9 In section 12.2 we've only done examples of mirrors with hollowed-out shapes (called concave mirrors). Now draw a ray diagram for a curved mirror that has a bulging outward shape (called a convex mirror). (a) How does the image’s distance from the mirror compare with the actual object’s distance from the mirror? From this comparison, determine whether the magnification is greater than or less than one. (b) Is the image real, or virtual? Could this mirror ever make the other type of image?

10 As discussed in question 9, there are two types of curved mirrors, concave and convex. Make a list of all the possible combinations of types of images (virtual or real) with types of mirrors (concave and convex). (Not all of the four combinations are physically possible.) Now for each one, use ray diagrams to determine whether increasing the distance of the object from the mirror leads to an increase or a decrease in the distance of the image from the mirror.

Draw BIG ray diagrams! Each diagram should use up about half a page of paper.

Some tips: To draw a ray diagram, you need two rays. For one of these, pick the ray that comes straight along the mirror’s axis, since its reflection is easy to draw. After you draw the two rays and locate the image for the original object position, pick a new object position that results in the same type of image, and start a new ray diagram, in a different color of pen, right on top of the first one. For the two new rays, pick the ones that just happen to hit the mirror at the same two places; this makes it much easier to get the result right without depending on extreme accuracy in your ability to draw the reflected rays.

11 If the user of an astronomical telescope moves her head closer to or farther away from the image she is looking at, does the magnification change? Does the angular magnification change? Explain. (For simplicity, assume that no eyepiece is being used.)

12 In figure g/2 in on page 756, only the image of my forehead was located by drawing rays. Either photocopy the figure or download the book and print out the relevant page. On this copy of the figure, make a new set of rays coming from my chin, and locate its image. To make it easier to judge the angles accurately, draw rays from the chin that happen to hit the mirror at the same points where the two rays from the forehead were shown hitting it. By comparing the locations of the chin’s image and the forehead’s image, verify that the image is actually upside-down, as shown in the original figure.

13 The figure shows four points where rays cross. Of these, which are image points? Explain.
Here’s a game my kids like to play. I sit next to a sunny window, and the sun reflects from the glass on my watch, making a disk of light on the wall or floor, which they pretend to chase as I move it around. Is the spot a disk because that’s the shape of the sun, or because it’s the shape of my watch? In other words, would a square watch make a square spot, or do we just have a circular image of the circular sun, which will be circular no matter what?

Apply the equation $M = d_i/d_o$ to the case of a flat mirror.

Use the method described in the text to derive the equation relating object distance to image distance for the case of a virtual image produced by a converging mirror.

Find the focal length of the mirror in problem 6.

Rank the focal lengths of the mirrors in the figure, from shortest to longest. Explain.

(a) A converging mirror with a focal length of 20 cm is used to create an image, using an object at a distance of 10 cm. Is the image real, or is it virtual? (b) How about $f = 20$ cm and $d_o = 30$ cm? (c) What if it was a diverging mirror with $f = 20$ cm and $d_o = 10$ cm? (d) A diverging mirror with $f = 20$ cm and $d_o = 30$ cm?

(a) Make up a numerical example of a virtual image formed by a converging mirror with a certain focal length, and determine the magnification. (You will need the result of problem 16.) Make sure to choose values of $d_o$ and $f$ that would actually produce a virtual image, not a real one. Now change the location of the object a little bit and redetermine the magnification, showing that it changes. At my local department store, the cosmetics department sells hand mirrors advertised as giving a magnification of 5 times. How would you interpret this?

(b) Suppose a Newtonian telescope is being used for astronomical observing. Assume for simplicity that no eyepiece is used, and assume a value for the focal length of the mirror that would be reasonable for an amateur instrument that is to fit in a closet. Is the angular magnification different for objects at different distances? For example, you could consider two planets, one of which is twice as far as the other.

(a) Find a case where the magnification of a curved mirror is infinite. Is the angular magnification infinite from any realistic viewing position? (b) Explain why an arbitrarily large magnification can’t be achieved by having a sufficiently small value of $d_o$. 

Problem 18.
22 A concave surface that reflects sound waves can act just like a converging mirror. Suppose that, standing near such a surface, you are able to find a point where you can place your head so that your own whispers are focused back on your head, so that they sound loud to you. Given your distance to the surface, what is the surface’s focal length?

