**Exercise 12C: How strong are your glasses?**

This exercise was created by Dan MacIsaac.

**Equipment:**
- eyeglasses
- diverging lenses for students who don’t wear glasses, or who use glasses with converging lenses
- rulers and metersticks
- scratch paper
- marking pens

Most people who wear glasses have glasses whose lenses are outbending, which allows them to focus on objects far away. Such a lens cannot form a real image, so its focal length cannot be measured as easily as that of a converging lens. In this exercise you will determine the focal length of your own glasses by taking them off, holding them at a distance from your face, and looking through them at a set of parallel lines on a piece of paper. The lines will be reduced (the lens’s magnification is less than one), and by adjusting the distance between the lens and the paper, you can make the magnification equal 1/2 exactly, so that two spaces between lines as seen through the lens fit into one space as seen simultaneously to the side of the lens. This object distance can be used in order to find the focal length of the lens.

1. Use a marker to draw three evenly spaced parallel lines on the paper. (A spacing of a few cm works well.)

2. Does this technique really measure magnification or does it measure angular magnification? What can you do in your experiment in order to make these two quantities nearly the same, so the math is simpler?

3. Before taking any numerical data, use algebra to find the focal length of the lens in terms of $d_o$, the object distance that results in a magnification of 1/2.

4. Measure the object distance that results in a magnification of 1/2, and determine the focal length of your lens.
**Exercise 12D: Double-Source Interference**

1. Two sources separated by a distance \( d = 2 \) cm make circular ripples with a wavelength of \( \lambda = 1 \) cm. On a piece of paper, make a life-size drawing of the two sources in the default setup, and locate the following points:

   A. The point that is 10 wavelengths from source #1 and 10 wavelengths from source #2.
   B. The point that is 10.5 wavelengths from #1 and 10.5 from #2.
   C. The point that is 11 wavelengths from #1 and 11 from #2.
   D. The point that is 10 wavelengths from #1 and 10.5 from #2.
   E. The point that is 11 wavelengths from #1 and 11.5 from #2.
   F. The point that is 10 wavelengths from #1 and 11 from #2.
   G. The point that is 11 wavelengths from #1 and 12 from #2.

   You can do this either using a compass or by putting the next page under your paper and tracing. It is not necessary to trace all the arcs completely, and doing so is unnecessarily time-consuming; you can fairly easily estimate where these points would lie, and just trace arcs long enough to find the relevant intersections.

   What do these points correspond to in the real wave pattern?

2. Make a fresh copy of your drawing, showing only point F and the two sources, which form a long, skinny triangle. Now suppose you were to change the setup by doubling \( d \), while leaving \( \lambda \) the same. It’s easiest to understand what’s happening on the drawing if you move both sources outward, keeping the center fixed. Based on your drawing, what will happen to the position of point F when you double \( d \)? How has the angle of point F changed?

3. What would happen if you doubled **both** \( \lambda \) and \( d \) compared to the standard setup?

4. Combining the ideas from parts 2 and 3, what do you think would happen to your angles if, starting from the standard setup, you doubled \( \lambda \) while leaving \( d \) the same?

5. Suppose \( \lambda \) was a millionth of a centimeter, while \( d \) was still as in the standard setup. What would happen to the angles? What does this tell you about observing diffraction of light?
Exercise 12E: Single-slit diffraction

Equipment:

- rulers
- computer with web browser

The following page is a diagram of a single slit and a screen onto which its diffraction pattern is projected. The class will make a numerical prediction of the intensity of the pattern at the different points on the screen. Each group will be responsible for calculating the intensity at one of the points. (Either 11 groups or six will work nicely – in the latter case, only points a, c, e, g, i, and k are used.) The idea is to break up the wavefront in the mouth of the slit into nine parts, each of which is assumed to radiate semicircular ripples as in Huygens’ principle. The wavelength of the wave is 1 cm, and we assume for simplicity that each set of ripples has an amplitude of 1 unit when it reaches the screen.

1. For simplicity, let’s imagine that we were only to use two sets of ripples rather than nine. You could measure the distance from each of the two points inside the slit to your point on the screen. Suppose the distances were both 25.0 cm. What would be the amplitude of the superimposed waves at this point on the screen?

    Suppose one distance was 24.0 cm and the other was 25.0 cm. What would happen?

    What if one was 24.0 cm and the other was 26.0 cm?

    What if one was 24.5 cm and the other was 25.0 cm?

    In general, what combinations of distances will lead to completely destructive and completely constructive interference?

    Can you estimate the answer in the case where the distances are 24.7 and 25.0 cm?

2. Although it is possible to calculate mathematically the amplitude of the sine wave that results from superimposing two sine waves with an arbitrary phase difference between them, the algebra is rather laborious, and it become even more tedious when we have more than two waves to superimpose. Instead, one can simply use a computer spreadsheet or some other computer program to add up the sine waves numerically at a series of points covering one complete cycle. This is what we will actually do. You just need to enter the relevant data into the computer, then examine the results and pick off the amplitude from the resulting list of numbers. You can run the software through a web interface at http://lightandmatter.com/cgi-bin/diffraction1.cgi.

