=10^{-27}$. The result is about 2500 years.

Example: ^{14}C Dating

Almost all the carbon on Earth is ^{12}C, but not quite. The isotope ^{14}C, with a half-life of 5600 years, is produced by cosmic rays in the atmosphere. It decays naturally, but is replenished at such a rate that the fraction of ^{14}C in the atmosphere remains constant, at 1.3×10^{-12}. Living plants and animals take in both ^{12}C and ^{14}C from the atmosphere and incorporate both into their bodies. Once the living organism dies, it no longer takes in C atoms from the atmosphere, and the proportion of ^{14}C gradually falls off as it undergoes radioactive decay. This effect can be used to find the age of dead organisms, or human artifacts made from plants or animals. The following graph shows the exponential decay curve of ^{14}C in various objects. Similar methods, using longer-lived isotopes, provided the first firm proof that the earth was billions of years old, not a few thousand as some had claimed on religious grounds.



Calibration of the ^{14}C dating method using tree rings and artifacts whose ages were known from other methods. Redrawn from Emilio Segrè, **Nuclei and Particles**, 1965.

Rate of decay

If you want to find how many radioactive decays occur within a time interval lasting from time t to time $t+\Delta t$, the most straightforward approach is to calculate it like this:

$$
\begin{aligned}
&(\text{number of decays between } t \text{ and } t+\Delta t) \\
&= N(t) - N(t+\Delta t) \\
&= N(0)\left[P_{\text{surv}}(t) - P_{\text{surv}}(t+\Delta t)\right] \\
&= N(0)\left[0.5^{t/t_{1/2}} - 0.5^{(t+\Delta t)/t_{1/2}}\right] \\
&= N(0)\left[1 - 0.5^{\Delta t/t_{1/2}}\right]0.5^{t/t_{1/2}}
\end{aligned}
$$

A problem arises when Δt is small compared to $t_{1/2}$. For instance, suppose you have a hunk of 10^{22} atoms of ^{235}U, with a half-life of 700 million years, which is $2.2\text{x}10^{16}$ s. You want to know how many decays will occur in $\Delta t=1$ s. Since we're specifying the current number of atoms, $t=0$. As you plug in to the formula above on your calculator, the quantity $0.5^{\Delta t/t_{1/2}}$ comes out on your calculator to equal one, so the final result is zero. That's incorrect, though. In reality, $0.5^{\Delta t/t_{1/2}}$ should equal 0.999999999999999968, but your calculator only gives eight digits of precision, so it rounded it off to one. In other words, the probability that a ^{235}U atom will survive for 1 s is very close to one, but not equal to one. The number of decays in one second is therefore $3.2\text{x}10^5$, not zero.

Well, my calculator only does eight digits of precision, just like yours, so how did I know the right answer? The way to do it is to use the following approximation:

$$a^b \approx 1 + b \ln a, \text{ if } b << 1$$

(The symbol << means “is much less than.”) Using it, we can find the following approximation:

$$
\begin{aligned}
&(\text{number of decays between } t \text{ and } t+\Delta t) \\
&= N(0)\left[1 - 0.5^{\Delta t/t_{1/2}}\right]0.5^{t/t_{1/2}} \\
&\approx N(0)\left[1 - \left(1 + \frac{\Delta t}{t_{1/2}} \ln 0.5\right)\right]0.5^{t/t_{1/2}}, \text{ if } \Delta t << t_{1/2} \\
&= (\ln 2) N(0) \left(0.5^{t/t_{1/2}}\right)\frac{\Delta t}{t_{1/2}}
\end{aligned}
$$

This also gives us a way to calculate the rate of decay, i.e. the number of decays per unit time. Dividing by Δt on both sides, we have

$$(\text{decays per unit time}) \approx \frac{(\ln 2) N(0)}{t_{1/2}} 0.5^{t/t_{1/2}}, \text{ if } \Delta t << t_{1/2}$$

Example: The hot potato

Question: A nuclear physicist with a demented sense of humor tosses you a cigar box, yelling "hot potato." The label on the box says "contains 10^{20} atoms of ^{17}F, half-life of 66 s, produced today in our reactor at 1 p.m." It takes you two seconds to read the label, after which you toss it behind some lead bricks and run away. The time is 1:40 p.m. Will you die?

Solution: The time elapsed since the radioactive fluorine was produced in the reactor was 40 minutes, or 2400 s. The number of elapsed half-lives is therefore $t / t_{1/2} = 36$. The initial number of atoms was $N(0)=10^{20}$. The number of decays per second is now about $10^7\ s^{-1}$, so it produced about $2x10^7$ high-energy electrons while you held it in your hands. Although twenty million electrons sounds like a lot, it is not really enough to be dangerous.

By the way, none of the equations we've derived so far was the actual probability distribution for the time at which a particular radioactive atom will decay. That probability distribution would be found by substituting $N(0)=1$ into the equation for the rate of decay.

If the sheer number of equations is starting to seem formidable, let's pause and think for a second. The simple equation for P_{surv} is something you can derive easily from the law of independent probabilities any time you need it. From that, you can quickly find the exact equation for the rate of decay. The derivation of the approximate equations for $\Delta t << t$ is a little hairier, but note that except for the factors of ln 2, everything in these equations can be found simply from considerations of logic and units. For instance, a longer half-life will obviously lead to a slower rate of decays, so it makes sense that we divide by it. As for the ln 2 factors, they are exactly the kind of thing that one looks up in a book when one needs to know them.

Discussion Questions

A. In the medical procedure involving ^{131}I, why is it the gamma rays that are detected, not the electrons or neutrinos that are also emitted?

B. For 1 s, Fred holds in his hands 1 kg of radioactive stuff with a half-life of 1000 years. Ginger holds 1 kg of a different substance, with a half-life of 1 min, for the same amount of time. Did they place themselves in equal danger, or not?

C. How would you interpret it if you calculated $N(t)$, and found it was less than one?

D. Does the half-life depend on how much of the substance you have? Does the expected time until the sample decays completely depend on how much of the substance you have?

3.5∫ Applications of Calculus

The area under the probability distribution is of course an integral. If we call the random number x and the probability distribution $D(x)$, then the probability that x lies in a certain range is given by

$$(\text{probability of } a \le x \le b) = \int_a^b D(x)\,dx \quad .$$

What about averages? If x had a finite number of equally probable values, we would simply add them up and divide by how many we had. If they weren't equally likely, we'd make the weighted average $x_1P_1+x_2P_2+\ldots$ But we need to generalize this to a variable x that can take on any of a continuum of values. The continuous version of a sum is an integral, so the average is

$$(\text{average value of } x) = \int x\,D(x)\,dx \quad ,$$

where the integral is over all possible values of x.

Example: Probability distribution for radioactive decay

Here is a rigorous justification for the statement in the previous section that the probability distribution for radioactive decay is found by substituting $N(0)=1$ into the equation for the rate of decay. We know that the probability distribution must be of the form

$$D(x) = k\,0.5^{t/t_{1/2}} \quad ,$$

where k is a constant that we need to determine. The atom is guaranteed to decay eventually, so normalization gives us

$$\begin{aligned}(\text{probability of } 0 \le t \le \infty)\\ &= 1\\ &= \int_0^\infty D(t)\,dt\end{aligned}$$

The integral is most easily evaluated by converting the function into an exponential with e as the base

$$\begin{aligned}D(x) &= k\exp\left[\ln\left(0.5^{t/t_{1/2}}\right)\right]\\ &= k\exp\left[\frac{t}{t_{1/2}}\ln(0.5)\right]\\ &= k\exp\left(-\frac{\ln 2}{t_{1/2}}t\right) \quad ,\end{aligned}$$

which give an integral of the familiar form $\int e^{cx}dx = \frac{1}{c}e^{cx}$. We thus have

$$1 = -\frac{k\,t_{1/2}}{\ln 2}\exp\left(-\frac{\ln 2}{t_{1/2}}t\right)\Bigg]_0^\infty \quad ,$$

which gives the desired result:

$$k = \frac{\ln 2}{t_{1/2}} \quad .$$

Example: Average lifetime

You might think that the half-life would also be the average lifetime of an atom, since half the atoms' lives are shorter and half longer. But the half whose lives are longer include some that survive for many half-lives, and these rare long-lived atoms skew the average. We can calculate the average lifetime as follows:

(average lifetime)

$$= \int_0^\infty t\, D(t)\, dt$$

Using the convenient base-e form again, we have

(average lifetime)

$$= \frac{\ln 2}{t_{1/2}} \int_0^\infty t \exp\left(-\frac{\ln 2}{t_{1/2}} t\right) dt \quad .$$

This integral is of a form that can either be attacked with integration by parts or by looking it up in a table. The result is $\int xe^{cx} dx = \frac{x}{c} e^{cx} - \frac{1}{c^2} e^{cx}$, and the first term can be ignored for our purposes because it equals zero at both limits of integration. We end up with

(average lifetime)

$$= \frac{\ln 2}{t_{1/2}} \left(\frac{t_{1/2}}{\ln 2}\right)^2$$

$$= \frac{t_{1/2}}{\ln 2}$$

$$= 1.443\, t_{1/2} \quad ,$$

which is, as expected, longer than one half-life.

Summary

Selected Vocabulary

probability	the likelihood that something will happen, expressed as a number between zero and one
normalization	the property of probabilities that the sum of the probabilities of all possible outcomes must equal one
independence	the lack of any relationship between two random events
probability distribution	a curve that specifies the probabilities of various random values of a variable; areas under the curve correspond to probabilities
FWHM	the full width at half-maximum of a probability distribution; a measure of the width of the distribution
half-life	the amount of time that a radioactive atom has a probability of 1/2 of surviving without decaying

Notation

P	probability
$t_{1/2}$	half-life
D	a probability distribution (used only in optional section 3.5)

Summary

Quantum physics differs from classical physics in many ways, the most dramatic of which is that certain processes at the atomic level, such as radioactive decay, are random rather than deterministic. There is a method to the madness, however: quantum physics still rules out any process that violates conservation laws, and it also offers methods for calculating probabilities numerically.

In this chapter we focused on certain generic methods of working with probabilities, without concerning ourselves with any physical details. Without knowing any of the details of radioactive decay, for example, we were still able to give a fairly complete treatment of the relevant probabilities. The most important of these generic methods is the law of independent probabilities, which states that if two random events are not related in any way, then the probability that they will both occur equals the product of the two probabilities,

$$\text{probability of A and B} = P_A P_B \text{, if A and B are independent} \quad .$$

The most important application is to radioactive decay. The time that a radioactive atom has a 50% chance of surviving is called the half-life, $t_{1/2}$. The probability of surviving for two half-lives is (1/2)(1/2)=1/4, and so on. In general, the probability of surviving a time t is given by

$$P_{surv} = 0.5^{t/t_{1/2}} \quad .$$

Related quantities such as the rate of decay and probability distribution for the time of decay are given by the same type of exponential function, but multiplied by certain constant factors.

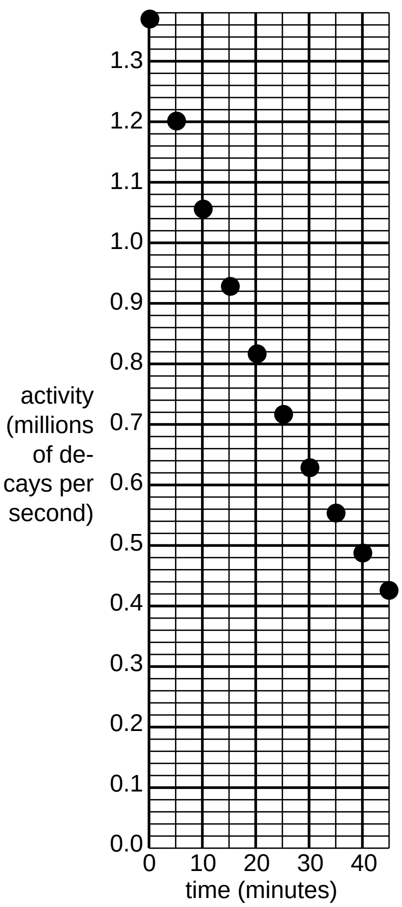
Homework Problems

1. If a radioactive substance has a half-life of one year, does this mean that it will be completely decayed after two years? Explain.

2. What is the probability of rolling a pair of dice and getting "snake eyes," i.e. both dice come up with ones?

3. Use a calculator to check the approximation that $a^b \approx 1 + b \ln a$, if $b \ll 1$ using some arbitrary numbers. See how good the approximation is for values of b that are not quite as small compared to one.

4. Make up an example of a numerical problem involving a rate of decay where $\Delta t \ll t_{1/2}$, but $0.5^{t/t_{1/2}}$ can still be evaluated on a calculator without getting something that rounds off to one. Check that you get approximately the same result using both methods to calculate the number of decays between t and $t+\Delta t$.



Problem 5.

5. (a) A nuclear physicist is studying a nuclear reaction caused in an accelerator experiment, with a beam of ions from the accelerator striking a thin metal foil and causing nuclear reactions when a nucleus from one of the beam ions happens to hit one of the nuclei in the target. After the experiment has been running for a few hours, a few billion radioactive atoms have been produced, embedded in the target. She does not know what nuclei are being produced, but she suspects they are an isotope of some heavy element such as Pb, Bi, Fr or U. Following one such experiment, she takes the target foil out of the accelerator, sticks it in front of a detector, measures the activity every 5 min, and makes a graph (figure). The isotopes she thinks may have been produced are:

isotope	half-life (minutes)
^{211}Pb	36.1
^{214}Pb	26.8
^{214}Bi	19.7
^{223}Fr	21.8
^{239}U	23.5

Which one is it?

(b) Having decided that the original experimental conditions produced

S A solution is given in the back of the book.
✓ A computerized answer check is available.
★ A difficult problem.
∫ A problem that requires calculus.

one specific isotope, she now tries using beams of ions traveling at several different speeds, which may cause different reactions. The following table gives the activity of the target 10, 20 and 30 minutes after the end of the experiment, for three different ion speeds.

	activity (millions of decays/s) after...		
	10 min	20 min	30 min
first ion speed	1.933	0.832	0.382
second ion speed	1.200	0.545	0.248
third ion speed	6.544	1.296	0.248

Since such a large number of decays is being counted, assume that the data are only inaccurate due to rounding off when writing down the table. Which are consistent with the production of a single isotope, and which imply that more than one isotope was being created?

6. Devise a method for testing experimentally the hypothesis that a gambler's chance of winning at craps is independent of her previous record of wins and losses.

7. Refer to the probability distribution for people's heights in section 3.3.

(a) Show that the graph is properly normalized.

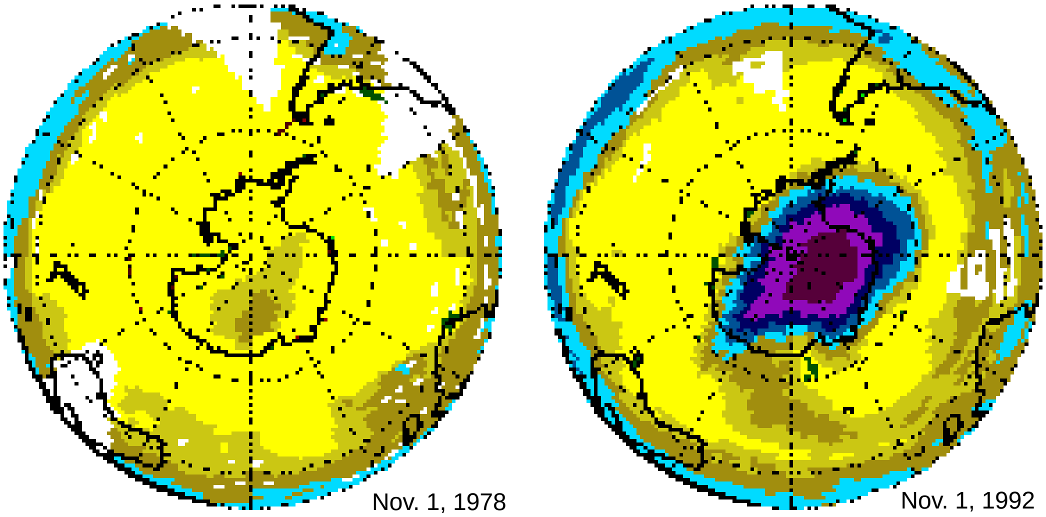
(b) Estimate the fraction of the population having heights between 140 and 150 cm.

8. ^{238}U decays be alpha emission, with a half-life of 4.5×10^9 years. The subsequent chain of alpha and electron (beta) decays involves much shorter half-lives, and terminates in the stable nucleus ^{206}Pb. Almost all natural uranium is ^{238}U. All helium on earth is from the decay chain that leads from ^{238}U to ^{206}Pb.

(a) How many alphas are emitted per decay chain? [Hint: Use conservation of mass.]

(b) How many electrons are emitted per decay chain? [Hint: Use conservation of charge.]

(c ✓) Each alpha particles ends up claiming two electrons and becoming a helium atom. If the original ^{238}U atom is in solid rock (as opposed to the earth's molten regions), the He atoms are unable to diffuse out of the rock. Suppose a geologist finds a sample of hardened lava, melts it in a furnace, and finds that it contains 1230 mg of uranium and 2.3 mg of helium. How long has it been since the lava originally hardened?



In recent decades, a huge hole in the ozone layer has spread out from Antarctica.

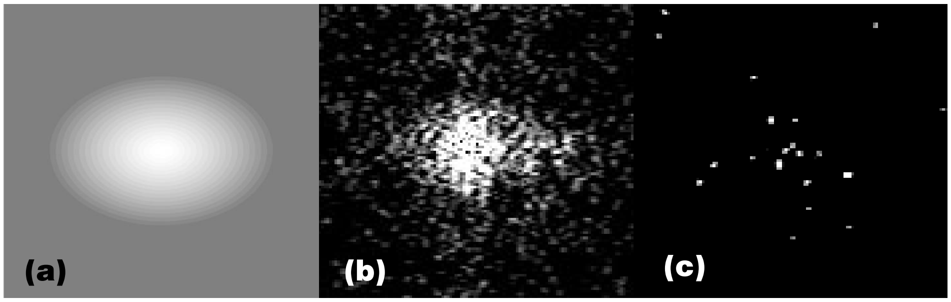
4 Light as a Particle

The only thing that interferes with my learning is my education.
Albert Einstein

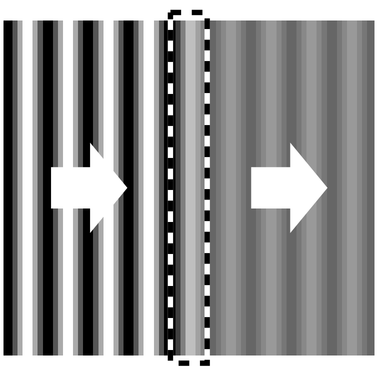
Radioactivity is random, but do the laws of physics exhibit randomness in other contexts besides radioactivity? Yes. Radioactive decay was just a good playpen to get us started with concepts of randomness, because all atoms of a given isotope are identical. By stocking the playpen with an unlimited supply of identical atom-toys, nature helped us to realize that their future behavior could be different regardless of their original identicality. We are now ready to leave the playpen, and see how randomness fits into the structure of physics at the most fundamental level.