23 The figure shows a device for constructing a realistic optical illusion. Two mirrors of equal focal length are put against each other with their silvered surfaces facing inward. A small object placed in the bottom of the cavity will have its image projected in the air above. The way it works is that the top mirror produces a virtual image, and the bottom mirror then creates a real image of the virtual image. (a) Show that if the image is to be positioned as shown, at the mouth of the cavity, then the focal length of the mirrors is related to the dimension $h$ via the equation

$$\frac{1}{f} = \frac{1}{h} + \frac{1}{h + \left(\frac{1}{h} - \frac{1}{f}\right)^{-1}}.$$  

(b) Restate the equation in terms of a single variable $x = h/f$, and show that there are two solutions for $x$. Which solution is physically consistent with the assumptions of the calculation?

24 (a) A converging mirror is being used to create a virtual image. What is the range of possible magnifications? (b) Do the same for the other types of images that can be formed by curved mirrors (both converging and diverging).

25 A diverging mirror of focal length $f$ is fixed, and faces down. An object is dropped from the surface of the mirror, and falls away from it with acceleration $g$. The goal of the problem is to find the maximum velocity of the image.  
(a) Describe the motion of the image verbally, and explain why we should expect there to be a maximum velocity.
(b) Use arguments based on units to determine the form of the solution, up to an unknown unitless multiplicative constant.
(c) Complete the solution by determining the unitless constant.

26 Diamond has an index of refraction of 2.42, and part of the reason diamonds sparkle is that this encourages a light ray to undergo many total internal reflections before it emerges. (a) Calculate the critical angle at which total internal reflection occurs in diamond. (b) Explain the interpretation of your result: Is it measured from the normal, or from the surface? Is it a minimum angle for total internal reflection, or is it a maximum? How would the critical angle have been different for a substance such as glass or plastic, with a lower index of refraction?
27 Suppose a converging lens is constructed of a type of plastic whose index of refraction is less than that of water. How will the lens’s behavior be different if it is placed underwater?

28 There are two main types of telescopes, refracting (using a lens) and reflecting (using a mirror, as in figure i on p. 758). (Some telescopes use a mixture of the two types of elements: the light first encounters a large curved mirror, and then goes through an eyepiece that is a lens. To keep things simple, assume no eyepiece is used.) What implications would the color-dependence of focal length have for the relative merits of the two types of telescopes? Describe the case where an image is formed of a white star. You may find it helpful to draw a ray diagram.

29 Based on Snell’s law, explain why rays of light passing through the edges of a converging lens are bent more than rays passing through parts closer to the center. It might seem like it should be the other way around, since the rays at the edge pass through less glass — shouldn’t they be affected less? In your answer:

- Include a ray diagram showing a huge, full-page, close-up view of the relevant part of the lens.

- Make use of the fact that the front and back surfaces aren’t always parallel; a lens in which the front and back surfaces are always parallel doesn’t focus light at all, so if your explanation doesn’t make use of this fact, your argument must be incorrect.

- Make sure your argument still works even if the rays don’t come in parallel to the axis.

30 When you take pictures with a camera, the distance between the lens and the film has to be adjusted, depending on the distance at which you want to focus. This is done by moving the lens. If you want to change your focus so that you can take a picture of something farther away, which way do you have to move the lens? Explain using ray diagrams. [Based on a problem by Eric Mazur.]

31 When swimming underwater, why is your vision made much clearer by wearing goggles with flat pieces of glass that trap air behind them? [Hint: You can simplify your reasoning by considering the special case where you are looking at an object far away, and along the optic axis of the eye.]

32 An object is more than one focal length from a converging lens. (a) Draw a ray diagram. (b) Using reasoning like that developed in section 12.3, determine the positive and negative signs in the equation \(1/f = \pm 1/d_i \pm 1/d_o\). (c) The images of the rose in
section 4.2 were made using a lens with a focal length of 23 cm. If the lens is placed 80 cm from the rose, locate the image. √

33 The figure shows four lenses. Lens 1 has two spherical surfaces. Lens 2 is the same as lens 1 but turned around. Lens 3 is made by cutting through lens 1 and turning the bottom around. Lens 4 is made by cutting a central circle out of lens 1 and recessing it.

