

STUDY MATERIAL FOR 19PHY101

This compilation is for personal use and doesn't violate copyright issues

Coulomb's Law

21-1 COULOMB'S LAW

Learning Objectives

After reading this module, you should be able to...

- 21.01** Distinguish between being electrically neutral, negatively charged, and positively charged and identify excess charge.
- 21.02** Distinguish between conductors, nonconductors (insulators), semiconductors, and superconductors.
- 21.03** Describe the electrical properties of the particles inside an atom.
- 21.04** Identify conduction electrons and explain their role in making a conducting object negatively or positively charged.
- 21.05** Identify what is meant by "electrically isolated" and by "grounding."
- 21.06** Explain how a charged object can set up induced charge in a second object.
- 21.07** Identify that charges with the same electrical sign repel each other and those with opposite electrical signs attract each other.
- 21.08** For either of the particles in a pair of charged particles, draw a free-body diagram, showing the electrostatic force (Coulomb force) on it and anchoring the tail of the force vector on that particle.
- 21.09** For either of the particles in a pair of charged particles, apply Coulomb's law to relate the magnitude of the electrostatic force, the charge magnitudes of the particles, and the separation between the particles.

Key Ideas

- The strength of a particle's electrical interaction with objects around it depends on its electric charge (usually represented as q), which can be either positive or negative. Particles with the same sign of charge repel each other, and particles with opposite signs of charge attract each other.
- An object with equal amounts of the two kinds of charge is electrically neutral, whereas one with an imbalance is electrically charged and has an excess charge.
- Conductors are materials in which a significant number of electrons are free to move. The charged particles in nonconductors (insulators) are not free to move.
- Electric current i is the rate dq/dt at which charge passes a point:

$$i = \frac{dq}{dt}.$$

- Coulomb's law describes the electrostatic force (or electric

- 21.10** Identify that Coulomb's law applies only to (point-like) particles and objects that can be treated as particles.
- 21.11** If more than one force acts on a particle, find the net force by adding all the forces as vectors, not scalars.
- 21.12** Identify that a shell of uniform charge attracts or repels a charged particle that is outside the shell as if all the shell's charge were concentrated as a particle at the shell's center.
- 21.13** Identify that if a charged particle is located inside a shell of uniform charge, there is no net electrostatic force on the particle from the shell.
- 21.14** Identify that if excess charge is put on a spherical conductor, it spreads out uniformly over the external surface area.
- 21.15** Identify that if two identical spherical conductors touch or are connected by conducting wire, any excess charge will be shared equally.
- 21.16** Identify that a nonconducting object can have any given distribution of charge, including charge at interior points.
- 21.17** Identify current as the rate at which charge moves through a point.
- 21.18** For current through a point, apply the relationship between the current, a time interval, and the amount of charge that moves through the point in that time interval.

force) between two charged particles. If the particles have charges q_1 and q_2 , are separated by distance r , and are at rest (or moving only slowly) relative to each other, then the magnitude of the force acting on each due to the other is given by

$$F = \frac{1}{4\pi\epsilon_0} \frac{|q_1||q_2|}{r^2} \quad (\text{Coulomb's law}),$$

where $\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$ is the permittivity constant. The ratio $1/4\pi\epsilon_0$ is often replaced with the electrostatic constant (or Coulomb constant) $k = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.

- The electrostatic force vector acting on a charged particle due to a second charged particle is either directly toward the second particle (opposite signs of charge) or directly away from it (same sign of charge).
- If multiple electrostatic forces act on a particle, the net force is the vector sum (not scalar sum) of the individual forces.

- Shell theorem 1: A charged particle outside a shell with charge uniformly distributed on its surface is attracted or repelled as if the shell's charge were concentrated as a particle at its center.
- Shell theorem 2: A charged particle inside a shell with

charge uniformly distributed on its surface has no net force acting on it due to the shell.

- Charge on a conducting spherical shell spreads uniformly over the (external) surface.

What Is Physics?

You are surrounded by devices that depend on the physics of electromagnetism, which is the combination of electric and magnetic phenomena. This physics is at the root of computers, television, radio, telecommunications, household lighting, and even the ability of food wrap to cling to a container. This physics is also the basis of the natural world. Not only does it hold together all the atoms and molecules in the world, it also produces lightning, auroras, and rainbows.

The physics of electromagnetism was first studied by the early Greek philosophers, who discovered that if a piece of amber is rubbed and then brought near bits of straw, the straw will jump to the amber. We now know that the attraction between amber and straw is due to an electric force. The Greek philosophers also discovered that if a certain type of stone (a naturally occurring magnet) is brought near bits of iron, the iron will jump to the stone. We now know that the attraction between magnet and iron is due to a magnetic force.

From these modest origins with the Greek philosophers, the sciences of electricity and magnetism developed separately for centuries—until 1820, in fact, when Hans Christian Oersted found a connection between them: an electric current in a wire can deflect a magnetic compass needle. Interestingly enough, Oersted made this discovery, a big surprise, while preparing a lecture demonstration for his physics students.

The new science of electromagnetism was developed further by workers in many countries. One of the best was Michael Faraday, a truly gifted experimenter with a talent for physical intuition and visualization. That talent is attested to by the fact that his collected laboratory notebooks do not contain a single equation. In the mid-nineteenth century, James Clerk Maxwell put Faraday's ideas into mathematical form, introduced many new ideas of his own, and put electromagnetism on a sound theoretical basis.

Our discussion of electromagnetism is spread through the next 16 chapters. We begin with electrical phenomena, and our first step is to discuss the nature of electric charge and electric force.

Electric Charge

Here are two demonstrations that seem to be magic, but our job here is to make sense of them. After rubbing a glass rod with a silk cloth (on a day when the humidity is low), we hang the rod by means of a thread tied around its center (Fig. 21-1a). Then we rub a second glass rod with the silk cloth and bring it near the hanging rod. The hanging rod magically moves away. We can see that a force repels it from the second rod, but how? There is no contact with that rod, no breeze to push on it, and no sound wave to disturb it.

In the second demonstration we replace the second rod with a plastic rod that has been rubbed with fur. This time, the hanging rod moves toward the nearby rod (Fig. 21-1b). Like the repulsion, this attraction occurs without any contact or obvious communication between the rods.

In the next chapter we shall discuss how the hanging rod knows of the presence of the other rods, but in this chapter let's focus on just the forces that are involved. In the first demonstration, the force on the hanging rod was *repulsive*, and

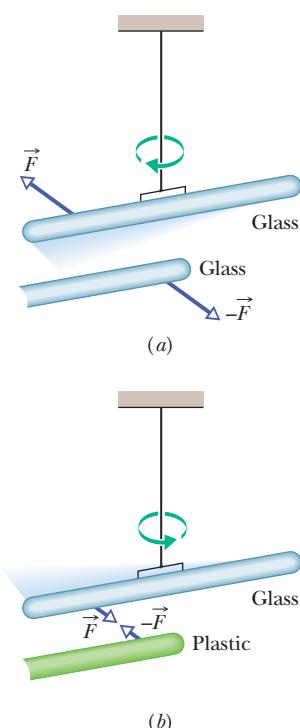


Figure 21-1 (a) The two glass rods were each rubbed with a silk cloth and one was suspended by thread. When they are close to each other, they repel each other. (b) The plastic rod was rubbed with fur. When brought close to the glass rod, the rods attract each other.

in the second, *attractive*. After a great many investigations, scientists figured out that the forces in these types of demonstrations are due to the *electric charge* that we set up on the rods when they are in contact with silk or fur. Electric charge is an intrinsic property of the fundamental particles that make up objects such as the rods, silk, and fur. That is, charge is a property that comes automatically with those particles wherever they exist.

Two Types. There are two types of electric charge, named by the American scientist and statesman Benjamin Franklin as positive charge and negative charge. He could have called them anything (such as cherry and walnut), but using algebraic signs as names comes in handy when we add up charges to find the net charge. In most everyday objects, such as a mug, there are about equal numbers of negatively charged particles and positively charged particles, and so the net charge is zero, the charge is said to be *balanced*, and the object is said to be *electrically neutral* (or just *neutral* for short).

Excess Charge. Normally you are approximately neutral. However, if you live in regions where the humidity is low, you know that the charge on your body can become slightly unbalanced when you walk across certain carpets. Either you gain negative charge from the carpet (at the points of contact between your shoes with the carpet) and become negatively charged, or you lose negative charge and become positively charged. Either way, the extra charge is said to be an *excess charge*. You probably don't notice it until you reach for a door handle or another person. Then, if your excess charge is enough, a spark leaps between you and the other object, eliminating your excess charge. Such *charging* and *discharging* does not happen in humid conditions because the water in the air *neutralizes* your excess charge about as fast as you acquire it.

Two of the grand mysteries in physics are (1) *why* does the universe have particles with electric charge (what is it, really?) and (2) *why* does electric charge come in two types (and not, say, one type or three types). We just do not know. Nevertheless, with lots of experiments similar to our two demonstrations scientists discovered that



Particles with the same sign of electrical charge repel each other, and particles with opposite signs attract each other.

In a moment we shall put this rule into quantitative form as Coulomb's law of *electrostatic force* (or *electric force*) between charged particles. The term *electrostatic* is used to emphasize that, relative to each other, the charges are either stationary or moving only very slowly.

Demos. Now let's get back to the demonstrations to understand the motions of the rod as being something other than just magic. When we rub the glass rod with a silk cloth, a small amount of negative charge moves from the rod to the silk (a transfer like that between you and a carpet), leaving the rod with a small amount of excess positive charge. (Which way the negative charge moves is not obvious and requires a lot of experimentation.) We rub the silk over the rod to increase the number of contact points and thus the amount, still tiny, of transferred charge. We hang the rod from the thread so as to *electrically isolate* it from its surroundings (so that the surroundings cannot neutralize the rod by giving it enough negative charge to rebalance its charge). When we rub the second rod with the silk cloth, it too becomes positively charged. So when we bring it near the first rod, the two rods repel each other (Fig. 21-2a).

Next, when we rub the plastic rod with fur, it gains excess negative charge from the fur. (Again, the transfer direction is learned through many experiments.) When we bring the plastic rod (with negative charge) near the hanging glass rod (with positive charge), the rods are attracted to each other (Fig. 21-2b). All this is subtle. You cannot see the charge or its transfer, only the results.

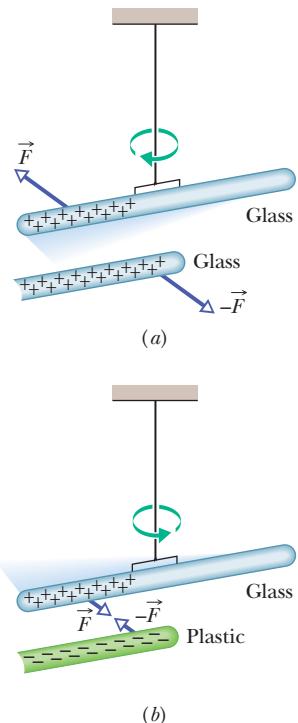


Figure 21-2 (a) Two charged rods of the same sign repel each other. (b) Two charged rods of opposite signs attract each other. Plus signs indicate a positive net charge, and minus signs indicate a negative net charge.

Conductors and Insulators

We can classify materials generally according to the ability of charge to move through them. **Conductors** are materials through which charge can move rather freely; examples include metals (such as copper in common lamp wire), the human body, and tap water. **Nonconductors**—also called **insulators**—are materials through which charge cannot move freely; examples include rubber (such as the insulation on common lamp wire), plastic, glass, and chemically pure water. **Semiconductors** are materials that are intermediate between conductors and insulators; examples include silicon and germanium in computer chips. **Superconductors** are materials that are *perfect* conductors, allowing charge to move without *any* hindrance. In these chapters we discuss only conductors and insulators.

Conducting Path. Here is an example of how conduction can eliminate excess charge on an object. If you rub a copper rod with wool, charge is transferred from the wool to the rod. However, if you are holding the rod while also touching a faucet, you cannot charge the rod in spite of the transfer. The reason is that you, the rod, and the faucet are all conductors connected, via the plumbing, to Earth's surface, which is a huge conductor. Because the excess charges put on the rod by the wool repel one another, they move away from one another by moving first through the rod, then through you, and then through the faucet and plumbing to reach Earth's surface, where they can spread out. The process leaves the rod electrically neutral.

In thus setting up a pathway of conductors between an object and Earth's surface, we are said to *ground* the object, and in neutralizing the object (by eliminating an unbalanced positive or negative charge), we are said to *discharge* the object. If instead of holding the copper rod in your hand, you hold it by an insulating handle, you eliminate the conducting path to Earth, and the rod can then be charged by rubbing (the charge remains on the rod), as long as you do not touch it directly with your hand.

Charged Particles. The properties of conductors and insulators are due to the structure and electrical nature of atoms. Atoms consist of positively charged *protons*, negatively charged *electrons*, and electrically neutral *neutrons*. The protons and neutrons are packed tightly together in a central *nucleus*.

The charge of a single electron and that of a single proton have the same magnitude but are opposite in sign. Hence, an electrically neutral atom contains equal numbers of electrons and protons. Electrons are held near the nucleus because they have the electrical sign opposite that of the protons in the nucleus and thus are attracted to the nucleus. Were this not true, there would be no atoms and thus no you.

When atoms of a conductor like copper come together to form the solid, some of their outermost (and so most loosely held) electrons become free to wander about within the solid, leaving behind positively charged atoms (*positive ions*). We call the mobile electrons *conduction electrons*. There are few (if any) free electrons in a nonconductor.

Induced Charge. The experiment of Fig. 21-3 demonstrates the mobility of charge in a conductor. A negatively charged plastic rod will attract either end of an isolated neutral copper rod. What happens is that many of the conduction electrons in the closer end of the copper rod are repelled by the negative charge on the plastic rod. Some of the conduction electrons move to the far end of the copper rod, leaving the near end depleted in electrons and thus with an unbalanced positive charge. This positive charge is attracted to the negative charge in the plastic rod. Although the copper rod is still neutral, it is said to have an *induced charge*, which means that some of its positive and negative charges have been separated due to the presence of a nearby charge.

Similarly, if a positively charged glass rod is brought near one end of a neutral copper rod, induced charge is again set up in the neutral copper rod but now the near end gains conduction electrons, becomes negatively charged, and is attracted to the glass rod, while the far end is positively charged.

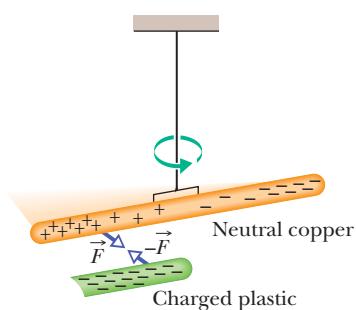


Figure 21-3 A neutral copper rod is electrically isolated from its surroundings by being suspended on a nonconducting thread. Either end of the copper rod will be attracted by a charged rod. Here, conduction electrons in the copper rod are repelled to the far end of that rod by the negative charge on the plastic rod. Then that negative charge attracts the remaining positive charge on the near end of the copper rod, rotating the copper rod to bring that near end closer to the plastic rod.

Note that only conduction electrons, with their negative charges, can move; positive ions are fixed in place. Thus, an object becomes positively charged only through the *removal of negative charges*.

Blue Flashes from a Wintergreen LifeSaver

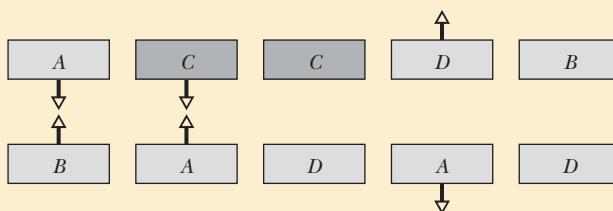
Indirect evidence for the attraction of charges with opposite signs can be seen with a wintergreen LifeSaver (the candy shaped in the form of a marine lifesaver). If you adapt your eyes to darkness for about 15 minutes and then have a friend chomp on a piece of the candy in the darkness, you will see a faint blue flash from your friend's mouth with each chomp. Whenever a chomp breaks a sugar crystal into pieces, each piece will probably end up with a different number of electrons. Suppose a crystal breaks into pieces *A* and *B*, with *A* ending up with more electrons on its surface than *B* (Fig. 21-4). This means that *B* has positive ions (atoms that lost electrons to *A*) on its surface. Because the electrons on *A* are strongly attracted to the positive ions on *B*, some of those electrons jump across the gap between the pieces.

As *A* and *B* move away from each other, air (primarily nitrogen, N_2) flows into the gap, and many of the jumping electrons collide with nitrogen molecules in the air, causing the molecules to emit ultraviolet light. You cannot see this type of light. However, the wintergreen molecules on the surfaces of the candy pieces absorb the ultraviolet light and then emit blue light, which you *can* see—it is the blue light coming from your friend's mouth.



Checkpoint 1

The figure shows five pairs of plates: *A*, *B*, and *D* are charged plastic plates and *C* is an electrically neutral copper plate. The electrostatic forces between the pairs of plates are shown for three of the pairs. For the remaining two pairs, do the plates repel or attract each other?



Coulomb's Law

Now we come to the equation for Coulomb's law, but first a caution. This equation works for only charged particles (and a few other things that can be treated as particles). For extended objects, with charge located in many different places, we need more powerful techniques. So, here we consider just charged particles and not, say, two charged cats.

If two charged particles are brought near each other, they each exert an **electrostatic force** on the other. The direction of the force vectors depends on the signs of the charges. If the particles have the same sign of charge, they repel each other. That means that the force vector on each is directly away from the other particle (Figs. 21-5*a* and *b*). If we release the particles, they accelerate away from each other. If, instead, the particles have opposite signs of charge, they attract each other. That means that the force vector on each is directly toward the other particle (Fig. 21-5*c*). If we release the particles, they accelerate toward each other.

The equation for the electrostatic forces acting on the particles is called **Coulomb's law** after Charles-Augustin de Coulomb, whose experiments in 1785 led him to it. Let's write the equation in vector form and in terms of the particles shown in Fig. 21-6, where particle 1 has charge q_1 and particle 2 has charge q_2 . (These symbols can represent either positive or negative charge.) Let's also focus on particle 1 and write the force acting on it in terms of a unit vector \hat{r} that points along a radial

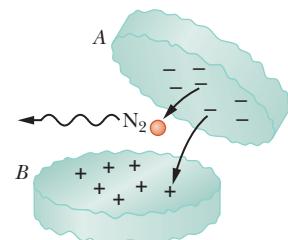


Figure 21-4 Two pieces of a wintergreen LifeSaver candy as they fall away from each other. Electrons jumping from the negative surface of piece *A* to the positive surface of piece *B* collide with nitrogen (N_2) molecules in the air.

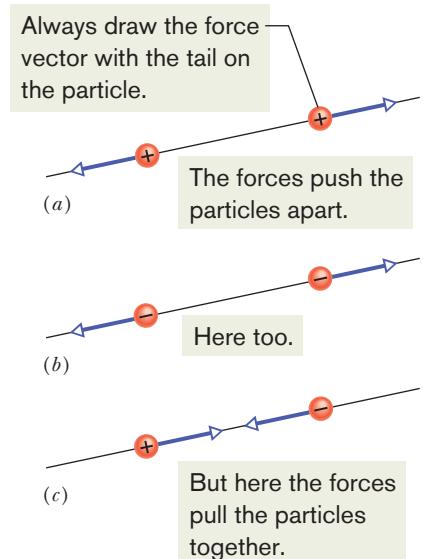


Figure 21-5 Two charged particles repel each other if they have the same sign of charge, either (a) both positive or (b) both negative. (c) They attract each other if they have opposite signs of charge.