The laws of physics describe light and matter, and the quantum revolution rewrote both descriptions. Radioactivity was a good example of matter's behaving in a way that was inconsistent with classical physics, but if we want to get under the hood and understand how nonclassical things happen, it will be easier to focus on light rather than matter. A radioactive atom such as uranium-235 is after all an extremely complex system, consisting of 92 protons, 143 neutrons, and 92 electrons. Light, however, can be a simple sine wave.

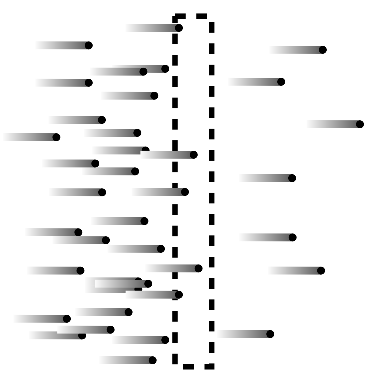
However successful the classical wave theory of light had been — allowing the creation of radio and radar, for example — it still failed to describe many important phenomena. An example that is currently of great interest is the way the ozone layer protects us from the dangerous short-wavelength ultraviolet part of the sun's spectrum. In the classical description, light is a wave. When a wave passes into and back out of a medium, its frequency is unchanged, and although its wavelength is altered while it is in the medium, it returns to its original value when the wave reemerges. Luckily for us, this is not at all what ultraviolet light does when it passes through the ozone layer, or the layer would offer no protection at all!



4.1 Evidence for Light as a Particle



(d)



(e)

For a long time, physicists tried to explain away the problems with the classical theory of light as arising from an imperfect understanding of atoms and the interaction of light with individual atoms and molecules. The ozone paradox, for example, could have been attributed to the incorrect assumption that one could think of the ozone layer as a smooth, continuous substance, when in reality it was made of individual ozone molecules. It wasn't until 1905 that Albert Einstein threw down the gauntlet, proposing that the problem had nothing to do with the details of light's interaction with atoms and everything to do with the fundamental nature of light itself.

In those days the data were sketchy, the ideas vague, and the experiments difficult to interpret; it took a genius like Einstein to cut through the thicket of confusion and find a simple solution. Today, however, we can get right to the heart of the matter with a piece of ordinary consumer electronics, the digital camera. Instead of film, a digital camera has a computer chip with its surface divided up into a grid of light-sensitive squares, called "pixels." Compared to a grain of the silver compound used to make regular photographic film, a digital camera pixel is activated by an amount of light energy orders of magnitude smaller. We can learn something new about light by using a digital camera to detect smaller and smaller amounts of light, as shown in figures (a) through (c) above. Figure (a) is fake, but (b) and (c) are real digital-camera images made by Prof. Lyman Page of Princeton University as a classroom demonstration. Figure (a) is what we would see if we used the digital camera to take a picture of a fairly dim source of light. In figures (b) and (c), the intensity of the light was drastically reduced by inserting semitransparent absorbers like the tinted plastic used in sunglasses. Going from (a) to (b) to (c), more and more light energy is being thrown away by the absorbers.

The results are drastically different from what we would expect based on the wave theory of light. If light was a wave and nothing but a wave, (d), then the absorbers would simply cut down the wave's amplitude across the whole wavefront. The digital camera's entire chip would be illuminated uniformly, and weakening the wave with an absorber would just mean that every pixel would take a long time to soak up enough energy to register a signal.

But figures (b) and (c) show that some pixels take strong hits while others pick up no energy at all. Instead of the wave picture, the image that is naturally evoked by the data is something more like a hail of bullets from a machine gun, (e). Each "bullet" of light apparently carries only a tiny

Einstein and Seurat: twins separated at birth.
Seine Grande Jatte by Georges Seurat (19th century)

amount of energy, which is why detecting them individually requires a sensitive digital camera rather than an eye or a piece of film.

Although Einstein was interpreting different observations, this is the conclusion he reached in his 1905 paper: that the pure wave theory of light is an oversimplification, and that the energy of a beam of light comes in finite chunks rather than being spread smoothly throughout a region of space.

We now think of these chunks as particles of light, and call them "photons," although Einstein avoided the word "particle," and the word "photon" was invented later. Regardless of words, the trouble was that waves and particles seemed like inconsistent categories. The reaction to Einstein's paper could be kindly described as vigorously skeptical. Even twenty years later, Einstein wrote, "There are therefore now two theories of light, both indispensable, and — as one must admit today despite twenty years of tremendous effort on the part of theoretical physicists — without any logical connection." In the remainder of this chapter we will learn how the seeming paradox was eventually resolved.

Discussion Questions

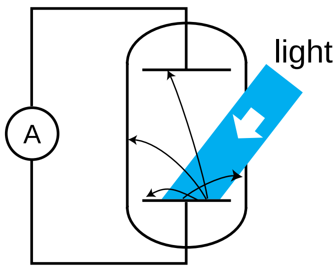
A. Suppose someone rebuts the digital camera data, claiming that the random pattern of dots occurs not because of anything fundamental about the nature of light but simply because the camera's pixels are not all exactly the same. How could we test this interpretation?

B. Discuss how the correspondence principle applies to the observations and concepts discussed so far.

4.2 How Much Light Is One Photon?

The photoelectric effect

We have seen evidence that light energy comes in little chunks, so the next question to be asked is naturally how much energy is in one chunk. The most straightforward experimental avenue for addressing this question is a phenomenon known as the photoelectric effect. The photoelectric effect occurs when a photon strikes the surface of a solid object and knocks out an electron. It occurs continually all around you. It is happening right now at the surface of your skin and on the paper or computer screen from which you are reading these words. It does not ordinarily lead to any observable electrical effect, however, because on the average free electrons are wandering back in just as frequently as they are being ejected. (If an object did somehow lose a significant number of electrons, its growing net positive charge would begin attracting the electrons back more and more strongly.)



(a) Apparatus for observing the photoelectric effect. A beam of light strikes a capacitor plate inside a vacuum tube, and electrons are ejected (black arrows).

Figure (a) shows a practical method for detecting the photoelectric effect. Two very clean parallel metal plates (the electrodes of a capacitor) are sealed inside a vacuum tube, and only one plate is exposed to light. Because there is a good vacuum between the plates, any ejected electron that happens to be headed in the right direction will almost certainly reach the other capacitor plate without colliding with any air molecules.

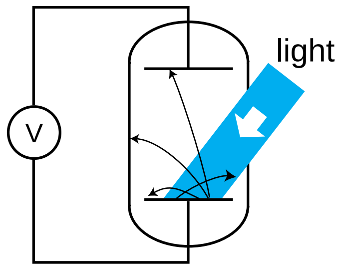
The illuminated (bottom) plate is left with a net positive charge, and the unilluminated (top) plate acquires a negative charge from the electrons deposited on it. There is thus an electric field between the plates, and it is because of this field that the electrons' paths are curved, as shown in the diagram. However, since vacuum is a good insulator, any electrons that reach the top plate are prevented from responding to the electrical attraction by jumping back across the gap. Instead they are forced to make their way around the circuit, passing through an ammeter. The ammeter allows a measurement of the strength of the photoelectric effect.

An unexpected dependence on frequency

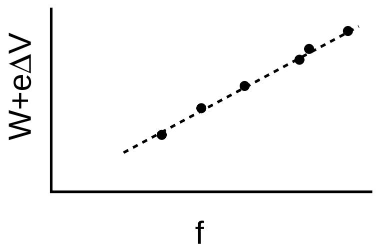
The photoelectric effect was discovered serendipitously by Heinrich Hertz in 1887, as he was experimenting with radio waves. He was not particularly interested in the phenomenon, but he did notice that the effect was produced strongly by ultraviolet light and more weakly by lower frequencies. Light whose frequency was lower than a certain critical value did not eject any electrons at all. (In fact this was all prior to Thomson's discovery of the electron, so Hertz would not have described the effect in terms of electrons — we are discussing everything with the benefit of hindsight.) This dependence on frequency didn't make any sense in terms of the classical wave theory of light. A light wave consists of electric and magnetic fields. The stronger the fields, i.e. the greater the wave's amplitude, the greater the forces that would be exerted on electrons that found themselves bathed in the light. It should have been amplitude (brightness) that was relevant, not frequency. The dependence on frequency not only proves that the wave model of light needs modifying, but with the proper interpretation it allows us to determine how much energy is in one photon, and it also leads to a connection between the wave and particle models that we need in order to reconcile them.

To make any progress, we need to consider the physical process by which a photon would eject an electron from the metal electrode. A metal contains electrons that are free to move around. Ordinarily, in the interior of the metal, such an electron feels attractive forces from atoms in every direction around it. The forces cancel out. But if the electron happens to find itself at the surface of the metal, the attraction from the interior side is not balanced out by any attraction from outside. Bringing the electron out through the surface therefore requires a certain amount of work, W, which depends on the type of metal used.

Suppose a photon strikes an electron, annihilating itself and giving up all its energy to the electron. (We now know that this is what always happens in the photoelectric effect, although it had not yet been established in 1905 whether or not the photon was completely annihilated.) The



(b) A different way of studying the photoelectric effect.



(c) The quantity $W+e\Delta V$ indicates the energy of one photon. It is found to be proportional to the frequency of the light.

Historical Note

What I'm presenting in this chapter is a simplified explanation of how the photon *could* have been discovered. The actual history is more complex.

Max Planck (1858-1947) began the photon saga with a theoretical investigation of the spectrum of light emitted by a hot, glowing object. He introduced quantization of the energy of light waves, in multiples of *hf*, purely as a mathematical trick that happened to produce the right results. Planck did not believe that his procedure could have any physical significance. In his 1905 paper Einstein took Planck's quantization as a description of reality, and applied it to various theoretical and experimental puzzles, including the photoelectric effect.

Millikan then subjected Einstein's ideas to a series of rigorous experimental tests. Although his results matched Einstein's predictions perfectly, Millikan was skeptical about photons, and his papers conspicuously omit any reference to them. Only in his autobiography did Millikan rewrite history and claim that he had given experimental proof for photons.

electron will (1) lose kinetic energy through collisions with other electrons as it plows through the metal on its way to the surface; (2) lose an amount of kinetic energy equal to W as it emerges through the surface; and (3) lose more energy on its way across the gap between the plates, due to the electric field between the plates. Even if the electron happens to be right at the surface of the metal when it absorbs the photon, and even if the electric field between the plates has not yet built up very much, W is the bare minimum amount of energy that it must receive from the photon if it is to contribute to a measurable current. The reason for using very clean electrodes is to minimize W and make it have a definite value characteristic of the metal surface, not a mixture of values due to the various types of dirt and crud that are present in tiny amounts on all surfaces in everyday life.

We can now interpret the frequency dependence of the photoelectric effect in a simple way: apparently the amount of energy possessed by a photon is related to its frequency. A low-frequency red or infrared photon has an energy less than W, so a beam of them will not produce any current. A high-frequency blue or violet photon, on the other hand, packs enough of a punch to allow an electron to make it to the other plate. At frequencies higher than the minimum, the photoelectric current continues to increase with the frequency of the light because of effects (1) and (3).

Numerical relationship between energy and frequency

Prompted by Einstein's photon paper, Robert Millikan (whom we encountered in book 4 of this series) figured out how to use the photoelectric effect to probe precisely the link between frequency and photon energy. Rather than going into the historical details of Millikan's actual experiments (a lengthy experimental program that occupied a large part of his professional career) we will describe a simple version, shown in figure (b), that is used sometimes in college laboratory courses. The idea is simply to illuminate one plate of the vacuum tube with light of a single wavelength and monitor the voltage difference between the two plates as they charge up. Since the resistance of a voltmeter is very high (much higher than the resistance of an ammeter), we can assume to a good approximation that electrons reaching the top plate are stuck there permanently, so the voltage will keep on increasing for as long as electrons are making it across the vacuum tube.

At a moment when the voltage difference has a reached a value ΔV, the minimum energy required by an electron to make it out of the bottom plate and across the gap to the other plate is $W+e\Delta V$. As ΔV increases, we eventually reach a point at which $W+e\Delta V$ equals the energy of one photon. No more electrons can cross the gap, and the reading on the voltmeter stops rising. The quantity $W+e\Delta V$ now tells us the energy of one photon. If we determine this energy for a variety of wavelengths, (c), we find the following simple relationship between the energy of a photon and the frequency of the light:

$$E = hf \quad ,$$

where h is a constant having a numerical value of 6.63×10^{-34} J·s. Note how the equation brings the wave and particle models of light under the same roof: the left side is the energy of one *particle* of light, while the right side is the frequency of the same light, interpreted as a *wave*. The constant h is known as Planck's constant (see historical note).

Self-Check

How would you extract h from the graph in figure (c)?

Since the energy of a photon is hf, a beam of light can only have energies of hf, $2hf$, $3hf$, etc. Its energy is quantized — there is no such thing as a fraction of a photon. Quantum physics gets its name from the fact that it quantizes quantities like energy, momentum, and angular momentum that had previously been thought to be smooth, continuous and infinitely divisible.

Example: number of photons emitted by a lightbulb per second
Question: Roughly how many photons are emitted by a 100-W lightbulb in 1 second?
Solution: People tend to remember wavelengths rather than frequencies for visible light. The bulb emits photons with a range of frequencies and wavelengths, but let's take 600 nm as a typical wavelength for purposes of estimation. The energy of a single photon is

$$E_{photon} = hf = hc/\lambda$$

A power of 100 W means 100 joules per second, so the number of photons is

$$(100\text{ J})/E_{photon} = (100\text{ J}) / (hc/\lambda) \approx 3\times10^{20}$$

Example: Momentum of a photon
Question: According to the theory of relativity, the momentum of a beam of light is given by $p=E/c$ (see ch. 2, homework problem #6). Apply this to find the momentum of a single photon in terms of its frequency, and in terms of its wavelength.
Solution: Combining the equations $p=E/c$ and $E=hf$, we find

$$p = E/c = \frac{h}{c}f \quad .$$

To reexpress this in terms of wavelength, we use $c=f\lambda$:

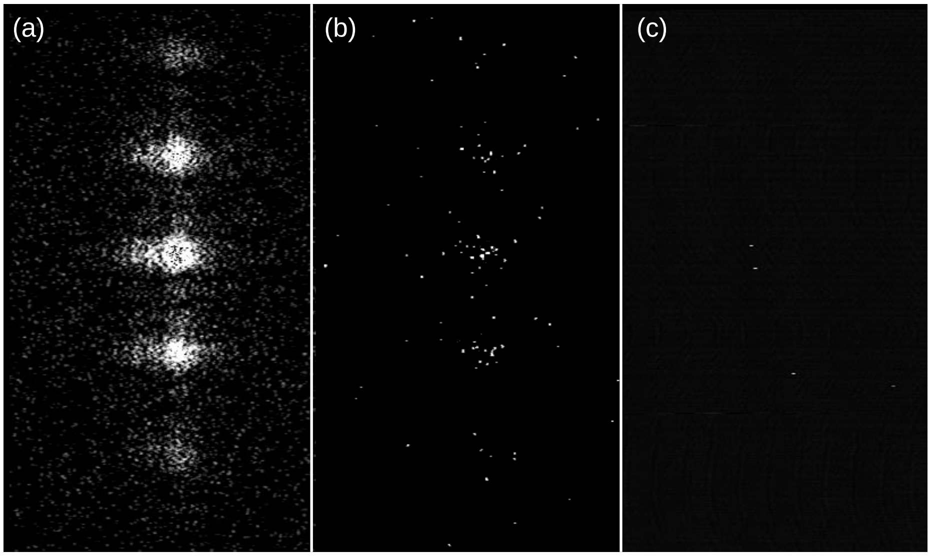
$$p = \frac{h}{c}\cdot\frac{c}{\lambda} = \frac{h}{\lambda}$$

The second form turns out to be simpler.

Discussion Questions

A. Only a very tiny percentage of the electrons available near the surface of an object is ever ejected by the photoelectric effect. How well does this agree the wave model of light, and how well with the particle model?
B. What is the significance of the fact that Planck's constant is numerically very small? How would our everyday experience of light be different if it was not so small?
C. How would the experiments described above be affected if electrons were likely to get hit by more than one photon?
D. Draw some representative trajectories of electrons for $\Delta V=0$, ΔV less than the maximum value, and ΔV greater than the maximum value.
E. Explain based on the photon theory of light why ultraviolet light would be more likely than visible or infrared light to cause cancer by damaging DNA molecules.
F. Does $E=hf$ imply that a photon changes its energy when it passes from one transparent material into another substance with a different index of refraction?

The axes of the graph are frequency and photon energy, so its slope is Planck's constant.

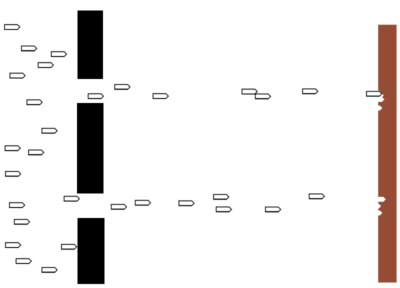


Wave interference patterns photographed by Prof. Lyman Page with a digital camera. Laser light with a single well-defined wavelength passed through a series of absorbers to cut down its intensity, then through a set of slits to produce interference, and finally into a digital camera chip. (A triple slit was actually used, but for conceptual simplicity we discuss the results in the main text as if it was a double slit.) In figure (b) the intensity has been reduced relative to (a), and even more so for figure (c).