(a) A parallel beam of light enters lens 1 from the left, parallel to its axis. Reasoning based on Snell’s law, will the beam emerging from the lens be bent inward, or outward, or will it remain parallel to the axis? Explain your reasoning. As part of your answer, make a huge drawing of one small part of the lens, and apply Snell’s law at both interfaces. Recall that rays are bent more if they come to the interface at a larger angle with respect to the normal.

(b) What will happen with lenses 2, 3, and 4? Explain. Drawings are not necessary.

34 The drawing shows the anatomy of the human eye, at twice life size. Find the radius of curvature of the outer surface of the cornea by measurements on the figure, and then derive the focal length of the air-cornea interface, where almost all the focusing of light occurs. You will need to use physical reasoning to modify the lensmaker’s equation for the case where there is only a single refracting surface. Assume that the index of refraction of the cornea is essentially that of water.

35 An object is less than one focal length from a converging lens. (a) Draw a ray diagram. (b) Using reasoning like that developed in section 12.3, determine the positive and negative signs in the equation $1/f = \pm 1/d_i \pm 1/d_o$. (c) The images of the rose in section 4.2 were made using a lens with a focal length of 23 cm. If the lens is placed 10 cm from the rose, locate the image. √

36 Nearsighted people wear glasses whose lenses are diverging. (a) Draw a ray diagram. For simplicity pretend that there is no eye behind the glasses. (b) Using reasoning like that developed in section 12.3, determine the positive and negative signs in the equation $1/f = \pm 1/d_i \pm 1/d_o$. (c) If the focal length of the lens is 50.0 cm, and the person is looking at an object at a distance of 80.0 cm, locate the image. √

37 (a) Light is being reflected diffusely from an object 1.000 m underwater. The light that comes up to the surface is refracted at the water-air interface. If the refracted rays all appear to come from the same point, then there will be a virtual image of the object in the water, above the object’s actual position, which will be visible to an observer above the water. Consider three rays, A, B and C, whose angles in the water with respect to the normal are $\theta_i = 0.000^\circ$, $1.000^\circ$ and $20.000^\circ$ respectively. Find the depth of the point at which
the refracted parts of A and B appear to have intersected, and do the same for A and C. Show that the intersections are at nearly the same depth, but not quite. [Check: The difference in depth should be about 4 cm.]

(b) Since all the refracted rays do not quite appear to have come from the same point, this is technically not a virtual image. In practical terms, what effect would this have on what you see?

(c) In the case where the angles are all small, use algebra and trig to show that the refracted rays do appear to come from the same point, and find an equation for the depth of the virtual image. Do not put in any numerical values for the angles or for the indices of refraction — just keep them as symbols. You will need the approximation \( \sin \theta \approx \tan \theta \approx \theta \), which is valid for small angles measured in radians.

38  Prove that the principle of least time leads to Snell’s law.

39  Two standard focal lengths for camera lenses are 50 mm (standard) and 28 mm (wide-angle). To see how the focal lengths relate to the angular size of the field of view, it is helpful to visualize things as represented in the figure. Instead of showing many rays coming from the same point on the same object, as we normally do, the figure shows two rays from two different objects. Although the lens will intercept infinitely many rays from each of these points, we have shown only the ones that pass through the center of the lens, so that they suffer no angular deflection. (Any angular deflection at the front surface of the lens is canceled by an opposite deflection at the back, since the front and back surfaces are parallel at the lens’s center.) What is special about these two rays is that they are aimed at the edges of one 35-mm-wide frame of film; that is, they show the limits of the field of view. Throughout this problem, we assume that \( d_o \) is much greater than \( d_i \). (a) Compute the angular width of the camera’s field of view when these two lenses are used. (b) Use small-angle approximations to find a simplified equation for the angular width of the field of view, \( \theta \), in terms of the focal length, \( f \), and the width of the film, \( w \). Your equation should not have any trig functions in it. Compare the results of this approximation with your answers from part a. (c) Suppose that we are holding constant the aperture (amount of surface area of the lens being used to collect light). When switching from a 50-mm lens to a 28-mm lens, how many times longer or shorter must the exposure be in order to make a properly developed picture, i.e., one that is not under- or overexposed? [Based on a problem by Arnold Arons.]