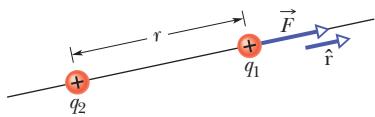


Figure 21-6 The electrostatic force on particle 1 can be described in terms of a unit vector \hat{r} along an axis through the two particles, radially away from particle 2.

axis extending through the two particles, radially away from particle 2. (As with other unit vectors, \hat{r} has a magnitude of exactly 1 and no unit; its purpose is to point, like a direction arrow on a street sign.) With these decisions, we write the electrostatic force as

$$\vec{F} = k \frac{q_1 q_2}{r^2} \hat{r} \quad (\text{Coulomb's law}), \quad (21-1)$$

where r is the separation between the particles and k is a positive constant called the *electrostatic constant* or the *Coulomb constant*. (We'll discuss k below.)

Let's first check the direction of the force on particle 1 as given by Eq. 21-1. If q_1 and q_2 have the same sign, then the product $q_1 q_2$ gives us a positive result. So, Eq. 21-1 tells us that the force on particle 1 is in the direction of \hat{r} . That checks, because particle 1 is being repelled from particle 2. Next, if q_1 and q_2 have opposite signs, the product $q_1 q_2$ gives us a negative result. So, now Eq. 21-1 tells us that the force on particle 1 is in the direction opposite \hat{r} . That checks because particle 1 is being attracted toward particle 2.

An Aside. Here is something that is very curious. The form of Eq. 21-1 is the same as that of Newton's equation (Eq. 13-3) for the gravitational force between two particles with masses m_1 and m_2 and separation r :

$$\vec{F} = G \frac{m_1 m_2}{r^2} \hat{r} \quad (\text{Newton's law}), \quad (21-2)$$

where G is the gravitational constant. Although the two types of forces are wildly different, both equations describe inverse square laws (the $1/r^2$ dependences) that involve a product of a property of the interacting particles—the charge in one case and the mass in the other. However, the laws differ in that gravitational forces are always attractive but electrostatic forces may be either attractive or repulsive, depending on the signs of the charges. This difference arises from the fact that there is only one type of mass but two types of charge.

Unit. The SI unit of charge is the **coulomb**. For practical reasons having to do with the accuracy of measurements, the coulomb unit is derived from the SI unit *ampere* for electric current i . We shall discuss current in detail in Chapter 26, but here let's just note that current i is the rate dq/dt at which charge moves past a point or through a region:

$$i = \frac{dq}{dt} \quad (\text{electric current}). \quad (21-3)$$

Rearranging Eq. 21-3 and replacing the symbols with their units (coulombs C, amperes A, and seconds s) we see that

$$1 \text{ C} = (1 \text{ A})(1 \text{ s}).$$

Force Magnitude. For historical reasons (and because doing so simplifies many other formulas), the electrostatic constant k in Eq. 21-1 is often written as $1/4\pi\epsilon_0$. Then the magnitude of the electrostatic force in Coulomb's law becomes

$$F = \frac{1}{4\pi\epsilon_0} \frac{|q_1||q_2|}{r^2} \quad (\text{Coulomb's law}). \quad (21-4)$$

The constants in Eqs. 21-1 and 21-4 have the value

$$k = \frac{1}{4\pi\epsilon_0} = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2. \quad (21-5)$$

The quantity ϵ_0 , called the **permittivity constant**, sometimes appears separately in equations and is

$$\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2. \quad (21-6)$$

Working a Problem. Note that the charge magnitudes appear in Eq. 21-4, which gives us the force magnitude. So, in working problems in this chapter, we use Eq. 21-4 to find the magnitude of a force on a chosen particle due to a second

particle and we separately determine the direction of the force by considering the charge signs of the two particles.

Multiple Forces. As with all forces in this book, the electrostatic force obeys the principle of superposition. Suppose we have n charged particles near a chosen particle called particle 1; then the net force on particle 1 is given by the vector sum

$$\vec{F}_{1,\text{net}} = \vec{F}_{12} + \vec{F}_{13} + \vec{F}_{14} + \vec{F}_{15} + \cdots + \vec{F}_{1n}, \quad (21-7)$$

in which, for example, \vec{F}_{14} is the force on particle 1 due to the presence of particle 4.

This equation is the key to many of the homework problems, so let's state it in words. If you want to know the net force acting on a chosen charged particle that is surrounded by other charged particles, first clearly identify that chosen particle and then find the force on it due to each of the other particles. Draw those force vectors in a free-body diagram of the chosen particle, with the tails anchored on the particle. (That may sound trivial, but failing to do so easily leads to errors.) Then add all those forces *as vectors* according to the rules of Chapter 3, not as scalars. (You cannot just willy-nilly add up their magnitudes.) The result is the net force (or resultant force) acting on the particle.

Although the vector nature of the forces makes the homework problems harder than if we simply had scalars, be thankful that Eq. 21-7 works. If two force vectors did not simply add but for some reason amplified each other, the world would be very difficult to understand and manage.

Shell Theories. Analogous to the shell theories for the gravitational force (Module 13-1), we have two shell theories for the electrostatic force:



Shell theory 1. A charged particle outside a shell with charge uniformly distributed on its surface is attracted or repelled as if the shell's charge were concentrated as a particle at its center.



Shell theory 2. A charged particle inside a shell with charge uniformly distributed on its surface has no net force acting on it due to the shell.

(In the first theory, we assume that the charge on the shell is much greater than the particle's charge. Thus the presence of the particle has negligible effect on the distribution of charge on the shell.)

Spherical Conductors

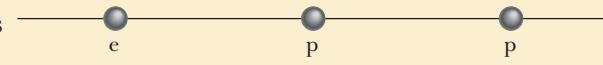
If excess charge is placed on a spherical shell that is made of conducting material, the excess charge spreads uniformly over the (external) surface. For example, if we place excess electrons on a spherical metal shell, those electrons repel one another and tend to move apart, spreading over the available surface until they are uniformly distributed. That arrangement maximizes the distances between all pairs of the excess electrons. According to the first shell theorem, the shell then will attract or repel an external charge as if all the excess charge on the shell were concentrated at its center.

If we remove negative charge from a spherical metal shell, the resulting positive charge of the shell is also spread uniformly over the surface of the shell. For example, if we remove n electrons, there are then n sites of positive charge (sites missing an electron) that are spread uniformly over the shell. According to the first shell theorem, the shell will again attract or repel an external charge as if all the shell's excess charge were concentrated at its center.



Checkpoint 2

The figure shows two protons (symbol p) and one electron (symbol e) on an axis. On the central proton, what is the direction of (a) the force due to the electron, (b) the force due to the other proton, and (c) the net force?



Sample Problem 21.01 Finding the net force due to two other particles

This sample problem actually contains three examples, to build from basic stuff to harder stuff. In each we have the same charged particle 1. First there is a single force acting on it (easy stuff). Then there are two forces, but they are just in opposite directions (not too bad). Then there are again two forces but they are in very different directions (ah, now we have to get serious about the fact that they are vectors). The key to all three examples is to draw the forces correctly *before* you reach for a calculator, otherwise you may be calculating nonsense on the calculator. (Figure 21-7 is available in *WileyPLUS* as an animation with voiceover.)

(a) Figure 21-7a shows two positively charged particles fixed in place on an x axis. The charges are $q_1 = 1.60 \times 10^{-19} \text{ C}$ and $q_2 = 3.20 \times 10^{-19} \text{ C}$, and the particle separation is $R = 0.0200 \text{ m}$. What are the magnitude and direction of the electrostatic force \vec{F}_{12} on particle 1 from particle 2?

KEY IDEAS

Because both particles are positively charged, particle 1 is repelled by particle 2, with a force magnitude given by Eq. 21-4. Thus, the direction of force \vec{F}_{12} on particle 1 is *away from* particle 2, in the negative direction of the x axis, as indicated in the free-body diagram of Fig. 21-7b.

Two particles: Using Eq. 21-4 with separation R substituted for r , we can write the magnitude F_{12} of this force as

$$\begin{aligned} F_{12} &= \frac{1}{4\pi\epsilon_0} \frac{|q_1||q_2|}{R^2} \\ &= (8.99 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2) \\ &\quad \times \frac{(1.60 \times 10^{-19} \text{ C})(3.20 \times 10^{-19} \text{ C})}{(0.0200 \text{ m})^2} \\ &= 1.15 \times 10^{-24} \text{ N}. \end{aligned}$$

Thus, force \vec{F}_{12} has the following magnitude and direction (relative to the positive direction of the x axis):

$$1.15 \times 10^{-24} \text{ N} \text{ and } 180^\circ. \quad (\text{Answer})$$

We can also write \vec{F}_{12} in unit-vector notation as

$$\vec{F}_{12} = -(1.15 \times 10^{-24} \text{ N})\hat{i}. \quad (\text{Answer})$$

(b) Figure 21-7c is identical to Fig. 21-7a except that particle 3 now lies on the x axis between particles 1 and 2. Particle 3 has charge $q_3 = -3.20 \times 10^{-19} \text{ C}$ and is at a distance $\frac{3}{4}R$ from particle 1. What is the net electrostatic force $\vec{F}_{1,\text{net}}$ on particle 1 due to particles 2 and 3?

KEY IDEA

The presence of particle 3 does not alter the electrostatic force on particle 1 from particle 2. Thus, force \vec{F}_{12} still acts on particle 1. Similarly, the force \vec{F}_{13} that acts on particle 1 due to particle 3 is not affected by the presence of particle 2. Because particles 1

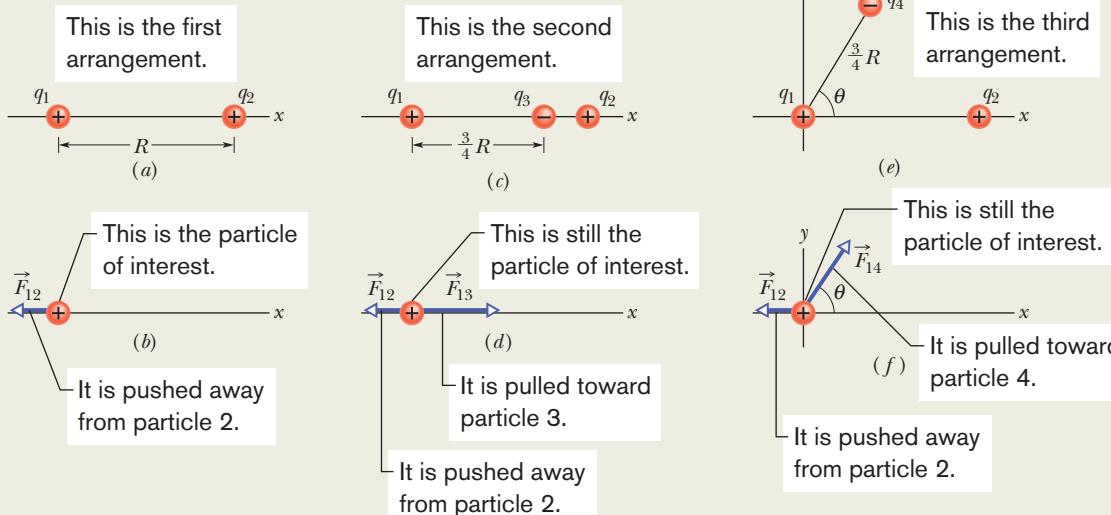


Figure 21-7 (a) Two charged particles of charges q_1 and q_2 are fixed in place on an x axis. (b) The free-body diagram for particle 1, showing the electrostatic force on it from particle 2. (c) Particle 3 included. (d) Free-body diagram for particle 1. (e) Particle 4 included. (f) Free-body diagram for particle 1.

and 3 have charge of opposite signs, particle 1 is attracted to particle 3. Thus, force \vec{F}_{13} is directed *toward* particle 3, as indicated in the free-body diagram of Fig. 21-7d.

Three particles: To find the magnitude of \vec{F}_{13} , we can rewrite Eq. 21-4 as

$$\begin{aligned} F_{13} &= \frac{1}{4\pi\epsilon_0} \frac{|q_1||q_3|}{(\frac{3}{4}R)^2} \\ &= (8.99 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2) \\ &\quad \times \frac{(1.60 \times 10^{-19} \text{ C})(3.20 \times 10^{-19} \text{ C})}{(\frac{3}{4})^2(0.0200 \text{ m})^2} \\ &= 2.05 \times 10^{-24} \text{ N}. \end{aligned}$$

We can also write \vec{F}_{13} in unit-vector notation:

$$\vec{F}_{13} = (2.05 \times 10^{-24} \text{ N})\hat{i}.$$

The net force $\vec{F}_{1,\text{net}}$ on particle 1 is the vector sum of \vec{F}_{12} and \vec{F}_{13} ; that is, from Eq. 21-7, we can write the net force $\vec{F}_{1,\text{net}}$ on particle 1 in unit-vector notation as

$$\begin{aligned} \vec{F}_{1,\text{net}} &= \vec{F}_{12} + \vec{F}_{13} \\ &= -(1.15 \times 10^{-24} \text{ N})\hat{i} + (2.05 \times 10^{-24} \text{ N})\hat{i} \\ &= (9.00 \times 10^{-25} \text{ N})\hat{i}. \end{aligned} \quad (\text{Answer})$$

Thus, $\vec{F}_{1,\text{net}}$ has the following magnitude and direction (relative to the positive direction of the x axis):

$$9.00 \times 10^{-25} \text{ N} \text{ and } 0^\circ. \quad (\text{Answer})$$

(c) Figure 21-7e is identical to Fig. 21-7a except that particle 4 is now included. It has charge $q_4 = -3.20 \times 10^{-19} \text{ C}$, is at a distance $\frac{3}{4}R$ from particle 1, and lies on a line that makes an angle $\theta = 60^\circ$ with the x axis. What is the net electrostatic force $\vec{F}_{1,\text{net}}$ on particle 1 due to particles 2 and 4?

KEY IDEA

The net force $\vec{F}_{1,\text{net}}$ is the vector sum of \vec{F}_{12} and a new force \vec{F}_{14} acting on particle 1 due to particle 4. Because particles 1 and 4 have charge of opposite signs, particle 1 is attracted to particle 4. Thus, force \vec{F}_{14} on particle 1 is directed *toward* particle 4, at angle $\theta = 60^\circ$, as indicated in the free-body diagram of Fig. 21-7f.

Four particles: We can rewrite Eq. 21-4 as

$$\begin{aligned} F_{14} &= \frac{1}{4\pi\epsilon_0} \frac{|q_1||q_4|}{(\frac{3}{4}R)^2} \\ &= (8.99 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2) \\ &\quad \times \frac{(1.60 \times 10^{-19} \text{ C})(3.20 \times 10^{-19} \text{ C})}{(\frac{3}{4})^2(0.0200 \text{ m})^2} \\ &= 2.05 \times 10^{-24} \text{ N}. \end{aligned}$$

Then from Eq. 21-7, we can write the net force $\vec{F}_{1,\text{net}}$ on particle 1 as

$$\vec{F}_{1,\text{net}} = \vec{F}_{12} + \vec{F}_{14}.$$

Because the forces \vec{F}_{12} and \vec{F}_{14} are not directed along the same axis, we *cannot* sum simply by combining their magnitudes. Instead, we must add them as vectors, using one of the following methods.

Method 1. *Summing directly on a vector-capable calculator.* For \vec{F}_{12} , we enter the magnitude 1.15×10^{-24} and the angle 180° . For \vec{F}_{14} , we enter the magnitude 2.05×10^{-24} and the angle 60° . Then we add the vectors.

Method 2. *Summing in unit-vector notation.* First we rewrite \vec{F}_{14} as

$$\vec{F}_{14} = (F_{14} \cos \theta)\hat{i} + (F_{14} \sin \theta)\hat{j}.$$

Substituting $2.05 \times 10^{-24} \text{ N}$ for F_{14} and 60° for θ , this becomes

$$\vec{F}_{14} = (1.025 \times 10^{-24} \text{ N})\hat{i} + (1.775 \times 10^{-24} \text{ N})\hat{j}.$$

Then we sum:

$$\begin{aligned} \vec{F}_{1,\text{net}} &= \vec{F}_{12} + \vec{F}_{14} \\ &= -(1.15 \times 10^{-24} \text{ N})\hat{i} \\ &\quad + (1.025 \times 10^{-24} \text{ N})\hat{i} + (1.775 \times 10^{-24} \text{ N})\hat{j} \\ &\approx (-1.25 \times 10^{-25} \text{ N})\hat{i} + (1.78 \times 10^{-24} \text{ N})\hat{j}. \end{aligned} \quad (\text{Answer})$$

Method 3. *Summing components axis by axis.* The sum of the x components gives us

$$\begin{aligned} F_{1,\text{net},x} &= F_{12,x} + F_{14,x} = F_{12} + F_{14} \cos 60^\circ \\ &= -1.15 \times 10^{-24} \text{ N} + (2.05 \times 10^{-24} \text{ N})(\cos 60^\circ) \\ &= -1.25 \times 10^{-25} \text{ N}. \end{aligned}$$

The sum of the y components gives us

$$\begin{aligned} F_{1,\text{net},y} &= F_{12,y} + F_{14,y} = 0 + F_{14} \sin 60^\circ \\ &= (2.05 \times 10^{-24} \text{ N})(\sin 60^\circ) \\ &= 1.78 \times 10^{-24} \text{ N}. \end{aligned}$$

The net force $\vec{F}_{1,\text{net}}$ has the magnitude

$$F_{1,\text{net}} = \sqrt{F_{1,\text{net},x}^2 + F_{1,\text{net},y}^2} = 1.78 \times 10^{-24} \text{ N}. \quad (\text{Answer})$$

To find the direction of $\vec{F}_{1,\text{net}}$, we take

$$\theta = \tan^{-1} \frac{F_{1,\text{net},y}}{F_{1,\text{net},x}} = -86.0^\circ.$$

However, this is an unreasonable result because $\vec{F}_{1,\text{net}}$ must have a direction between the directions of \vec{F}_{12} and \vec{F}_{14} . To correct θ , we add 180° , obtaining

$$-86.0^\circ + 180^\circ = 94.0^\circ. \quad (\text{Answer})$$



Additional examples, video, and practice available at WileyPLUS

**Checkpoint 3**

The figure here shows three arrangements of an electron e and two protons p . (a) Rank the arrangements according to the magnitude of the net electrostatic force on the electron due to the protons, largest first. (b) In situation c , is the angle between the net force on the electron and the line labeled d less than or more than 45° ?

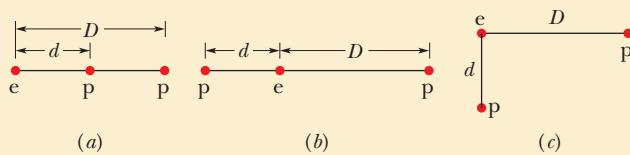
**Sample Problem 21.02 Equilibrium of two forces on a particle**

Figure 21-8a shows two particles fixed in place: a particle of charge $q_1 = +8q$ at the origin and a particle of charge $q_2 = -2q$ at $x = L$. At what point (other than infinitely far away) can a proton be placed so that it is in *equilibrium* (the net force on it is zero)? Is that equilibrium *stable* or *unstable*? (That is, if the proton is displaced, do the forces drive it back to the point of equilibrium or drive it farther away?)