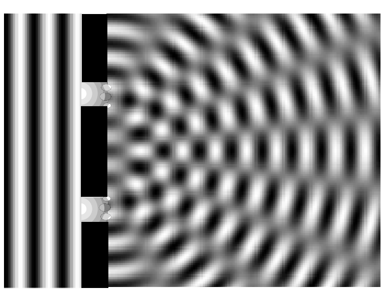
4.3 Wave-Particle Duality

How can light be both a particle and a wave? We are now ready to resolve this seeming contradiction. Often in science when something seems paradoxical, it's because we (1) don't define our terms carefully, or (2) don't test our ideas against any specific real-world situation. Let's define particles and waves as follows:

Waves exhibit superposition, and specifically interference phenomena.
Particles can only exist in whole numbers, not fractions



(d) Bullets pass through a double slit.



(e) A water wave passes through a double slit.

As a real-world check on our philosophizing, there is one particular experiment that works perfectly. We set up a double-slit interference experiment that we know will produce a diffraction pattern if light is an honest-to-goodness wave, but we detect the light with a detector that is capable of sensing individual photons, e.g. a digital camera. To make it possible to pick out individual dots from individual photons, we must use filters to cut down the intensity of the light to a very low level, just as in the photos by Prof. Page in section 4.1. The whole thing is sealed inside a light-tight box. The results are shown in figures (a), (b), and (c) above. (In fact, the similar figures in section 4.1 are simply cutouts from these figures.)

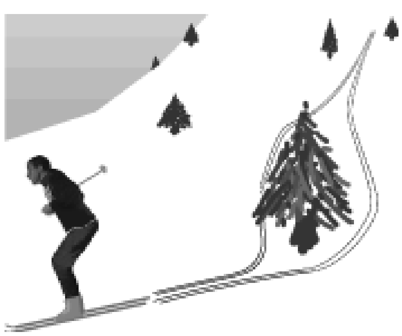
Neither the pure wave theory nor the pure particle theory can explain the results. If light was only a particle and not a wave, there would be no interference effect. The result of the experiment would be like firing a hail of bullets through a double slit, (d). Only two spots directly behind the slits would be hit.

If, on the other hand, light was only a wave and not a particle, we would get the same kind of diffraction pattern that would happen with a water wave, (e). There would be no discrete dots in the photo, only a diffraction pattern that shaded smoothly between light and dark.

Applying the definitions to this experiment, light must be both a particle and a wave. It is a wave because it exhibits interference effects. At the same time, the fact that the photographs contain discrete dots is a direct demonstration that light refuses to be split into units of less than a single photon. There can only be whole numbers of photons: four photons in figure (c), for example.

A wrong interpretation: photons interfering with each other

One possible interpretation of wave-particle duality that occurred to physicists early in the game was that perhaps the interference effects came from photons interacting with each other. By analogy, a water wave consists of moving water molecules, and interference of water waves results ultimately from all the mutual pushes and pulls of the molecules. This interpretation was conclusively disproved by G.I. Taylor, a student at Cambridge. The demonstration by Prof. Page that we've just been discussing is essentially a modernized version of Taylor's work. Taylor reasoned that if interference effects came from photons interacting with each other, a bare minimum of two photons would have to be present at the same time to produce interference. By making the light source extremely dim, we can be virtually certain that there are never two photons in the box at the same time. In figure (c), however, the intensity of the light has been cut down so much by the absorbers that if it was in the open, the average separation between photons would be on the order of a kilometer! At any given moment, the number of photons in the box is most likely to be zero. It is virtually certain that there were never two photons in the box at once.



A single photon can go through both slits.

The concept of a photon's path is undefined.

If a single photon can demonstrate double-slit interference, then which slit did it pass through? The unavoidable answer must be that it passes through both! This might not seem so strange if we think of the photon as a wave, but it is highly counterintuitive if we try to visualize it as a particle. The moral is that we should not think in terms of the path of a photon. Like the fully human and fully divine Jesus of Christian theology, a photon is supposed to be 100% wave and 100% particle. If a photon had a well defined path, then it would not demonstrate wave superposition and interference effects, contradicting its wave nature. (In the next chapter we will discuss the Heisenberg uncertainty principle, which gives a numerical way of approaching this issue.)

Another wrong interpretation: the pilot wave hypothesis

A second possible explanation of wave-particle duality was taken seriously in the early history of quantum mechanics. What if the photon *particle* is like a surfer riding on top of its accompanying *wave*? As the wave travels along, the particle is pushed, or "piloted" by it. Imagining the particle and the wave as two separate entities allows us to avoid the seemingly paradoxical idea that a photon is both at once. The wave happily does its wave tricks, like superposition and interference, and the particle acts like a respectable particle, resolutely refusing to be in two different places at once. If the wave, for instance, undergoes destructive interference, becoming nearly zero in a particular region of space, then the particle simply is not guided into that region.

The problem with the pilot wave interpretation is that the only way it can be experimentally tested or verified is if someone manages to detach the particle from the wave, and show that there really are two entities involved, not just one. Part of the scientific method is that hypotheses are supposed to be experimentally testable. Since nobody has ever managed to separate the wavelike part of a photon from the particle part, the interpretation is not useful or meaningful in a scientific sense.

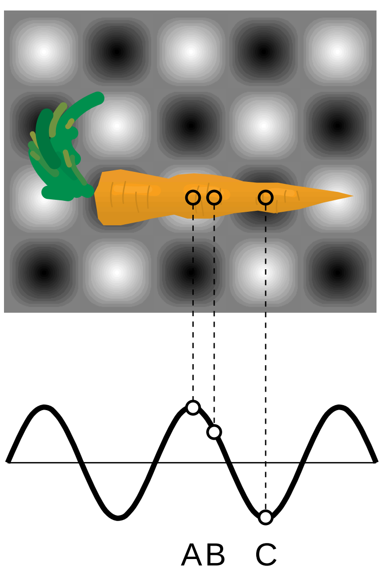
The probability interpretation

The correct interpretation of wave-particle duality is suggested by the random nature of the experiment we've been discussing: even though every photon wave/particle is prepared and released in the same way, the location at which it is eventually detected by the digital camera is different every time. The idea of the probability interpretation of wave-particle duality is that the location of the photon-particle is random, but the probability that it is in a certain location is higher where the photon-wave's amplitude is greater.

More specifically, the probability distribution of the particle must be proportional to the *square* of the wave's amplitude,

$$(\text{probability distribution}) \propto (\text{amplitude})^2 \quad .$$

This follows from the correspondence principle and from the fact that a wave's energy density is proportional to the square of its amplitude. If we run the double-slit experiment for a long enough time, the pattern of dots fills in and becomes very smooth as would have been expected in classical physics. To preserve the correspondence between classical and quantum physics, the amount of energy deposited in a given region of the picture over the long run must be proportional to the square of the wave's amplitude. The amount of energy deposited in a certain area depends on the number of photons picked up, which is proportional to the probability of finding any given photon there.



Example: a microwave oven

Question: The figure shows two-dimensional (top) and one-dimensional (bottom) representations of the standing wave inside a microwave oven. Gray represents zero field, and white and black signify the strongest fields, with white being a field that is in the opposite direction compared to black. Compare the probabilities of detecting a microwave photon at points A, B, and C.

Solution: A and C are both extremes of the wave, so the probabilities of detecting a photon at A and C are equal. It doesn't matter that we have represented C as negative and A as positive, because it is the square of the amplitude that is relevant. The amplitude at B is about 1/2 as much as the others, so the probability of detecting a photon there is about 1/4 as much.

The probability interpretation was disturbing to physicists who had spent their previous careers working in the deterministic world of classical physics, and ironically the most strenuous objections against it were raised by Einstein, who had invented the photon concept in the first place. The probability interpretation has nevertheless passed every experimental test, and is now as well established as any part of physics.

An aspect of the probability interpretation that has made many people uneasy is that the process of detecting and recording the photon's position seems to have a magical ability to get rid of the wavelike side of the photon's personality and force it to decide for once and for all where it really wants to be. But detection or measurement is after all only a physical process like any other, governed by the same laws of physics. We will postpone a detailed discussion of this issue until the following chapter, since a measuring device like a digital camera is made of matter, but we have so far only discussed how quantum mechanics relates to light.

Example: What is the proportionality constant?

Question: What is the proportionality constant that would make an actual equation out of (probability distribution)$\propto$(amplitude)^2?

Solution: The probability that the photon is in a certain small region of volume v should equal the fraction of the wave's energy that is within that volume:

$$P = \frac{\text{energy in volume } v}{\text{energy of photon}} = \frac{\text{energy in volume } v}{hf}$$

We assume v is small enough so that the electric and magnetic fields are nearly constant throughout it. We then have

$$P = \frac{\left(\frac{1}{8\pi k}|\mathbf{E}|^2 + \frac{1}{2\mu_o}|\mathbf{B}|^2\right)v}{hf} \quad .$$

We can simplify this formidable looking expression by recognizing that in an electromagnetic wave, $|\mathbf{E}|$ and $|\mathbf{B}|$ are related by $|\mathbf{E}|=c|\mathbf{B}|$. With some algebra, it turns out that the electric and magnetic fields each contribute half the total energy (see book 4, ch. 6, homework problem #5), so we can simplify this to

$$P = \frac{2\left(\frac{1}{8\pi k}|\mathbf{E}|^2\right)v}{hf} = \frac{v}{4\pi khf}|\mathbf{E}|^2 \quad .$$

As advertised, the probability is proportional to the square of the wave's amplitude.

Discussion Questions

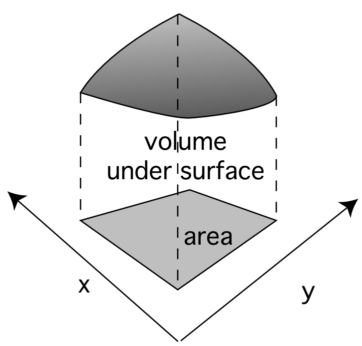
A. Referring back to the example of the carrot in the microwave oven, show that it would be nonsensical to have probability be proportional to the field itself, rather than the square of the field.

B. Einstein did not try to reconcile the wave and particle theories of light, and did not say much about their apparent inconsistency. Einstein basically visualized a beam of light as a stream of bullets coming from a machine gun. In the photoelectric effect, a photon "bullet" would only hit one atom, just as a real bullet would only hit one person. Suppose someone reading his 1905 paper wanted to interpret it by saying that Einstein's so-called particles of light were simply short wave-trains that only occupy a small region of space. Comparing the wavelength of visible light (a few hundred nm) to the size of an atom (on the order of 0.1 nm), explain why this poses a difficulty for reconciling the particle and wave theories.

C. Can a white photon exist?

D. In double-slit diffraction of photons, would you get the same pattern of dots on the photo if you covered one slit? Why should it matter whether you give the photon two choices or only one?

4.4 Photons in Three Dimensions



Up until now I've been sneaky and avoided a full discussion of the three-dimensional aspects of the probability interpretation. The example of the carrot in the microwave oven, for example, reduced to a one-dimensional situation because we were considering three points along the same line and because we were only comparing ratios of probabilities. The purpose of bringing it up now is to head off any feeling that you've been cheated conceptually rather than to prepare you for mathematical problem solving in three dimensions, which would not be appropriate for the level of this course.

A typical example of a probability distribution in chapter 3 was the distribution of heights of human beings. The thing that varied randomly, height, h, had units of meters, and the probability distribution was a graph of a function $D(h)$. The units of the probability distribution had to be m^{-1} (inverse meters) so that areas under the curve, interpreted as probabilities, would be unitless (area = width x height = m x m^{-1}).

Now suppose we have a two-dimensional problem, e.g. the probability distribution for the place on the surface of a digital camera chip where a photon will be detected. The point where it is detected would be described with two variables, x and y, each having units of meters. The probability distribution will be a function of both variables, $D(x,y)$. A probability is now visualized as the volume under the surface described by the function $D(x,y)$, as shown in the figure. The units of D must be m^{-2} so that probabilities will be unitless (area = width x depth x height = m x m x m^{-2}).

Generalizing finally to three dimensions, we find by analogy that the probability distribution will be a function of all three coordinates, $D(x,y,z)$, and will have units of m^{-3}. It is unfortunately impossible to visualize the graph unless you are a mutant with a natural feel for life in four dimensions. If the probability distribution is nearly constant within a certain volume of space v, the probability that the photon is in that volume is simply vD. If you know enough calculus, it should be clear that this can be generalized to $P=\int D\,dx\,dy\,dz$ if D is not constant.

Summary

Selected Vocabulary

photon .... a particle of light
photoelectric effect .... the ejection, by a photon, of an electron from the surface of an object
wave-particle duality.......... the idea that light is both a wave and a particle

Summary

Around the turn of the twentieth century, experiments began to show problems with the classical wave theory of light. In any experiment sensitive enough to detect very small amounts of light energy, it becomes clear that light energy cannot be divided into chunks smaller than a certain amount. Measurements involving the photoelectric effect demonstrate that this smallest unit of light energy equals hf, where f is the frequency of the light and h is a number known as Planck's constant. We say that light energy is quantized in units of hf, and we interpret this quantization as evidence that light has particle properties as well as wave properties. Particles of light are called photons.

The only method of reconciling the wave and particle natures of light that has stood the test of experiment is the probability interpretation. It states that the probability that the particle is at a given location is proportional to the square of the amplitude of the wave at that location.

One important consequence of wave-particle duality is that we must abandon the concept of the path the particle takes through space. To hold on to this concept, we would have to contradict the well established wave nature of light, since a wave can spread out in every direction simultaneously.

Homework Problems

1. When light is reflected from a mirror, perhaps only 80% of the energy comes back. One could try to explain this in two different ways: (1) 80% of the photons are reflected, or (2) all the photons are reflected, but each loses 20% of its energy. Based on your everyday knowledge about mirrors, how can you tell which interpretation is correct? [Based on a problem from PSSC Physics.]

2. Suppose we want to build an electronic light sensor using an apparatus like the one described in the section on the photoelectric effect. How would its ability to detect different parts of the spectrum depend on the type of metal used in the capacitor plates?

3. The photoelectric effect can occur not just for metal cathodes but for any substance, including living tissue. Ionization of DNA molecules in can cause cancer or birth defects. If the energy required to ionize DNA is on the same order of magnitude as the energy required to produce the photoelectric effect in a metal, which of these types of electromagnetic waves might pose such a hazard? Explain.

60 Hz waves from power lines
100 MHz FM radio
microwaves from a microwave oven
visible light
ultraviolet light
x-rays

4✓. The beam of a 100-W overhead projector covers an area of 1 m x 1 m when it hits the screen 3 m away. Estimate the number of photons that are in flight at any given time. (Since this is only an estimate, we can ignore the fact that the beam is not parallel.)

5✓. In the photoelectric effect, electrons are observed with virtually no time delay (~10 ns), even when the light source is very weak. (A weak light source does however only produce a small number of ejected electrons.) The purpose of this problem is to show that the lack of a significant time delay contradicted the classical wave theory of light, so throughout this problem you should put yourself in the shoes of a classical physicist and pretend you don't know about photons at all. At that time, it was thought that the electron might have a radius on the order of 10^{-15} m. (Recent experiments have shown that if the electron has any finite size at all, it is far smaller.)

(a) Estimate the power that would be soaked up by a single electron in a beam of light with an intensity of 1 mW/m^2.

(b) The energy, *W*, required for the electron to escape through the surface of the cathode is on the order of 10^{-19} J. Find how long it would take the electron to absorb this amount of energy, and explain why your result constitutes strong evidence that there is something wrong with the classical theory.

S A solution is given in the back of the book. ★ A difficult problem.
✓ A computerized answer check is available. ∫ A problem that requires calculus.

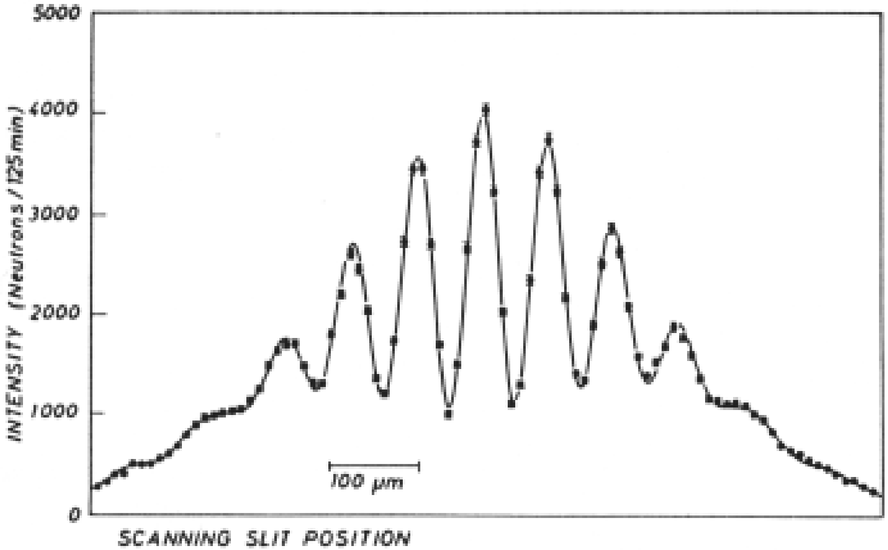
5 Matter as a Wave

> [In] a few minutes I shall be all melted... I have been wicked in my day, but I never thought a little girl like you would ever be able to melt me and end my wicked deeds. Look out — here I go!
>
> The Wicked Witch of the West

As the Wicked Witch learned the hard way, losing molecular cohesion can be unpleasant. That's why we should be very grateful that the concepts of quantum physics apply to matter as well as light. If matter obeyed the laws of classical physics, molecules wouldn't exist.

Consider, for example, the simplest atom, hydrogen. Why does one hydrogen atom form a chemical bond with another hydrogen atom? Roughly speaking, we'd expect a neighboring pair of hydrogen atoms, A and B, to exert no force on each other at all, attractive or repulsive: there are two repulsive interactions (proton A with proton B and electron A with electron B) and two attractive interactions (proton A with electron B and electron A with proton B). Thinking a little more precisely, we should even expect that once the two atoms got close enough, the interaction would be repulsive. For instance, if you squeezed them so close together that the two protons were almost on top of each other, there would be a tremendously strong repulsion between them due to the $1/r^2$ nature of the electrical force. The repulsion between the electrons would not be as strong, because each electron ranges over a large area, and is not likely to be found right on top of the other electron. Thus hydrogen molecules should not exist according to classical physics.

Quantum physics to the rescue! As we'll see shortly, the whole problem is solved by applying the same quantum concepts to electrons that we have already used for photons.