▷ Solution, p. 944
A近視的人是其眼睛聚焦光強度太強，且因此無法適應其眼睛內的晶體，使其無法在視野中形成物體的像。這就是近視。

(a) 在未校正的視野中，畫出當人試圖聚焦在無窮遠處時的情況。

(b) 她的玻璃鏡片是什麼類型的？解釋。

(c) 畫出當她戴上鏡片時的情況。標記兩張照片，分別是鏡片所形成的像和最終的像。

(d) 如果她有時使用接觸鏡片而不是鏡片，她的接觸鏡片的焦距必須是小於，等於，還是大於於普通鏡片的焦距？解釋。

Fred的的眼睛可以聚焦在距離5.0厘米的地方。Fred把一個焦距為3.0厘米的放大鏡放在距平板蟲2.0厘米高的位置。 (a) 標記像的位置，並找出放大率。(b) 沒有用放大鏡的情況下，Fred會希望從多遠的位置來看平板蟲來看得更清楚？有放大鏡的情況下？(c) 計算視角放大率。
Panel 1 of the figure shows the optics inside a pair of binoculars. They are essentially a pair of telescopes, one for each eye. But to make them more compact, and allow the eyepieces to be the right distance apart for a human face, they incorporate a set of eight prisms, which fold the light path. In addition, the prisms make the image upright. Panel 2 shows one of these prisms, known as a Porro prism. The light enters along a normal, undergoes two total internal reflections at angles of 45 degrees with respect to the back surfaces, and exits along a normal. The image of the letter R has been flipped across the horizontal. Panel 3 shows a pair of these prisms glued together. The image will be flipped across both the horizontal and the vertical, which makes it oriented the right way for the user of the binoculars.

(a) Find the minimum possible index of refraction for the glass used in the prisms.
(b) For a material of this minimal index of refraction, find the fraction of the incoming light that will be lost to reflection in the four Porro prisms on each side of a pair of binoculars. (See section 6.2.) In real, high-quality binoculars, the optical surfaces of the prisms have antireflective coatings, but carry out your calculation for the case where there is no such coating.
(c) Discuss the reasons why a designer of binoculars might or might not want to use a material with exactly the index of refraction found in part a.

It would be annoying if your eyeglasses produced a magnified or reduced image. Prove that when the eye is very close to a lens, and the lens produces a virtual image, the angular magnification is always approximately equal to 1 (regardless of whether the lens is diverging or converging).
44. The figure shows a diffraction pattern made by a double slit, along with an image of a meter stick to show the scale. Sketch the diffraction pattern from the figure on your paper. Now consider the four variables in the equation $\frac{\lambda}{d} = \frac{\sin \theta}{m}$. Which of these are the same for all five fringes, and which are different for each fringe? Which variable would you naturally use in order to label which fringe was which? Label the fringes on your sketch using the values of that variable.

45. Match gratings A-C with the diffraction patterns 1-3 that they produce. Explain.

A

B

C

1

2

3
46 The figure below shows two diffraction patterns. The top one was made with yellow light, and the bottom one with red. Could the slits used to make the two patterns have been the same?

47 The figure on p. 809 shows a diffraction pattern made by a double slit, along with an image of a meter stick to show the scale. The slits were 146 cm away from the screen on which the diffraction pattern was projected. The spacing of the slits was 0.050 mm. What was the wavelength of the light?

48 Why would blue or violet light be the best for microscopy?

49 The figure below shows two diffraction patterns, both made with the same wavelength of red light. (a) What type of slits made the patterns? Is it a single slit, double slits, or something else? Explain. (b) Compare the dimensions of the slits used to make the top and bottom pattern. Give a numerical ratio, and state which way the ratio is, i.e., which slit pattern was the larger one. Explain.

50 When white light passes through a diffraction grating, what is the smallest value of \( m \) for which the visible spectrum of order \( m \) overlaps the next one, of order \( m + 1 \)? (The visible spectrum runs from about 400 nm to about 700 nm.)
51. For star images such as the ones in figure y, estimate the angular width of the diffraction spot due to diffraction at the mouth of the telescope. Assume a telescope with a diameter of 10 meters (the largest currently in existence), and light with a wavelength in the middle of the visible range. Compare with the actual angular size of a star of diameter $10^9$ m seen from a distance of $10^{17}$ m. What does this tell you?