3. Measure all nine distances to your group’s point on the screen, and write them on the board - that way everyone can see everyone else’s data, and the class can try to make sense of why the results came out the way they did. Determine the amplitude of the combined wave, and write it on the board as well.

The class will discuss why the results came out the way they did.
Exercise 12F: Diffraction of Light

Equipment:

slit patterns, lasers, straight-filament bulbs

station 1
You have a mask with a bunch of different double slits cut out of it. The values of w and d are as follows:

- pattern A  w=0.04 mm  d=.250 mm
- pattern B  w=0.04 mm  d=.500 mm
- pattern C  w=0.08 mm  d=.250 mm
- pattern D  w=0.08 mm  d=.500 mm

Predict how the patterns will look different, and test your prediction. The easiest way to get the laser to point at different sets of slits is to stick folded up pieces of paper in one side or the other of the holders.

station 2
This is just like station 1, but with single slits:

- pattern A  w=0.02 mm
- pattern B  w=0.04 mm
- pattern C  w=0.08 mm
- pattern D  w=0.16 mm

Predict what will happen, and test your predictions. If you have time, check the actual numerical ratios of the w values against the ratios of the sizes of the diffraction patterns.

station 3
This is like station 1, but the only difference among the sets of slits is how many slits there are:

- pattern A  double slit
- pattern B  3 slits
- pattern C  4 slits
- pattern D  5 slits

station 4
Hold the diffraction grating up to your eye, and look through it at the straight-filament light bulb. If you orient the grating correctly, you should be able to see the m = 1 and m = −1 diffraction patterns off the left and right. If you have it oriented the wrong way, they’ll be above and below the bulb instead, which is inconvenient because the bulb’s filament is vertical. Where is the m = 0 fringe? Can you see m = 2, etc.?

Station 5 has the same equipment as station 4. If you’re assigned to station 5 first, you should actually do activity 4 first, because it’s easier.

station 5
Use the transformer to increase and decrease the voltage across the bulb. This allows you to control the filament’s temperature. Sketch graphs of intensity as a function of wavelength for various temperatures. The inability of the wave model of light to explain the mathematical shapes of these curves was historically one of the reasons for creating a new model, in which light is both a particle and a wave.
Chapter 13
Quantum Physics

13.1 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 decay 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 would...
energy, instead of being spread out over eons, would all be released at one instant, blowing our whole planet to Kingdom Come.¹

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. This chapter is a basic introduction to quantum physics.

**Discussion Question**

A I said “Pick two identical atoms of a radioactive isotope.” Are two atoms really identical? If their electrons are orbiting the nucleus, can we distinguish each atom by the particular arrangement of its electrons at some instant in time?

13.1.1 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, remind 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

¹This is under the assumption that all the uranium atoms were created at the same time. In reality, we have only a general idea of the processes that might have created the heavy elements in the nebula from which our solar system condensed. Some portion of them may have come from nuclear reactions in supernova explosions in that particular nebula, but some may have come from previous supernova explosions throughout our galaxy, or from exotic events like collisions of white dwarf stars.
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**

A 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?

13.1.2 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 than 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 \( \frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{1000} = 0.001 \). 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.

This can be taken as the definition of statistical independence.

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 slot-machine game described on page 859, 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 ($10,000 \times 0.001 = 10$), 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
Foundations of probability

1. **Positivity:** Probabilities are positive.

2. **Normalization:** The total probability is 1.

3. **Additivity:** Mutually exclusive probabilities are additive.

4. **Independence:** Independent systems obey the definition of statistical independence, i.e., their probabilities multiply.

5. **The weak law of large numbers:** In the limit of a large number of trials, the frequency of a certain event converges to its probability.

Statements 1-3 are called the Kolmogorov axioms. In 3, for technical reasons, “additive” is usually taken to include infinite sums, such as \(1 + 1/2 + 1/4 + \ldots\), but not continuous sums such as integrals. Statement 4 is prediction about experiment. Statement 5 can be considered to be either a theorem to be proved from axioms such as 1-3, or an operational definition of probability, or a prediction about experiments.

---

**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 1212121212121212 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 a friend came to visit him in a cab, and remarked that the number of the cab, 1729, seemed relatively uninteresting. Ramanujan replied that on the contrary, it was very interesting because it was the smallest number that could be represented in two different ways as the sum of two cubes. 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 in-
clude 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 the lyrics to a set of incomparably beautiful love songs written in a language that never existed. In view of these examples, what does it really mean to say that something is random?

13.1.3 Probability distributions

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, d. 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.

Figure e 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: 2.9999999999999996876876587658465436, for example. It doesn’t usually make sense, therefore, to talk about the probability

Section 13.1 Rules of randomness
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.

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

$$\text{(unitless area of a square)} = (\text{width of square with time units}) \times (\text{height of square}).$$

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 f shows another example, a probability distribution for people’s height. This kind of bell-shaped curve is quite common.