KEY IDEA

If \vec{F}_1 is the force on the proton due to charge q_1 and \vec{F}_2 is the force on the proton due to charge q_2 , then the point we seek is where $\vec{F}_1 + \vec{F}_2 = 0$. Thus,

$$\vec{F}_1 = -\vec{F}_2. \quad (21-8)$$

This tells us that at the point we seek, the forces acting on the proton due to the other two particles must be of equal magnitudes,

$$F_1 = F_2, \quad (21-9)$$

and that the forces must have opposite directions.

Reasoning: Because a proton has a positive charge, the proton and the particle of charge q_1 are of the same sign, and force \vec{F}_1 on the proton must point away from q_1 . Also, the proton and the particle of charge q_2 are of opposite signs, so force \vec{F}_2 on the proton must point toward q_2 . “Away from q_1 ” and “toward q_2 ” can be in opposite directions only if the proton is located on the x axis.

If the proton is on the x axis at any point between q_1 and q_2 , such as point P in Fig. 21-8b, then \vec{F}_1 and \vec{F}_2 are in the same direction and not in opposite directions as required. If the proton is at any point on the x axis to the left of q_1 , such as point S in Fig. 21-8c, then \vec{F}_1 and \vec{F}_2 are in opposite directions. However, Eq. 21-4 tells us that \vec{F}_1 and \vec{F}_2 cannot have equal magnitudes there: F_1 must be greater than F_2 , because F_1 is produced by a closer charge (with lesser r) of greater magnitude ($8q$ versus $2q$).

Finally, if the proton is at any point on the x axis to the right of q_2 , such as point R in Fig. 21-8d, then \vec{F}_1 and \vec{F}_2 are again in opposite directions. However, because now the charge of greater magnitude (q_1) is *farther* away from the proton than the charge of lesser magnitude, there is a point at which F_1 is equal to F_2 . Let x be the coordinate of this point, and let q_p be the charge of the proton.

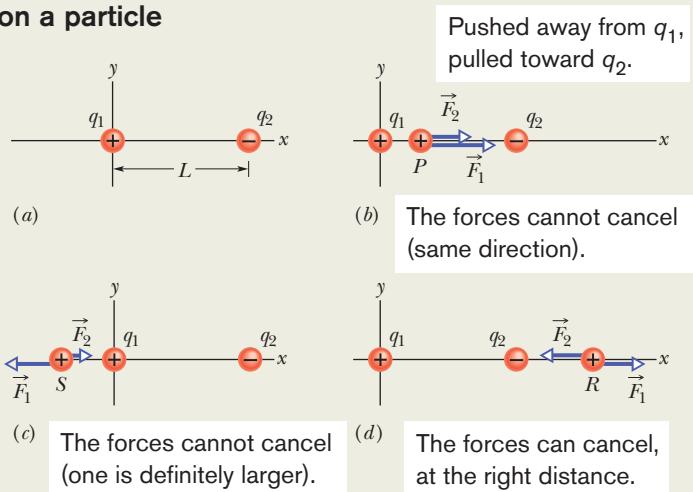


Figure 21-8 (a) Two particles of charges q_1 and q_2 are fixed in place on an x axis, with separation L . (b)–(d) Three possible locations P , S , and R for a proton. At each location, \vec{F}_1 is the force on the proton from particle 1 and \vec{F}_2 is the force on the proton from particle 2.

Calculations: With Eq. 21-4, we can now rewrite Eq. 21-9:

$$\frac{1}{4\pi\epsilon_0} \frac{8qq_p}{x^2} = \frac{1}{4\pi\epsilon_0} \frac{2qq_p}{(x-L)^2}. \quad (21-10)$$

(Note that only the charge magnitudes appear in Eq. 21-10. We already decided about the directions of the forces in drawing Fig. 21-8d and do not want to include any positive or negative signs here.) Rearranging Eq. 21-10 gives us

$$\left(\frac{x-L}{x}\right)^2 = \frac{1}{4}.$$

After taking the square roots of both sides, we find

$$\frac{x-L}{x} = \frac{1}{2}$$

and

$$x = 2L. \quad (\text{Answer})$$

The equilibrium at $x = 2L$ is unstable; that is, if the proton is displaced leftward from point R , then F_1 and F_2 both increase but F_2 increases more (because q_2 is closer than q_1), and a net force will drive the proton farther leftward. If the proton is displaced rightward, both F_1 and F_2 decrease but F_2 decreases more, and a net force will then drive the proton farther rightward. In a stable equilibrium, if the proton is displaced slightly, it returns to the equilibrium position.



Additional examples, video, and practice available at WileyPLUS

Sample Problem 21.03 Charge sharing by two identical conducting spheres

In Fig. 21-9a, two identical, electrically isolated conducting spheres *A* and *B* are separated by a (center-to-center) distance *a* that is large compared to the spheres. Sphere *A* has a positive charge of $+Q$, and sphere *B* is electrically neutral. Initially, there is no electrostatic force between the spheres. (The large separation means there is no induced charge.)

(a) Suppose the spheres are connected for a moment by a conducting wire. The wire is thin enough so that any net charge on it is negligible. What is the electrostatic force between the spheres after the wire is removed?

KEY IDEAS

(1) Because the spheres are identical, connecting them means that they end up with identical charges (same sign and same amount). (2) The initial sum of the charges (including the signs of the charges) must equal the final sum of the charges.

Reasoning: When the spheres are wired together, the (negative) conduction electrons on *B*, which repel one another, have a way to move away from one another (along the wire to positively charged *A*, which attracts them—Fig. 21-9b). As *B* loses negative charge, it becomes positively charged, and as *A* gains negative charge, it becomes *less* positively charged. The transfer of charge stops when the charge on *B* has increased to $+Q/2$ and the charge on *A* has decreased to $+Q/2$, which occurs when $-Q/2$ has shifted from *B* to *A*.

After the wire has been removed (Fig. 21-9c), we can assume that the charge on either sphere does not disturb the uniformity of the charge distribution on the other sphere, because the spheres are small relative to their separation. Thus, we can apply the first shell theorem to each sphere. By Eq. 21-4 with $q_1 = q_2 = Q/2$ and $r = a$,

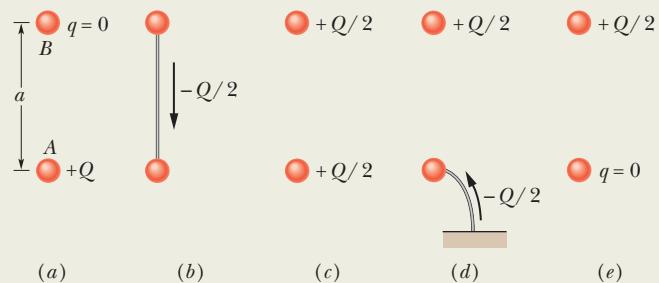


Figure 21-9 Two small conducting spheres *A* and *B*. (a) To start, sphere *A* is charged positively. (b) Negative charge is transferred from *B* to *A* through a connecting wire. (c) Both spheres are then charged positively. (d) Negative charge is transferred through a grounding wire to sphere *A*. (e) Sphere *A* is then neutral.

$$F = \frac{1}{4\pi\epsilon_0} \frac{(Q/2)(Q/2)}{a^2} = \frac{1}{16\pi\epsilon_0} \left(\frac{Q}{a}\right)^2. \quad (\text{Answer})$$

The spheres, now positively charged, repel each other.

(b) Next, suppose sphere *A* is grounded momentarily, and then the ground connection is removed. What now is the electrostatic force between the spheres?

Reasoning: When we provide a conducting path between a charged object and the ground (which is a huge conductor), we neutralize the object. Were sphere *A* negatively charged, the mutual repulsion between the excess electrons would cause them to move from the sphere to the ground. However, because sphere *A* is positively charged, electrons with a total charge of $-Q/2$ move *from* the ground up onto the sphere (Fig. 21-9d), leaving the sphere with a charge of 0 (Fig. 21-9e). Thus, the electrostatic force is again zero.



Additional examples, video, and practice available at WileyPLUS

21-2 CHARGE IS QUANTIZED

Learning Objectives

After reading this module, you should be able to . . .

21.19 Identify the elementary charge.

Key Ideas

- Electric charge is quantized (restricted to certain values).
- The charge of a particle can be written as ne , where n is a positive or negative integer and e is the elementary charge,

21.20 Identify that the charge of a particle or object must be a positive or negative integer times the elementary charge.

which is the magnitude of the charge of the electron and proton ($\approx 1.602 \times 10^{-19}$ C).

Charge Is Quantized

In Benjamin Franklin's day, electric charge was thought to be a continuous fluid—an idea that was useful for many purposes. However, we now know that

fluids themselves, such as air and water, are not continuous but are made up of atoms and molecules; matter is discrete. Experiment shows that “electrical fluid” is also not continuous but is made up of multiples of a certain elementary charge. Any positive or negative charge q that can be detected can be written as

$$q = ne, \quad n = \pm 1, \pm 2, \pm 3, \dots, \quad (21-11)$$

in which e , the **elementary charge**, has the approximate value

$$e = 1.602 \times 10^{-19} \text{ C.} \quad (21-12)$$

The elementary charge e is one of the important constants of nature. The electron and proton both have a charge of magnitude e (Table 21-1). (Quarks, the constituent particles of protons and neutrons, have charges of $\pm e/3$ or $\pm 2e/3$, but they apparently cannot be detected individually. For this and for historical reasons, we do not take their charges to be the elementary charge.)

You often see phrases—such as “the charge on a sphere,” “the amount of charge transferred,” and “the charge carried by the electron”—that suggest that charge is a substance. (Indeed, such statements have already appeared in this chapter.) You should, however, keep in mind what is intended: *Particles* are the substance and charge happens to be one of their properties, just as mass is.

When a physical quantity such as charge can have only discrete values rather than any value, we say that the quantity is **quantized**. It is possible, for example, to find a particle that has no charge at all or a charge of $+10e$ or $-6e$, but not a particle with a charge of, say, 3.57e.

The quantum of charge is small. In an ordinary 100 W lightbulb, for example, about 10^{19} elementary charges enter the bulb every second and just as many leave. However, the graininess of electricity does not show up in such large-scale phenomena (the bulb does not flicker with each electron).



Checkpoint 4

Initially, sphere A has a charge of $-50e$ and sphere B has a charge of $+20e$. The spheres are made of conducting material and are identical in size. If the spheres then touch, what is the resulting charge on sphere A?

Sample Problem 21.04 Mutual electric repulsion in a nucleus

The nucleus in an iron atom has a radius of about $4.0 \times 10^{-15} \text{ m}$ and contains 26 protons.

(a) What is the magnitude of the repulsive electrostatic force between two of the protons that are separated by $4.0 \times 10^{-15} \text{ m}$?

KEY IDEA

The protons can be treated as charged particles, so the magnitude of the electrostatic force on one from the other is given by Coulomb's law.

Calculation: Table 21-1 tells us that the charge of a proton is $+e$. Thus, Eq. 21-4 gives us

$$\begin{aligned} F &= \frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} \\ &= \frac{(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2)(1.602 \times 10^{-19} \text{ C})^2}{(4.0 \times 10^{-15} \text{ m})^2} \\ &= 14 \text{ N.} \end{aligned} \quad (\text{Answer})$$

No explosion: This is a small force to be acting on a macroscopic object like a cantaloupe, but an enormous force to be acting on a proton. Such forces should explode the nucleus of any element but hydrogen (which has only one proton in its nucleus). However, they don't, not even in nuclei with a great many protons. Therefore, there must be some enormous attractive force to counter this enormous repulsive electrostatic force.

(b) What is the magnitude of the gravitational force between those same two protons?

KEY IDEA

Because the protons are particles, the magnitude of the gravitational force on one from the other is given by Newton's equation for the gravitational force (Eq. 21-2).

Calculation: With $m_p (= 1.67 \times 10^{-27} \text{ kg})$ representing the

mass of a proton, Eq. 21-2 gives us

$$F = G \frac{m_p^2}{r^2}$$

$$= \frac{(6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2)(1.67 \times 10^{-27} \text{ kg})^2}{(4.0 \times 10^{-15} \text{ m})^2}$$

$$= 1.2 \times 10^{-35} \text{ N.} \quad (\text{Answer})$$

Weak versus strong: This result tells us that the (attractive) gravitational force is far too weak to counter the repulsive electrostatic forces between protons in a nucleus. Instead, the protons are bound together by an enormous force called

(aptly) the *strong nuclear force*—a force that acts between protons (and neutrons) when they are close together, as in a nucleus.

Although the gravitational force is many times weaker than the electrostatic force, it is more important in large-scale situations because it is always attractive. This means that it can collect many small bodies into huge bodies with huge masses, such as planets and stars, that then exert large gravitational forces. The electrostatic force, on the other hand, is repulsive for charges of the same sign, so it is unable to collect either positive charge or negative charge into large concentrations that would then exert large electrostatic forces.



Additional examples, video, and practice available at WileyPLUS

21-3 CHARGE IS CONSERVED

Learning Objectives

After reading this module, you should be able to . . .

21.21 Identify that in any isolated physical process, the net charge cannot change (the net charge is always conserved).

Key Ideas

- The net electric charge of any isolated system is always conserved.
- If two charged particles undergo an annihilation process,

21.22 Identify an annihilation process of particles and a pair production of particles.

21.23 Identify mass number and atomic number in terms of the number of protons, neutrons, and electrons.

they have opposite signs of charge.

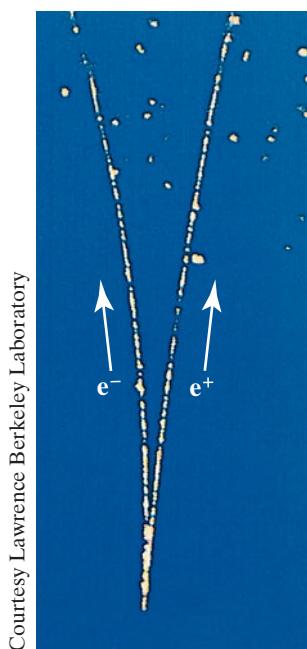
- If two charged particles appear as a result of a pair production process, they have opposite signs of charge.

Charge Is Conserved

If you rub a glass rod with silk, a positive charge appears on the rod. Measurement shows that a negative charge of equal magnitude appears on the silk. This suggests that rubbing does not create charge but only transfers it from one body to another, upsetting the electrical neutrality of each body during the process. This hypothesis of **conservation of charge**, first put forward by Benjamin Franklin, has stood up under close examination, both for large-scale charged bodies and for atoms, nuclei, and elementary particles. No exceptions have ever been found. Thus, we add electric charge to our list of quantities—including energy and both linear momentum and angular momentum—that obey a conservation law.

Important examples of the conservation of charge occur in the *radioactive decay* of nuclei, in which a nucleus transforms into (becomes) a different type of nucleus. For example, a uranium-238 nucleus (^{238}U) transforms into a thorium-234 nucleus (^{234}Th) by emitting an *alpha particle*. Because that particle has the same makeup as a helium-4 nucleus, it has the symbol ^4He . The number used in the name of a nucleus and as a superscript in the symbol for the nucleus is called the *mass number* and is the total number of the protons and neutrons in the nucleus. For example, the total number in ^{238}U is 238. The number of protons in a nucleus is the *atomic number Z*, which is listed for all the elements in Appendix F. From that list we find that in the decay





Courtesy Lawrence Berkeley Laboratory

Figure 21-10 A photograph of trails of bubbles left in a bubble chamber by an electron and a positron. The pair of particles was produced by a gamma ray that entered the chamber directly from the bottom. Being electrically neutral, the gamma ray did not generate a telltale trail of bubbles along its path, as the electron and positron did.

the *parent* nucleus ^{238}U contains 92 protons (a charge of $+92e$), the *daughter* nucleus ^{234}Th contains 90 protons (a charge of $+90e$), and the emitted alpha particle ^4He contains 2 protons (a charge of $+2e$). We see that the total charge is $+92e$ before and after the decay; thus, charge is conserved. (The total number of protons and neutrons is also conserved: 238 before the decay and $234 + 4 = 238$ after the decay.)

Another example of charge conservation occurs when an electron e^- (charge $-e$) and its antiparticle, the *positron* e^+ (charge $+e$), undergo an *annihilation process*, transforming into two *gamma rays* (high-energy light):

$$e^- + e^+ \rightarrow \gamma + \gamma \quad (\text{annihilation}). \quad (21-14)$$

In applying the conservation-of-charge principle, we must add the charges algebraically, with due regard for their signs. In the annihilation process of Eq. 21-14 then, the net charge of the system is zero both before and after the event. Charge is conserved.

In *pair production*, the converse of annihilation, charge is also conserved. In this process a gamma ray transforms into an electron and a positron:

$$\gamma \rightarrow e^- + e^+ \quad (\text{pair production}). \quad (21-15)$$

Figure 21-10 shows such a pair-production event that occurred in a bubble chamber. (This is a device in which a liquid is suddenly made hotter than its boiling point. If a charged particle passes through it, tiny vapor bubbles form along the particle's trail.) A gamma ray entered the chamber from the bottom and at one point transformed into an electron and a positron. Because those new particles were charged and moving, each left a trail of bubbles. (The trails were curved because a magnetic field had been set up in the chamber.) The gamma ray, being electrically neutral, left no trail. Still, you can tell exactly where it underwent pair production—at the tip of the curved V, which is where the trails of the electron and positron begin.

Review & Summary

Electric Charge The strength of a particle's electrical interaction with objects around it depends on its **electric charge** (usually represented as q), which can be either positive or negative. Particles with the same sign of charge repel each other, and particles with opposite signs of charge attract each other. An object with equal amounts of the two kinds of charge is electrically neutral, whereas one with an imbalance is electrically charged and has an excess charge.

Conductors are materials in which a significant number of electrons are free to move. The charged particles in **nonconductors (insulators)** are not free to move.

Electric current i is the rate dq/dt at which charge passes a point:

$$i = \frac{dq}{dt} \quad (\text{electric current}). \quad (21-3)$$

Coulomb's Law Coulomb's law describes the electrostatic force (or electric force) between two charged particles. If the particles have charges q_1 and q_2 , are separated by distance r , and are at rest (or moving only slowly) relative to each other, then the magnitude of the force acting on each due to the other is given by

$$F = \frac{1}{4\pi\epsilon_0} \frac{|q_1| |q_2|}{r^2} \quad (\text{Coulomb's law}), \quad (21-4)$$

where $\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$ is the **permittivity constant**. The ratio $1/4\pi\epsilon_0$ is often replaced with the **electrostatic constant** (or **Coulomb constant**) $k = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.

The electrostatic force vector acting on a charged particle due to a second charged particle is either directly toward the second particle (opposite signs of charge) or directly away from it (same sign of charge). As with other types of forces, if multiple electrostatic forces act on a particle, the net force is the vector sum (not scalar sum) of the individual forces.

The two shell theories for electrostatics are

Shell theorem 1: A charged particle outside a shell with charge uniformly distributed on its surface is attracted or repelled as if the shell's charge were concentrated as a particle at its center.