A double-slit interference pattern made with neutrons. (A. Zeilinger, R. Gähler, C.G. Shull, W. Treimer, and W. Mampe, *Reviews of Modern Physics*, Vol. 60, 1988.)

5.1 Electrons as Waves

We started our journey into quantum physics by studying the random behavior of *matter* in radioactive decay, and then asked how randomness could be linked to the basic laws of nature governing *light*. The probability interpretation of wave-particle duality was strange and hard to accept, but it provided such a link. It is now natural to ask whether the same explanation could be applied to matter. If the fundamental building block of light, the photon, is a particle as well as a wave, is it possible that the basic units of matter, such as electrons, are waves as well as particles?

A young French aristocrat studying physics, Louis de Broglie (pronounced "broylee"), made exactly this suggestion in his 1923 Ph.D. thesis. His idea had seemed so farfetched that there was serious doubt about whether to grant him the degree. Einstein was asked for his opinion, and with his strong support, de Broglie got his degree.

Only two years later, American physicists C.J. Davisson and L. Germer confirmed de Broglie's idea by accident. They had been studying the scattering of electrons from the surface of a sample of nickel, made of many small crystals. (One can often see such a crystalline pattern on a brass doorknob that has been polished by repeated handling.) An accidental explosion occurred, and when they put their apparatus back together they observed something entirely different: the scattered electrons were now creating an interference pattern! This dramatic proof of the wave nature of matter came about because the nickel sample had been melted by the explosion and then resolidified as a single crystal. The nickel atoms, now nicely arranged in the regular rows and columns of a crystalline lattice, were acting as the lines of a diffraction grating. The new crystal was analogous to the type of ordinary diffraction grating in which the lines are etched on the surface of a mirror (a reflection grating) rather than the kind in which the light passes through the transparent gaps between the lines (a transmission grating).

Although we will concentrate on the wave-particle duality of electrons because it is important in chemistry and the physics of atoms, all the other "particles" of matter you've learned about show wave properties as well. The figure above, for instance, shows a wave interference pattern of neutrons.

It might seem as though all our work was already done for us, and there would be nothing new to understand about electrons: they have the same kind of funny wave-particle duality as photons. That's almost true, but not quite. There are some important ways in which electrons differ significantly from photons:

(1) Electrons have mass, and photons don't.

(2) Photons always move at the speed of light, but electrons can move at any speed less than c.

(3) Photons don't have electric charge, but electrons do, so electric forces can act on them. The most important example is the atom, in which the electrons are held by the electric force of the nucleus.

(4) Electrons cannot be absorbed or emitted as photons are. Destroying an electron or creating one out of nothing would violate conservation of charge.

(In chapter 6 we will learn of one more fundamental way in which electrons differ from photons, for a total of five.)

Because electrons are different from photons, it is not immediately obvious which of the photon equations from the previous chapter can be applied to electrons as well. A particle property, the energy of one photon, is related to its wave properties via $E=hf$ or, equivalently, $E=hc/\lambda$. The momentum of a photon was given by $p=hf/c$ or $p=h/\lambda$. Ultimately it was a matter of experiment to determine which of these equations, if any, would work for electrons, but we can make a quick and dirty guess simply by noting that some of the equations involve c, the speed of light, and some do not. Since c is irrelevant in the case of an electron, we might guess that the equations of general validity are those that do not have c in them:

$$E = hf$$
$$p = h/\lambda$$

This is essentially the reasoning that de Broglie went through, and experiments have confirmed these two equations for all the fundamental building blocks of light and matter, not just for photons and electrons.

The second equation, which I soft-pedaled in the previous chapter, takes on a greater important for electrons. This is first of all because the momentum of matter is more likely to be significant than the momentum of light under ordinary conditions, and also because force is the transfer of momentum, and electrons are affected by electrical forces.

Discussion Question

Frequency is oscillations per second, whereas wavelength is meters per oscillation. How could the equations $E = hf$ and $p = h/\lambda$ be made to look more alike by using quantities that were more closely analogous? How would this more symmetric treatment relate to incorporating relativity into quantum mechanics?

Example: the wavelength of an elephant

Question: What is the wavelength of a trotting elephant?

Solution: One may doubt whether the equation should be applied to an elephant, which is not just a single particle but a rather large collection of them. Throwing caution to the wind, however, we estimate the elephant's mass at 10^3 kg and its trotting speed at 10 m/s. Its wavelength is therefore roughly

$$\begin{aligned}
\lambda &= h/p \\
&= \frac{h}{mv} \\
&= \frac{6.63\times10^{-34}\,\text{J}\cdot\text{s}}{(10^3\ \text{kg})(10\ \text{m/s})} \\
&\sim 10^{-37}\frac{\left(\text{kg}\cdot\text{m}^2/\text{s}^2\right)\cdot\text{s}}{\text{kg}\cdot\text{m/s}} \\
&= 10^{-37}\ \text{m}
\end{aligned}$$

The wavelength found in this example is so fantastically small that we can be sure we will never observe any measurable wave phenomena with elephants or any other human-scale objects. The result is numerically small because Planck's constant is so small, and as in some examples encountered previously, this smallness is in accord with the correspondence principle.

Although a smaller mass in the equation $\lambda=h/mv$ does result in a longer wavelength, the wavelength is still quite short even for individual electrons under typical conditions, as shown in the following example.

Example: the typical wavelength of an electron

Question: Electrons in circuits and in atoms are typically moving through potential differences on the order of 1 V, so that a typical energy is (e)(1 V), which is on the order of 10^{-19} J. What is the wavelength of an electron with this amount of kinetic energy?

Solution: This energy is nonrelativistic, since it is much less than mc^2. Momentum and energy are therefore related by the nonrelativistic equation $KE=p^2/2m$. Solving for p and substituting in to the equation for the wavelength, we find

$$\begin{aligned}
\lambda &= \frac{h}{\sqrt{2m\cdot KE}} \\
&= 1.6\text{x}10^{-9}\ \text{m}
\end{aligned}$$

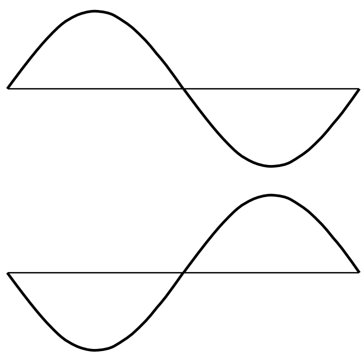
This is on the same order of magnitude as the size of an atom, which is no accident: as we will discuss in the next chapter in more detail, an electron in an atom can be interpreted as a standing wave. The smallness of the wavelength of a typical electron also helps to explain why the wave nature of electrons wasn't discovered until a hundred years after the wave nature of light. To scale the usual wave-optics devices such as diffraction gratings down to the size needed to work with electrons at ordinary energies, we need to make them so small that their parts are comparable in size to individual atoms. This is essentially what Davisson and Germer did with their nickel crystal.

Self-Check

These remarks about the inconvenient smallness of electron wavelengths apply only under the assumption that the electrons have typical energies. What kind of energy would an electron have to have in order to have a longer wavelength that might be more convenient to work with?

What kind of wave is it?

If a sound wave is a vibration of matter, and a photon is a vibration of electric and magnetic fields, what kind of a wave is an electron made of? The disconcerting answer is that there is no experimental "observable," i.e. directly measurable quantity, to correspond to the electron wave itself. In other words, there are devices like microphones that detect the oscillations of air pressure in a sound wave, and devices such as radio receivers that measure the oscillation of the electric and magnetic fields in a light wave, but nobody has ever found any way to measure the electron wave directly.



These two electron waves are not distinguishable by any measuring device.

We can of course detect the energy (or momentum) possessed by an electron just as we could detect the energy of a photon using a digital camera. (In fact I'd imagine that an unmodified digital camera chip placed in a vacuum chamber would detect electrons just as handily as photons.) But this only allows us to determine where the wave carries high probability and where it carries low probability. Probability is proportional to the square of the wave's amplitude, but measuring its square is not the same as measuring the wave itself. In particular, we get the same result by squaring either a positive number or its negative, so there is no way to determine the positive or negative sign of an electron wave.

Most physicists tend toward the school of philosophy known as operationalism, which says that a concept is only meaningful if we can define some set of operations for observing, measuring, or testing it. According to a strict operationalist, then, the electron wave itself is a meaningless concept. Nevertheless, it turns out to be one of those concepts like love or humor that is impossible to measure and yet very useful to have around. We therefore give it a symbol, Ψ (the capital Greek letter psi), and a special name, the electron *wavefunction* (because it is a function of the coordinates x, y, and z that specify where you are in space). It would be impossible, for example, to calculate the shape of the electron wave in a hydrogen atom without having some symbol for the wave. But when the calculation produces a result that can be compared directly to experiment, the final algebraic result will turn out to involve only Ψ^2, which is what is observable, not Ψ itself.

Since Ψ, unlike $\boldsymbol{E}$ and $\boldsymbol{B}$, is not directly measurable, we are free to make the probability equations have a simple form: instead of having the probability density equal to some funny constant multiplied by Ψ^2, we simply define Ψ so that the constant of proportionality is one:

$$(\text{probability density}) = \Psi^2 \quad .$$

Since the probability density has units of m^{-3}, the units of Ψ must be $\text{m}^{-3/2}$.

Wavelength is inversely proportional to momentum, so to produce a large wavelength we would need to use electrons with very *small* momenta and energies. (In practical terms, this isn't very easy to do, since ripping an electron out of an object is a violent process, and it's not so easy to calm the electrons down afterward.)

5.2*∫ Dispersive Waves

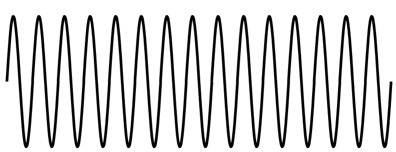
A colleague of mine who teaches chemistry loves to tell the story about an exceptionally bright student who, when told of the equation $p=h/\lambda$, protested, "But when I derived it, it had a factor of 2!" The issue that's involved is a real one, albeit one that could be glossed over (and is, in most textbooks) without raising any alarms in the mind of the average student. The present optional section addresses this point; it is intended for the student who wishes to delve a little deeper.

Here's how the now-legendary student was presumably reasoning. We start with the equation $v=f\lambda$, which is valid for any sine wave, whether it's quantum or classical. Let's assume we already know $E=hf$, and are trying to derive the relationship between wavelength and momentum:

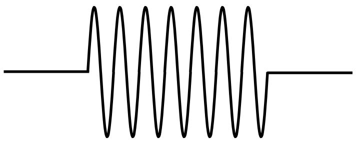
$$\begin{aligned}
\lambda &= v/f \\
&= \frac{vh}{E} \\
&= \frac{vh}{\frac{1}{2}mv^2} \\
&= \frac{2h}{mv} \\
&= \frac{2h}{p} \quad .
\end{aligned}$$

The reasoning seems valid, but the result does contradict the accepted one, which is after solidly based on experiment.

The mistaken assumption is that we can figure everything out in terms of pure sine waves. Mathematically, the only wave that has a perfectly well defined wavelength and frequency is a sine wave, and not just any sine wave but an infinitely long sine wave, (a). The unphysical thing about such a wave is that it has no leading or trailing edge, so it can never be said to enter or leave any particular region of space. Our derivation made use of the velocity, v, and if velocity is to be a meaningful concept, it must tell us how quickly stuff (mass, energy, momentum,...) is transported from one region of space to another. Since an infinitely long sine wave doesn't remove any stuff from one region and take it to another, the "velocity of its stuff" is not a well defined concept.

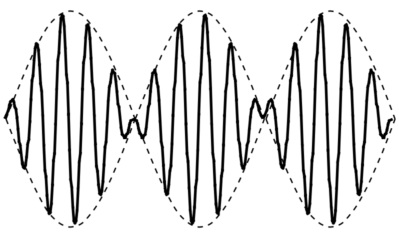

(a) Part of an infinite sine wave.

Of course the individual wave peaks do travel through space, and one might think that it would make sense to associate their speed with the "speed of stuff," but as we will see, the two velocities are in general unequal when a wave's velocity depends on wavelength. Such a wave is called a *dispersive* wave, because a wave pulse consisting of a superposition of waves of different wavelengths will separate (disperse) into its separate wavelengths as the waves move through space at different speeds. Nearly all the waves we have encountered have been nondispersive. For instance, sound waves and light waves (in a vacuum) have speeds independent of wavelength. A water wave is one good example of a dispersive wave. Long-wavelength water waves travel faster, so a ship at sea that encounters a storm typically sees the long-wavelength parts of the wave first. When dealing with dispersive waves, we need symbols and words to distinguish the two speeds. The speed at which wave peaks move is called the phase velocity, v_p, and the speed at which "stuff" moves is called the group velocity, v_g.



(b) A finite-length sine wave.

An infinite sine wave can only tell us about the phase velocity, not the group velocity, which is really what we would be talking about when we refer to the speed of an electron. If an infinite sine wave is the simplest possible wave, what's the next best thing? We might think the runner up in simplicity would be a wave train consisting of a chopped-off segment of a sine wave, (b). However, this kind of wave has kinks in it at the end. A simple wave should be one that we can build by superposing a small number of infinite sine waves, but a kink can never be produced by superposing any number of infinitely long sine waves.



(c) A beat pattern created by superimposing two sine waves with slightly different wavelengths.

Actually the simplest wave that transports stuff from place to place is the pattern shown in figure (c). Called a beat pattern, it is formed by superposing two sine waves whose wavelengths are similar but not quite the same. If you have ever heard the pulsating howling sound of musicians in the process of tuning their instruments to each other, you have heard a beat pattern. The beat pattern gets stronger and weaker as the two sine waves go in and out of phase with each other. The beat pattern has more "stuff" (energy, for example) in the areas where constructive interference occurs, and less in the regions of cancellation. As the whole pattern moves through space, stuff is transported from some regions and into other ones.

If the frequency of the two sine waves differs by 10%, for instance, then ten periods will be occur between times when they are in phase. Another way of saying it is that the sinusoidal "envelope" (the dashed lines in figure (c)) has a frequency equal to the difference in frequency between the two waves. For instance, if the waves had frequencies of 100 Hz and 110 Hz, the frequency of the envelope would be 10 Hz.

To apply similar reasoning to the wavelength, we must define a quantity $z=1/\lambda$ that relates to wavelength in the same way that frequency relates to period. In terms of this new variable, the z of the envelope equals the difference between the z's of the two sine waves.

The group velocity is the speed at which the envelope moves through space. Let Δf and Δz be the differences between the frequencies and z's of the two sine waves, which means that they equal the frequency and z of the envelope. The group velocity is $v_g = f_{envelope} \cdot \lambda_{envelope} = \Delta f/\Delta z$. If Δf and Δz are sufficiently small, we can approximate this expression as a derivative,

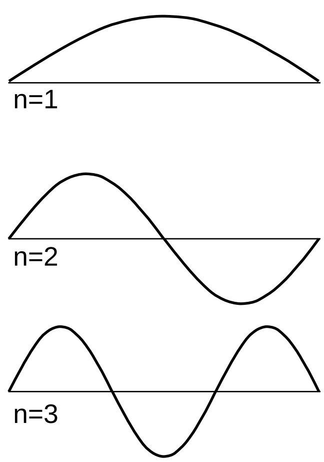
$$v_g = \frac{df}{dz} \quad .$$

This expression is usually taken as the definition of the group velocity for wave patterns that consist of a superposition of sine waves having a narrow range of frequencies and wavelengths. In quantum mechanics, with $f=E/h$ and $z=p/h$, we have $v_g=dE/dp$. In the case of a nonrelativistic electron the relationship between energy and momentum is $E=p^2/2m$, so the group velocity is $dE/dp=p/m=v$, exactly what it should be. It is only the phase velocity that is different from what we would have expected by a factor of two, but the phase velocity is not what is physically important.

5.3 Bound States

Electrons are at their most interesting when they're in atoms, that is, when they are bound within a small region of space. We can understand a great deal about atoms and molecules based on simple arguments about such bound states, without going into any of the realistic details of atom. The simplest model of a bound state is known as the particle in a box: like a ball on a pool table, the electron feels zero force while in the interior, but when it reaches an edge it encounters a wall that pushes back inward on it with a large force. In particle language, we would describe the electron as bouncing off of the wall, but this incorrectly assumes that the electron has a certain path through space. It is more correct to describe the electron as a wave that undergoes 100% reflection at the boundaries of the box.

Like a generation of physics students before me, I rolled my eyes when initially introduced to the unrealistic idea of putting a particle in a box. It seemed completely impractical, an artificial textbook invention. Today, however, it has become routine to study electrons in rectangular boxes in actual laboratory experiments. The "box" is actually just an empty cavity within a solid piece of silicon, amounting in volume to a few hundred atoms. The methods for creating these electron-in-a-box setups (known as "quantum dots") were a by-product of the development of technologies for fabricating computer chips.



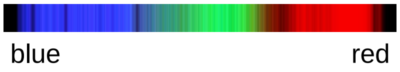
For simplicity let's imagine a one-dimensional electron in a box, i.e. we assume that the electron is only free to move along a line. The resulting standing wave patterns, of which the first three are shown in the figure, are just like some of the patterns we encountered with sound waves in musical instruments. The wave patterns must be zero at the ends of the box, because we are assuming the walls are impenetrable, and there should therefore be zero probability of finding the electron outside the box. Each wave pattern is labeled according to n, the number of peaks and valleys it has. In quantum physics, these wave patterns are referred to as "states" of the particle-in-the-box system.

The following seemingly innocuous observations about the particle in the box lead us directly to the solutions to some of the most vexing failures of classical physics:

The particle's energy is quantized (can only have certain values). Each wavelength corresponds to a certain momentum, and a given momentum implies a definite kinetic energy, $E=p^2/2m$. (This is the second type of energy quantization we have encountered. The type we studied previously had to do with restricting the number of particles to a whole number, while assuming some specific wavelength and energy for each particle. This type of quantization refers to the energies that a single particle can have. Both photons and matter particles demonstrate both types of quantization under the appropriate circumstances.)

The particle has a minimum kinetic energy. Long wavelengths correspond to low momenta and low energies. There can be no state with an energy lower than that of the n=1 state, called the ground state.

The smaller the space in which the particle is confined, the higher its kinetic energy must be. Again, this is because long wavelengths give lower energies.



The spectrum of the light from the star Sirius.
Photograph by the author.