**self-check B**

Compare the number of people with heights in the range of 130-135 cm to the number in the range 135-140. ▶ Answer, p. 1067

Looking for tall basketball players example 1

▶ 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 feet 4 inches) in height? Figure g gives a close-up of the “tail” of the distribution shown previously in figure f.

▶ The shaded area under the curve represents the probability that a given person is tall enough. Each rectangle represents a probability of $0.2 \times 10^{-7} \text{ cm}^{-1} \times 1 \text{ cm} = 2 \times 10^{-8}$. There are about 35 rectangles covered by the shaded area, so the probability of having a height greater than 225 cm is $7 \times 10^{-7}$, or just under one in a million. Using the rule for calculating averages, the average, or expected number of people this tall is $(10^8) \times (7 \times 10^{-7}) = 70$.

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.

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 figure i.

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.

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).

13.1.4 Exponential decay and half-life

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 $^{131}\text{I}$, 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 $^{127}\text{I}$, 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 $^{131}\text{I}$ 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 $P_{\text{surv}}(t)$ 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

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

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,

$$P_{\text{surv}}(16 \text{ hr}) = 0.50 \times 0.50 = 0.25.$$

Similarly we have

$$P_{\text{surv}}(24 \text{ hr}) = 0.50 \times 0.5 \times 0.5 = 0.125.$$

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

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

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

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

which can be rearranged to give

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

This is exactly what we would have found simply by plugging in $P_{\text{surv}}(t) = 0.5^{t/8 \text{ hr}}$ 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 Equation**

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

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.

\section*{Radioactive contamination at Chernobyl example 2}

One of the most dangerous radioactive isotopes released by the Chernobyl disaster in 1986 was $^{90}\text{Sr}$, 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?

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

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

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

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

so

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

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.

\section*{C Dating example 3}

Almost all the carbon on Earth is $^{12}\text{C}$, but not quite. The isotope $^{14}\text{C}$, 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 $^{14}\text{C}$ in the atmosphere remains constant, at $1.3 \times 10^{-12}$. Living plants and animals take in both $^{12}\text{C}$ and $^{14}\text{C}$ 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 $^{14}\text{C}$ gradually falls off as it undergoes radioactive decay. This effect can be used to find the
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:

$$(\text{number of decays between } t \text{ and } t + \Delta t) = N(t) - N(t + \Delta t)$$

Usually we’re interested in the case where $\Delta t$ is small compared to $t_{1/2}$, and in this limiting case the calculation starts to look exactly like the limit that goes into the definition of the derivative $dN/dt$. It is therefore more convenient to talk about the *rate* of decay $-dN/dt$ rather than the *number* of decays in some finite time interval. Doing calculus on the function $e^x$ is also easier than with $0.5^x$, so we rewrite...
the function \( N(t) \) as

\[
N = N(0)e^{-t/\tau},
\]

where \( \tau = t_{1/2}/\ln 2 \) is shown in example 6 on p. 871 to be the average time of survival. The rate of decay is then

\[
\frac{dN}{dt} = \frac{N(0)}{\tau} e^{-t/\tau}.
\]

Mathematically, differentiating an exponential just gives back another exponential. Physically, this is telling us that as \( N \) falls off exponentially, the rate of decay falls off at the same exponential rate, because a lower \( N \) means fewer atoms that remain available to decay.

self-check C

Check that both sides of the equation for the rate of decay have units of \( s^{-1} \), i.e., decays per unit time. 

\[ \text{Answer, p. 1067} \]

The hot potato example 4

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 \( ^{17}\text{F} \), 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?

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 \( 2 \times 10^7 \) 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.

Discussion Questions

A In the medical procedure involving \( ^{131}\text{I} \), 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?
13.1.5 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

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

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

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

where the integral is over all possible values of $x$.

**Probability distribution for radioactive decay example 5**

Here is a rigorous justification for the statement in subsection 13.1.4 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

$$D(t) = k0.5^{t/t_{1/2}},$$

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

$$(\text{probability of } 0 \leq t < \infty) = 1 = \int_0^\infty D(t) \, dt.$$ 

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

$$D(t) = k \exp \left[ \ln \left( 0.5^{t/t_{1/2}} \right) \right]$$

$$= k \exp \left[ \frac{t}{t_{1/2}} \ln 0.5 \right]$$

$$= k \exp \left( -\frac{\ln 2}{t_{1/2}} \right),$$

which gives an integral of the familiar form $\int e^{cx} \, dx = (1/c)e^{cx}$. We thus have

$$1 = -\frac{kt_{1/2}}{\ln 2} \exp \left( -\frac{\ln 2}{t_{1/2}} t \right) \bigg|_0^\infty,$$

which gives the desired result:

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

\[
(\text{average lifetime}) = \int_0^\infty t \, D(t) \, dt
\]