52. The figure below shows three diffraction patterns. All were made under identical conditions, except that a different set of double slits was used for each one. The slits used to make the top pattern had a center-to-center separation $d = 0.50$ mm, and each slit was $w = 0.04$ mm wide. (a) Determine $d$ and $w$ for the slits used to make the pattern in the middle. (b) Do the same for the slits used to make the bottom pattern.
53 The beam of a laser passes through a diffraction grating, fans out, and illuminates a wall that is perpendicular to the original beam, lying at a distance of 2.0 m from the grating. The beam is produced by a helium-neon laser, and has a wavelength of 694.3 nm. The grating has 2000 lines per centimeter. (a) What is the distance on the wall between the central maximum and the maxima immediately to its right and left? (b) How much does your answer change when you use the small-angle approximations $\theta \approx \sin \theta \approx \tan \theta$?

54 Ultrasound, i.e., sound waves with frequencies too high to be audible, can be used for imaging fetuses in the womb or for breaking up kidney stones so that they can be eliminated by the body. Consider the latter application. Lenses can be built to focus sound waves, but because the wavelength of the sound is not all that small compared to the diameter of the lens, the sound will not be concentrated exactly at the geometrical focal point. Instead, a diffraction pattern will be created with an intense central spot surrounded by fainter rings. About 85% of the power is concentrated within the central spot. The angle of the first minimum (surrounding the central spot) is given by $\sin \theta = \lambda/b$, where $b$ is the diameter of the lens. This is similar to the corresponding equation for a single slit, but with a factor of 1.22 in front which arises from the circular shape of the aperture. Let the distance from the lens to the patient’s kidney stone be $L = 20$ cm. You will want $f > 20$ kHz, so that the sound is inaudible. Find values of $b$ and $f$ that would result in a usable design, where the central spot is small enough to lie within a kidney stone 1 cm in diameter.

55 Under what circumstances could one get a mathematically undefined result by solving the double-slit diffraction equation for $\theta$? Give a physical interpretation of what would actually be observed.

56 When ultrasound is used for medical imaging, the frequency may be as high as 5-20 MHz. Another medical application of ultrasound is for therapeutic heating of tissues inside the body; here, the frequency is typically 1-3 MHz. What fundamental physical reasons could you suggest for the use of higher frequencies for imaging?
57  Suppose we have a polygonal room whose walls are mirrors, and there a pointlike light source in the room. In most such examples, every point in the room ends up being illuminated by the light source after some finite number of reflections. A difficult mathematical question, first posed in the middle of the last century, is whether it is ever possible to have an example in which the whole room is \textit{not} illuminated. (Rays are assumed to be absorbed if they strike exactly at a vertex of the polygon, or if they pass exactly through the plane of a mirror.)

The problem was finally solved in 1995 by G.W. Tokarsky, who found an example of a room that was not illuminable from a certain point. Figure 57 shows a slightly simpler example found two years later by D. Castro. If a light source is placed at either of the locations shown with dots, the other dot remains unilluminated, although every other point is lit up. It is not straightforward to prove rigorously that Castro’s solution has this property. However, the plausibility of the solution can be demonstrated as follows.

Suppose the light source is placed at the right-hand dot. Locate all the images formed by single reflections. Note that they form a regular pattern. Convince yourself that none of these images illuminates the left-hand dot. Because of the regular pattern, it becomes plausible that even if we form images of images, images of images of images, etc., none of them will ever illuminate the other dot.

There are various other versions of the problem, some of which remain unsolved. The book by Klee and Wagon gives a good introduction to the topic, although it predates Tokarsky and Castro’s work.

References:

58  A mechanical linkage is a device that changes one type of motion into another. The most familiar example occurs in a gasoline car’s engine, where a connecting rod changes the linear motion of the piston into circular motion of the crankshaft. The top panel of the figure shows a mechanical linkage invented by Peaucellier in 1864, and independently by Lipkin around the same time. It consists of six rods joined by hinges, the four short ones forming a rhombus. Point O is fixed in space, but the apparatus is free to rotate about O. Motion at P is transformed into a different motion at P’ (or vice versa).