Example: spectra of thin gases

A fact that was inexplicable by classical physics was that thin gases absorb and emit light only at certain wavelengths. This was observed both in earthbound laboratories and in the spectra of stars. The figure on the left shows the example of the spectrum of the star Sirius, in which there are "gap teeth" at certain wavelengths. Taking this spectrum as an example, we can give a straightforward explanation using quantum physics.

Energy is released in the dense interior of the star, but the outer layers of the star are thin, so the atoms are far apart and electrons are confined within individual atoms. Although their standing-wave patterns are not as simple as those of the particle in the box, their energies are quantized.

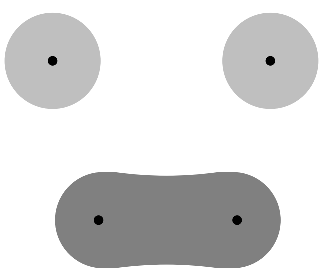
When a photon is on its way out through the outer layers, it can be absorbed by an electron in an atom, but only if the amount of energy it carries happens to be the right amount to kick the electron from one of the allowed energy levels to one of the higher levels. The photon energies that are missing from the spectrum are the ones that equal the difference in energy between two electron energy levels. (The most prominent of the absorption lines in Sirius's spectrum are absorption lines of the hydrogen atom.)

Example: the stability of atoms

In many Star Trek episodes the Enterprise, in orbit around a planet, suddenly lost engine power and began spiraling down toward the planet's surface. This was utter nonsense, of course, due to conservation of energy: the ship had no way of getting rid of energy, so it did not need the engines to replenish it.

Consider, however, the electron in an atom as it orbits the nucleus. The electron *does* have a way to release energy: it has an acceleration due to its continuously changing direction of motion, and according to classical physics, any accelerating charged particle emits electromagnetic waves. According to classical physics, atoms should collapse!

The solution lies in the observation that a bound state has a minimum energy. An electron in one of the higher-energy atomic states can and does emit photons and hop down step by step in energy. But once it is in the ground state, it cannot emit a photon because there is no lower-energy state for it to go to.



Two hydrogen atoms bond to form an H_2 molecule. In the molecule, the two electrons' wave patterns overlap, and are about twice as wide.

Example: chemical bonds

I began this chapter with a classical argument that chemical bonds, as in an H_2 molecule, should not exist. Quantum physics explains why this type of bonding does in fact occur. When the atoms are next to each other, the electrons are shared between them. The "box" is about twice as wide, and a larger box allows a smaller energy. Energy is required in order to separate the atoms. (A qualitatively different type of bonding is discussed in section 6.6.)

Discussion Question

A. Neutrons attract each other via the strong nuclear force, so according to classical physics it should be possible to form nuclei out of clusters of two or more neutrons, with no protons at all. Experimental searches, however, have failed to turn up evidence of a stable two-neutron system (dineutron) or larger stable clusters. Explain based on quantum physics why a dineutron might spontaneously fly apart.

B. The following table shows the energy gap between the ground state and the first excited state for four nuclei in units of picojoules. (The nuclei have been chosen to be ones that have similar structures, e.g. they are all spherical nuclei.)

nucleus	energy gap
^{4}He	3.234 pJ
^{16}O	0.968
^{40}Ca	0.536
^{208}Pb	0.418

Explain the trend in the data.

5.4 The Uncertainty Principle and Measurement

The uncertainty principle

Eliminating randomness through measurement?

A common reaction to quantum physics, among both early-twentieth-century physicists and modern students, is that we should be able to get rid of randomness through accurate measurement. If I say, for example, that it is meaningless to discuss the path of a photon or an electron, one might suggest that we simply measure the particle's position and velocity many times in a row. This series of snapshots would amount to a description of its path.

A practical objection to this plan is that the process of measurement will have an effect on the thing we are trying to measure. This may not be of much concern, for example, when a traffic cop measure's your car's motion with a radar gun, because the energy and momentum of the radar pulses are insufficient to change the car's motion significantly. But on the subatomic scale it is a very real problem. Making a videotape through a microscope of an electron orbiting a nucleus is not just difficult, it is theoretically impossible. The video camera makes pictures of things using light that has bounced off them and come into the camera. If even a single photon of visible light was to bounce off of the electron we were trying to study, the electron's recoil would be enough to change its behavior completely.

The Heisenberg uncertainty principle

This insight, that measurement changes the thing being measured, is the kind of idea that clove-cigarette-smoking intellectuals outside of the physical sciences like to claim they knew all along. If only, they say, the physicists had made more of a habit of reading literary journals, they could have saved a lot of work. The anthropologist Margaret Mead has recently been accused of inadvertently encouraging her teenaged Samoan informants to exaggerate the freedom of youthful sexual experimentation in their society. If this is considered a damning critique of her work, it is because she could have done better: other anthropologists claim to have been able to eliminate the observer-as-participant problem and collect untainted data.

The German physicist Werner Heisenberg, however, showed that in quantum physics, *any* measuring technique runs into a brick wall when we try to improve its accuracy beyond a certain point. Heisenberg showed that the limitation is a matter of *what there is to be known*, even in principle, about the system itself, not of the ability or inability of a specific measuring device to ferret out information that is knowable but not previously hidden.

Suppose, for example, that we have constructed an electron in a box (quantum dot) setup in our laboratory, and we are able adjust the length L of the box as desired. All the standing wave patterns pretty much fill the box, so our knowledge of the electron's position is of limited accuracy. If we write Δx for the range of uncertainty in our knowledge of its position, then Δx is roughly the same as the length of the box:

$$\Delta x \approx L \qquad (1)$$

If we wish to know its position more accurately, we can certainly squeeze it into a smaller space by reducing L, but this has an unintended side-effect. A standing wave is really a superposition of two traveling waves going in opposite directions. The equation $p=h/\lambda$ really only gives the magnitude of the momentum vector, not its direction, so we should really interpret the wave as a 50/50 mixture of a right-going wave with momentum $p=h/\lambda$ and a left-going one with momentum $p=-h/\lambda$. The uncertainty in our knowledge of the electron's momentum is $\Delta p=2h/\lambda$, covering the range between these two values. Even if we make sure the electron is in the ground state, whose wavelength $\lambda=2L$ is the longest possible, we have an uncertainty in momentum of $\Delta p=h/L$. In general, we find

$$\Delta p \geq h/L \quad , \qquad (2)$$

with equality for the ground state and inequality for the higher-energy states. Thus if we reduce L to improve our knowledge of the electron's position, we do so at the cost of knowing less about its momentum. This trade-off is neatly summarized by multiplying equations (1) and (2) to give

$$\Delta p \, \Delta x \gtrsim h$$

Although we have derived this in the special case of a particle in a box, it is an example of a principle of more general validity:

> **The Heisenberg uncertainty principle:**
> It is not possible, even in principle, to know the momentum and the position of a particle simultaneously and with perfect accuracy. The uncertainties in these two quantities are always such that $\Delta p \, \Delta x \gtrsim h$.

(This approximation can be made into a strict inequality, $\Delta p \, \Delta x \geq h/4\pi$, but only with more careful definitions, which we will not bother with.)

Note that although I encouraged you to think of this derivation in terms of a specific real-world system, the quantum dot, no reference was ever made to any specific laboratory equipment or procedures. The argument is simply that we cannot *know* the particle's position very accurately unless it *has* a very well defined position, it cannot have a very well defined position unless its wave-pattern covers only a very small amount of space, and its wave-pattern cannot be thus compressed without giving it a short wavelength and a correspondingly uncertain momentum. The uncertainty principle is therefore a restriction on how much there is to know about a particle, not just on what we can know about it with a certain technique.

Example: an estimate for electrons in atoms

Question: A typical energy for an electron in an atom is on the order of 1 volt · *e*, which corresponds to a speed of about 1% of the speed of light. If a typical atom has a size on the order of 0.1 nm, how close are the electrons to the limit imposed by the uncertainty principle?

Solution: If we assume the electron moves in all directions with equal probability, the uncertainty in its momentum is roughly twice its typical momentum. This only an order-of-magnitude estimate, so we take Δp to be the same as a typical momentum:

$$\begin{aligned}\Delta p\,\Delta x &= p_{typical}\,\Delta x \\ &= (m_{electron})\,(0.01c)\,(0.1\text{x}10^{-9}\text{ m}) \\ &= 3\text{x}10^{-34}\text{ J}\cdot\text{s}\end{aligned}$$

This is on the same order of magnitude as Planck's constant, so evidently the electron is "right up against the wall." (The fact that it is somewhat less than *h* is of no concern since this was only an estimate, and we have not stated the uncertainty principle in its most exact form.)

Self-Check

If we were to apply the uncertainty principle to human-scale objects, what would be the significance of the small numerical value of Planck's constant?

Measurement and Schrödinger's cat

In the previous chapter I briefly mentioned an issue concerning measurement that we are now ready to address carefully. If you hang around a laboratory where quantum-physics experiments are being done and secretly record the physicists' conversations, you'll hear them say many things that assume the probability interpretation of quantum mechanics. Usually they will speak as though the randomness of quantum mechanics enters the picture when something is measured. In the digital camera experiments of the previous chapter, for example, they would casually describe the detection of a photon at one of the pixels as if the moment of detection was when the photon was forced to "make up its mind." Although this mental cartoon usually works fairly well as a description of things they experience in the lab, it cannot ultimately be correct, because it attributes a special role to measurement, which is really just a physical process like all other physical processes.

If we are to find an interpretation that avoids giving any special role to measurement processes, then we must think of the entire laboratory, including the measuring devices and the physicists themselves, as one big quantum-mechanical system made out of protons, neutrons, electrons, and photons. In other words, we should take quantum physics seriously as a description not just of microscopic objects like atoms but of human-scale ("macroscopic") things like the apparatus, the furniture, and the people.

The most celebrated example is called the Schrödinger's cat experiment. Luckily for the cat, there probably was no actual experiment — it was

Under the ordinary circumstances of life, the accuracy with which we can measure position and momentum of an object doesn't result in a value of $\Delta p \Delta x$ that is anywhere near the tiny order of magnitude of Planck's constant. We run up against the ordinary limitations on the accuracy of our measuring techniques long before the uncertainty principle becomes an issue.

simply a "thought experiment" that the physicist the German theorist Schrödinger discussed with his colleagues. Schrödinger wrote:

> One can even construct quite burlesque cases. A cat is shut up in a steel container, together with the following diabolical apparatus (which one must keep out of the direct clutches of the cat): In a Geiger tube [radiation detector] there is a tiny mass of radioactive substance, so little that in the course of an hour perhaps one atom of it disintegrates, but also with equal probability not even one; if it does happen, the counter [detector] responds and ... activates a hammer that shatters a little flask of prussic acid [filling the chamber with poison gas]. If one has left this entire system to itself for an hour, then one will say to himself that the cat is still living, if in that time no atom has disintegrated. The first atomic disintegration would have poisoned it.

Now comes the strange part. Quantum mechanics describes the particles the cat is made of as having wave properties, including the property of superposition. Schrödinger describes the wavefunction of the box's contents at the end of the hour:

> The wavefunction of the entire system would express this situation by having the living and the dead cat mixed ... in equal parts [50/50 proportions]. The uncertainty originally restricted to the atomic domain has been transformed into a macroscopic uncertainty...

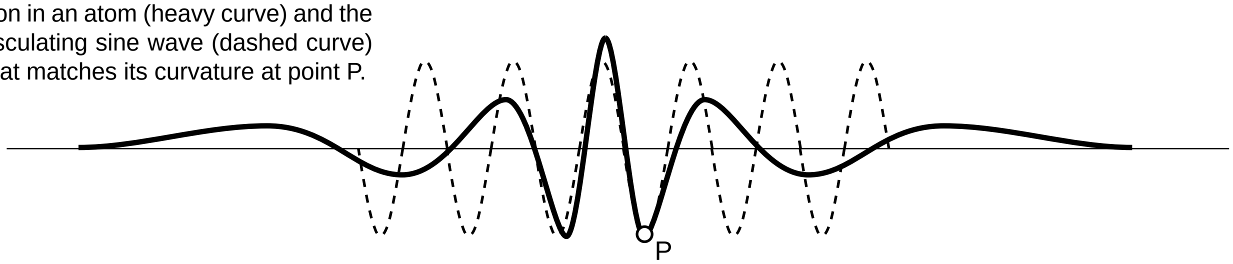
At first Schrödinger's description seems like nonsense. When you opened the box, would you see two ghostlike cats, as in a doubly exposed photograph, one dead and one alive? Obviously not. You would have a single, fully material cat, which would either be dead or very, very upset. But Schrödinger has an equally strange and logical answer for that objection. In the same way that the quantum randomness of the radioactive atom spread to the cat and made the its wavefunction a random mixture of life and death, the randomness spreads wider once you open the box, and your own wavefunction becomes a mixture of a person who has just killed a cat and a person who hasn't.

Discussion Question

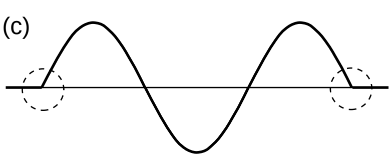
A. Compare Δp and Δx for the two loest energy levels of the one-dimensional particle in a box, and discuss how this relates to the uncertainty principle.

B. On a graph of Δp versus Δx, sketch the regions that are allowed and forbidden by the Heisenberg uncertainty principle. Interpret the graph: Where does an atom lie on it? An elephant? Can either p or x be measured with perfect accuracy if we don't care about the other?

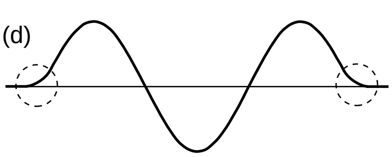
(b) A typical wavefunction of an electron in an atom (heavy curve) and the osculating sine wave (dashed curve) that matches its curvature at point P.



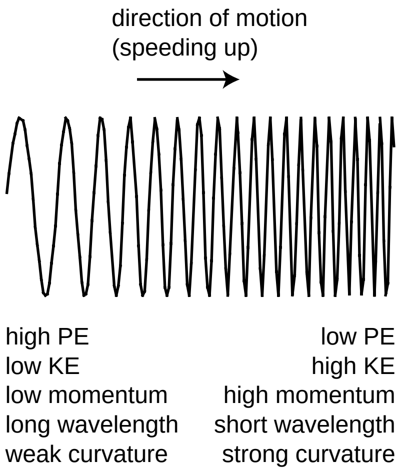
(c)



(d)



5.5 Electrons in Electric Fields



(a) An electron in a gentle electric field gradually shortens its wavelength as it gains energy.

So far the only electron wave patterns we've considered have been simple sine waves, but whenever an electron finds itself in an electric field, it must have a more complicated wave pattern. Let's consider the example of an electron being accelerated by the electron gun at the back of a TV tube. Newton's laws are not useful, because they implicitly assume that the path taken by the particle is a meaningful concept. Conservation of energy is still valid in quantum physics, however. In terms of energy, the electron is moving from a region of low voltage into a region of higher voltage. Since its charge is negative, it loses PE by moving to a higher voltage, so its KE increases. As its potential energy goes down, its kinetic energy goes up by an equal amount, keeping the total energy constant. Increasing kinetic energy implies a growing momentum, and therefore a shortening wavelength, (a).

The wavefunction as a whole does not have a single well-defined wavelength, but the wave changes so gradually that if you only look at a small part of it you can still pick out a wavelength and relate it to the momentum and energy. (The picture actually exaggerates by many orders of magnitude the rate at which the wavelength changes.)

But what if the electric field was stronger? The electric field in a TV is only ~10^5 N/C, but the electric field within an atom is more like 10^{12} N/C. In figure (b), the wavelength changes so rapidly that there is nothing that looks like a sine wave at all. We could get a rough idea of the wavelength in a given region by measuring the distance between two peaks, but that would only be a rough approximation. Suppose we want to know the wavelength at point P. The trick is to construct a sine wave, like the one shown with the dashed line, which matches the curvature of the actual wavefunction as closely as possible near P. The sine wave that matches as well as possible is called the "osculating" curve, from a Latin word meaning "to kiss." The wavelength of the osculating curve is the wavelength that will relate correctly to conservation of energy.

Tunneling

We implicitly assumed that the particle-in-a-box wavefunction would cut off abruptly at the sides of the box, (c), but that would be unphysical. A kink has infinite curvature, and curvature is related to energy, so it can't be infinite. A physically realistic wavefunction must always "tail off" gradually, (d). In classical physics, a particle can never enter a region in which its potential energy would be greater than the amount of energy it has available. But in quantum physics the wavefunction will always have a tail that reaches into the classically forbidden region. If it was not for this effect, called tunneling, the fusion reactions that power the sun would not occur due to the high potential energy nuclei need in order to get close together! Tunneling is discussed in more detail in the following section.

5.6*∫ The Schrödinger Equation

In the previous section we were able to apply conservation of energy to an electron's wavefunction, but only by using the clumsy graphical technique of osculating sine waves as a measure of the wave's curvature. You have learned a more convenient measure of curvature in calculus: the second derivative. To relate the two approaches, we take the second derivative of a sine wave:

$$\begin{aligned}
&\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sin\left(2\pi x/\lambda\right) \\
&\quad = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{2\pi}{\lambda}\cos\left(2\pi x/\lambda\right)\right] \\
&\quad = -\left(\frac{2\pi}{\lambda}\right)^2\sin\left(2\pi x/\lambda\right)
\end{aligned}$$

Taking the second derivative gives us back the same function, but with a minus sign and a constant out in front that is related to the wavelength. We can thus relate the second derivative to the osculating wavelength:

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} = -\left(\frac{2\pi}{\lambda}\right)^2\Psi \qquad (1)$$

This could be solved for λ in terms of Ψ, but it will turn out below to be more convenient to leave it in this form.

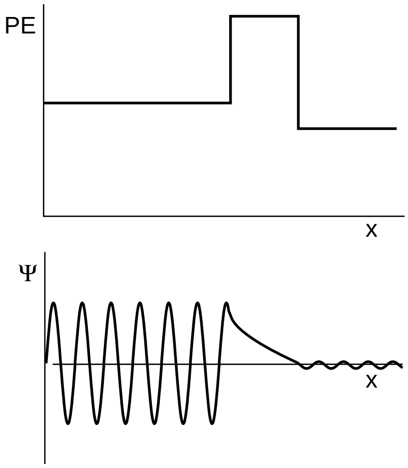
Applying this to conservation of energy, we have

$$\begin{aligned}
E &= KE + PE \\
&= \frac{p^2}{2m} + PE \\
&= \frac{\left(h/\lambda\right)^2}{2m} + PE \qquad (2)
\end{aligned}$$

Note that both equation (1) and equation (2) have λ^2 in the denominator. We can simplify our algebra by multiplying both sides of equation (2) by Ψ to make it look more like equation (1):

$$\begin{aligned}
E\cdot\Psi &= \frac{\left(h/\lambda\right)^2}{2m}\Psi + PE\cdot\Psi \\
&= \frac{1}{2m}\left(\frac{h}{2\pi}\right)^2\left(\frac{2\pi}{\lambda}\right)^2\Psi + PE\cdot\Psi \\
&= -\frac{1}{2m}\left(\frac{h}{2\pi}\right)^2\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} + PE\cdot\Psi
\end{aligned}$$

No. The equation $KE=p^2/2m$ is nonrelativistic, so it can't be applied to an electron moving at relativistic speeds. Photons always move at relativistic speeds, so it can't be applied to them either.