Using the convenient base-\(e\) form again, we have

\[
(\text{average lifetime}) = \frac{\ln 2}{t_{1/2}} \int_0^\infty t \exp\left(\frac{-\ln 2}{t_{1/2}} t\right) \, dt.
\]

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 \(\int e^{cx} \, dx = \frac{1}{c} e^{cx} - \frac{1}{c^2} e^{cx}\), and the first term can be ignored for our purposes because it equals zero at both limits of integration. We end up with

\[
(\text{average lifetime}) = \frac{\ln 2}{t_{1/2}} \left(\frac{t_{1/2}}{\ln 2}\right)^2
\]

\[
= \frac{t_{1/2}}{\ln 2}
\]

\[
= 1.443 \, t_{1/2},
\]

which is, as expected, longer than one half-life.
In recent decades, a huge hole in the ozone layer has spread out from Antarctica. Left: November 1978. Right: November 1992

13.2 Light as a particle

The only thing that interferes with my learning is my education.  
\textit{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!

13.2.1 Evidence for light as a particle

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.

![Digital camera images of dimmer and dimmer sources of light. The dots are records of individual photons.](image)

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 figure a. Figure a/1 is fake, but a/2 and a/3 are real digital-camera images made by Prof. Lyman Page of Princeton University as a classroom demonstration. Figure a/1 is what we would see if we used the digital camera to take a picture of a fairly
dim source of light. In figures a/2 and a/3, the intensity of the light was drastically reduced by inserting semitransparent absorbers like the tinted plastic used in sunglasses. Going from a/1 to a/2 to a/3, 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, b, 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 a/2 and a/3 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, c. Each “bullet” of light apparently carries only a tiny 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 section we will learn how the seeming paradox was eventually resolved.

**Discussion Questions**

**A** Suppose someone rebuts the digital camera data in figure a, 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 — some are just more sensitive than others. How could we test this interpretation?

**B** Discuss how the correspondence principle applies to the observations and concepts discussed in this section.
13.2.2 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.)

Figure e 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...
The hamster in her hamster ball is like an electron emerging from the metal (tiled kitchen floor) into the surrounding vacuum (wood floor). The wood floor is higher than the tiled floor, so as she rolls up the step, the hamster will lose a certain amount of kinetic energy, analogous to $E_s$. If her kinetic energy is too small, she won’t even make it up the step.

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. In popping out through the surface the electron therefore loses some amount of energy $E_s$, 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 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 $E_s$ 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, $E_s$ 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 $E_s$ 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 $E_s$, 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**

Figure g shows an experiment that is used sometimes in college laboratory courses to probe the relationship between the energy and frequency of a photon. 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 reached a value $\Delta V$, the minimum energy required by an electron to make it out of the bottom plate and across the gap to the other plate is $E_s + e\Delta V$. As $\Delta V$ increases, we eventually reach a point at which $E_s + e\Delta V$ equals the energy of one photon. No more electrons can cross the gap, and the reading on the voltmeter stops rising. The quantity $E_s + e\Delta V$ now tells us the energy of one photon. If we determine this energy for a variety of wavelengths, $h$, we find the following simple relationship between the energy of a photon and the frequency of the light:

$$E = hf,$$

where $h$ is a constant with the value $6.63 \times 10^{-34}$ J $\cdot$ 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, for historical reasons explained in the footnote beginning on the preceding page.

**self-check D**

How would you extract $h$ from the graph in figure h? What if you didn’t even know $E_s$ in advance, and could only graph $e\Delta V$ versus $f$? 

Answer, p. 1067

Since the energy of a photon is $hf$, a beam of light can only have energies of $hf, 2hf, 3hf, \text{etc.}$ 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.
Photons from a lightbulb example 7
>
Roughly how many photons are emitted by a 100 watt lightbulb in 1 second?
>
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

\[ E_{\text{photon}} = hf = \frac{hc}{\lambda} \]

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

\[ \frac{(100 \text{ J})}{E_{\text{photon}}} = \frac{(100 \text{ J})}{(hc/\lambda)} \approx 3 \times 10^{20} \]

This hugeness of this number is consistent with the correspondence principle. The experiments that established the classical theory of optics weren’t wrong. They were right, within their domain of applicability, in which the number of photons was so large as to be indistinguishable from a continuous beam.

Measuring the wave example 8
>
When surfers are out on the water waiting for their chance to catch a wave, they’re interested in both the height of the waves and when the waves are going to arrive. In other words, they observe both the amplitude and phase of the waves, and it doesn’t matter to them that the water is granular at the molecular level. The correspondence principle requires that we be able to do the same thing for electromagnetic waves, since the classical theory of electricity and magnetism was all stated and verified experimentally in terms of the fields \( E \) and \( B \), which are the amplitude of an electromagnetic wave. The phase is also necessary, since the induction effects predicted by Maxwell’s equation would flip their signs depending on whether an oscillating field is on its way up or on its way back down.