Geometrically, the linkage is a mechanical implementation of
the ancient problem of inversion in a circle. Considering the case in which the rhombus is folded flat, let the \( k \) be the distance from \( O \) to the point where \( P \) and \( P' \) coincide. Form the circle of radius \( k \) with its center at \( O \). As \( P \) and \( P' \) move in and out, points on the inside of the circle are always mapped to points on its outside, such that \( rr' = k^2 \). That is, the linkage is a type of analog computer that exactly solves the problem of finding the inverse of a number \( r \). Inversion in a circle has many remarkable geometrical properties, discussed in H.S.M. Coxeter, *Introduction to Geometry*, Wiley, 1961. If a pen is inserted through a hole at \( P \), and \( P' \) is traced over a geometrical figure, the Peaucellier linkage can be used to draw a kind of image of the figure.

A related problem is the construction of pictures, like the one in the bottom panel of the figure, called anamorphs. The drawing of the column on the paper is highly distorted, but when the reflecting cylinder is placed in the correct spot on top of the page, an undistorted image is produced inside the cylinder. (Wide-format movie technologies such as Cinemascope are based on similar principles.)

Show that the Peaucellier linkage does not convert correctly between an image and its anamorph, and design a modified version of the linkage that does. Some knowledge of analytic geometry will be helpful.

59 The figure shows a lens with surfaces that are curved, but whose thickness is constant along any horizontal line. Use the lens-maker’s equation to prove that this “lens” is not really a lens at all.

\( \triangleright \) Solution, p. 944

60 Under ordinary conditions, gases have indices of refraction only a little greater than that of vacuum, i.e., \( n = 1 + \epsilon \), where \( \epsilon \) is some small number. Suppose that a ray crosses a boundary between a region of vacuum and a region in which the index of refraction is \( 1 + \epsilon \). Find the maximum angle by which such a ray can ever be deflected, in the limit of small \( \epsilon \).

\( \triangleright \) Hint, p. 925

61 A converging mirror has focal length \( f \). An object is located at a distance \( (1 + \epsilon)f \) from the mirror, where \( \epsilon \) is small. Find the distance of the image from the mirror, simplifying your result as much as possible by using the assumption that \( \epsilon \) is small.

\( \triangleright \) Answer, p. 934

---

*Key to symbols:*

- easy
- typical
- challenging
- difficult
- very difficult

✓ An answer check is available at www.lightandmatter.com.
Exercises

Exercise 12A: Exploring Images With a Curved Mirror

Equipment:
- concave mirrors with deep curvature
- concave mirrors with gentle curvature
- convex mirrors

1. Obtain a curved mirror from your instructor. If it is silvered on both sides, make sure you're working with the concave side, which bends light rays inward. Look at your own face in the mirror. Now change the distance between your face and the mirror, and see what happens. Explore the full range of possible distances between your face and the mirror.

In these observations you've been changing two variables at once: the distance between the object (your face) and the mirror, and the distance from the mirror to your eye. In general, scientific experiments become easier to interpret if we practice isolation of variables, i.e., only change one variable while keeping all the others constant. In parts 2 and 3 you'll form an image of an object that's not your face, so that you can have independent control of the object distance and the point of view.

2. With the mirror held far away from you, observe the image of something behind you, over your shoulder. Now bring your eye closer and closer to the mirror. Can you see the image with your eye very close to the mirror? See if you can explain your observation by drawing a ray diagram.
3. Now imagine the following new situation, but *don’t actually do it yet*. Suppose you lay the mirror face-up on a piece of tissue paper, put your finger a few cm above the mirror, and look at the image of your finger. As in part 2, you can bring your eye closer and closer to the mirror. Will you be able to see the image with your eye very close to the mirror? Draw a ray diagram to help you predict what you will observe.

Prediction:________________________

Now test your prediction. If your prediction was incorrect, see if you can figure out what went wrong, or ask your instructor for help.

4. For parts 4 and 5, it’s more convenient to use concave mirrors that are more gently curved; obtain one from your instructor. Lay the mirror on the tissue paper, and use it to create an image of the overhead lights on a piece of paper above it and a little off to the side. What do you have to do in order to make the image clear? Can you explain this observation using a ray diagram?
5. Now imagine the following experiment, but *don’t do it yet*. What will happen to the image on the paper if you cover half of the mirror with your hand?