Some simplification is achieved by using the symbol $\hbar$ (h with a slash through it, read "h-bar") as an abbreviation for $h/2\pi$. We then have the important equation known as the Schrödinger equation:

$$E \cdot \Psi = -\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + PE \cdot \Psi$$

(Actually this is a simplified version of the Schrödinger equation, applying only to standing waves in one dimension.) Physically it is a statement of conservation of energy. The total energy E must be constant, so the equation tells us that a change in potential energy must be accompanied by a change in the curvature of the wavefunction. This change in curvature relates to a change in wavelength, which corresponds to a change in momentum and kinetic energy.

Self-Check

Considering the assumptions that were made in deriving the Schrödinger equation, would it be correct to apply it to a photon? To an electron moving at relativistic speeds?

Usually we know right off the bat how PE depends on x, so the basic mathematical problem of quantum physics is to find a function $\Psi(x)$ that satisfies the Schrödinger equation for a given potential-energy function $PE(x)$. An equation, such as the Schrödinger equation, that specifies a relationship between a function and its derivatives is known as a differential equations.

The study of differential equations in general is beyond the mathematical level of this book, but we can gain some important insights by considering the easiest version of the Schrödinger equation, in which the potential energy is constant. We can then rearrange the Schrödinger equation as follows:

$$\frac{d^2\Psi}{dx^2} = \frac{2m\left(PE - E\right)}{\hbar^2}\Psi \quad ,$$

which boils down to

$$\frac{d^2\Psi}{dx^2} = a\Psi \quad ,$$

where, according to our assumptions, a is independent of x. We need to find a function whose second derivative is the same as the original function except for a multiplicative constant. The only functions with this property are sine waves and exponentials:

$$\frac{d^2}{dx^2}\left[q \sin\left(rx + s\right)\right] = -qr^2 \sin\left(rx + s\right)$$

$$\frac{d^2}{dx^2}\left[qe^{rx+s}\right] = qr^2 e^{rx+s}$$

Dividing by Planck's constant, a small number, gives a large negative result inside the exponential, so the probability will be very small.

The sine wave gives negative values of a, $a=-r^2$, and the exponential gives positive ones, $a=r^2$. The former applies to the classically allowed region with $PE<E$, the latter to the classical forbidden region with $PE>E$.

This leads us to a quantitative calculation of the tunneling effect discussed briefly in the previous section. The wavefunction evidently tails off exponentially in the classically forbidden region. Suppose, as shown in the figure, a wave-particle traveling to the right encounters a barrier that it is classically forbidden to enter. Although the form of the Schrödinger equation we're using technically does not apply to traveling waves (because it makes no reference to time), it turns out that we can still use it to make a reasonable calculation of the probability that the particle will make it through the barrier. If we let the barrier's width be w, then the ratio of the wavefunction on the left side of the barrier to the wavefunction on the right is

$$\frac{qe^{rx+s}}{qe^{r(x+w)+s}} = e^{-rw} \quad .$$

Probabilities are proportional to the squares of wavefunctions, so the probability of making it through the barrier is

$$\begin{aligned} P &= e^{-2rw} \\ &= \exp\left(-\frac{2w}{\hbar}\sqrt{2m\left(PE-E\right)}\right) \end{aligned}$$

Self-Check

If we were to apply this equation to find the probability that a person can walk through a wall, what would the small value of Planck's constant imply?

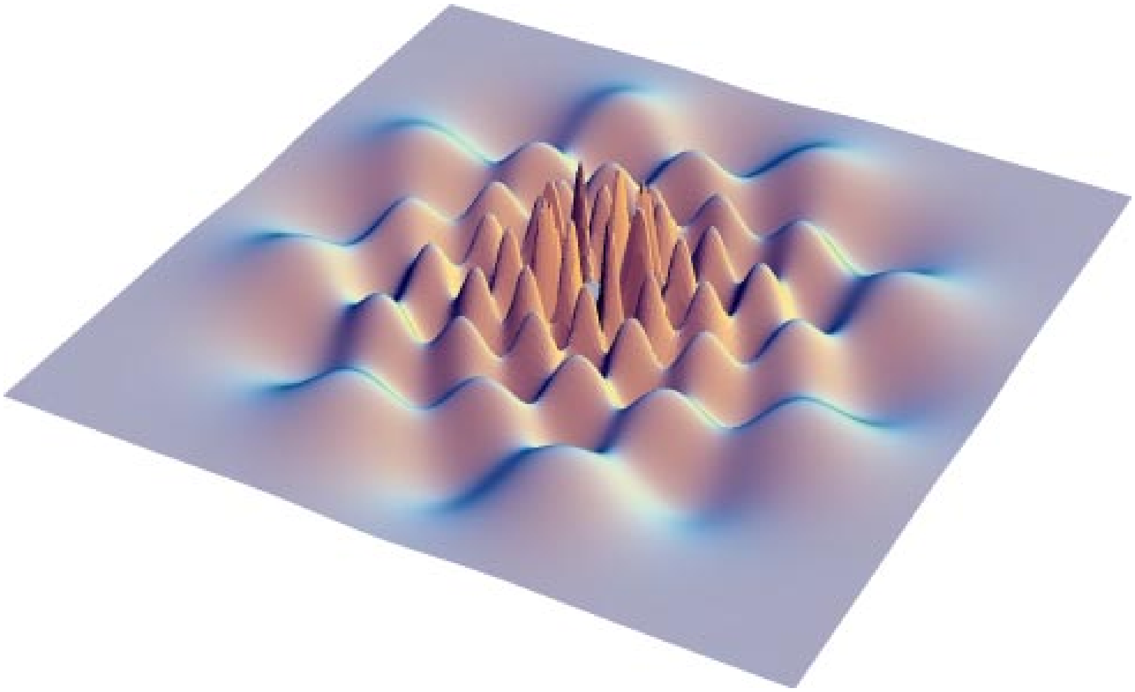
Use of complex numbers

In a classically forbidden region, a particle's total energy, $PE+KE$, is less than its PE, so its KE must be negative. If we want to keep believing in the equation $KE=p^2/2m$, then apparently the momentum of the particle is the square root of a negative number. This is a symptom of the fact that the Schrödinger equation fails to describe all of nature unless the wavefunction and various other quantities are allowed to be complex numbers. In particular it is not possible to describe traveling waves correctly without using complex wavefunctions.

This may seem like nonsense, since real numbers are the only ones that are, well, real! Quantum mechanics can always be related to the real world, however, because its structure is such that the results of measurements always come out to be real numbers. For example, we may describe an electron as having non-real momentum in classically forbidden regions, but its average momentum will always come out to be real (the imaginary parts average out to zero), and it can never transfer a non-real quantity of momentum to another particle.

A complete investigation of these issues is beyond the scope of this book, and this is why we have normally limited ourselves to standing waves, which can be described with real-valued wavefunctions.

S A solution is given in the back of the book.
✓ A computerized answer check is available.
★ A difficult problem.
∫ A problem that requires calculus.



6 The Atom

You can learn a lot by taking a car engine apart, but you will have learned a lot more if you can put it all back together again and make it run. Half the job of reductionism is to break nature down into its smallest parts and understand the rules those parts obey. The second half is to show how those parts go together, and that is our goal in this chapter. We have seen how certain features of all atoms can be explained on a generic basis in terms of the properties of bound states, but this kind of argument clearly cannot tell us any details of the behavior of an atom or explain why one atom acts differently from another.

The biggest embarrassment for reductionists is that the job of putting things back together job is usually much harder than the taking them apart. Seventy years after the fundamentals of atomic physics were solved, it is only beginning to be possible to calculate accurately the properties of atoms that have many electrons. Systems consisting of many atoms are even harder. Supercomputer manufacturers point to the folding of large protein molecules as a process whose calculation is just barely feasible with their fastest machines. The goal of this chapter is to give a gentle and visually oriented guide to some of the simpler results about atoms.

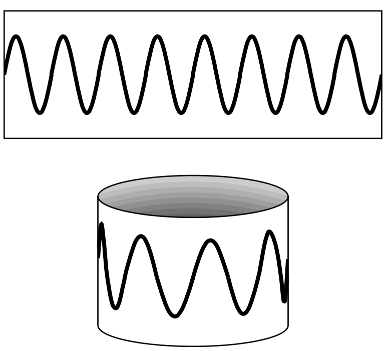
6.1 Classifying States

We'll focus our attention first on the simplest atom, hydrogen, with one proton and one electron. We know in advance a little of what we should expect for the structure of this atom. Since the electron is bound to the proton by electrical forces, it should display a set of discrete energy states, each corresponding to a certain standing wave pattern. We need to understand what states there are and what their properties are.

What properties should we use to classify the states? The most sensible approach is to used conserved quantities. Energy is one conserved quantity, and we already know to expect each state to have a specific energy. It turns out, however, that energy alone is not sufficient. Different standing wave patterns of the atom can have the same energy.

Momentum is also a conserved quantity, but it is not particularly appropriate for classifying the states of the electron in a hydrogen atom. The reason is that the force between the electron and the proton results in the continual exchange of momentum between them. (Why wasn't this a problem for energy as well? Kinetic energy and momentum are related by $KE=p^2/2m$, so the much more massive proton never has very much kinetic energy. We are making an approximation by assuming all the kinetic energy is in the electron, but it is quite a good approximation.)

Angular momentum does help with classification. There is no transfer of angular momentum between the proton and the electron, since the force between them is a center-to-center force, producing no torque.



Eight wavelengths fit around this circle (ℓ =8).

Like energy, angular momentum is quantized in quantum physics. As an example, consider a quantum wave-particle confined to a circle, like a wave in a circular moat surrounding a castle. A sine wave in such a "quantum moat" cannot have any old wavelength, because an integer number of wavelengths must fit around the circumference, C, of the moat. The larger this integer is, the shorter the wavelength, and a shorter wavelength relates to greater momentum and angular momentum. Since this integer is related to angular momentum, we use the symbol ℓ for it:

$$\lambda = C/\ell$$

Its angular momentum is

$$L = rp \quad .$$

Here, $r=C/2\pi$, and $p = h/\lambda = h\ell/C$, so

$$L = \frac{C}{2\pi}\cdot\frac{h\ell}{C}$$

$$= \frac{h}{2\pi}\ell$$

In the example of the quantum moat, angular momentum is quantized in units of $h/2\pi$. This makes $h/2\pi$ a pretty important number, so we define the abbreviation $\hbar = h/2\pi$. This symbol is read "h-bar."

In fact, this is a completely general fact in quantum physics, not just a fact about the quantum moat:

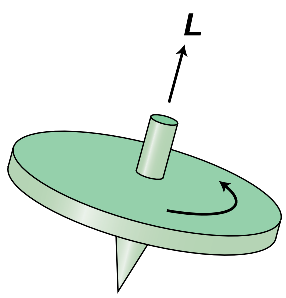
Quantization of angular momentum
The angular momentum of a particle due to its motion through space is quantized in units of $\hbar$.

Self-Check

What is the angular momentum of the wavefunction shown at the beginning of the chapter?

6.2 Angular Momentum in Three Dimensions

Up until now we've only worked with angular momentum in the context of rotation in a plane, for which we could simply use positive and negative signs to indicate clockwise and counterclockwise directions of rotation. A hydrogen atom, however, is unavoidably three-dimensional. Let's first consider the generalization of angular momentum to three dimensions in the classical case, and then consider how it carries over into quantum physics.



(a) The angular momentum vector of a spinning top.

Three-dimensional angular momentum in classical physics

If we are to completely specify the angular momentum of a classical object like a top, (a), in three dimensions, it's not enough to say whether the rotation is clockwise or counterclockwise. We must also give the orientation of the plane of rotation or, equivalently, the direction of the top's axis. The convention is to specify the direction of the axis. There are two possible directions along the axis, and as a matter of convention we use the direction such that if we sight along it, the rotation appears clockwise.

Angular momentum can, in fact, be defined as a vector pointing along this direction. This might seem like a strange definition, since nothing actually moves in that direction, but it wouldn't make sense to define the angular momentum vector as being in the direction of motion, because every part of the top has a different direction of motion. Ultimately it's not just a matter of picking a definition that is convenient and unambiguous: the definition we're using is the only one that makes the total angular momentum of a system a conserved quantity if we let "total" mean the vector sum.

As with rotation in one dimension, we cannot define what we mean by angular momentum in a particular situation unless we pick a point as an axis. This is really a different use of the word "axis" than the one in the previous paragraphs. Here we simply mean a point from which we measure the distance r. In the hydrogen atom, the nearly immobile proton provides a natural choice of axis.

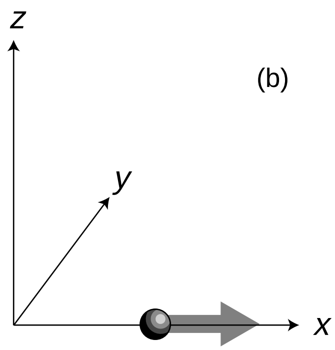
If you trace a circle going around the center, you run into a series of eight complete wavelengths. Its angular momentum is $8\hbar$.

Three-dimensional angular momentum in quantum physics

Once we start to think more carefully about the role of angular momentum in quantum physics, it may seem that there is a basic problem: the angular momentum of the electron in a hydrogen atom depends on both its distance from the proton and its momentum, so in order to know its angular momentum precisely it would seem we would need to know both its position and its momentum simultaneously with good accuracy. This, however, might seem to be forbidden by the Heisenberg uncertainty principle.

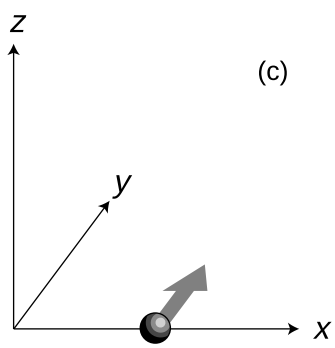
Actually the uncertainty principle does place limits on what can be known about a particle's angular momentum vector, but it does not prevent us from knowing its magnitude as an exact integer multiple of $\hbar$. The reason is that in three dimensions, there are really three separate uncertainty principles:

$$\Delta p_x \, \Delta x \gtrsim \hbar$$
$$\Delta p_y \, \Delta y \gtrsim \hbar$$
$$\Delta p_z \, \Delta z \gtrsim \hbar$$



(b)

Now consider a particle, (b), that is moving along the x axis at position x and with momentum p_x. We may not be able to know both x and p_x with unlimited accurately, but we can still know the particle's angular momentum about the origin exactly: it is zero, because the particle is moving directly away from the origin.



(c)

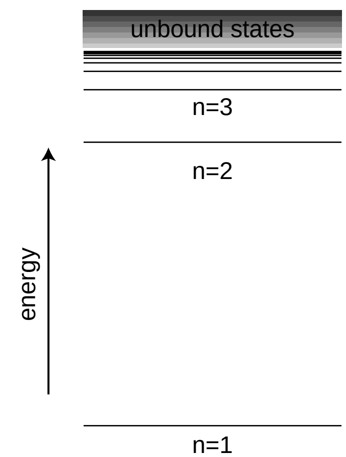
Suppose, on the other hand, a particle finds itself, (c), at a position x along the x axis, and it is moving parallel to the y axis with momentum p_y. It has angular momentum xp_y about the z axis, and again we can know its angular momentum with unlimited accuracy, because the uncertainty principle on relates x to p_x and y to p_y. It does not relate x to p_y.

As shown by these examples, the uncertainty principle does not restrict the accuracy of our knowledge of angular momenta as severely as might be imagined. However, it does prevent us from knowing all three components of an angular momentum vector simultaneously. The most general statement about this is the following theorem, which we present without proof:

The angular momentum vector in quantum physics

The most the can be known about an angular momentum vector is its magnitude and one of its three vector components. Both are quantized in units of $\hbar$.

6.3 The Hydrogen Atom

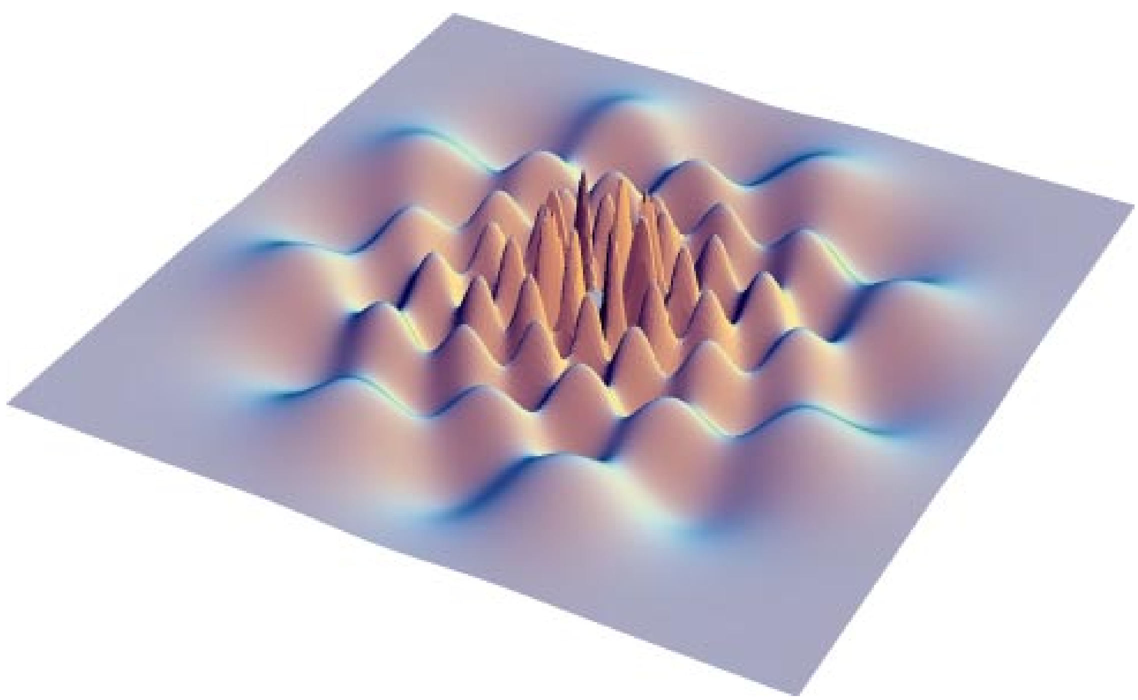


The energy of a state in the hydrogen atom depends only on its *n* quantum number.