This is a more demanding application of the correspondence principle than the one in example 7, since amplitudes and phases constitute more detailed information than the over-all intensity of a beam of light. Eyeball measurements can’t detect this type of information, since the eye is much bigger than a wavelength, but for example an AM radio receiver can do it with radio waves, since the wavelength for a station at 1000 kHz is about 300 meters, which is much larger than the antenna. The correspondence principle demands that we be able to explain this in terms of the photon theory, and this requires not just that we have a large number
of photons emitted by the transmitter per second, as in example 7, but that even by the time they spread out and reach the receiving antenna, there should be many photons overlapping each other within a space of one cubic wavelength. Problem 47 on p. 952 verifies that the number is in fact extremely large.

\[ \text{Momentum of a photon} \]

\[ \text{Example 9} \]

According to the theory of relativity, the momentum of a beam of light is given by \( p = E/c \). Apply this to find the momentum of a single photon in terms of its frequency, and in terms of its wavelength.

Combining the equations \( p = E/c \) and \( E = hf \), we find

\[ p = \frac{E}{c} = \frac{hf}{c}. \]

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** The photoelectric effect only ever ejects a very tiny percentage of the electrons available near the surface of an object. How well does this agree with the wave model of light, and how well with the particle model? Consider the two different distance scales involved: the wavelength of the light, and the size of an atom, which is on the order of \( 10^{-10} \) or \( 10^{-9} \) m.

**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 a single electron was likely to get hit by more than one photon?

**D** Draw some representative trajectories of electrons for \( \Delta V = 0 \), \( \Delta V \) less than the maximum value, and \( \Delta V \) 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. How does this relate to discussion question C?

**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?
13.2.3 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.

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 due to 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 on p. 873. The whole thing is sealed inside a light-tight box. The results are shown in figure i. (In fact, the similar figures in on page 873 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, j. 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
Bullets pass through a double slit.

A water wave passes through a double slit.

A single photon can go through both slits.

with a water wave, k. 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 i/3, 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 has been conclusively disproved by forming interference patterns with light so dim that no more than one photon is in flight at a time. In figure i/3, for example, 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! Although most light sources tend to emit photons in bunches, experiments have been done with light sources that really do emit single photons at wide time intervals, and the same type of interference pattern is observed, showing that a single photon can interfere with itself.

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 sec. 13.3.4 we will discuss the Heisenberg uncertainty principle, which gives a numerical way of approaching this issue.)

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,

\[
\text{(probability distribution)} \propto (\text{amplitude})^2.
\]

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.

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 are 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 digital camera image if you covered one slit? Why should it matter whether you give the photon two choices or only one?

13.2.4 Nonlocality and entanglement

Nonlocality

People sometimes say that quantum mechanics is the set of rules for describing the world of the very small, but this is a false generalization, like saying that terriers are untrainable. How do we define our measure of how small is small? The only distance scales we’ve discussed have been wavelengths, and there is no upper limit on wavelengths. The wavelength of an FM radio photon is bigger than my terrier, who is very obedient to Newton’s laws. The only scale built into the structure of quantum mechanics is Planck’s constant, and Planck’s constant has units of joules per hertz, not meters, so it can’t be converted into a distance. Quantum mechanics is, as far as we can tell, a valid tool for describing systems at scales from quarks to galaxies.

So quantum behavior can occur at any scale, even large ones. For an example that may be a little disturbing, consider the arrangement shown in figure n. A single photon comes in from the left and encounters a diagonal piece of glass. The glass reflects half the light and transmits half of it. The photon is a wave, and this is expected wave behavior. But the photon is also a particle, and we can’t have half a particle. Therefore either camera A will detect a whole photon and B will see none, or it will be the other way around. If we repeat the experiment many times, we might come up with a list of results like this:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

An instant before the moment of detection, the photon is a wave pattern that just happens to consist of two widely separated pieces, each carrying half the energy. The situation seems perfectly symmetric, but then a moment later we find that B has detected the photon and A hasn’t. If B’s detection of the photon is random, then how does the information get to A that it had better not detect it? This seems as though there is some sort of conspiracy being carried out over arbitrarily large distances and with no time delay. It’s as though the two parts of the wave are a pair of criminal suspects who would like to line up their stories but are being kept in separate jail.
cells so that they can’t communicate. If the part of the wave at B is going to be detected (at full strength, carrying 100% of the energy \( E = hf \)), how does the part at A get the message that it should fade away like the Cheshire cat? This coordination would have to occur over very large distances — real-world experiments of this type have been done over distances of a thousand kilometers, with the photons traveling either through outer space or through fiber-optic cables. Einstein derisively referred to this apparent coordination as “spooky action at a distance.”

Niels Bohr and two collaborators proposed in 1924 the seemingly reasonable solution that there can’t be any such coordination. Then the random detection of the photon by camera A and camera B would be independent. Independent probabilities multiply, so there would be a probability of \( (1/2)(1/2) = 1/4 \) that both cameras would see photons. This would violate conservation of energy, since the original energy \( E = hf \) would have been detected twice, and the universe would have gained \( 1hf \) worth of total energy. But Bohr pointed out that there would also be the same probability that neither camera would detect a photon, in which case the change in the universe’s energy would be \( -hf \). On the average, energy would be conserved. According to Bohr’s theory, conservation of energy and momentum would not be absolute laws of physics but only rules that would be true on the average.