Prediction:_________________________

Test your prediction. If your prediction was incorrect, can you explain what happened?

6. Now imagine forming an image with a convex mirror (one that bulges outward), and that therefore bends light rays away from the central axis (i.e., is diverging). Draw a typical ray diagram.

Is the image real, or virtual? Will there be more than one type of image?

Prediction:_________________________

Test your prediction.
Exercise 12B: Object and Image Distances

Equipment:
- optical benches
- converging mirrors
- illuminated objects

1. Set up the optical bench with the mirror at zero on the centimeter scale. Set up the illuminated object on the bench as well.

2. Each group will locate the image for their own value of the object distance, by finding where a piece of paper has to be placed in order to see the image on it. (The instructor will do one point as well.) Note that you will have to tilt the mirror a little so that the paper on which you project the image doesn’t block the light from the illuminated object.

Is the image real or virtual? How do you know? Is it inverted, or uninverted?

Draw a ray diagram.

3. Measure the image distance and write your result in the table on the board. Do the same for the magnification.

4. What do you notice about the trend of the data on the board? Draw a second ray diagram with a different object distance, and show why this makes sense. Some tips for doing this correctly: (1) For simplicity, use the point on the object that is on the mirror’s axis. (2) You need to trace two rays to locate the image. To save work, don’t just do two rays at random angles. You can either use the on-axis ray as one ray, or do two rays that come off at the same angle, one above and one below the axis. (3) Where each ray hits the mirror, draw the normal line, and make sure the ray is at equal angles on both sides of the normal.

5. We will find the mirror’s focal length from the instructor’s data-point. Then, using this focal length, calculate a theoretical prediction of the image distance, and write it on the board next to the experimentally determined image distance.
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, 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

---

\(^1\)This 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 827, 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
Why are dice random?

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

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

**self-check A**

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

**Discussion Questions**

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

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

C The sequence of digits 121212121212121212 seems clearly nonrandom, and 41592653589793 seems random. The latter sequence, however, is the decimal form of pi, starting with the third digit. There is a story about the Indian mathematician Ramanujan, a self-taught prodigy, that 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 include a book containing only the repeated letter “a”; all the ancient Greek tragedies known today, all the lost Greek tragedies, and millions of Greek tragedies that were never actually written; your own life story, and various incorrect versions of your own life story; and countless anthologies containing a short story called “The Library of Babel.” Of course, if you picked a book from the shelves of the library, it would almost certainly look like a
nonsensical sequence of letters and punctuation, but it’s always possible that the seemingly meaningless book would be a science-fiction screenplay written in the language of a Neanderthal tribe, or 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: 2.9999999999999999687687658465436, for example. It doesn’t usually make sense, therefore, to talk about the probability of a single numerical result, but it does make sense to talk about the probability of a certain range of results. For instance, the probability that an atom will last more than 3 and less than 4 seconds is a perfectly reasonable thing to discuss. We can still summarize the probability information on a graph, and we can still interpret areas under the curve as probabilities.
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. 933

**Looking for tall basketball players**

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 mea-
sure 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.

\[\text{Radioactive contamination at Chernobyl example 2}\]

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.

\[\text{\(^{14}\text{C Dating example 3}\)}\]

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 age of dead organisms, or human artifacts made from plants or animals. Figure j on page 836 shows the exponential decay curve of \(^{14}\text{C}\) in various objects. Similar methods, using longer-lived isotopes, provided the first firm proof that the earth was billions of years old, not a few thousand as some had claimed on religious grounds.
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 \( \frac{dN}{dt} \). It is therefore more convenient to talk about the rate of decay \( \frac{dN}{dt} \) 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. 839 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.  
▶ Answer, p. 933

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 \(^{17}\)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 \(^{131}\)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 \( x_1 P_1 + x_2 P_2 + \ldots \). But we need to generalize this to a variable \( x \) that can take on any of a continuum of values. The continuous version of a sum is an integral, so the average is

\[
\text{average value of } x = \int x D(x) \, dx,
\]

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

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

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

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

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

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

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

\[
(\text{average lifetime}) = \frac{\ln 2}{t_{1/2}} \left( \frac{t_{1/2}}{\ln 2} \right)^2 = \frac{t_{1/2}}{\ln 2} = 1.443 \, t_{1/2},
\]

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.