Shell theorem 2: A charged particle inside a shell with charge uniformly distributed on its surface has no net force acting on it due to the shell.

Charge on a conducting spherical shell spreads uniformly over the (external) surface.

The Elementary Charge Electric charge is quantized (restricted to certain values). The charge of a particle can be written as ne , where n is a positive or negative integer and e is the elementary charge, which is the magnitude of the charge of the electron and proton ($\approx 1.602 \times 10^{-19} \text{ C}$).

Conservation of Charge The net electric charge of any isolated system is always conserved.

Questions

- 1** Figure 21-11 shows four situations in which five charged particles are evenly spaced along an axis. The charge values are indicated except for the central particle, which has the same charge in all four situations. Rank the situations according to the magnitude of the net electrostatic force on the central particle, greatest first.

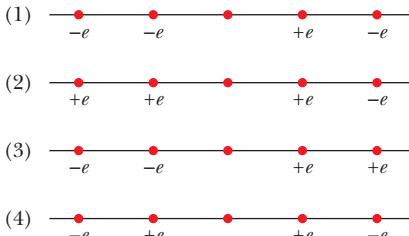


Figure 21-11 Question 1.

- 2** Figure 21-12 shows three pairs of identical spheres that are to be touched together and then separated. The initial charges on them are indicated. Rank the pairs according to (a) the magnitude of the charge transferred during touching and (b) the charge left on the positively charged sphere, greatest first.

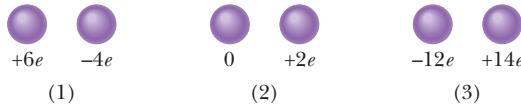


Figure 21-12 Question 2.

- 3** Figure 21-13 shows four situations in which charged particles are fixed in place on an axis. In which situations is there a point to the left of the particles where an electron will be in equilibrium?

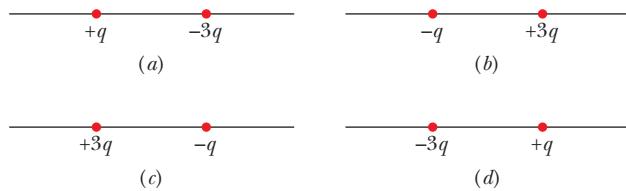


Figure 21-13 Question 3.

- 4** Figure 21-14 shows two charged particles on an axis. The charges are free to move. However, a third charged particle can be placed at a certain point such that all three particles are in equilibrium. (a) Is that point to the left of the first two particles, to their right, or between them? (b) Should the third particle be positively or negatively charged? (c) Is the equilibrium stable or unstable?



Figure 21-14 Question 4.

- 5** In Fig. 21-15, a central particle of charge $-q$ is surrounded by two circular rings of charged particles. What are the magnitude and direction of the net electrostatic force on the central particle due to the other particles? (Hint: Consider symmetry.)

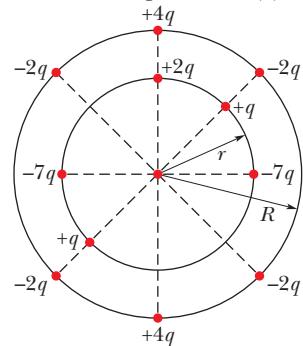


Figure 21-15 Question 5.

- 6** A positively charged ball is brought close to an electrically neutral isolated conductor. The conductor is then grounded while the ball is kept close. Is the conductor charged positively, charged negatively, or neutral if (a) the ball is first taken away and then the

ground connection is removed and (b) the ground connection is first removed and then the ball is taken away?

- 7** Figure 21-16 shows three situations involving a charged particle and a uniformly charged spherical shell. The charges are given, and the radii of the shells are indicated. Rank the situations according to the magnitude of the force on the particle due to the presence of the shell, greatest first.

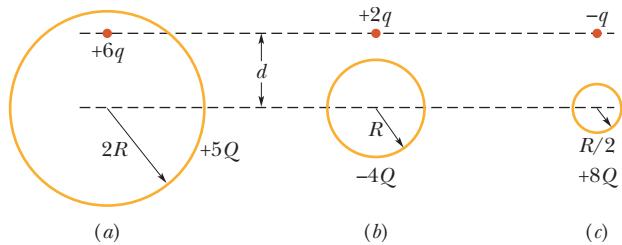


Figure 21-16 Question 7.

- 8** Figure 21-17 shows four arrangements of charged particles. Rank the arrangements according to the magnitude of the net electrostatic force on the particle with charge $+Q$, greatest first.

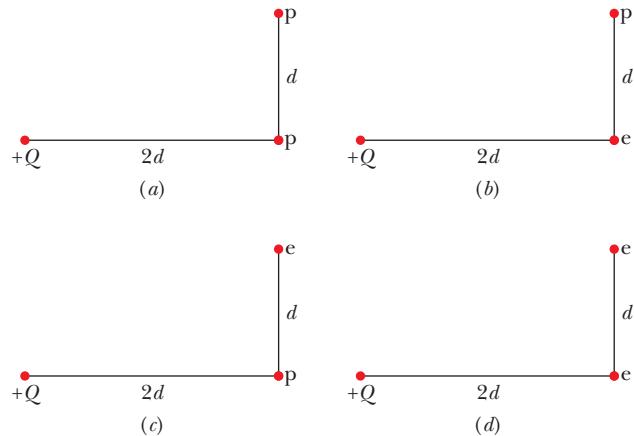


Figure 21-17 Question 8.

- 9** Figure 21-18 shows four situations in which particles of charge $+q$ or $-q$ are fixed in place. In each situation, the parti-

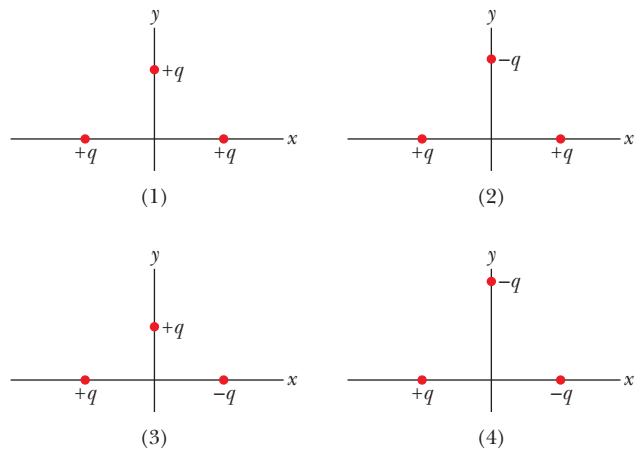


Figure 21-18 Question 9.

cles on the x axis are equidistant from the y axis. First, consider the middle particle in situation 1; the middle particle experiences an electrostatic force from each of the other two particles. (a) Are the magnitudes F of those forces the same or different? (b) Is the magnitude of the net force on the middle particle equal to, greater than, or less than $2F$? (c) Do the x components of the two forces add or cancel? (d) Do their y components add or cancel? (e) Is the direction of the net force on the middle particle that of the canceling components or the adding components? (f) What is the direction of that net force? Now consider the remaining situations: What is the direction of the net force on the middle particle in (g) situation 2, (h) situation 3, and (i) situation 4? (In each situation, consider the symmetry of the charge distribution and determine the canceling components and the adding components.)

10 In Fig. 21-19, a central particle of charge $-2q$ is surrounded by a square array of charged particles, separated by either distance d or $d/2$ along the perimeter of the square. What are the magnitude and direction of the net electrostatic force on the central particle due to the other particles? (*Hint:* Consideration of symmetry can greatly reduce the amount of work required here.)

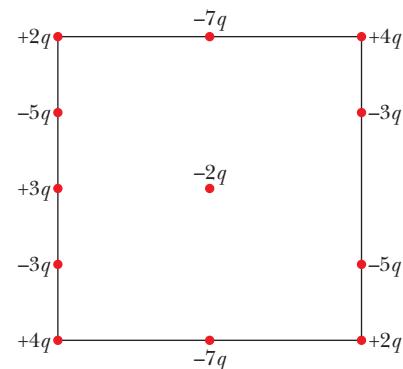


Figure 21-19 Question 10.

11 Figure 21-20 shows three identical conducting bubbles A , B , and C floating in a con-

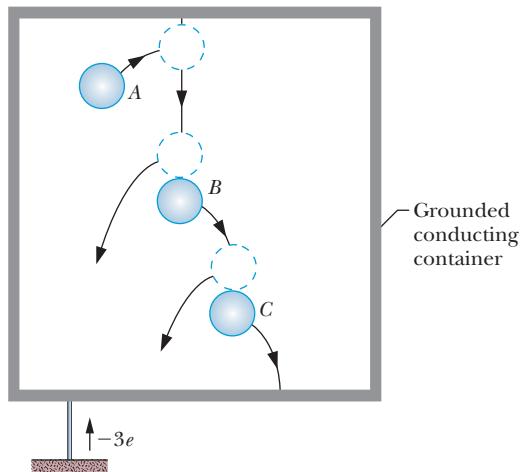


Figure 21-20 Question 11.

ducting container that is grounded by a wire. The bubbles initially have the same charge. Bubble A bumps into the container's ceiling and then into bubble B . Then bubble B bumps into bubble C , which then drifts to the container's floor. When bubble C reaches the floor, a charge of $-3e$ is transferred upward through the wire, from the ground to the container, as indicated. (a) What was the initial charge of each bubble? When (b) bubble A and (c) bubble B reach the floor, what is the charge transfer through the wire? (d) During this whole process, what is the total charge transfer through the wire?

12 Figure 21-21 shows four situations in which a central proton is partially surrounded by protons or electrons fixed in place along a half-circle. The angles θ are identical; the angles ϕ are also. (a) In each situation, what is the direction of the net force on the central proton due to the other particles? (b) Rank the four situations according to the magnitude of that net force on the central proton, greatest first.

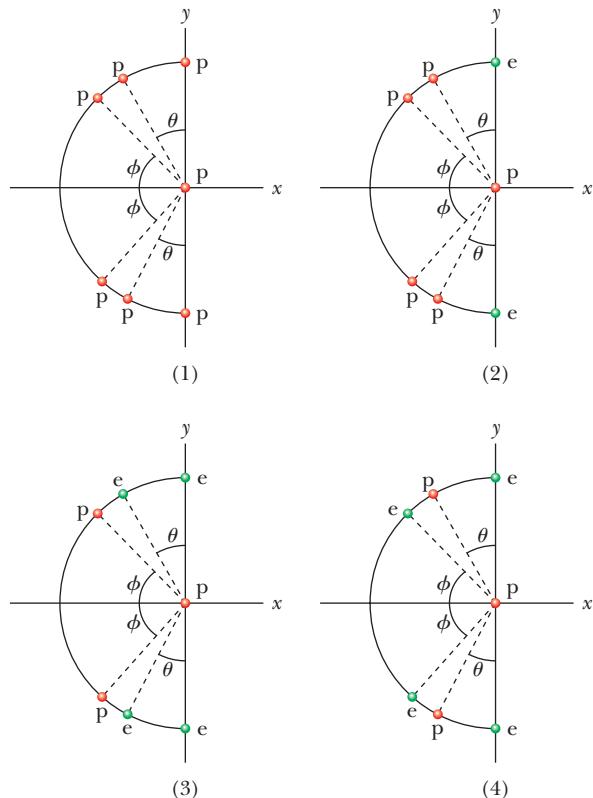


Figure 21-21 Question 12.

Problems



Tutoring problem available (at instructor's discretion) in *WileyPLUS* and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com

WWW Worked-out solution is at

ILW Interactive solution is at

<http://www.wiley.com/college/halliday>

Module 21-1 Coulomb's Law

- **SSM** **ILW** Of the charge Q initially on a tiny sphere, a portion q is to be transferred to a second, nearby sphere. Both spheres

can be treated as particles and are fixed with a certain separation. For what value of q/Q will the electrostatic force between the two spheres be maximized?

- 2** Identical isolated conducting spheres 1 and 2 have equal charges and are separated by a distance that is large compared with their diameters (Fig. 21-22a). The electrostatic force acting on sphere 2 due to sphere 1 is \vec{F} . Suppose now that a third identical sphere 3, having an insulating handle and initially neutral, is touched first to sphere 1 (Fig. 21-22b), then to sphere 2 (Fig. 21-22c), and finally removed (Fig. 21-22d). The electrostatic force that now acts on sphere 2 has magnitude F' . What is the ratio F'/F ?

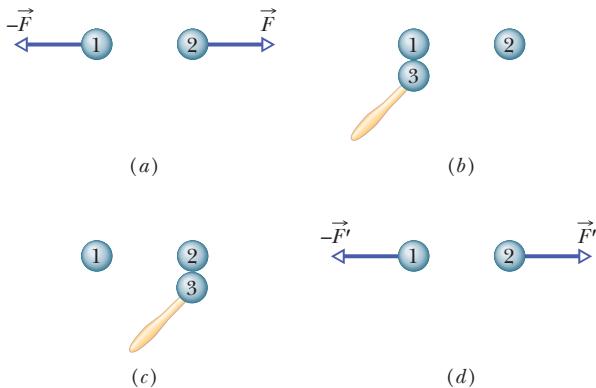


Figure 21-22 Problem 2.

- 3 SSM** What must be the distance between point charge $q_1 = 26.0 \mu\text{C}$ and point charge $q_2 = -47.0 \mu\text{C}$ for the electrostatic force between them to have a magnitude of 5.70 N?

- 4** In the return stroke of a typical lightning bolt, a current of $2.5 \times 10^4 \text{ A}$ exists for $20 \mu\text{s}$. How much charge is transferred in this event?

- 5** A particle of charge $+3.00 \times 10^{-6} \text{ C}$ is 12.0 cm distant from a second particle of charge $-1.50 \times 10^{-6} \text{ C}$. Calculate the magnitude of the electrostatic force between the particles.

- 6 ILW** Two equally charged particles are held $3.2 \times 10^{-3} \text{ m}$ apart and then released from rest. The initial acceleration of the first particle is observed to be 7.0 m/s^2 and that of the second to be 9.0 m/s^2 . If the mass of the first particle is $6.3 \times 10^{-7} \text{ kg}$, what are (a) the mass of the second particle and (b) the magnitude of the charge of each particle?

- 7** In Fig. 21-23, three charged particles lie on an x axis. Particles 1 and 2 are fixed in place. Particle 3 is free to move, but the net electrostatic force on it from particles 1 and 2 happens to be zero. If $L_{23} = L_{12}$, what is the ratio q_1/q_2 ?

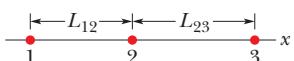


Figure 21-23 Problems 7 and 40.

- 8** In Fig. 21-24, three identical conducting spheres initially have the following charges: sphere A, $4Q$; sphere B, $-6Q$; and sphere C, 0. Spheres A and B are fixed in place, with a center-to-center separation that is much larger than the spheres. Two experiments are conducted. In experiment 1, sphere C is touched to sphere A and then (separately) to sphere B, and then it is removed. In experiment 2, starting with the same initial states, the procedure is reversed: Sphere C is touched to sphere B and then (separately) to sphere A, and then it is removed. What is the ratio of the electro-

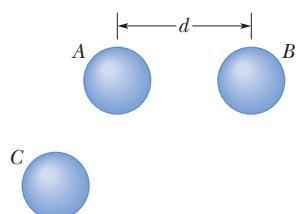


Figure 21-24
Problems 8 and 65.

static force between A and B at the end of experiment 2 to that at the end of experiment 1?

- 9 SSM WWW** Two identical conducting spheres, fixed in place, attract each other with an electrostatic force of 0.108 N when their center-to-center separation is 50.0 cm . The spheres are then connected by a thin conducting wire. When the wire is removed, the spheres repel each other with an electrostatic force of 0.0360 N . Of the initial charges on the spheres, with a positive net charge, what was (a) the negative charge on one of them and (b) the positive charge on the other?

- 10 GO** In Fig. 21-25, four particles form a square. The charges are $q_1 = q_4 = Q$ and $q_2 = q_3 = q$. (a) What is Q/q if the net electrostatic force on particles 1 and 4 is zero? (b) Is there any value of q that makes the net electrostatic force on each of the four particles zero? Explain.

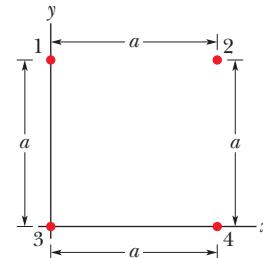


Figure 21-25
Problems 10, 11, and 70.

- 11 ILW** In Fig. 21-25, the particles have charges $q_1 = -q_2 = 100 \text{ nC}$ and $q_3 = -q_4 = 200 \text{ nC}$, and distance $a = 5.0 \text{ cm}$. What are the (a) x and (b) y components of the net electrostatic force on particle 3?

- 12** Two particles are fixed on an x axis. Particle 1 of charge $40 \mu\text{C}$ is located at $x = -2.0 \text{ cm}$; particle 2 of charge Q is located at $x = 3.0 \text{ cm}$. Particle 3 of charge magnitude $20 \mu\text{C}$ is released from rest on the y axis at $y = 2.0 \text{ cm}$. What is the value of Q if the initial acceleration of particle 3 is in the positive direction of (a) the x axis and (b) the y axis?

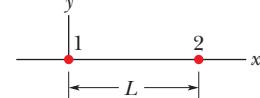


Figure 21-26 Problems 13,
19, 30, 58, and 67.

- 13 GO** In Fig. 21-26, particle 1 of charge $+1.0 \mu\text{C}$ and particle 2 of charge $-3.0 \mu\text{C}$ are held at separation $L = 10.0 \text{ cm}$ on an x axis. If particle 3 of unknown charge q_3 is to be located such that the net electrostatic force on it from particles 1 and 2 is zero, what must be the (a) x and (b) y coordinates of particle 3?

- 14** Three particles are fixed on an x axis. Particle 1 of charge q_1 is at $x = -a$, and particle 2 of charge q_2 is at $x = +a$. If their net electrostatic force on particle 3 of charge $+Q$ is to be zero, what must be the ratio q_1/q_2 when particle 3 is at (a) $x = +0.500a$ and (b) $x = +1.50a$?

- 15 GO** The charges and coordinates of two charged particles held fixed in an xy plane are $q_1 = +3.0 \mu\text{C}$, $x_1 = 3.5 \text{ cm}$, $y_1 = 0.50 \text{ cm}$, and $q_2 = -4.0 \mu\text{C}$, $x_2 = -2.0 \text{ cm}$, $y_2 = 1.5 \text{ cm}$. Find the (a) magnitude and (b) direction of the electrostatic force on particle 2 due to particle 1. At what (c) x and (d) y coordinates should a third particle of charge $q_3 = +4.0 \mu\text{C}$ be placed such that the net electrostatic force on particle 2 due to particles 1 and 3 is zero?

- 16 GO** In Fig. 21-27a, particle 1 (of charge q_1) and particle 2 (of charge q_2) are fixed in place on an x axis, 8.00 cm apart. Particle 3 (of

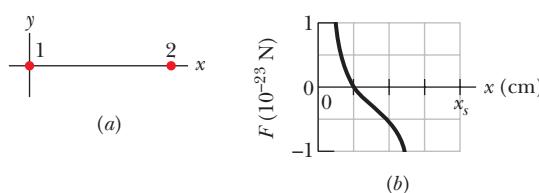


Figure 21-27 Problem 16.

charge $q_3 = +8.00 \times 10^{-19}$ C) is to be placed on the line between particles 1 and 2 so that they produce a net electrostatic force $\vec{F}_{3,\text{net}}$ on it. Figure 21-27b gives the x component of that force versus the coordinate x at which particle 3 is placed. The scale of the x axis is set by $x_s = 8.0$ cm. What are (a) the sign of charge q_1 and (b) the ratio q_2/q_1 ?