The experimentalists Geiger and Bothe immediately set out to test this prediction. They performed an experiment analogous to the one in figure n, but with x-rays rather than visible light. Their results, published in 1926, showed that if one detector saw the x-ray photon, the other did not, so that energy was always conserved at the microscopic level, not just on the average. We never observe an outcome in which both A and B detect a photon, or one in which neither detects it. That is, the occurrence of event A (camera A sees a photon) and event B (camera B sees one) are both random, but they are not independent.

Entanglement

At a 1927 conference in Brussels, Einstein protested that this was a problem, because the two detectors could in principle make their observations simultaneously, and it would then seem that some influence or communication was being transmitted between them faster than the speed of light. “It seems to me,” he complained, “that this difficulty cannot be overcome unless the description of the process in terms of the . . . wave is supplemented by some detailed specification of the [trajectory of the particle]. . . . If one works only with . . . waves, the interpretation . . . , I think, contradicts the postulate of relativity.”

The experimental fact ends up being that the spooky action at a distance exists, and it does go faster than light. In 2012, Guerreiro
et al.\textsuperscript{2} carried out a very direct and conceptually simple enactment of exactly the experiment in figure n, with electronic timing precise enough to prove that the detection events at A and B were separated from each other by too great a distance to have been linked by any influence traveling at \( \leq c \). These findings are summarized by saying that quantum mechanics is \textit{nonlocal}. A single wave-particle can be spread out over an arbitrarily large region of space, but its interactions that transfer energy and momentum are always correlated over these distances in such a way that the conservation laws are maintained.

What Einstein had not originally appreciated was that these correlations do not violate relativity because they do not actually transport any energy, or even any information, between A and B. For example, if Alice is at detector A, and Bob is at B, a million kilometers away, Alice can detect the photon and know immediately that Bob did not detect it. She learns something seemingly instantaneously about Bob — Bob is probably sad and disappointed right now. But because Bob does not have any control over the result, he cannot use this fact to send a message to Alice, so there is no transmission of information. Alice and Bob’s states are said to be \textit{entangled}.

By analogy, suppose that you head off to work on a winter day in New York. As you step out of the subway station into the cold air, you reach into your pockets for your gloves, but you find that you only have your left glove. Oh, you must have dropped your right glove on the floor while you were petting your adorable terrier on the way out the door. The presence of your left glove tells you that your right glove must be at home. But there has been no spooky action at a distance. You have simply recovered some information about a region of space that lies at some distance from you.

Einstein and Bohr had strong physical intuitions that led them to incorrect predictions about experiments, and these predictions

\footnote{\url{arxiv.org/abs/1204.1712}. The paper is very readable.}
were the fruits of inappropriate mental pictures of what was going on. If we take the principles of quantum mechanics seriously, then the correct picture is the following. Before the photon in figure n hits the glass diagonal, the state of things is the following.

\[
\text{A photon is headed to the right.}
\]

Our photon is then partially reflected and partially transmitted. Now we have a superposition of two wave patterns:

\[
\begin{align*}
\text{The photon has been} & \quad + \quad \text{The photon has continued to the right.} \\
\text{reflected upward.} & \quad \text{continue to the right.}
\end{align*}
\]

where the amplitudes \(c\) and \(c'\) are equal in absolute value.\(^3\)

Let’s say that the cameras are at equal distances from the glass diagonal, so that their chances to detect the photon occur simultaneously.\(^4\) After detection, we have this:

\[
\begin{align*}
\text{Camera A detected a} & \quad + \quad \text{B detected a photon} \\
\text{photon and B didn’t.} & \quad \text{and A didn’t.}
\end{align*}
\]

Here we have made the nontrivial assumption that material objects like cameras obey the same wave-superposition rules as photons. This turns out to be true. Cameras are made out of things like electrons, and as we’ll see in section 13.3, things like electrons are also wave-particles, and they obey all the same wave-particle rules as photons. The states of the two cameras are now entangled.

You can see where this is going. Alice had been standing by camera A, watching anxiously, while Bob, a million kilometers away, was breathlessly observing camera B.

\[
\begin{align*}
\text{Alice saw a photon} & \quad + \quad \text{Bob saw a photon and} \\
\text{and Bob didn’t. They} & \quad \text{Alice didn’t. They} \\
\text{consider this result to} & \quad \text{consider this result to} \\
\text{have been random.} & \quad \text{have been random.}
\end{align*}
\]

It doesn’t \textit{seem} to Alice and Bob as though their brains are in a superposition of two states. They \textit{feel} as though they have only experienced the one possibility that actually happened, not a mixture of both at the same time. And yet this picture of the physics

\[^3\text{Conservation of energy requires } c^2 = 1/2 \text{ and } c'^2 = 1/2, \text{ even in classical physics. We could have, for example, } c = 1/\sqrt{2} \text{ and } c' = -1/\sqrt{2}. \text{ Such a possible difference in signs wouldn’t concern us in this example. It would only be relevant if there were some later opportunity for the two parts of the wave to recombine and superimpose on one another, producing interference effects.}\]

\[^4\text{According to special relativity, this simultaneity holds only in one frame of reference, say the lab frame. But if simultaneity does hold in one frame, then we can also say that in } \textit{all} \text{ frames, the distance between the two events is “spacelike,” i.e., they are too far apart to have been connected by any causal influence propagating at } \leq c.\]
explains very nicely how the deterministic laws of physics produce a result that seems to them to have been random.