Deriving the wavefunctions of the states of the hydrogen atom from first principles would be mathematically too complex for this book, but it's not hard to understand the logic behind such a wavefunction in visual terms. Consider the wavefunction from the beginning of the chapter, which is reproduced below. Although the graph looks three-dimensional, it is really only a representation of the part of the wavefunction lying within a two-dimensional plane. The third (up-down) dimension of the plot represents the value of the wavefunction at a given point, not the third dimension of space. The plane chosen for the graph is the one perpendicular to the angular momentum vector.

Each ring of peaks and valleys has eight wavelengths going around in a circle, so this state has $L=8\hbar$, i.e. we label it $\ell=8$. The wavelength is shorter near the center, and this makes sense because when the electron is close to the nucleus it has a lower *PE*, a higher *KE*, and a higher momentum.

Between each ring of peaks in this wavefunction is a nodal circle, i.e. a circle on which the wavefunction is zero. The full three-dimensional wavefunction has nodal spheres: a series of nested spherical surfaces on which it is zero. The number of radii at which nodes occur, including $r=\infty$, is called n, and n turns out to be closely related to energy. The ground state has $n=1$ (a single node only at $r=\infty$), and higher-energy states have higher n values. There is a simple equation relating n to energy, which we will discuss in section 6.4.



The numbers n and ℓ, which identify the state, are called its quantum numbers. A state of a given n and ℓ can be oriented in a variety of directions in space. We might try to indicate the orientation using the three quantum numbers $\ell_x = L_x/\hbar$, $\ell_y = L_y/\hbar$, and $\ell_z = L_z/\hbar$. But we have already seen that it is impossible to know all three of these simultaneously. To give the most complete possible description of a state, we choose an arbitrary axis, say the z axis, and label the state according to n, ℓ, and ℓ_z.

Angular momentum requires motion, and motion implies kinetic energy. Thus it is not possible to have a given amount of angular momentum without having a certain amount of kinetic energy as well. Since energy relates to the n quantum number, this means that for a given n value there will be a maximum possible ℓ. It turns out that this maximum value of ℓ equals n–1.

In general, we can list the possible combinations of quantum numbers as follows:

n can equal 1, 2, 3, ...
ℓ can range from 0 to n–1, in steps of 1
ℓ_z can range from $-\ell$ to ℓ, in steps of 1

Applying these, rules, we have the following list of states:

$n=1$, $\ell=0$, $\ell_z=0$	one state
$n=2$, $\ell=0$, $\ell_z=0$	one state
$n=2$, $\ell=1$, $\ell_z=-1$, 0, or 1	three states
etc.	

Self-Check

Continue the list for n=3.

The figures on the facing page show the lowest-energy states of the hydrogen atom. The left-hand column of graphs displays the wavefunctions in the x-y plane, and the right-hand column shows the probability density in a three-dimensional representation.

Discussion Questions

A. The quantum number n is defined as the number of radii at which the wavefunction is zero, including $r=\infty$. Relate this to the features of the figures on the facing page.

B. Based on the definition of n, why can't there be any such thing as an n=0 state?

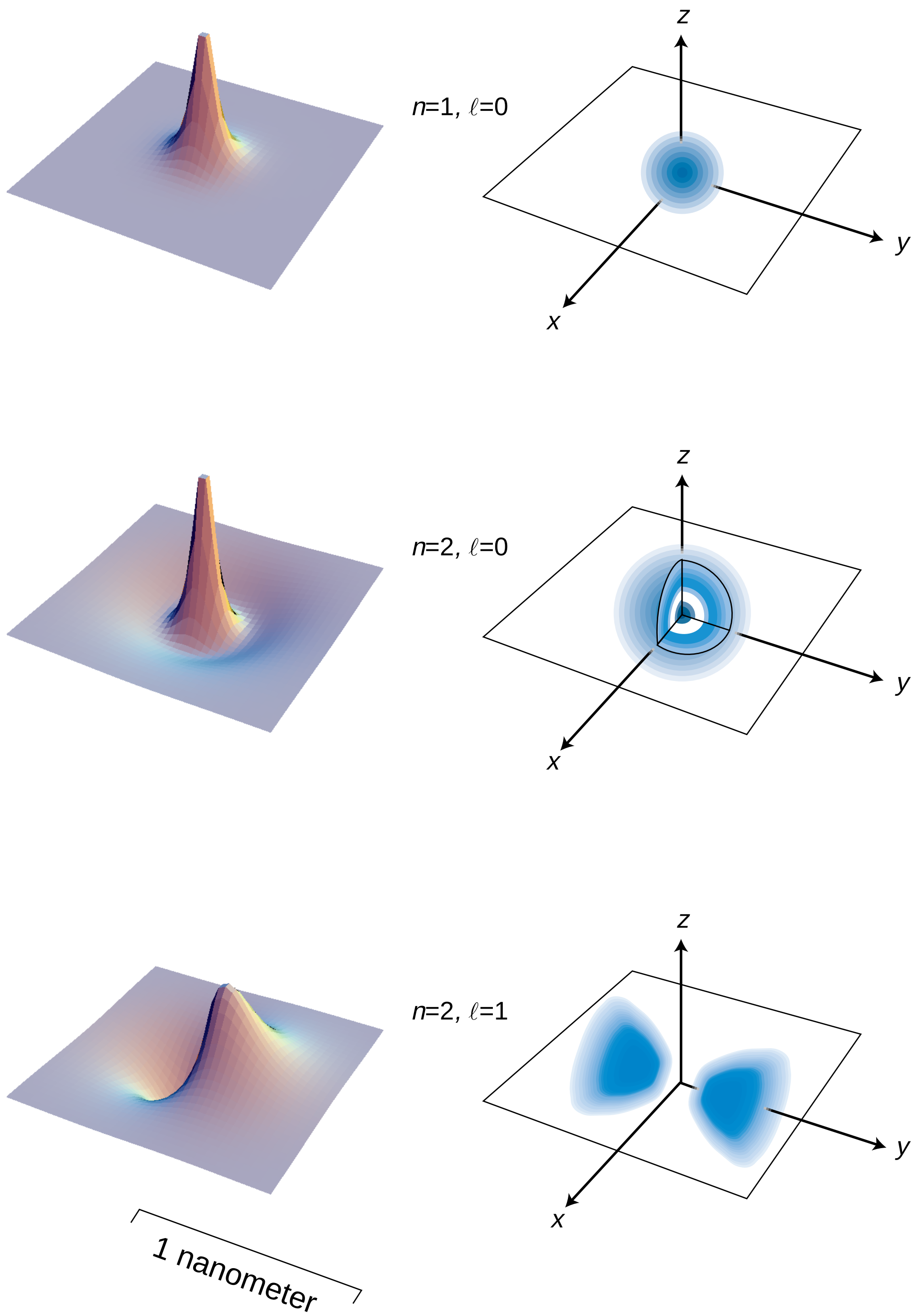
C. Relate the features of the wavefunction plots on the facing page to the corresponding features of the probability density pictures.

D. How can you tell from the wavefunction plots on the right which ones have which angular momenta?

E. Criticize the following incorrect statement: "The $\ell=8$ wavefunction on the previous page has a shorter wavelength in the center because in the center the electron is in a higher energy level."

F. Discuss the implications of the fact that the probability cloud in of the n=2, $\ell=1$ state is split into two parts.

$n=3$, $\ell=0$, $\ell_z=0$: one state; $n=3$, $\ell=1$, $\ell_z=-1$, 0, or 1: three states; $n=3$, $\ell=2$, $\ell_z=-2$, –1, 0, 1, or 2: five states


z
n=1, ℓ=0
y
x
z
n=2, ℓ=0
y
x
z
n=2, ℓ=1
y
x
1 nanometer

6.4* Energies of States in Hydrogen

The experimental technique for measuring the energy levels of an atom accurately is spectroscopy: the study of the spectrum of light emitted (or absorbed) by the atom. Only photons with certain energies can be emitted or absorbed by a hydrogen atom, for example, since the amount of energy gained or lost by the atom must equal the difference in energy between the atom's initial and final states. Spectroscopy had actually become a highly developed art several decades before Einstein even proposed the photon, and the Swiss spectroscopist Johann Balmer determined in 1885 that there was a simple equation that gave all the wavelengths emitted by hydrogen. In modern terms, we think of the photon wavelengths merely as indirect evidence about the underlying energy levels of the atom, and we rework Balmer's result into an equation for these atomic energy levels:

$$E_n = -\frac{2.2\times10^{-18}\ \mathrm{J}}{n^2} \quad ,$$

where we have made use of the electron-volt (1 volt multiplied by the charge of the electron) as a convenient energy unit on the atomic scale. This energy includes both the kinetic energy of the electron and the electrical potential energy. The zero-level of the potential energy scale is chosen to be the energy of an electron and a proton that are infinitely far apart. With this choice, negative energies correspond to bound states and positive energies to unbound ones.

Where does the mysterious numerical factor of $2.2\text{x}10^{-18}$ J come from? In 1913 the Danish theorist Niels Bohr realized that it was exactly numerically equal to a certain combination of fundamental physical constants:

$$E_n = -\frac{mk^2e^4}{2\hbar^2}\cdot\frac{1}{n^2} \quad ,$$

where m is the mass of the electron, k is the Coulomb force constant for electric forces, and $\hbar$ is the abbreviation for $h/2\pi$ that we have already mentioned in passing.

Bohr was able to cook up a derivation of this equation based on the incomplete version of quantum physics that had been developed by that time, but his derivation is today mainly of historical interest. It assumes that the electron follows a circular path, whereas the whole concept of a path for a particle is considered meaningless in our more complete modern version of quantum physics. Although Bohr was able to produce the right equation for the energy levels, his model also gave various wrong results, such as predicting that the atom would be flat, and that the ground state would have $\ell=1$ rather than the correct $\ell=0$.

A full and correct treatment is impossible at the mathematical level of this book, but we can provide a straightforward explanation for the form of the equation using approximate arguments. A typical standing-wave pattern for the electron consists of a central oscillating area surrounded by a region in which the wavefunction tails off. As discussed in section 5.5, the oscillating type of pattern is typically encountered in the classically allowed region, while the tailing off occurs in the classically forbidden region where the electron has insufficient kinetic energy to penetrate according to classical physics. We use the symbol r for the radius of the spherical boundary between the classically allowed and classically forbidden regions.

When the electron is at the distance r from the proton, it has zero kinetic energy — in classical terms, this would be the distance at which the electron would have to stop, turn around, and head back toward the proton. Thus when the electron is at distance r, its energy is purely potential:

$$E = -\frac{ke^2}{r} \qquad (1)$$

Now comes the approximation. In reality, the electron's wavelength cannot be constant in the classically allowed region, but we pretend that it is. Since n is the number of nodes in the wavefunction, we can interpret it approximately as the number of wavelengths that fit across the diameter $2r$. We are not even attempting a derivation that would produce all the correct numerical factors like 2 and π and so on, so we simply make the approximation

$$\lambda \sim \frac{r}{n} \quad . \qquad (2)$$

Finally we assume that the typical kinetic energy of the electron is on the same order of magnitude as the absolute value of its total energy. (This is true to within a factor of two for a typical classical system like a planet in a circular orbit around the sun.) We then have

$$\begin{aligned} &\text{absolute value of total energy} \\ &= \frac{ke^2}{r} \\ &\sim KE \\ &= p^2/2m \\ &= (h/\lambda)^2 / 2m \\ &\sim hn^2 / 2mr^2 \end{aligned} \qquad (3)$$

We now solve the equation $ke^2/r \sim hn^2 / 2mr^2$ for r and throw away numerical factors we can't hope to have gotten right, yielding

$$r \sim \frac{h^2 n^2}{mke^2} \qquad (4)$$

Plugging n=1 into this equation gives r=2 nm, which is indeed on the right order of magnitude. Finally we combine equations (4) and (1) to find

$$E \sim -\frac{mk^2 e^4}{h^2 n^2} \quad , \qquad (5)$$

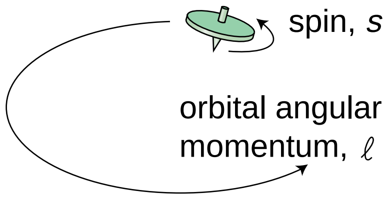
which is correct except for the numerical factors we never aimed to find.

Discussion Questions

A. States of hydrogen with n greater than about 10 are never observed in the sun. Why might this be?

B. Sketch graphs of r and E versus n for the hydrogen, and compare with analogous graphs for the one-dimensional particle in a box.

6.5 Electron Spin



The top has angular momentum both because of the motion of its center of mass through space and due to its internal rotation. Electron spin is roughly analogous to the intrinsic spin of the top.

It's disconcerting to the novice ping-pong player to encounter for the first time a more skilled player who can put spin on the ball. Even though you can't see that the ball is spinning, you can tell something is going on by the way it interacts with other objects in its environment. In the same way, we can tell from the way electrons interact with other things that they have an intrinsic spin of their own. Experiments show that even when an electron is not moving through space, it still has angular momentum amounting to $\hbar/2$.

This may seem paradoxical because the quantum moat, for instance, gave only angular momenta that were integer multiples of $\hbar$, not half-units, and I claimed that angular momentum was always quantized in units of $\hbar$, not just in the case of the quantum moat. That whole discussion, however, assumed that the angular momentum would come from the motion of a particle through space. The $\hbar/2$ angular momentum of the electron is simply a property of the particle, like its charge or its mass. It has nothing to do with whether the electron is moving or not, and it does not come from any internal motion within the electron. Nobody has ever succeeded in finding any internal structure inside the electron, and even if there was internal structure, it would be mathematically impossible for it to result in a half-unit of angular momentum.

We simply have to accept this $\hbar/2$ angular momentum, called the "spin" of the electron, as an experimentally proven fact. Protons and neutrons have the same $\hbar/2$ spin, while photons have an intrinsic spin of $\hbar$.

As was the case with ordinary angular momentum, we can describe spin angular momentum in terms of its magnitude, and its component along a given axis. The usual notation for these quantities, in units of $\hbar$, are s and s_z, so an electron has $s=1/2$ and $s_z=+1/2$ or $-1/2$.

Taking electron spin into account, we need a total of four quantum numbers to label a state of an electron in the hydrogen atom: n, ℓ, ℓ_z, and s_z. (We omit s because it always has the same value.) The symbols ℓ and ℓ_z include only the angular momentum the electron has because it is moving through space, not its spin angular momentum. The availability of two possible spin states of the electron leads to a doubling of the numbers of states:

$n=1$, $\ell=0$, $\ell_z=0$, $s_z=+1/2$ or $-1/2$	two states
$n=2$, $\ell=0$, $\ell_z=0$, $s_z=+1/2$ or $-1/2$	two states
$n=2$, $\ell=1$, $\ell_z=-1$, 0, or 1, $s_z=+1/2$ or $-1/2$	six states
...	

6.6 Atoms With More Than One Electron

What about other atoms besides hydrogen? It would seem that things would get much more complex with the addition of a second electron. A hydrogen atom only has one particle that moves around much, since the nucleus is so heavy and nearly immobile. Helium, with two, would be a mess. Instead of a wavefunction whose square tells us the probability of finding a single electron at any given location in space, a helium atom would need to have a wavefunction whose square would tell us the probability of finding two electrons at any given combination of points. Ouch! In addition, we would have the extra complication of the electrical interaction between the two electrons, rather than being able to imagine everything in terms of an electron moving in a static field of force created by the nucleus alone.

Despite all this, it turns out that we can get a surprisingly good description of many-electron atoms simply by assuming the electrons can occupy the same standing-wave patterns that exist in a hydrogen atom. The ground state of helium, for example, would have both electrons in states that are very similar to the n=1 states of hydrogen. The second-lowest-energy state of helium would have one electron in an n=1 state, and the other in an n=2 states. The relatively complex spectra of elements heavier than hydrogen can be understood as arising from the great number of possible combinations of states for the electrons.

A surprising thing happens, however, with lithium, the three-electron atom. We would expect the ground state of this atom to be one in which all three electrons settle down into n=1 states. What really happens is that two electrons go into n=1 states, but the third stays up in an n=2 state. This is a consequence of a new principle of physics:

The Pauli Exclusion Principle
Only one electron can ever occupy a given state.

There are two n=1 states, one with s_z=+1/2 and one with s_z=–1/2, but there is no third n=1 state for lithium's third electron to occupy, so it is forced to go into an n=2 state.

It can be proven mathematically that the Pauli exclusion principle applies to any type of particle that has half-integer spin. Thus two neutrons can never occupy the same state, and likewise for two protons. Photons, however, are immune to the exclusion principle because their spin is an integer.

I	II	III	IV	V	VI	VII	0
1 H							2 He
3 Li	4 Be	5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	...						

The beginning of the periodic table.

Hydrogen is highly reactive.

Deriving the periodic table

We can now account for the structure of the periodic table, which seemed so mysterious even to its inventor Mendeleev. The first row consists of atoms with electrons only in the n=1 states:

H 1 electron in an n=1 state
He 2 electrons in the two n=1 states

The next row is built by filling the n=2 energy levels:

Li 2 electrons in n=1 states, 1 electron in an n=2 state
Be 2 electrons in n=1 states, 2 electrons in n=2 states
...
O 2 electrons in n=1 states, 6 electrons in n=2 states
F 2 electrons in n=1 states, 7 electrons in n=2 states
Ne 2 electrons in n=1 states, 8 electrons in n=2 states

In the third row we start in on the n=3 levels:

Na 2 electrons in n=1 states, 8 electrons in n=2 states, 1 electron in an n=3 state
...

We can now see a logical link between the filling of the energy levels and the structure of the periodic table. Column 0, for example, consists of atoms with the right number of electrons to fill all the available states up to a certain value of n. Column I contains atoms like lithium that have just one electron more than that.