••17 In Fig. 21-28a, particles 1 and 2 have charge $20.0 \mu\text{C}$ each and are held at separation distance $d = 1.50$ m. (a) What is the magnitude of the electrostatic force on particle 1 due to particle 2? In Fig. 21-28b, particle 3 of charge $20.0 \mu\text{C}$ is positioned so as to complete an equilateral triangle. (b) What is the magnitude of the net electrostatic force on particle 1 due to particles 2 and 3?

••18 In Fig. 21-29a, three positively charged particles are fixed on an x axis. Particles B and C are so close to each other that they can be considered to be at the same distance from particle A . The net force on particle A due to particles B and C is 2.014×10^{-23} N in the negative direction of the x axis. In Fig. 21-29b, particle B has been moved to the opposite side of A but is still at the same distance from it. The net force on A is now 2.877×10^{-24} N in the negative direction of the x axis. What is the ratio q_C/q_B ?

••19 SSM WWW In Fig. 21-26, particle 1 of charge $+q$ and particle 2 of charge $+4.00q$ are held at separation $L = 9.00$ cm on an x axis. If particle 3 of charge q_3 is to be located such that the three particles remain in place when released, what must be the (a) x and (b) y coordinates of particle 3, and (c) the ratio q_3/q ?

••20 Figure 21-30a shows an arrangement of three charged particles separated by distance d . Particles A and C are fixed on the x axis, but particle B can be moved along a circle centered on particle A . During the movement, a radial line between A and B makes an angle θ relative to the positive direction of the x axis (Fig. 21-30b). The curves in Fig. 21-30c give, for two situations, the magnitude F_{net} of the net electrostatic force on particle A due to the other particles. That net force is given as a function of angle θ and as a multiple of a basic amount F_0 . For example on curve 1, at $\theta = 180^\circ$, we see that $F_{\text{net}} = 2F_0$. (a) For the situation corresponding to curve 1, what is the ratio of the charge of particle C to that of particle B (including sign)? (b) For the situation corresponding to curve 2, what is that ratio?

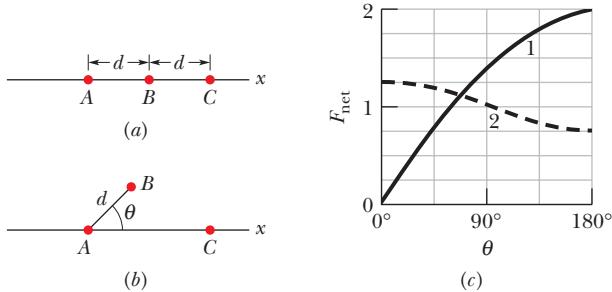


Figure 21-30 Problem 20.

••21 GO A nonconducting spherical shell, with an inner radius of 4.0 cm and an outer radius of 6.0 cm, has charge spread nonuniformly through its volume between its inner and outer surfaces. The *volume charge density* ρ is the charge per unit volume, with the unit coulomb per cubic meter. For this shell $\rho = b/r$, where r is the distance in meters from the center of the shell and $b = 3.0 \mu\text{C}/\text{m}^3$. What is the net charge in the shell?

••22 GO Figure 21-31 shows an arrangement of four charged particles, with angle $\theta = 30.0^\circ$ and distance $d = 2.00$ cm. Particle 2 has charge $q_2 = +8.00 \times 10^{-19}$ C; particles 3 and 4 have charges $q_3 = q_4 = -1.60 \times 10^{-19}$ C. (a) What is distance D between the origin and particle 2 if the net electrostatic force on particle 1 due to the other particles is zero? (b) If particles 3 and 4 were moved closer to the x axis but maintained their symmetry about that axis, would the required value of D be greater than, less than, or the same as in part (a)?

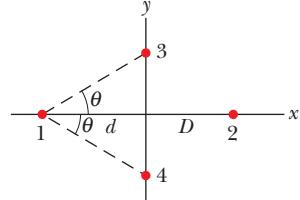


Figure 21-31 Problem 22.

••23 GO In Fig. 21-32, particles 1 and 2 of charge $q_1 = q_2 = +3.20 \times 10^{-19}$ C are on a y axis at distance $d = 17.0$ cm from the origin. Particle 3 of charge $q_3 = +6.40 \times 10^{-19}$ C is moved gradually along the x axis from $x = 0$ to $x = +5.0$ m. At what values of x will the magnitude of the electrostatic force on the third particle from the other two particles be (a) minimum and (b) maximum? What are the (c) minimum and (d) maximum magnitudes?

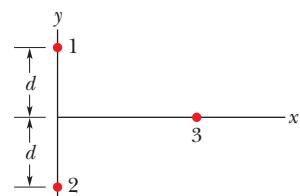


Figure 21-32 Problem 23.

Module 21-2 Charge Is Quantized

•24 Two tiny, spherical water drops, with identical charges of -1.00×10^{-16} C, have a center-to-center separation of 1.00 cm. (a) What is the magnitude of the electrostatic force acting between them? (b) How many excess electrons are on each drop, giving it its charge imbalance?

•25 ILW How many electrons would have to be removed from a coin to leave it with a charge of $+1.0 \times 10^{-7}$ C?

•26 What is the magnitude of the electrostatic force between a singly charged sodium ion (Na^+ , of charge $+e$) and an adjacent singly charged chlorine ion (Cl^- , of charge $-e$) in a salt crystal if their separation is 2.82×10^{-10} m?

•27 SSM The magnitude of the electrostatic force between two identical ions that are separated by a distance of 5.0×10^{-10} m is 3.7×10^{-9} N. (a) What is the charge of each ion? (b) How many electrons are “missing” from each ion (thus giving the ion its charge imbalance)?

•28 A current of 0.300 A through your chest can send your heart into fibrillation, ruining the normal rhythm of heartbeat and disrupting the flow of blood (and thus oxygen) to your brain. If that current persists for 2.00 min, how many conduction electrons pass through your chest?

•29 GO In Fig. 21-33, particles 2 and 4, of charge $-e$, are fixed in place on a y axis, at $y_2 = -10.0$ cm

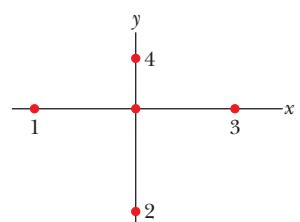


Figure 21-33 Problem 29.

and $y_4 = 5.00 \text{ cm}$. Particles 1 and 3, of charge $-e$, can be moved along the x axis. Particle 5, of charge $+e$, is fixed at the origin. Initially particle 1 is at $x_1 = -10.0 \text{ cm}$ and particle 3 is at $x_3 = 10.0 \text{ cm}$. (a) To what x value must particle 1 be moved to rotate the direction of the net electric force \vec{F}_{net} on particle 5 by 30° counter-clockwise? (b) With particle 1 fixed at its new position, to what x value must you move particle 3 to rotate \vec{F}_{net} back to its original direction?

••30 In Fig. 21-26, particles 1 and 2 are fixed in place on an x axis, at a separation of $L = 8.00 \text{ cm}$. Their charges are $q_1 = +e$ and $q_2 = -27e$. Particle 3 with charge $q_3 = +4e$ is to be placed on the line between particles 1 and 2, so that they produce a net electrostatic force $\vec{F}_{3,\text{net}}$ on it. (a) At what coordinate should particle 3 be placed to minimize the magnitude of that force? (b) What is that minimum magnitude?

••31 ILW Earth's atmosphere is constantly bombarded by *cosmic ray protons* that originate somewhere in space. If the protons all passed through the atmosphere, each square meter of Earth's surface would intercept protons at the average rate of 1500 protons per second. What would be the electric current intercepted by the total surface area of the planet?

••32 GO Figure 21-34a shows charged particles 1 and 2 that are fixed in place on an x axis. Particle 1 has a charge with a magnitude of $|q_1| = 8.00e$. Particle 3 of charge $q_3 = +8.00e$ is initially on the x axis near particle 2. Then particle 3 is gradually moved in the positive direction of the x axis. As a result, the magnitude of the net electrostatic force $\vec{F}_{2,\text{net}}$ on particle 2 due to particles 1 and 3 changes. Figure 21-34b gives the x component of that net force as a function of the position x of particle 3. The scale of the x axis is set by $x_s = 0.80 \text{ m}$. The plot has an asymptote of $F_{2,\text{net}} = 1.5 \times 10^{-25} \text{ N}$ as $x \rightarrow \infty$. As a multiple of e and including the sign, what is the charge q_2 of particle 2?

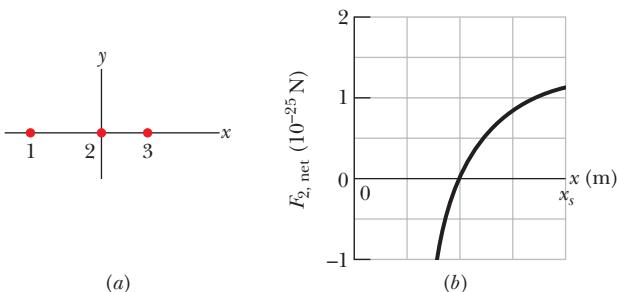


Figure 21-34 Problem 32.

••33 Calculate the number of coulombs of positive charge in 250 cm^3 of (neutral) water. (*Hint:* A hydrogen atom contains one proton; an oxygen atom contains eight protons.)

••34 GO Figure 21-35 shows electrons 1 and 2 on an x axis and charged ions 3 and 4 of identical charge $-q$ and at identical angles θ . Electron 2 is free to move; the other three particles are fixed in place at horizontal distances R from electron 2 and are intended to hold electron 2 in place. For physically possible values of $q \leq 5e$, what are the (a) smallest, (b) second smallest, and (c) third smallest values of θ for which electron 2 is held in place?

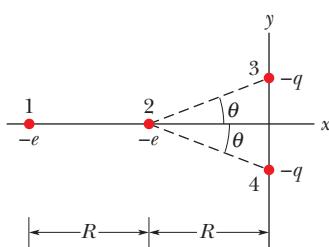


Figure 21-35 Problem 34.

••35 SSM In crystals of the salt cesium chloride, cesium ions Cs^+ form the eight corners of a cube and a chlorine ion Cl^- is at the cube's center (Fig. 21-36). The edge length of the cube is 0.40 nm . The Cs^+ ions are each deficient by one electron (and thus each has a charge of $+e$), and the Cl^- ion has one excess electron (and thus has a charge of $-e$). (a) What is the magnitude of the net electrostatic force exerted on the Cl^- ion by the eight Cs^+ ions at the corners of the cube? (b) If one of the Cs^+ ions is missing, the crystal is said to have a *defect*; what is the magnitude of the net electrostatic force exerted on the Cl^- ion by the seven remaining Cs^+ ions?

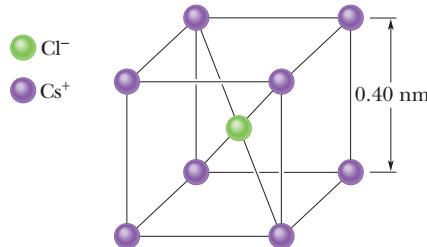


Figure 21-36 Problem 35.

Module 21-3 Charge Is Conserved

•36 Electrons and positrons are produced by the nuclear transformations of protons and neutrons known as *beta decay*. (a) If a proton transforms into a neutron, is an electron or a positron produced? (b) If a neutron transforms into a proton, is an electron or a positron produced?

•37 SSM Identify X in the following nuclear reactions: (a) ${}^1\text{H} + {}^9\text{Be} \rightarrow \text{X} + \text{n}$; (b) ${}^{12}\text{C} + {}^1\text{H} \rightarrow \text{X}$; (c) ${}^{15}\text{N} + {}^1\text{H} \rightarrow {}^4\text{He} + \text{X}$. Appendix F will help.

Additional Problems

38 GO Figure 21-37 shows four identical conducting spheres that are actually well separated from one another. Sphere W (with an initial charge of zero) is touched to sphere A and then they are separated. Next, sphere W is touched to sphere B (with an initial charge of $-32e$) and then they are separated. Finally, sphere W is touched to sphere C (with an initial charge of $+48e$), and then they are separated. The final charge on sphere W is $+18e$. What was the initial charge on sphere A?

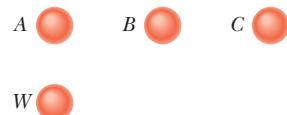


Figure 21-37 Problem 38.

39 SSM In Fig. 21-38, particle 1 of charge $+4e$ is above a floor by distance $d_1 = 2.00 \text{ mm}$ and particle 2 of charge $+6e$ is on the floor, at distance $d_2 = 6.00 \text{ mm}$ horizontally from particle 1. What is the x component of the electrostatic force on particle 2 due to particle 1?

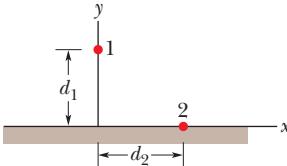


Figure 21-38 Problem 39.

40 In Fig. 21-23, particles 1 and 2 are fixed in place, but particle 3 is free to move. If the net electrostatic force on particle 3 due to particles 1 and 2 is zero and $L_{23} = 2.00L_{12}$, what is the ratio q_1/q_2 ?

41 (a) What equal positive charges would have to be placed on Earth and on the Moon to neutralize their gravitational attraction? (b) Why don't you need to know the lunar distance to solve this problem? (c) How many kilograms of hydrogen ions (that is, protons) would be needed to provide the positive charge calculated in (a)?

42 In Fig. 21-39, two tiny conducting balls of identical mass m and identical charge q hang from nonconducting threads of length L . Assume that θ is so small that $\tan \theta$ can be replaced by its approximate equal, $\sin \theta$. (a) Show that

$$x = \left(\frac{q^2 L}{2\pi\epsilon_0 mg} \right)^{1/3}$$

gives the equilibrium separation x of the balls. (b) If $L = 120$ cm, $m = 10$ g, and $x = 5.0$ cm, what is $|q|$?

43 (a) Explain what happens to the balls of Problem 42 if one of them is discharged (loses its charge q to, say, the ground). (b) Find the new equilibrium separation x , using the given values of L and m and the computed value of $|q|$.

44 SSM How far apart must two protons be if the magnitude of the electrostatic force acting on either one due to the other is equal to the magnitude of the gravitational force on a proton at Earth's surface?

45 How many megacoulombs of positive charge are in 1.00 mol of neutral molecular-hydrogen gas (H_2)?

46 In Fig. 21-40, four particles are fixed along an x axis, separated by distances $d = 2.00$ cm. The charges are $q_1 = +2e$, $q_2 = -e$, $q_3 = +e$, and $q_4 = +4e$, with $e = 1.60 \times 10^{-19}$ C. In unit-vector notation, what is the net electrostatic force on (a) particle 1 and (b) particle 2 due to the other particles?

47 GO Point charges of $+6.0 \mu\text{C}$ and $-4.0 \mu\text{C}$ are placed on an x axis, at $x = 8.0$ m and $x = 16$ m, respectively. What charge must be placed at $x = 24$ m so that any charge placed at the origin would experience no electrostatic force?

48 In Fig. 21-41, three identical conducting spheres form an equilateral triangle of side length $d = 20.0$ cm. The sphere radii are much smaller than d , and the sphere charges are $q_A = -2.00$ nC, $q_B = -4.00$ nC, and $q_C = +8.00$ nC. (a) What is the magnitude of the electrostatic force between spheres A and C ? The following steps are then taken: A and B are connected by a thin wire and then disconnected; B is grounded by the wire, and the wire is then removed; B and C are connected by the wire and then disconnected. What now are the magnitudes of the electrostatic force (b) between spheres A and C and (c) between spheres B and C ?

49 A neutron consists of one "up" quark of charge $+2e/3$ and two "down" quarks each having charge $-e/3$. If we assume that the down quarks are 2.6×10^{-15} m apart inside the neutron, what is the magnitude of the electrostatic force between them?

50 Figure 21-42 shows a long, nonconducting, massless rod of length L , pivoted at its center and balanced with a block of weight W at a distance x from the left end. At the left and right ends of the rod are attached small conducting spheres with positive charges q and $2q$, respectively. A distance h directly beneath each of these spheres is a fixed sphere with positive charge Q . (a) Find the distance x when the rod is horizontal and balanced. (b)

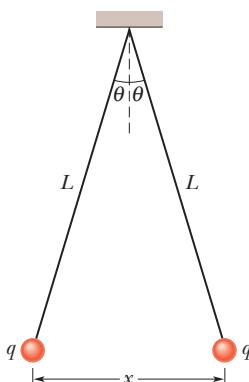


Figure 21-39
Problems 42 and 43.

What value should h have so that the rod exerts no vertical force on the bearing when the rod is horizontal and balanced?

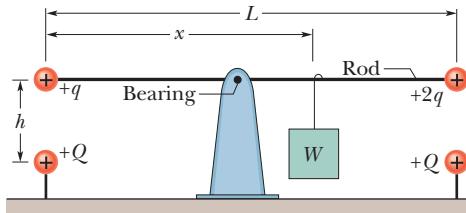


Figure 21-42 Problem 50.

51 A charged nonconducting rod, with a length of 2.00 m and a cross-sectional area of 4.00 cm^2 , lies along the positive side of an x axis with one end at the origin. The *volume charge density* ρ is charge per unit volume in coulombs per cubic meter. How many excess electrons are on the rod if ρ is (a) uniform, with a value of $-4.00 \mu\text{C}/\text{m}^3$, and (b) nonuniform, with a value given by $\rho = bx^2$, where $b = -2.00 \mu\text{C}/\text{m}^5$?

52 A particle of charge Q is fixed at the origin of an xy coordinate system. At $t = 0$ a particle ($m = 0.800$ g, $q = 4.00 \mu\text{C}$) is located on the x axis at $x = 20.0$ cm, moving with a speed of 50.0 m/s in the positive y direction. For what value of Q will the moving particle execute circular motion? (Neglect the gravitational force on the particle.)

53 What would be the magnitude of the electrostatic force between two 1.00 C point charges separated by a distance of (a) 1.00 m and (b) 1.00 km if such point charges existed (they do not) and this configuration could be set up?

54 A charge of $6.0 \mu\text{C}$ is to be split into two parts that are then separated by 3.0 mm. What is the maximum possible magnitude of the electrostatic force between those two parts?

55 Of the charge Q on a tiny sphere, a fraction α is to be transferred to a second, nearby sphere. The spheres can be treated as particles. (a) What value of α maximizes the magnitude F of the electrostatic force between the two spheres? What are the (b) smaller and (c) larger values of α that put F at half the maximum magnitude?

56 If a cat repeatedly rubs against your cotton slacks on a dry day, the charge transfer between the cat hair and the cotton can leave you with an excess charge of $-2.00 \mu\text{C}$. (a) How many electrons are transferred between you and the cat?