If Alice and Bob have been split into two ghostlike halves of themselves, then conceivably these half-selves could undergo interference, as in the double-slit experiment. But there are practical reasons why we cannot actually detect such interference effects. For one thing, Alice and Bob are macroscopic objects, with energies $E$ on the order of many joules. Because Planck’s constant is small, their wave frequencies $f = E/h$ are extremely high, and their wavelengths incredibly short (on the order of $10^{-34} \text{ m}$!). We have seen that diffraction becomes undetectable when wavelengths are too short. Furthermore, there is a phenomenon called decoherence, discussed further in sec. 14.9.2, p. 1002, in which interactions with the environment tend to rapidly randomize the wave-phases of large objects. When phases are randomized, interference and diffraction effects become undetectable.

Historically, it seemed absurd to the originators of quantum mechanics to imagine a macroscopic object in a superposition of states. The most celebrated example is called the Schrödinger’s cat experiment. Luckily for the cat, there probably was no actual experiment — it was simply a “thought experiment” that 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.

It seemed ridiculous to Schrödinger that at the end of the hour, “The uncertainty originally restricted to the atomic domain has been transformed into a macroscopic uncertainty...,” and the cat would be in a superposed state.

In modern language, people like Einstein and Schrödinger didn’t feel comfortable with nonlocality, or with entanglement of subatomic particles, and they felt even less comfortable with applying these concepts to macroscopic objects. Today, entanglement has been
demonstrated using objects that clearly deserve to be called macroscopic. For example, in 2012, K.C. Lee et al. created a version of the experiment in figure n in which the cameras were replaced by small diamonds, about 1 mm in size. They were separated by 15 cm, which is a macroscopic distance. When a photon hit one of the diamonds, it produced a vibration in the crystal lattice. This vibration was localized to a relatively small region within the diamond, but this region was still large enough that one has to admit that it qualifies as macroscopic. Its atoms had a total weight of about 0.1 nanograms, which is a quantity big enough to weigh on a state-of-the-art balance, and the region was about 0.01 mm in size, which would make it visible with a magnifying glass.

The quantum states of the two diamonds became entangled: if one had detected the photon, the other hadn’t. This entangled state was maintained for only about 7 picoseconds before decoherence destroyed the phase relationship between one diamond and the other. But Lee was able to use additional photons to “read out” the quantum states in only 0.5 ps, before decoherence occurred, and verify that there were wave interference effects in which one diamond’s quantum-mechanical wave had a definite phase relationship with the other’s. Although these experiments are difficult, they suggest that there is no obstruction in principle to observing quantum-mechanical effects such as superposition in arbitrarily large objects.

Entanglement is discussed in more mathematical detail in sec. 14.11, p. 1007.

The Copenhagen and many-worlds approximations

When we last saw Alice and Bob, they were in this superposition of states,

\[
|c\rangle + |c'\rangle,
\]

with

\[
|c| = |c'|.
\]

Let’s focus on Bob number one — the sad Bob — who didn’t see a photon. This is just one of the disappointments that Bob has experienced in his life, which include breaking up with his college crush and failing to summit Kilimanjaro due to altitude sickness. But Bob is a sane, normal person, and he’s not going to spend the rest of his life obsessing over how things might have been, in another world. Like a banker writing off a bad debt, Bob decides to stop maintaining all the bookkeeping that is, to him, irrelevant going forward. He now rewrites history and says that

\[
|c| = 1 \quad \text{and} \quad c' = 0.
\]
Technically speaking, this is wrong, because it is in principle still possible to have wave interference effects between sad Bob and happy Bob. But such effects are impractical to observe, due to effects like short wavelengths and decoherence, so what Bob is doing by “clearing the books” is an extremely good approximation. We will refer to this approximation by two different names, the Copenhagen approximation and the many-worlds approximation, for the following historical and psychological reasons.

In the early years of quantum mechanics, the school of physicists centering on Niels Bohr in Copenhagen were horribly confused about how to interpret quantum mechanics. They had all kinds of wrong ideas, such as the idea that quantum mechanics applied to individual atoms but not to light or to macroscopic objects. They didn’t know about decoherence. They thought there was a clear dividing line between microscopic things and macroscopic things (there isn’t), and they hypothesized that quantum mechanics only applied to microscopic ones (it applies to both). They claimed that clearing the books was an actual physical process, which they described as the “collapse” of the wave. This has traditionally been referred to as the Copenhagen “interpretation,” but we can now see that it is an approximation. There are cases where it is a bad approximation, e.g., at $t = 3$ ps during the experiment by Lee et al. (p. 888), when decoherence had started to happen but was only about half-way complete.