This shows that the columns relate to the filling of energy levels, but why does that have anything to do with chemistry? Why, for example, are the elements in columns I and VII dangerously reactive? Consider, for example, the element sodium (Na), which is so reactive that it may burst into flames when exposed to air. The electron in the n=3 state has an unusually high energy. If we let a sodium atom come in contact with an oxygen atom, energy can be released by transferring the n=3 electron from the sodium to one of the vacant lower-energy n=2 states in the oxygen. This energy is transformed into heat. Any atom in column I is highly reactive for the same reason: it can release energy by giving away the electron that has an unusually high energy.

Column VII is spectacularly reactive for the opposite reason: these atoms have a single vacancy in a low-energy state, so energy is released when these atoms steal an electron from another atom.

It might seem as though these arguments would only explain reactions of atoms that are in different rows of the periodic table, because only in these reactions can a transferred electron move from a higher-n state to a lower-n state. This is incorrect. An n=2 electron in fluorine (F), for example, would have a different energy than an n=2 electron in lithium (Li), due to the different number of protons and electrons with which it is interacting. Roughly speaking, the n=2 electron in fluorine is more tightly bound (lower in energy) because of the larger number of protons attracting it. The effect of the increased number of attracting protons is only partly counteracted by the increase in the number of repelling electrons, because the forces exerted on an electron by the other electrons are in many different directions and cancel out partially.

Summary

Selected Vocabulary

quantum number .... a numerical label used to classify a quantum state
spin .... the built-in angular momentum possessed by a particle even when at rest

Notation

n .... the number of radial nodes in the wavefunction, including the one at $r=\infty$
$\hbar$ .... $h/2\pi$
$\mathbf{L}$ .... the angular momentum vector of a particle, not including its spin
ℓ .... the magnitude of the $\mathbf{L}$ vector, divided by $\hbar$
ℓ_z .... the z component of the $\mathbf{L}$ vector, divided by $\hbar$; this is the standard notation in nuclear physics, but not in atomic physics
s .... the magnitude of the spin angular momentum vector, divided by $\hbar$
s_z .... the z component of the spin angular momentum vector, divided by $\hbar$; this is the standard notation in nuclear physics, but not in atomic physics

Notation Used in Other Books

m_ℓ .... a less obvious notation for ℓ_z , standard in atomic physics
m_s .... a less obvious notation for s_z, standard in atomic physics

Summary

Hydrogen, with one proton and one electron, is the simplest atom, and more complex atoms can often be analyzed to a reasonably good approximation by assuming their electrons occupy states that have the same structure as the hydrogen atom's. The electron in a hydrogen atom exchanges very little energy or angular momentum with the proton, so its energy and angular momentum are nearly constant, and can be used to classify its states. The energy of a hydrogen state depends only on its n quantum number.

In quantum physics, the angular momentum of a particle moving in a plane is quantized in units of $\hbar$. Atoms are three-dimensional, however, so the question naturally arises of how to deal with angular momentum in three dimensions. In three dimensions, angular momentum is a vector in the direction perpendicular to the plane of motion, such that the motion appears clockwise if viewed along the direction of the vector. Since angular momentum depends on both position and momentum, the Heisenberg uncertainty principle limits the accuracy with which one can know it. The most the can be known about an angular momentum vector is its magnitude and one of its three vector components, both of which are quantized in units of $\hbar$.

In addition to the angular momentum that an electron carries by virtue of its motion through space, it possesses an intrinsic angular momentum with a magnitude of $\hbar$ /2. Protons and neutrons also have spins of $\hbar$ /2, while the photon has a spin equal to $\hbar$.

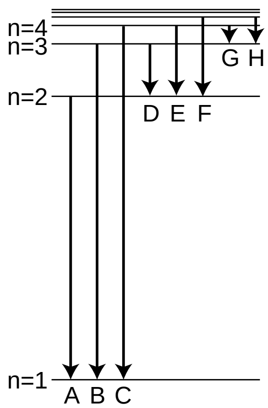
Particles with half-integer spin obey the Pauli exclusion principle: only one such particle can exist is a given state, i.e. with a given combination of quantum numbers.

We can enumerate the lowest-energy states of hydrogen as follows:

n=1, ℓ =0, ℓ_z =0, s_z=+1/2 or –1/2	two states
n=2, ℓ =0, ℓ_z =0, s_z=+1/2 or –1/2	two states
n=2, ℓ =1, ℓ_z =–1, 0, or 1, s_z=+1/2 or –1/2	six states
...	

The periodic table can be understood in terms of the filling of these states. The nonreactive noble gases are those atoms in which the electrons are exactly sufficient to fill all the states up to a given n value. The most reactive elements are those with one more electron than a noble gas element, which can release a great deal of energy by giving away their high-energy electron, and those with one electron fewer than a noble gas, which release energy by accepting an electron.

Homework Problems



Problem 2.

1. (a) A distance scale is shown below the wavefunctions and probability densities illustrated in section 6.3. Compare this with the order-of-magnitude estimate derived in section 6.4 for the radius r at which the wavefunction begins tailing off. Was the estimate in section 6.4 on the right order of magnitude? (b) Although we normally say the moon orbits the earth, actually they both orbit around their common center of mass, which is below the earth's surface but not at its center. The same is true of the hydrogen atom. Does the center of mass lie inside the proton or outside it?

2. The figure shows eight of the possible ways in which an electron in a hydrogen atom could drop from a higher energy state to a state of lower energy, releasing the difference in energy as a photon. Of these eight transitions, only D, E, and F produce photons with wavelengths in the visible spectrum. (a) Which of the visible transitions would be closest to the violet end of the spectrum, and which would be closest to the red end? Explain. (b) In what part of the electromagnetic spectrum would the photons from transitions A, B, and C lie? What about G and H? Explain. (c) Is there an upper limit to the wavelengths that could be emitted by a hydrogen atom going from one bound state to another bound state? Is there a lower limit? Explain.

3. Before the quantum theory, experimentalists noted that in many cases, they would find three lines in the spectrum of the same atom that satisfied the following mysterious rule: $1/\lambda_1=1/\lambda_2+1/\lambda_3$. Explain why this would occur. Do not use reasoning that only works for hydrogen — such combinations occur in the spectra of all elements. [Hint: Restate the equation in terms of the energies of photons.]

4. Find an equation for the wavelength of the photon emitted when the electron in a hydrogen atom makes a transition from energy level n_1 to level n_2. [You will need to have read optional section 6.4.]

5. (a) Verify that Planck's constant has the same units as angular momentum. (b) Estimate the angular momentum of a spinning basketball, in units of $\hbar$.

6. Assume that the kinetic energy of an electron the n=1 state of a hydrogen atom is on the same order of magnitude as the absolute value of its total energy, and estimate a typical speed at which it would be moving. (It cannot really have a single, definite speed, because its kinetic and potential energy trade off at different distances from the proton, but this is just a rough estimate of a typical speed.) Based on this speed, were we justified in assuming that the electron could be described nonrelativistically?

S A solution is given in the back of the book.	★ A difficult problem.
✓ A computerized answer check is available.	∫ A problem that requires calculus.

7. The wavefunction of the electron in the ground state of a hydrogen atom is

$$\Psi = 2\, a^{-3/2}\, e^{-r/a}$$

where r is the distance from the proton, and $a=\hbar^2/kme^2=5.3\text{x}10^{-11}$ m is a constant that sets the size of the wave.

(a) Calculate symbolically, without plugging in numbers, the probability that at any moment, the electron is inside the proton. Assume the proton is a sphere with a radius of b=0.5 fm. [Hint: Does it matter if you plug in r=0 or $r=b$ in the equation for the wavefunction?]

(b) Calculate the probability numerically.

(c) Based on the equation for the wavefunction, is it valid to think of a hydrogen atom as having a finite size? Can a be interpreted as the size of the atom, beyond which there is nothing? Or is there any limit on how far the electron can be from the proton?

8 ★. Use physical reasoning to explain how the equation for the energy levels of hydrogen,

$$E_n = -\frac{mk^2e^4}{2\hbar^2}\cdot\frac{1}{n^2},$$

should be generalized to the case of a heavier atom with atomic number Z that has had all its electrons stripped away except for one.

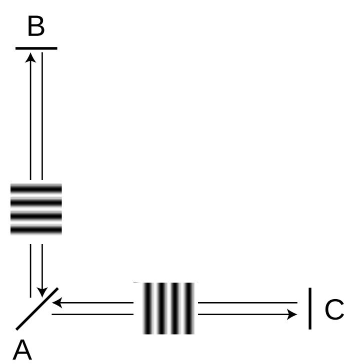
9. This question requires that you read optional section 6.4. A muon is a subatomic particle that acts exactly like an electron except that its mass is 207 times greater. Muons can be created by cosmic rays, and it can happen that one of an atom's electrons is displaced by a muon, forming a muonic atom. If this happens to a hydrogen atom, the resulting system consists simply of a proton plus a muon. (a) How would the size of a muonic hydrogen atom in its ground state compare with the size of the normal atom? (b) If you were searching for muonic atoms in the sun or in the earth's atmosphere by spectroscopy, in what part of the electromagnetic spectrum would you expect to find the absorption lines?

Exercises

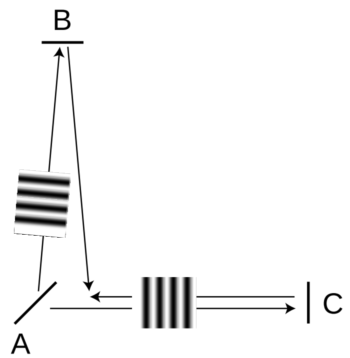
Ex. 1A: The Michelson-Morley Experiment

In this exercise you will analyze the Michelson-Morley experiment, and find what the results should have been according to Galilean relativity and Einstein's theory of relativity. A beam of light coming from the west (not shown) comes to the half-silvered mirror A. Half the light goes through to the east, is reflected by mirror C, and comes back to A. The other half is reflected north by A, is reflected by B, and also comes back to A. When the beams reunite at A, part of each ends up going south, and these parts interfere with one another. If the time taken for a round trip differs by, for example, half the period of the wave, there will be destructive interference.

The point of the experiment was to search for a difference in the experimental results between the daytime, when the laboratory was moving west relative to the sun, and the nighttime, when the laboratory was moving east relative to the sun. Galilean relativity and Einstein's theory of relativity make different predictions about the results. According to Galilean relativity, the speed of light cannot be the same in all reference frames, so it is assumed that there is one special reference frame, perhaps the sun's, in which light travels at the same speed in all directions; in other frames, Galilean relativity predicts that the speed of light will be different in different directions, e.g. slower if the observer is chasing a beam of light. There are four different ways to analyze the experiment:



laboratory's x,t frame of reference



sun's x',t' frame (lab moving to the right)

1. *Laboratory's frame of reference, Galilean relativity*. This is not a useful way to analyze the experiment, since one does not know how fast light will travel in various directions.

2. *Sun's frame of reference, Galilean relativity*. We assume that in this special frame of reference, the speed of light is the same in all directions: we call this speed c. In this frame, the laboratory moves with velocity v, and mirrors A, B, and C move while the light beam is in flight.

3. *Laboratory's frame of reference, Einstein's theory of relativity*. The analysis is extremely simple. Let the length of each arm be L. Then the time required to get from A to either mirror is L/c, so each beam's round-trip time is $2L/c$.

4. *Sun's frame of reference, Einstein's theory of relativity*. We analyze this case by starting with the laboratory's frame of reference and then transforming to the sun's frame.

Groups 1-4 work in the sun's frame of reference according to Galilean relativity.

Group 1 finds time AC. Group 2 finds time CA. Group 3 finds time AB. Group 4 finds time BA.

Groups 5 and 6 transform the lab-frame results into the sun's frame according to Einstein's theory.

Group 5 transforms the x and t when ray ACA gets back to A into the sun's frame of reference, and group 6 does the same for ray ABA.

Discussion:

Michelson and Morley found no change in the interference of the waves between day and night. Which version of relativity is consistent with their results?

What does each theory predict if v approaches c?

What if the arms are not exactly equal in length?

Does it matter if the "special" frame is some frame other than the sun's?

Ex. 2A: Sports in Slowlightland

In Slowlightland, the speed of light is 20 mi/hr = 32 km/hr = 9 m/s. Think of an example of how relativistic effects would work in sports. Things can get very complex very quickly, so try to think of a simple example that focuses on just one of the following effects:

- relativistic momentum
- relativistic kinetic energy
- relativistic addition of velocities
- time dilation and length contraction
- Doppler shifts of light
- equivalence of mass and energy
- time it takes for light to get to an athlete’s eye
- deflection of light rays by gravity

Ex. 6A: Quantum Versus Classical Randomness

1. Imagine the *classical* version of the particle in a one-dimensional box. Suppose you insert the particle in the box and give it a known, predetermined energy, but a random initial position and a random direction of motion. You then pick a random later moment in time to see where it is. Sketch the resulting probability distribution by shading on top of a line segment. Does the probability distribution depend on energy?

2. Do similar sketches for the first few energy levels of the *quantum mechanical* particle in a box, and compare with 1.

3. Do the same thing as in 1, but for a classical hydrogen atom in two dimensions, which acts just like a miniature solar system. Assume you're always starting out with the same fixed values of energy and angular momentum, but a position and direction of motion that are otherwise random. Do this for L=0, and compare with a real L=0 probability distribution for the hydrogen atom.

4. Repeat 3 for a nonzero value of L, say L=$\hbar$.

5. Summarize: Are the classical probability distributions accurate? What qualitative features are possessed by the classical diagrams but not by the quantum mechanical ones, or vice-versa?

Glossary

FWHM. The full width at half-maximum of a probability distribution; a measure of the width of the distribution.

Half-life. The amount of time that a radioactive atom has a probability of 1/2 of surviving without decaying.

Independence. The lack of any relationship between two random events.

Invariant. A quantity that does not change when transformed.

Lorentz transformation. The transformation between frames in relative motion.

Mass. What some books mean by "mass" is our mg.

Normalization. The property of probabilities that the sum of the probabilities of all possible outcomes must equal one.

Photon. A particle of light.

Photoelectric effect. The ejection, by a photon, of an electron from the surface of an object.

Probability. The likelihood that something will happen, expressed as a number between zero and one.

Probability distribution. A curve that specifies the probabilities of various random values of a variable; areas under the curve correspond to probabilities.

Quantum number. A numerical label used to classify a quantum state.

Rest mass. Referred to as mass in this book; written as m_0 in some books.

Spin. The built-in angular momentum possessed by a particle even when at rest.

Transformation. The mathematical relationship between the variables such as x and t, as observed in different frames of reference.

Wave-particle duality. The idea that light is both a wave and a particle.

Wavefunction. The numerical measure of an electron wave, or in general of the wave corresponding to any quantum mechanical particle.

Index

Photo Credits

All photographs are by Benjamin Crowell, except as noted below. In some cases I have used historical photographs for which I know the date the picture was taken but not the photographer; in these cases I have simply listed the dates, which indicate the copyrights have expired. I will be grateful for any information that helps me to credit the photographers properly.

Chapter 1
Einstein: ca. 1905

Chapter 2
Eclipse: 1919
Large Hadron Collider: Courtesy of CERN.

Chapter 3
Mount St. Helens: Public-domain image by Austin Post, USGS.
Pu'u O'o: Public-domain image by Lyn Topinka, USGS.

Chapter 4
Ozone maps: NASA/GSFC TOMS Team.
Photon interference photos: Lyman Page.

Chapter 5
Wicked witch: W.W. Denslow, 1900. Quote from The Wizard of Oz, by L. Frank Baum, 1900.

Chapter 6
Hindenburg: Arthur Cofod Jr., U.S. Air Force, 1937, courtesy of the National Air and Space Museum Archives.

Useful Data

Metric Prefixes

M-	mega-	10^6
k-	kilo-	10^3
m-	milli-	10^{-3}
μ- (Greek mu)	micro-	10^{-6}
n-	nano-	10^{-9}
p-	pico-	10^{-12}
f-	femto-	10^{-15}

(Centi-, 10^{-2}, is used only in the centimeter.)

Notation and Units

quantity	**unit**	**symbol**
distance	meter, m	x, Δx
time	second, s	t, Δt
mass	kilogram, kg	m
density	kg/m^3	ρ
force	newton, 1 N=1 kg·m/s^2	$\boldsymbol{F}$
velocity	m/s	$\boldsymbol{v}$
acceleration	m/s^2	$\boldsymbol{a}$
energy	joule, J	E
momentum	kg·m/s	$\boldsymbol{p}$
angular momentum	kg·m^2/s	L
period	s	T
wavelength	m	λ
frequency	s^{-1} or Hz	f
focal length	m	f
magnification	unitless	M
index of refraction	unitless	n

Fundamental Constants

gravitational constant	$G=6.67\times10^{-11}$ N·m^2/kg^2
Coulomb constant	$k=8.99\times10^{9}$ N·m^2/C^2
quantum of charge	$e=1.60\times10^{-19}$ C
speed of light	$c=3.00\times10^{8}$ m/s
Planck's constant	$h=6.63\times10^{-34}$ J·s

Conversions

Conversions between SI and other units:

1 inch	=	2.54 cm (exactly)
1 mile	=	1.61 km
1 pound	=	4.45 N
(1 kg)·g	=	2.2 lb
1 gallon	=	3.78×10^3 cm^3
1 horsepower	=	746 W
1 kcal*	=	4.18×10^3 J

*When speaking of food energy, the word "Calorie" is used to mean 1 kcal, i.e. 1000 calories. In writing, the capital C may be used to indicate 1 Calorie=1000 calories.

Conversions between U.S. units:

1 foot	=	12 inches
1 yard	=	3 feet
1 mile	=	5280 ft

Some Indices of Refraction

substance	**index of refraction**
vacuum	0 by definition
air	1.0003
water	1.3
glass	1.5 to 1.9
diamond	2.4

Note that all indices of refraction depend on wavelength. These values are about right for the middle of the visible spectrum (yellow).

Subatomic Particles

particle	**mass (kg)**	**charge**	**radius (fm)**
electron	9.109×10^{-31}	$-e$	<~0.01
proton	1.673×10^{-27}	$+e$	~1.1
neutron	1.675×10^{-27}	0	~1.1
neutrino	~10^{-39} kg?	0	?

The radii of protons and neutrons can only be given approximately, since they have fuzzy surfaces. For comparison, a typical atom is about a million fm in radius.