You will gradually discharge via the floor, but if instead of waiting, you immediately reach toward a faucet, a painful spark can suddenly appear as your fingers near the faucet. (b) In that spark, do electrons flow from you to the faucet or vice versa? (c) Just before the spark appears, do you induce positive or negative charge in the faucet? (d) If, instead, the cat reaches a paw toward the faucet, which way do electrons flow in the resulting spark? (e) If you stroke a cat with a bare hand on a dry day, you should take care not to bring your fingers near the cat's nose or you will hurt it with a spark. Considering that cat hair is an insulator, explain how the spark can appear.

57 We know that the negative charge on the electron and the positive charge on the proton are equal. Suppose, however, that these magnitudes differ from each other by 0.00010%. With what force would two copper coins, placed 1.0 m apart, repel each other? Assume that each coin contains 3×10^{22} copper atoms. (Hint: A neutral copper atom contains 29 protons and 29 electrons.) What do you conclude?

58 In Fig. 21-26, particle 1 of charge $-80.0 \mu\text{C}$ and particle 2 of charge $+40.0 \mu\text{C}$ are held at separation $L = 20.0 \text{ cm}$ on an x axis. In unit-vector notation, what is the net electrostatic force on particle 3, of charge $q_3 = 20.0 \mu\text{C}$, if particle 3 is placed at (a) $x = 40.0 \text{ cm}$ and (b) $x = 80.0 \text{ cm}$? What should be the (c) x and (d) y coordinates of particle 3 if the net electrostatic force on it due to particles 1 and 2 is zero?

59 What is the total charge in coulombs of 75.0 kg of electrons?

60 In Fig. 21-43, six charged particles surround particle 7 at radial distances of either $d = 1.0 \text{ cm}$ or $2d$, as drawn. The charges are $q_1 = +2e, q_2 = +4e, q_3 = +e, q_4 = +4e, q_5 = +2e, q_6 = +8e, q_7 = +6e$, with $e = 1.60 \times 10^{-19} \text{ C}$. What is the magnitude of the net electrostatic force on particle 7?

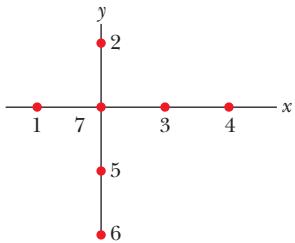


Figure 21-43 Problem 60.

61 Three charged particles form a triangle: particle 1 with charge $Q_1 = 80.0 \text{ nC}$ is at xy coordinates $(0, 3.00 \text{ mm})$, particle 2 with charge Q_2 is at $(0, -3.00 \text{ mm})$, and particle 3 with charge $q = 18.0 \text{ nC}$ is at $(4.00 \text{ mm}, 0)$. In unit-vector notation, what is the electrostatic force on particle 3 due to the other two particles if Q_2 is equal to (a) 80.0 nC and (b) -80.0 nC ?

62 In Fig. 21-44, what are the (a) magnitude and (b) direction of the net electrostatic force on particle 4 due to the other three particles? All four particles are fixed in the xy plane, and $q_1 = -3.20 \times 10^{-19} \text{ C}, q_2 = +3.20 \times 10^{-19} \text{ C}, q_3 = +6.40 \times 10^{-19} \text{ C}, q_4 = +3.20 \times 10^{-19} \text{ C}, \theta_1 = 35.0^\circ, d_1 = 3.00 \text{ cm}$, and $d_2 = d_3 = 2.00 \text{ cm}$.

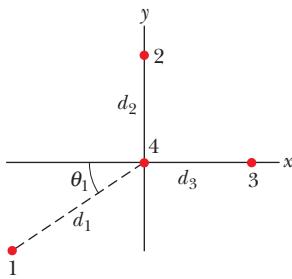


Figure 21-44 Problem 62.

63 Two point charges of 30 nC and -40 nC are held fixed on an x axis, at the origin and at $x = 72 \text{ cm}$, respectively. A particle with a charge of $42 \mu\text{C}$ is released from rest at $x = 28 \text{ cm}$. If the initial acceleration of the particle has a magnitude of 100 km/s^2 , what is the particle's mass?

64 Two small, positively charged spheres have a combined charge of $5.0 \times 10^{-5} \text{ C}$. If each sphere is repelled from the other by an electrostatic force of 1.0 N when the spheres are 2.0 m apart, what is the charge on the sphere with the smaller charge?

65 The initial charges on the three identical metal spheres in Fig. 21-24 are the following: sphere A, Q ; sphere $B, -Q/4$; and sphere $C, Q/2$, where $Q = 2.00 \times 10^{-14} \text{ C}$. Spheres A and B are fixed in place, with a center-to-center separation of $d = 1.20 \text{ m}$, which is much larger than the spheres. Sphere C is touched first to sphere A and then to sphere B and is then removed. What then is the magnitude of the electrostatic force between spheres A and B ?

66 An electron is in a vacuum near Earth's surface and located at $y = 0$ on a vertical y axis. At what value of y should a second electron be placed such that its electrostatic force on the first electron balances the gravitational force on the first electron?

67 In Fig. 21-26, particle 1 of charge $-5.00q$ and particle 2 of charge $+2.00q$ are held at separation L on an x axis. If particle 3 of unknown charge q_3 is to be located such that the net electrostatic force on it from particles 1 and 2 is zero, what must be the (a) x and (b) y coordinates of particle 3?

68 Two engineering students, John with a mass of 90 kg and Mary with a mass of 45 kg , are 30 m apart. Suppose each has a 0.01% imbalance in the amount of positive and negative charge, one student being positive and the other negative. Find the order of magnitude of the electrostatic force of attraction between them by replacing each student with a sphere of water having the same mass as the student.

69 In the radioactive decay of Eq. 21-13, a ^{238}U nucleus transforms to ^{234}Th and an ejected ^4He . (These are nuclei, not atoms, and thus electrons are not involved.) When the separation between ^{234}Th and ^4He is $9.0 \times 10^{-15} \text{ m}$, what are the magnitudes of (a) the electrostatic force between them and (b) the acceleration of the ^4He particle?

70 In Fig. 21-25, four particles form a square. The charges are $q_1 = +Q, q_2 = q_3 = q$, and $q_4 = -2.00Q$. What is q/Q if the net electrostatic force on particle 1 is zero?

71 In a spherical metal shell of radius R , an electron is shot from the center directly toward a tiny hole in the shell, through which it escapes. The shell is negatively charged with a *surface charge density* (charge per unit area) of $6.90 \times 10^{-13} \text{ C/m}^2$. What is the magnitude of the electron's acceleration when it reaches radial distances (a) $r = 0.500R$ and (b) $2.00R$?

72 An electron is projected with an initial speed $v_i = 3.2 \times 10^5 \text{ m/s}$ directly toward a very distant proton that is at rest. Because the proton mass is large relative to the electron mass, assume that the proton remains at rest. By calculating the work done on the electron by the electrostatic force, determine the distance between the two particles when the electron instantaneously has speed $2v_i$.

73 In an early model of the hydrogen atom (the *Bohr model*), the electron orbits the proton in uniformly circular motion. The radius of the circle is restricted (*quantized*) to certain values given by

$$r = n^2 a_0, \quad \text{for } n = 1, 2, 3, \dots,$$

where $a_0 = 52.92 \text{ pm}$. What is the speed of the electron if it orbits in (a) the smallest allowed orbit and (b) the second smallest orbit? (c) If the electron moves to larger orbits, does its speed increase, decrease, or stay the same?

74 A 100 W lamp has a steady current of 0.83 A in its filament. How long is required for 1 mol of electrons to pass through the lamp?

75 The charges of an electron and a positron are $-e$ and $+e$. The mass of each is $9.11 \times 10^{-31} \text{ kg}$. What is the ratio of the electrical force to the gravitational force between an electron and a positron?

Electric Fields

22-1 THE ELECTRIC FIELD

Learning Objectives

After reading this module, you should be able to...

22.01 Identify that at every point in the space surrounding a charged particle, the particle sets up an electric field \vec{E} , which is a vector quantity and thus has both magnitude and direction.

22.02 Identify how an electric field \vec{E} can be used to explain how a charged particle can exert an electrostatic force \vec{F}

on a second charged particle even though there is no contact between the particles.

22.03 Explain how a small positive test charge is used (in principle) to measure the electric field at any given point.

22.04 Explain electric field lines, including where they originate and terminate and what their spacing represents.

Key Ideas

- A charged particle sets up an electric field (a vector quantity) in the surrounding space. If a second charged particle is located in that space, an electrostatic force acts on it due to the magnitude and direction of the field at its location.

- The electric field \vec{E} at any point is defined in terms of the electrostatic force \vec{F} that would be exerted on a positive test charge q_0 placed there:

$$\vec{E} = \frac{\vec{F}}{q_0}$$

- Electric field lines help us visualize the direction and magnitude of electric fields. The electric field vector at any point is tangent to the field line through that point. The density of field lines in that region is proportional to the magnitude of the electric field there. Thus, closer field lines represent a stronger field.

- Electric field lines originate on positive charges and terminate on negative charges. So, a field line extending from a positive charge must end on a negative charge.

What Is Physics?

Figure 22-1 shows two positively charged particles. From the preceding chapter we know that an electrostatic force acts on particle 1 due to the presence of particle 2. We also know the force direction and, given some data, we can calculate the force magnitude. However, here is a leftover nagging question. How does particle 1 “know” of the presence of particle 2? That is, since the particles do not touch, how can particle 2 push on particle 1—how can there be such an *action at a distance*?

One purpose of physics is to record observations about our world, such as the magnitude and direction of the push on particle 1. Another purpose is to provide an explanation of what is recorded. Our purpose in this chapter is to provide such an explanation to this nagging question about electric force at a distance.

The explanation that we shall examine here is this: Particle 2 sets up an **electric field** at all points in the surrounding space, even if the space is a vacuum. If we place particle 1 at any point in that space, particle 1 knows of the presence of particle 2 because it is affected by the electric field particle 2 has already set up at that point. Thus, particle 2 pushes on particle 1 not by touching it as you would push on a coffee mug by making contact. Instead, particle 2 pushes by means of the electric field it has set up.



Figure 22-1 How does charged particle 2 push on charged particle 1 when they have no contact?

Our goals in this chapter are to (1) define electric field, (2) discuss how to calculate it for various arrangements of charged particles and objects, and (3) discuss how an electric field can affect a charged particle (as in making it move).

The Electric Field

A lot of different fields are used in science and engineering. For example, a *temperature field* for an auditorium is the distribution of temperatures we would find by measuring the temperature at many points within the auditorium. Similarly, we could define a *pressure field* in a swimming pool. Such fields are examples of *scalar fields* because temperature and pressure are scalar quantities, having only magnitudes and not directions.

In contrast, an electric field is a *vector field* because it is responsible for conveying the information for a force, which involves both magnitude and direction. This field consists of a distribution of electric field vectors \vec{E} , one for each point in the space around a charged object. In principle, we can define \vec{E} at some point near the charged object, such as point P in Fig. 22-2a, with this procedure: At P , we place a particle with a small positive charge q_0 , called a *test charge* because we use it to test the field. (We want the charge to be small so that it does not disturb the object's charge distribution.) We then measure the electrostatic force \vec{F} that acts on the test charge. The electric field at that point is then

$$\vec{E} = \frac{\vec{F}}{q_0} \quad (\text{electric field}). \quad (22-1)$$

Because the test charge is positive, the two vectors in Eq. 22-1 are in the same direction, so the direction of \vec{E} is the direction we measure for \vec{F} . The magnitude of \vec{E} at point P is F/q_0 . As shown in Fig. 22-2b, we always represent an electric field with an arrow with its tail anchored on the point where the measurement is made. (This may sound trivial, but drawing the vectors any other way usually results in errors. Also, another common error is to mix up the terms *force* and *field* because they both start with the letter f. Electric force is a push or pull. Electric field is an abstract property set up by a charged object.) From Eq. 22-1, we see that the SI unit for the electric field is the newton per coulomb (N/C).

We can shift the test charge around to various other points, to measure the electric fields there, so that we can figure out the distribution of the electric field set up by the charged object. That field exists independent of the test charge. It is something that a charged object sets up in the surrounding space (even vacuum), independent of whether we happen to come along to measure it.

For the next several modules, we determine the field around charged particles and various charged objects. First, however, let's examine a way of visualizing electric fields.

Electric Field Lines

Look at the space in the room around you. Can you visualize a field of vectors throughout that space—vectors with different magnitudes and directions? As impossible as that seems, Michael Faraday, who introduced the idea of electric fields in the 19th century, found a way. He envisioned lines, now called **electric field lines**, in the space around any given charged particle or object.

Figure 22-3 gives an example in which a sphere is uniformly covered with negative charge. If we place a positive test charge at any point near the sphere (Fig. 22-3a), we find that an electrostatic force pulls on it toward the center of the sphere. Thus at every point around the sphere, an electric field vector points radially inward toward the sphere. We can represent this electric field with

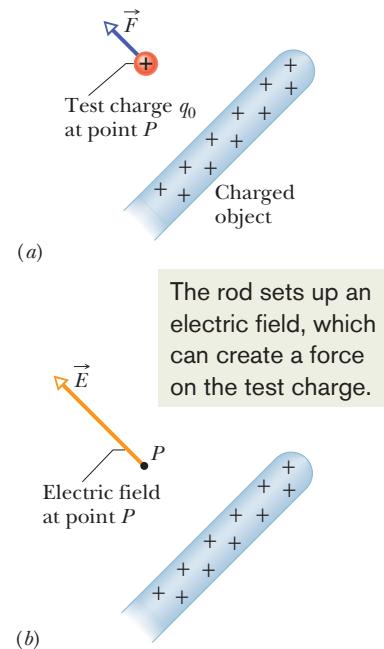


Figure 22-2 (a) A positive test charge q_0 placed at point P near a charged object. An electrostatic force \vec{F} acts on the test charge. (b) The electric field \vec{E} at point P produced by the charged object.

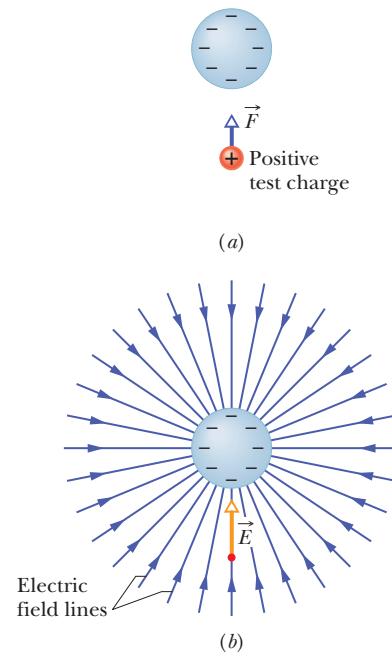


Figure 22-3 (a) The electrostatic force \vec{F} acting on a positive test charge near a sphere of uniform negative charge. (b) The electric field vector \vec{E} at the location of the test charge, and the electric field lines in the space near the sphere. The field lines extend toward the negatively charged sphere. (They originate on distant positive charges.)

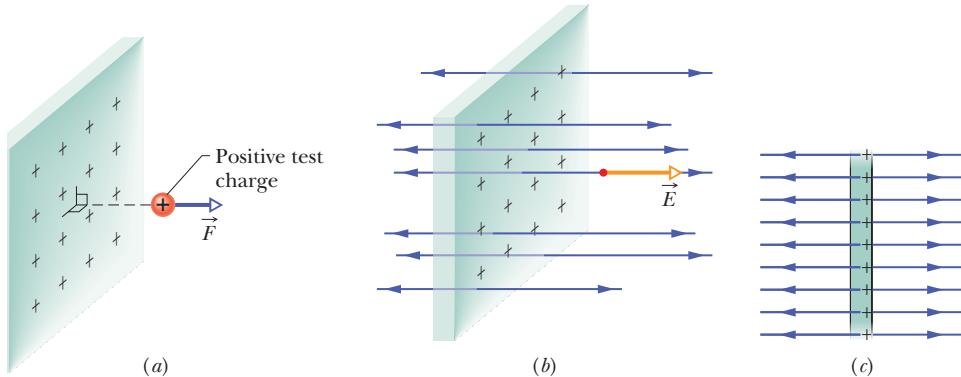


Figure 22-4 (a) The force on a positive test charge near a very large, nonconducting sheet with uniform positive charge on one side. (b) The electric field vector \vec{E} at the test charge's location, and the nearby electric field lines, extending away from the sheet. (c) Side view.

electric field lines as in Fig. 22-3b. At any point, such as the one shown, the direction of the field line through the point matches the direction of the electric vector at that point.

The rules for drawing electric fields lines are these: (1) At any point, the electric field vector must be tangent to the electric field line through that point and in the same direction. (This is easy to see in Fig. 22-3 where the lines are straight, but we'll see some curved lines soon.) (2) In a plane perpendicular to the field lines, the relative density of the lines represents the relative magnitude of the field there, with greater density for greater magnitude.

If the sphere in Fig. 22-3 were uniformly covered with positive charge, the electric field vectors at all points around it would be radially outward and thus so would the electric field lines. So, we have the following rule:



Electric field lines extend away from positive charge (where they originate) and toward negative charge (where they terminate).

In Fig. 22-3b, they originate on distant positive charges that are not shown.

For another example, Fig. 22-4a shows part of an infinitely large, nonconducting *sheet* (or plane) with a uniform distribution of positive charge on one side. If we place a positive test charge at any point near the sheet (on either side), we find that the electrostatic force on the particle is outward and perpendicular to the sheet. The perpendicular orientation is reasonable because any force component that is, say, upward is balanced out by an equal component that is downward. That leaves only outward, and thus the electric field vectors and the electric field lines must also be outward and perpendicular to the sheet, as shown in Figs. 22-4b and c.

Because the charge on the sheet is uniform, the field vectors and the field lines are also. Such a field is a *uniform electric field*, meaning that the electric field has the same magnitude and direction at every point within the field. (This is a lot easier to work with than a *nonuniform field*, where there is variation from point to point.) Of course, there is no such thing as an infinitely large sheet. That is just a way of saying that we are measuring the field at points close to the sheet relative to the size of the sheet and that we are not near an edge.

Figure 22-5 shows the field lines for two particles with equal positive charge. Now the field lines are curved, but the rules still hold: (1) the electric field vector at any given point must be tangent to the field line at that point and in the same direction, as shown for one vector, and (2) a closer spacing means a larger field magnitude. To imagine the full three-dimensional pattern of field lines around the particles, mentally rotate the pattern in Fig. 22-5 around the *axis of symmetry*, which is a vertical line through both particles.

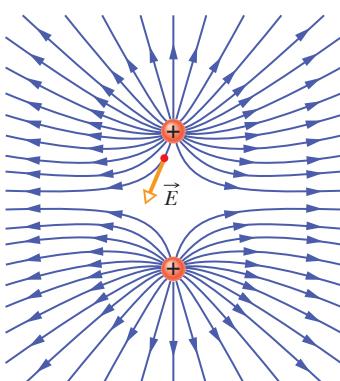


Figure 22-5 Field lines for two particles with equal positive charge. Doesn't the pattern itself suggest that the particles repel each other?

22-2 THE ELECTRIC FIELD DUE TO A CHARGED PARTICLE

Learning Objectives

After reading this module, you should be able to . . .