The many-worlds approximation came along a little later. It consists of making the same “clearing the books” approximation, but recognizing that there is no physical process of collapse.

Many physicists are philosophically attached to one or the other of these approximations, and would object to my description of them as approximations. My main purpose in writing this explanation is to immunize you against the impression, which can be mistakenly picked up from many descriptions of quantum mechanics, that a particular point of view on these topics (often the Copenhagen approximation) is somehow “standard.”

13.2.5 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.

A typical example of a probability distribution in section 13.1

---

5 It was originally proposed in a 1957 PhD thesis by Hugh Everett, who called it the relative state interpretation of quantum mechanics. Later it began to be referred to as the many-worlds interpretation.
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: 

\[(\text{area}) = (\text{height})(\text{width}) = m^{-1} \cdot m.\]

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 figure q. The units of $D$ must be $m^{-2}$ so that probabilities will be unitless: 

\[(\text{probability}) = (\text{depth})(\text{length})(\text{width}) = m^{-2} \cdot m \cdot m.\]

In terms of calculus, we have 

\[P = \int D \, dx \, dy.\]

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 not, then we can use an integral, 

\[P = \int D \, dx \, dy \, dz.\]
13.3 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. A more detailed calculation using classical physics gives an extremely weak binding, about 1/17 the strength of what we actually observe, which is far too weak to make the bond hold together.

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

### 13.3.1 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. Figure a, 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 section 13.4 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 chapter 11 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 importance 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.

\[\text{The wavelength of an elephant example 11} \]

\[\text{What is the wavelength of a trotting elephant?}\]

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

\[
\lambda = \frac{h}{p} = \frac{h}{mv} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{(10^3 \text{ kg})(10 \text{ m/s})} \sim 10^{-37} \text{ (kg} \cdot \text{m}^2/\text{s}^2) \cdot \text{s} \quad \text{kg} \cdot \text{m/s} = 10^{-37} \text{ m}
\]

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.

\[\text{The typical wavelength of an electron example 12} \]

\[\text{Electrons in circuits and in atoms are typically moving through}\]
These two electron waves are not distinguishable by any measuring device.

Voltage differences on the order of 1 V, so that a typical energy is \( e(1 \text{ V}) \), which is on the order of \( 10^{-19} \) J. What is the wavelength of an electron with this amount of kinetic energy?

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

\[
\lambda = \frac{h}{\sqrt{2mK}} = 1.6 \times 10^{-9} \text{ m.}
\]

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 E**

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?

**Answer, p. 1067**

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.

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

b / These two electron waves are not distinguishable by any measuring device.
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. This unobservability of the phase of the wavefunction is discussed in more detail on p. 917.

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, \( \Psi \) (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 \( \Psi^2 \), which is what is observable, not \( \Psi \) itself.

Since \( \Psi \), unlike \( E \) and \( B \), 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 \( \Psi^2 \), we simply define \( \Psi \) so that the constant of proportionality is one:

\[
(\text{probability distribution}) = |\Psi|^2.
\]

Since the probability distribution has units of \( \text{m}^{-3} \), the units of \( \Psi \) must be \( \text{m}^{-3/2} \). The square of a negative number is still positive, so the absolute value signs may seem unnecessary, but as we’ll see on p. 913 in sec. 13.3.6, the wavefunction may in general be a complex number. In fact, only standing waves, not traveling waves, can really be represented by real numbers, although we will often cheat and draw pictures of traveling waves as if they were real-valued functions.

**Discussion Question**

A 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? (This more symmetric treatment makes it easier to incorporate relativity into quantum mechanics, since relativity says that space and time are not entirely separate.)

13.3.2 **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
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:

\[
\lambda = \frac{v}{f} = \frac{vh}{E} = \frac{vh}{\frac{1}{2}mv^2} = \frac{2h}{mv} = \frac{2h}{p}.
\]

The reasoning seems valid, but the result does contradict the accepted one, which is after all 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, \( c \). 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.

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, \( v_g \).

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, \( d \). 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.

Actually the simplest wave that transports stuff from place to place is the pattern shown in figure e. 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 e) 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 \( \Delta f \) and \( \Delta z \) 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

\[
v_g = \frac{\Delta f}{\Delta z}
\]

If \( \Delta f \) and \( \Delta z \) are sufficiently
small, we can approximate this expression as a derivative,

\[ v_g = \frac{df}{dz}. \]

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 differs by a factor of two from what we would have expected, but the phase velocity is not the physically important thing.

### 13.3.3 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 generations 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, \( E = \frac{p^2}{2m} \). (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.

**Spectra of thin gases example 13**

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.)