*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 chapter 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 fields. The stronger the fields, i.e., the greater the wave’s amplitude, the greater the forces that would be exerted on electrons that
found themselves bathed in the light. It should have been amplitude (brightness) that was relevant, not frequency. The dependence on frequency not only proves that the wave model of light needs modifying, but with the proper interpretation it allows us to determine how much energy is in one photon, and it also leads to a connection between the wave and particle models that we need in order to reconcile them.

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

Prompted by Einstein’s photon paper, Robert Millikan (whom we first encountered in chapter 8) figured out how to use the photoelectric effect to probe precisely the link between frequency and photon energy. Rather than going into the historical details of Millikan’s actual experiments (a lengthy experimental program that occupied a large part of his professional career) we will describe a simple version, shown in figure g, that is used sometimes in college laboratory courses. The idea is simply to illuminate one plate of the vacuum tube with light of a single wavelength and monitor the voltage difference between the two plates as they charge up. Since the resistance of a voltmeter is very high (much higher than the resistance of an ammeter), we can assume to a good approximation that electrons reaching the top plate are stuck there permanently, so the voltage will keep on increasing for as long as electrons are making it across the vacuum tube.

At a moment when the voltage difference has 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.

---

2What I’m presenting in this chapter is a simplified explanation of how the photon could have been discovered. The actual history is more complex. Max Planck (1858-1947) began the photon saga with a theoretical investigation of the spectrum of light emitted by a hot, glowing object. He introduced quantization of the energy of light waves, in multiples of $hf$, purely as a mathematical trick that happened to produce the right results. Planck did not believe that his procedure could have any physical significance. In his 1905 paper Einstein took Planck’s quantization as a description of reality, and applied it to various theoretical and experimental puzzles, including the photoelectric effect. Millikan then subjected Einstein’s ideas to a series of rigorous experimental tests. Although his results matched Einstein’s predictions perfectly, Millikan was skeptical about photons, and his papers conspicuously omit any reference to them. Only in his autobiography did Millikan rewrite history and claim that he had given experimental proof for photons.
**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. 933

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

**Number of photons emitted by a lightbulb per second example 7**

Roughly how many photons are emitted by a 100-W 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

$$\frac{(100 \text{ J})}{E_{\text{photon}}} = \frac{(100 \text{ J})}{\left(\frac{hc}{\lambda}\right)}$$

$$\approx 3 \times 10^{20}$$

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 $\mathbf{E}$ and $\mathbf{B}$, 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. 907 verifies that the number is in fact extremely large.

\[ \text{Momentum of a photon example 9} \]

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

Combining the equations \( p = \frac{E}{c} \) 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. 841. 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 841 are simply cutouts from these figures.)

![Wave interference patterns photographed by Prof. Lyman Page with a digital camera. Laser light with a single well-defined wavelength passed through a series of absorbers to cut down its intensity, then through a set of slits to produce interference, and finally into a digital camera chip. (A triple slit was actually used, but for conceptual simplicity we discuss the results in the main text as if it was a double slit.) In panel 2 the intensity has been reduced relative to 1, and even more so for panel 3.](image)
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 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 was conclusively disproved by G.I. Taylor, a student at Cambridge. The demonstration by Prof. Page that we’ve just been discussing is essentially a modernized version of Taylor’s work. Taylor reasoned that if interference effects came from photons interacting with each other, a bare minimum of two photons would have to be present at the same time to produce interference. By making the light source extremely dim, we can be virtually certain that there are never two photons in the box at the same time. In figure i/3, however, the intensity of the light has been cut down so much by the absorbers that if it was in the open, the average separation between photons would be on the order of a kilometer! At any given moment, the number of photons in the box is most likely to be zero. It is virtually certain that there were never two photons in the box at once.

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
Bullets pass through a double slit.

A water wave passes through a double slit.

A single photon can go through both slits.

would not demonstrate wave superposition and interference effects, contradicting its wave nature. (In subsection 13.3.4 we will discuss the Heisenberg uncertainty principle, which gives a numerical way of approaching this issue.)

Another wrong interpretation: the pilot wave hypothesis

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

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

The probability interpretation

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

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

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

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