- 22.05** In a sketch, draw a charged particle, indicate its sign, pick a nearby point, and then draw the electric field vector \vec{E} at that point, with its tail anchored on the point.
- 22.06** For a given point in the electric field of a charged particle, identify the direction of the field vector \vec{E} when the particle is positively charged and when it is negatively charged.
- 22.07** For a given point in the electric field of a charged particle, apply the relationship between the field

magnitude E , the charge magnitude $|q|$, and the distance r between the point and the particle.

- 22.08** Identify that the equation given here for the magnitude of an electric field applies only to a particle, not an extended object.
- 22.09** If more than one electric field is set up at a point, draw each electric field vector and then find the net electric field by adding the individual electric fields as vectors (not as scalars).

Key Ideas

- The magnitude of the electric field \vec{E} set up by a particle with charge q at distance r from the particle is

$$E = \frac{1}{4\pi\epsilon_0} \frac{|q|}{r^2}.$$

- The electric field vectors set up by a positively charged particle all point directly away from the particle. Those set up

by a negatively charged particle all point directly toward the particle.

- If more than one charged particle sets up an electric field at a point, the net electric field is the vector sum of the individual electric fields—electric fields obey the superposition principle.

The Electric Field Due to a Point Charge

To find the electric field due to a charged particle (often called a *point charge*), we place a positive test charge at any point near the particle, at distance r . From Coulomb's law (Eq. 21-4), the force on the test charge due to the particle with charge q is

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \frac{qq_0}{r^2} \hat{r}.$$

As previously, the direction of \vec{F} is directly away from the particle if q is positive (because q_0 is positive) and directly toward it if q is negative. From Eq. 22-1, we can now write the electric field set up by the particle (at the location of the test charge) as

$$\vec{E} = \frac{\vec{F}}{q_0} = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{r} \quad (\text{charged particle}). \quad (22-2)$$

Let's think through the directions again. The direction of \vec{E} matches that of the force on the positive test charge: directly away from the point charge if q is positive and directly toward it if q is negative.

So, if given another charged particle, we can immediately determine the directions of the electric field vectors near it by just looking at the sign of the charge q . We can find the magnitude at any given distance r by converting Eq. 22-2 to a magnitude form:

$$E = \frac{1}{4\pi\epsilon_0} \frac{|q|}{r^2} \quad (\text{charged particle}). \quad (22-3)$$

We write $|q|$ to avoid the danger of getting a negative E when q is negative, and then thinking the negative sign has something to do with direction. Equation 22-3 gives magnitude E only. We must think about the direction separately.

Figure 22-6 gives a number of electric field vectors at points around a positively charged particle, but be careful. Each vector represents the vector

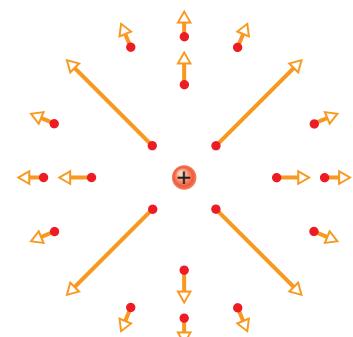


Figure 22-6 The electric field vectors at various points around a positive point charge.

quantity at the point where the tail of the arrow is anchored. The vector is not something that stretches from a “here” to a “there” as with a displacement vector.

In general, if several electric fields are set up at a given point by several charged particles, we can find the net field by placing a positive test particle at the point and then writing out the force acting on it due to each particle, such as \vec{F}_{01} due to particle 1. Forces obey the principle of superposition, so we just add the forces as vectors:

$$\vec{F}_0 = \vec{F}_{01} + \vec{F}_{02} + \cdots + \vec{F}_{0n}.$$

To change over to electric field, we repeatedly use Eq. 22-1 for each of the individual forces:

$$\begin{aligned}\vec{E} &= \frac{\vec{F}_0}{q_0} = \frac{\vec{F}_{01}}{q_0} + \frac{\vec{F}_{02}}{q_0} + \cdots + \frac{\vec{F}_{0n}}{q_0} \\ &= \vec{E}_1 + \vec{E}_2 + \cdots + \vec{E}_n.\end{aligned}\quad (22-4)$$

This tells us that electric fields also obey the principle of superposition. If you want the net electric field at a given point due to several particles, find the electric field due to each particle (such as \vec{E}_1 due to particle 1) and then sum the fields as vectors. (As with electrostatic forces, you cannot just willy-nilly add up the magnitudes.) This addition of fields is the subject of many of the homework problems.



Checkpoint 1

The figure here shows a proton p and an electron e on an x axis. What is the direction of the electric field due to the electron at (a) point S and (b) point R? What is the direction of the net electric field at (c) point R and (d) point S?



Sample Problem 22.01 Net electric field due to three charged particles

Figure 22-7a shows three particles with charges $q_1 = +2Q$, $q_2 = -2Q$, and $q_3 = -4Q$, each a distance d from the origin. What net electric field \vec{E} is produced at the origin?

KEY IDEA

Charges q_1 , q_2 , and q_3 produce electric field vectors \vec{E}_1 , \vec{E}_2 , and \vec{E}_3 , respectively, at the origin, and the net electric field is the vector sum $\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3$. To find this sum, we first must find the magnitudes and orientations of the three field vectors.

Magnitudes and directions: To find the magnitude of \vec{E}_1 , which is due to q_1 , we use Eq. 22-3, substituting d for r and $2Q$ for q and obtaining

$$E_1 = \frac{1}{4\pi\epsilon_0} \frac{2Q}{d^2}.$$

Similarly, we find the magnitudes of \vec{E}_2 and \vec{E}_3 to be

$$E_2 = \frac{1}{4\pi\epsilon_0} \frac{2Q}{d^2} \quad \text{and} \quad E_3 = \frac{1}{4\pi\epsilon_0} \frac{4Q}{d^2}.$$

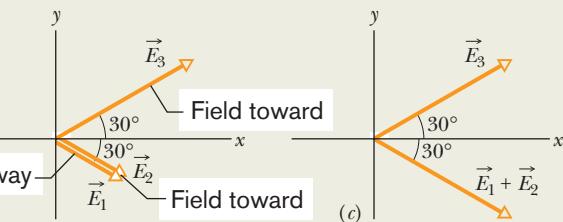
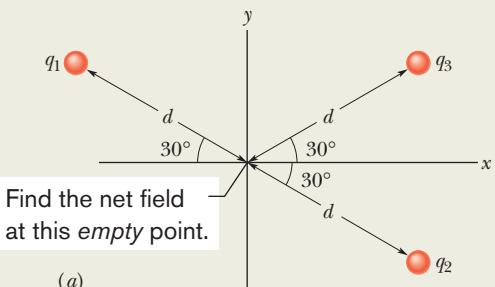


Figure 22-7 (a) Three particles with charges q_1 , q_2 , and q_3 are at the same distance d from the origin. (b) The electric field vectors \vec{E}_1 , \vec{E}_2 , and \vec{E}_3 , at the origin due to the three particles. (c) The electric field vector \vec{E}_3 and the vector sum $\vec{E}_1 + \vec{E}_2$ at the origin.

We next must find the orientations of the three electric field vectors at the origin. Because q_1 is a positive charge, the field vector it produces points directly *away* from it, and because q_2 and q_3 are both negative, the field vectors they produce point directly *toward* each of them. Thus, the three electric fields produced at the origin by the three charged particles are oriented as in Fig. 22-7b. (*Caution:* Note that we have placed the tails of the vectors at the point where the fields are to be evaluated; doing so decreases the chance of error. Error becomes very probable if the tails of the field vectors are placed on the particles creating the fields.)

Adding the fields: We can now add the fields vectorially just as we added force vectors in Chapter 21. However, here we can use symmetry to simplify the procedure. From Fig. 22-7b, we see that electric fields \vec{E}_1 and \vec{E}_2 have the same direction. Hence, their vector sum has that direction and has the magnitude

$$\begin{aligned} E_1 + E_2 &= \frac{1}{4\pi\epsilon_0} \frac{2Q}{d^2} + \frac{1}{4\pi\epsilon_0} \frac{2Q}{d^2} \\ &= \frac{1}{4\pi\epsilon_0} \frac{4Q}{d^2}, \end{aligned}$$

which happens to equal the magnitude of field \vec{E}_3 .

We must now combine two vectors, \vec{E}_3 and the vector sum $\vec{E}_1 + \vec{E}_2$, that have the same magnitude and that are oriented symmetrically about the x axis, as shown in Fig. 22-7c. From the symmetry of Fig. 22-7c, we realize that the equal y components of our two vectors cancel (one is upward and the other is downward) and the equal x components add (both are rightward). Thus, the net electric field \vec{E} at the origin is in the positive direction of the x axis and has the magnitude

$$\begin{aligned} E &= 2E_{3x} = 2E_3 \cos 30^\circ \\ &= (2) \frac{1}{4\pi\epsilon_0} \frac{4Q}{d^2} (0.866) = \frac{6.93Q}{4\pi\epsilon_0 d^2}. \end{aligned} \quad (\text{Answer})$$



Additional examples, video, and practice available at WileyPLUS

22-3 THE ELECTRIC FIELD DUE TO A DIPOLE

Learning Objectives

After reading this module, you should be able to . . .

- 22.10 Draw an electric dipole, identifying the charges (sizes and signs), dipole axis, and direction of the electric dipole moment.
- 22.11 Identify the direction of the electric field at any given point along the dipole axis, including between the charges.
- 22.12 Outline how the equation for the electric field due to an electric dipole is derived from the equations for the electric field due to the individual charged particles that form the dipole.
- 22.13 For a single charged particle and an electric dipole, compare the rate at which the electric field magnitude

decreases with increase in distance. That is, identify which drops off faster.

- 22.14 For an electric dipole, apply the relationship between the magnitude p of the dipole moment, the separation d between the charges, and the magnitude q of either of the charges.

- 22.15 For any distant point along a dipole axis, apply the relationship between the electric field magnitude E , the distance z from the center of the dipole, and either the dipole moment magnitude p or the product of charge magnitude q and charge separation d .

Key Ideas

- An electric dipole consists of two particles with charges of equal magnitude q but opposite signs, separated by a small distance d .
- The electric dipole moment \vec{p} has magnitude qd and points from the negative charge to the positive charge.
- The magnitude of the electric field set up by an electric dipole at a distant point on the dipole axis (which runs through both particles) can be written in terms of either the product qd or the magnitude p of the dipole moment:

$$E = \frac{1}{2\pi\epsilon_0} \frac{qd}{z^3} = \frac{1}{2\pi\epsilon_0} \frac{p}{z^3},$$

where z is the distance between the point and the center of the dipole.

- Because of the $1/z^3$ dependence, the field magnitude of an electric dipole decreases more rapidly with distance than the field magnitude of either of the individual charges forming the dipole, which depends on $1/r^2$.

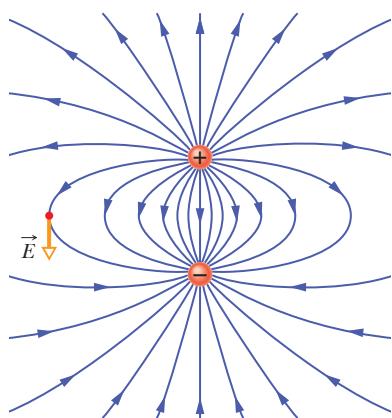


Figure 22-8 The pattern of electric field lines around an electric dipole, with an electric field vector \vec{E} shown at one point (tangent to the field line through that point).

The Electric Field Due to an Electric Dipole

Figure 22-8 shows the pattern of electric field lines for two particles that have the same charge magnitude q but opposite signs, a very common and important arrangement known as an **electric dipole**. The particles are separated by distance d and lie along the *dipole axis*, an axis of symmetry around which you can imagine rotating the pattern in Fig. 22-8. Let's label that axis as a z axis. Here we restrict our interest to the magnitude and direction of the electric field \vec{E} at an arbitrary point P along the dipole axis, at distance z from the dipole's midpoint.

Figure 22-9a shows the electric fields set up at P by each particle. The nearer particle with charge $+q$ sets up field $E_{(+)}$ in the positive direction of the z axis (directly away from the particle). The farther particle with charge $-q$ sets up a smaller field $E_{(-)}$ in the negative direction (directly toward the particle). We want the net field at P , as given by Eq. 22-4. However, because the field vectors are along the same axis, let's simply indicate the vector directions with plus and minus signs, as we commonly do with forces along a single axis. Then we can write the magnitude of the net field at P as

$$\begin{aligned} E &= E_{(+)} - E_{(-)} \\ &= \frac{1}{4\pi\epsilon_0} \frac{q}{r_{(+)}^2} - \frac{1}{4\pi\epsilon_0} \frac{q}{r_{(-)}^2} \\ &= \frac{q}{4\pi\epsilon_0(z - \frac{1}{2}d)^2} - \frac{q}{4\pi\epsilon_0(z + \frac{1}{2}d)^2}. \end{aligned} \quad (22-5)$$

After a little algebra, we can rewrite this equation as

$$E = \frac{q}{4\pi\epsilon_0 z^2} \left(\frac{1}{\left(1 - \frac{d}{2z}\right)^2} - \frac{1}{\left(1 + \frac{d}{2z}\right)^2} \right). \quad (22-6)$$

After forming a common denominator and multiplying its terms, we come to

$$E = \frac{q}{4\pi\epsilon_0 z^2} \frac{2d/z}{\left(1 - \left(\frac{d}{2z}\right)^2\right)^2} = \frac{q}{2\pi\epsilon_0 z^3} \frac{d}{\left(1 - \left(\frac{d}{2z}\right)^2\right)^2}. \quad (22-7)$$

We are usually interested in the electrical effect of a dipole only at distances that are large compared with the dimensions of the dipole—that is, at distances such that $z \gg d$. At such large distances, we have $d/2z \ll 1$ in Eq. 22-7. Thus, in our approximation, we can neglect the $d/2z$ term in the denominator, which leaves us with

$$E = \frac{1}{2\pi\epsilon_0} \frac{qd}{z^3}. \quad (22-8)$$

The product qd , which involves the two intrinsic properties q and d of the dipole, is the magnitude p of a vector quantity known as the **electric dipole moment** \vec{p} of the dipole. (The unit of \vec{p} is the coulomb-meter.) Thus, we can write Eq. 22-8 as

$$E = \frac{1}{2\pi\epsilon_0} \frac{p}{z^3} \quad (\text{electric dipole}). \quad (22-9)$$

The direction of \vec{p} is taken to be from the negative to the positive end of the dipole, as indicated in Fig. 22-9b. We can use the direction of \vec{p} to specify the orientation of a dipole.

Equation 22-9 shows that, if we measure the electric field of a dipole only at distant points, we can never find q and d separately; instead, we can find only their product. The field at distant points would be unchanged if, for example, q

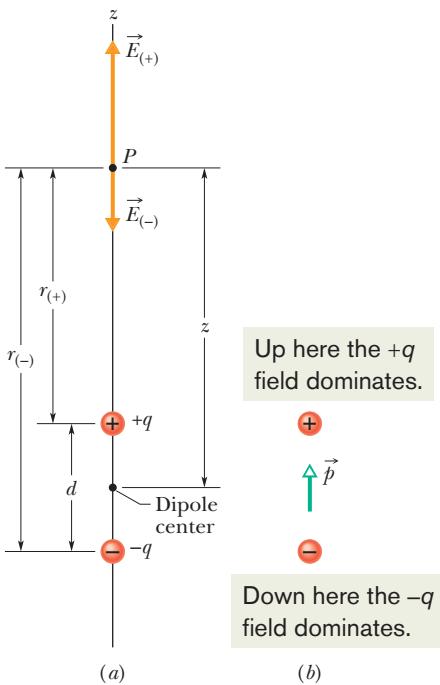


Figure 22-9 (a) An electric dipole. The electric field vectors $\vec{E}_{(+)}$ and $\vec{E}_{(-)}$ at point P on the dipole axis result from the dipole's two charges. Point P is at distances $r_{(+)}$ and $r_{(-)}$ from the individual charges that make up the dipole. (b) The dipole moment \vec{p} of the dipole points from the negative charge to the positive charge.

were doubled and d simultaneously halved. Although Eq. 22-9 holds only for distant points along the dipole axis, it turns out that E for a dipole varies as $1/r^3$ for all distant points, regardless of whether they lie on the dipole axis; here r is the distance between the point in question and the dipole center.

Inspection of Fig. 22-9 and of the field lines in Fig. 22-8 shows that the direction of \vec{E} for distant points on the dipole axis is always the direction of the dipole moment vector \vec{p} . This is true whether point P in Fig. 22-9a is on the upper or the lower part of the dipole axis.

Inspection of Eq. 22-9 shows that if you double the distance of a point from a dipole, the electric field at the point drops by a factor of 8. If you double the distance from a single point charge, however (see Eq. 22-3), the electric field drops only by a factor of 4. Thus the electric field of a dipole decreases more rapidly with distance than does the electric field of a single charge. The physical reason for this rapid decrease in electric field for a dipole is that from distant points a dipole looks like two particles that almost—but not quite—coincide. Thus, because they have charges of equal magnitude but opposite signs, their electric fields at distant points almost—but not quite—cancel each other.

Sample Problem 22.02 Electric dipole and atmospheric sprites

Sprites (Fig. 22-10a) are huge flashes that occur far above a large thunderstorm. They were seen for decades by pilots flying at night, but they were so brief and dim that most pilots figured they were just illusions. Then in the 1990s sprites were captured on video. They are still not well understood but are believed to be produced when especially powerful lightning occurs between the ground and storm clouds, particularly when the lightning transfers a huge amount of negative charge $-q$ from the ground to the base of the clouds (Fig. 22-10b).

Just after such a transfer, the ground has a complicated distribution of positive charge. However, we can model the electric field due to the charges in the clouds and the ground by assuming a vertical electric dipole that has charge $-q$ at cloud height h and charge $+q$ at below-ground depth h (Fig. 22-10c). If $q = 200 \text{ C}$ and $h = 6.0 \text{ km}$, what is the magnitude of the dipole's electric field at altitude $z_1 = 30 \text{ km}$ somewhat above the clouds and altitude $z_2 = 60 \text{ km}$ somewhat above the stratosphere?



(a) Courtesy NASA

KEY IDEA

We can approximate the magnitude E of an electric dipole's electric field on the dipole axis with Eq. 22-8.

Calculations: We write that equation as

$$E = \frac{1}{2\pi\epsilon_0} \frac{q(2h)}{z^3},$$

where $2h$ is the separation between $-q$ and $+q$ in Fig. 22-10c. For the electric field at altitude $z_1 = 30 \text{ km}$, we find

$$E = \frac{1}{2\pi\epsilon_0} \frac{(200 \text{ C})(2)(6.0 \times 10^3 \text{ m})}{(30 \times 10^3 \text{ m})^3} \\ = 1.6 \times 10^3 \text{ N/C.} \quad (\text{Answer})$$

Similarly, for altitude $z_2 = 60 \text{ km}$, we find

$$E = 2.0 \times 10^2 \text{ N/C.} \quad (\text{Answer})$$

As we discuss in Module 22-6, when the magnitude of

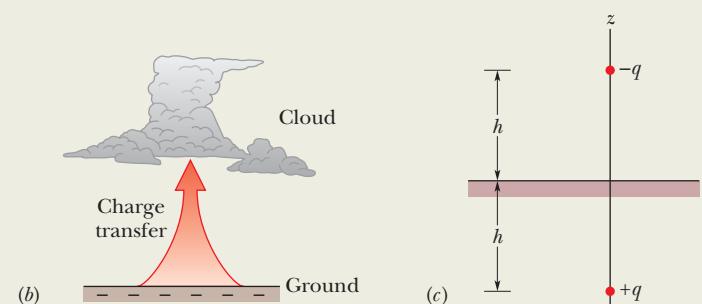


Figure 22-10 (a) Photograph of a sprite. (b) Lightning in which a large amount of negative charge is transferred from ground to cloud base. (c) The cloud–ground system modeled as a vertical electric dipole.

an electric field exceeds a certain critical value E_c , the field can pull electrons out of atoms (ionize the atoms), and then the freed electrons can run into other atoms, causing those atoms to emit light. The value of E_c depends on the density of the air in which the electric field exists. At altitude $z_2 = 60$ km the density of the air is so low that

$E = 2.0 \times 10^2$ N/C exceeds E_c , and thus light is emitted by the atoms in the air. That light forms sprites. Lower down, just above the clouds at $z_1 = 30$ km, the density of the air is much higher, $E = 1.6 \times 10^3$ N/C does not exceed E_c , and no light is emitted. Hence, sprites occur only far above storm clouds.



Additional examples, video, and practice available at WileyPLUS

22-4 THE ELECTRIC FIELD DUE TO A LINE OF CHARGE

Learning Objectives

After reading this module, you should be able to . . .

22.16 For a uniform distribution of charge, find the linear charge density λ for charge along a line, the surface charge density σ for charge on a surface, and the volume charge density ρ for charge in a volume.

22.17 For charge that is distributed uniformly along a line, find the net electric field at a given point near the line by

splitting the distribution up into charge elements dq and then summing (by integration) the electric field vectors $d\vec{E}$ set up at the point by each element.

22.18 Explain how symmetry can be used to simplify the calculation of the electric field at a point near a line of uniformly distributed charge.

Key Ideas

- The equation for the electric field set up by a particle does not apply to an extended object with charge (said to have a continuous charge distribution).
- To find the electric field of an extended object at a point, we first consider the electric field set up by a charge element dq in the object, where the element is small enough for us to apply

the equation for a particle. Then we sum, via integration, components of the electric fields $d\vec{E}$ from all the charge elements.

- Because the individual electric fields $d\vec{E}$ have different magnitudes and point in different directions, we first see if symmetry allows us to cancel out any of the components of the fields, to simplify the integration.

The Electric Field Due to a Line of Charge

So far we have dealt with only charged particles, a single particle or a simple collection of them. We now turn to a much more challenging situation in which a thin (approximately one-dimensional) object such as a rod or ring is charged with a huge number of particles, more than we could ever even count. In the next module, we consider two-dimensional objects, such as a disk with charge spread over a surface. In the next chapter we tackle three-dimensional objects, such as a sphere with charge spread through a volume.

Heads Up. Many students consider this module to be the most difficult in the book for a variety of reasons. There are lots of steps to take, a lot of vector features to keep track of, and after all that, we set up and then solve an integral. The worst part, however, is that the procedure can be different for different arrangements of the charge. Here, as we focus on a particular arrangement (a charged ring), be aware of the general approach, so that you can tackle other arrangements in the homework (such as rods and partial circles).

Figure 22-11 shows a thin ring of radius R with a uniform distribution of positive charge along its circumference. It is made of plastic, which means that the charge is fixed in place. The ring is surrounded by a pattern of electric field lines, but here we restrict our interest to an arbitrary point P on the central axis (the axis through the ring's center and perpendicular to the plane of the ring), at distance z from the center point.

The charge of an extended object is often conveyed in terms of a charge density rather than the total charge. For a line of charge, we use the *linear charge*

density λ (the charge per unit length), with the SI unit of coulomb per meter. Table 22-1 shows the other charge densities that we shall be using for charged surfaces and volumes.

First Big Problem. So far, we have an equation for the electric field of a particle. (We can combine the field of several particles as we did for the electric dipole to generate a special equation, but we are still basically using Eq. 22-3). Now take a look at the ring in Fig. 22-11. That clearly is not a particle and so Eq. 22-3 does not apply. So what do we do?

The answer is to mentally divide the ring into differential elements of charge that are so small that we can treat them as though they *are* particles. Then we *can* apply Eq. 22-3.

Second Big Problem. We now know to apply Eq. 22-3 to each charge element dq (the front d emphasizes that the charge is very small) and can write an expression for its contribution of electric field $d\vec{E}$ (the front d emphasizes that the contribution is very small). However, each such contributed field vector at P is in its own direction. How can we add them to get the net field at P ?

The answer is to split the vectors into components and then separately sum one set of components and then the other set. However, first we check to see if one set simply all cancels out. (Canceling out components saves lots of work.)

Third Big Problem. There is a huge number of dq elements in the ring and thus a huge number of $d\vec{E}$ components to add up, even if we can cancel out one set of components. How can we add up more components than we could even count? The answer is to add them by means of integration.

Do It. Let's do all this (but again, be aware of the general procedure, not just the fine details). We arbitrarily pick the charge element shown in Fig. 22-11. Let ds be the arc length of that (or any other) dq element. Then in terms of the linear density λ (the charge per unit length), we have

$$dq = \lambda ds. \quad (22-10)$$

An Element's Field. This charge element sets up the differential electric field $d\vec{E}$ at P , at distance r from the element, as shown in Fig. 22-11. (Yes, we are introducing a new symbol that is not given in the problem statement, but soon we shall replace it with "legal symbols.") Next we rewrite the field equation for a particle (Eq. 22-3) in terms of our new symbols dE and dq , but then we replace dq using Eq. 22-10. The field magnitude due to the charge element is

$$dE = \frac{1}{4\pi\epsilon_0} \frac{dq}{r^2} = \frac{1}{4\pi\epsilon_0} \frac{\lambda ds}{r^2}. \quad (22-11)$$

Notice that the illegal symbol r is the hypotenuse of the right triangle displayed in Fig. 22-11. Thus, we can replace r by rewriting Eq. 22-11 as

$$dE = \frac{1}{4\pi\epsilon_0} \frac{\lambda ds}{(z^2 + R^2)}. \quad (22-12)$$

Because every charge element has the same charge and the same distance from point P , Eq. 22-12 gives the field magnitude contributed by each of them. Figure 22-11 also tells us that each contributed $d\vec{E}$ leans at angle θ to the central axis (the z axis) and thus has components perpendicular and parallel to that axis.

Cancelling Components. Now comes the neat part, where we eliminate one set of those components. In Fig. 22-11, consider the charge element on the opposite side of the ring. It too contributes the field magnitude dE but the field vector leans at angle θ in the opposite direction from the vector from our first charge

Table 22-1 Some Measures of Electric Charge

Name	Symbol	SI Unit
Charge	q	C
Linear charge density	λ	C/m
Surface charge density	σ	C/m ²
Volume charge density	ρ	C/m ³

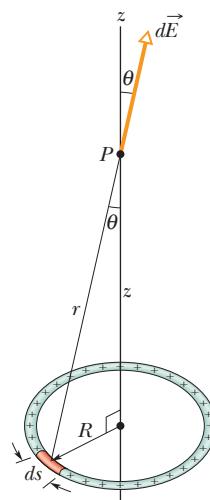


Figure 22-11 A ring of uniform positive charge. A differential element of charge occupies a length ds (greatly exaggerated for clarity). This element sets up an electric field $d\vec{E}$ at point P .

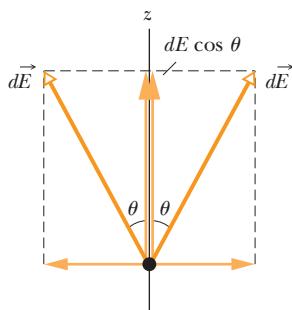


Figure 22-12 The electric fields set up at P by a charge element and its symmetric partner (on the opposite side of the ring). The components perpendicular to the z axis cancel; the parallel components add.

element, as indicated in the side view of Fig. 22-12. Thus the two perpendicular components cancel. All around the ring, this cancellation occurs for every charge element and its *symmetric partner* on the opposite side of the ring. So we can neglect all the perpendicular components.

Adding Components. We have another big win here. All the remaining components are in the positive direction of the z axis, so we can just add them up as scalars. Thus we can already tell the direction of the net electric field at P : directly away from the ring. From Fig. 22-12, we see that the parallel components each have magnitude $dE \cos \theta$, but θ is another illegal symbol. We can replace $\cos \theta$ with legal symbols by again using the right triangle in Fig. 22-11 to write

$$\cos \theta = \frac{z}{r} = \frac{z}{(z^2 + R^2)^{1/2}}. \quad (22-13)$$

Multiplying Eq. 22-12 by Eq. 22-13 gives us the parallel field component from each charge element:

$$dE \cos \theta = \frac{1}{4\pi\epsilon_0} \frac{z\lambda}{(z^2 + R^2)^{3/2}} ds. \quad (22-14)$$

Integrating. Because we must sum a huge number of these components, each small, we set up an integral that moves along the ring, from element to element, from a starting point (call it $s = 0$) through the full circumference ($s = 2\pi R$). Only the quantity s varies as we go through the elements; the other symbols in Eq. 22-14 remain the same, so we move them outside the integral. We find

$$\begin{aligned} E &= \int dE \cos \theta = \frac{z\lambda}{4\pi\epsilon_0(z^2 + R^2)^{3/2}} \int_0^{2\pi R} ds \\ &= \frac{z\lambda(2\pi R)}{4\pi\epsilon_0(z^2 + R^2)^{3/2}}. \end{aligned} \quad (22-15)$$

This is a fine answer, but we can also switch to the total charge by using $\lambda = q/(2\pi R)$:

$$E = \frac{qz}{4\pi\epsilon_0(z^2 + R^2)^{3/2}} \quad (\text{charged ring}). \quad (22-16)$$

If the charge on the ring is negative, instead of positive as we have assumed, the magnitude of the field at P is still given by Eq. 22-16. However, the electric field vector then points toward the ring instead of away from it.

Let us check Eq. 22-16 for a point on the central axis that is so far away that $z \gg R$. For such a point, the expression $z^2 + R^2$ in Eq. 22-16 can be approximated as z^2 , and Eq. 22-16 becomes

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{z^2} \quad (\text{charged ring at large distance}). \quad (22-17)$$

This is a reasonable result because from a large distance, the ring “looks like” a point charge. If we replace z with r in Eq. 22-17, we indeed do have the magnitude of the electric field due to a point charge, as given by Eq. 22-3.

Let us next check Eq. 22-16 for a point at the center of the ring—that is, for $z = 0$. At that point, Eq. 22-16 tells us that $E = 0$. This is a reasonable result because if we were to place a test charge at the center of the ring, there would be no net electrostatic force acting on it; the force due to any element of the ring would be canceled by the force due to the element on the opposite side of the ring. By Eq. 22-1, if the force at the center of the ring were zero, the electric field there would also have to be zero.

Sample Problem 22.03 Electric field of a charged circular rod

Figure 22-13a shows a plastic rod with a uniform charge $-Q$. It is bent in a 120° circular arc of radius r and symmetrically placed across an x axis with the origin at the center of curvature P of the rod. In terms of Q and r , what is the electric field \vec{E} due to the rod at point P ?

KEY IDEA

Because the rod has a continuous charge distribution, we must find an expression for the electric fields due to differential elements of the rod and then sum those fields via calculus.

An element: Consider a differential element having arc length ds and located at an angle θ above the x axis (Figs. 22-13b and c). If we let λ represent the linear charge density of the rod, our element ds has a differential charge of magnitude

$$dq = \lambda ds. \quad (22-18)$$

The element's field: Our element produces a differential electric field $d\vec{E}$ at point P , which is a distance r from the element. Treating the element as a point charge, we can

rewrite Eq. 22-3 to express the magnitude of $d\vec{E}$ as

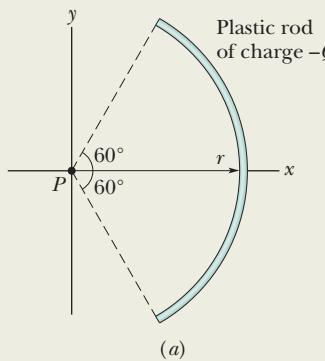
$$dE = \frac{1}{4\pi\epsilon_0} \frac{dq}{r^2} = \frac{1}{4\pi\epsilon_0} \frac{\lambda ds}{r^2}. \quad (22-19)$$

The direction of $d\vec{E}$ is toward ds because charge dq is negative.

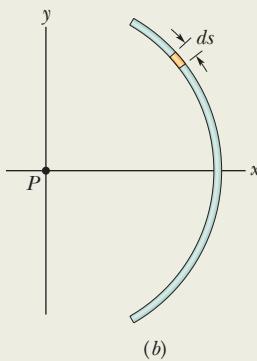
Symmetric partner: Our element has a symmetrically located (mirror image) element ds' in the bottom half of the rod. The electric field $d\vec{E}'$ set up at P by ds' also has the magnitude given by Eq. 22-19, but the field vector points toward ds' as shown in Fig. 22-13d. If we resolve the electric field vectors of ds and ds' into x and y components as shown in Figs. 22-13e and f, we see that their y components cancel (because they have equal magnitudes and are in opposite directions). We also see that their x components have equal magnitudes and are in the same direction.

Summing: Thus, to find the electric field set up by the rod, we need sum (via integration) only the x components of the differential electric fields set up by all the differential elements of the rod. From Fig. 22-13f and Eq. 22-19, we can write

This negatively charged rod is obviously not a particle.



But we can treat this element as a particle.



Here is the field the element creates.

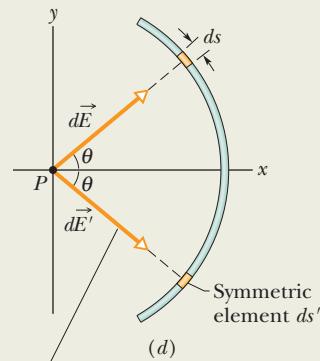
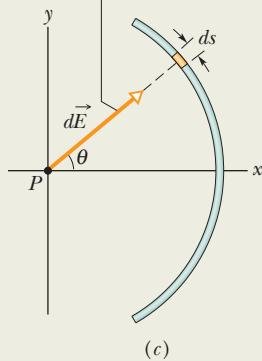
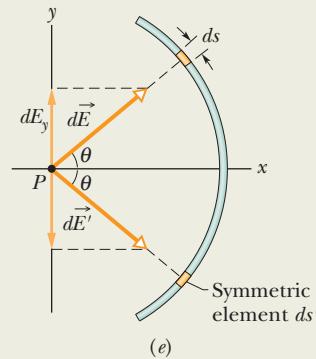
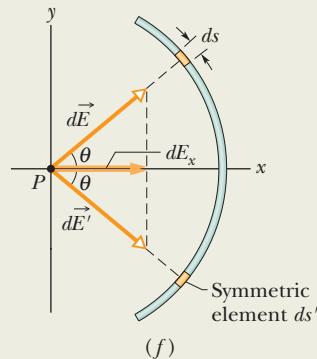


Figure 22-13 Available in WileyPLUS as an animation with voiceover. (a) A plastic rod of charge $-Q$ is a circular section of radius r and central angle 120° ; point P is the center of curvature of the rod. (b)–(c) A differential element in the top half of the rod, at an angle θ to the x axis and of arc length ds , sets up a differential electric field $d\vec{E}$ at P . (d) An element ds' , symmetric to ds about the x axis, sets up a field $d\vec{E}'$ at P with the same magnitude. (e)–(f) The field components. (g) Arc length ds makes an angle $d\theta$ about point P .

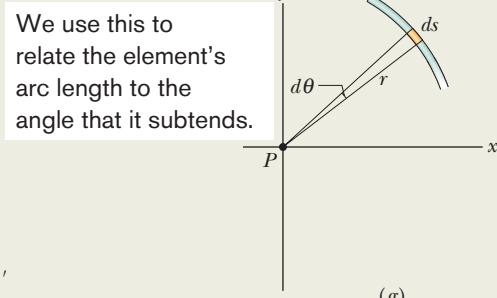
These y components just cancel, so neglect them.



These x components add. Our job is to add all such components.



Here is the field created by the symmetric element, same size and angle.



the component dE_x set up by ds as

$$dE_x = dE \cos \theta = \frac{1}{4\pi\epsilon_0} \frac{\lambda}{r^2} \cos \theta ds. \quad (22-20)$$

Equation 22-20 has two variables, θ and s . Before we can integrate it, we must eliminate one variable. We do so by replacing ds , using the relation

$$ds = r d\theta,$$

in which $d\theta$ is the angle at P that includes arc length ds (Fig. 22-13g). With this replacement, we can integrate Eq. 22-20 over the angle made by the rod at P , from $\theta = -60^\circ$ to $\theta = 60^\circ$; that will give us the field magnitude at P :

$$\begin{aligned} E &= \int dE_x = \int_{-60^\circ}^{60^\circ} \frac{1}{4\pi\epsilon_0} \frac{\lambda}{r^2} \cos \theta r d\theta \\ &= \frac{\lambda}{4\pi\epsilon_0 r} \int_{-60^\circ}^{60^\circ} \cos \theta d\theta = \frac{\lambda}{4\pi\epsilon_0 r} \left[\sin \theta \right]_{-60^\circ}^{60^\circ} \\ &= \frac{\lambda}{4\pi\epsilon_0 r} [\sin 60^\circ - \sin(-60^\circ)] \\ &= \frac{1.73\lambda}{4\pi\epsilon_0 r}. \end{aligned} \quad (22-21)$$

(If we had reversed the limits on the integration, we would have gotten the same result but with a minus sign. Since the integration gives only the magnitude of \vec{E} , we would then have discarded the minus sign.)

Charge density: To evaluate λ , we note that the full rod subtends an angle of 120° and so is one-third of a full circle. Its arc length is then $2\pi r/3$, and its linear charge density must be

$$\lambda = \frac{\text{charge}}{\text{length}} = \frac{Q}{2\pi r/3} = \frac{0.477Q}{r}.$$

Substituting this into Eq. 22-21 and simplifying give us

$$\begin{aligned} E &= \frac{(1.73)(0.477Q)}{4\pi\epsilon_0 r^2} \\ &= \frac{0.83Q}{4\pi\epsilon_0 r^2}. \end{aligned} \quad (\text{Answer})$$

The direction of \vec{E} is toward the rod, along the axis of symmetry of the charge distribution. We can write \vec{E} in unit-vector notation as

$$\vec{E} = \frac{0.83Q}{4\pi\epsilon_0 r^2} \hat{i}.$$

Problem-Solving Tactics A Field Guide for Lines of Charge

Here is a generic guide for finding the electric field \vec{E} produced at a point P by a line of uniform charge, either circular or straight. The general strategy is to pick out an element dq of the charge, find $d\vec{E}$ due to that element, and integrate $d\vec{E}$ over the entire line of charge.

Step 1. If the line of charge is circular, let ds be the arc length of an element of the distribution. If the line is straight, run an x axis along it and let dx be the length of an element. Mark the element on a sketch.

Step 2. Relate the charge dq of the element to the length of the element with either $dq = \lambda ds$ or $dq = \lambda dx$. Consider dq and λ to be positive, even if the charge is actually negative. (The sign of the charge is used in the next step.)

Step 3. Express the field $d\vec{E}$ produced at P by dq with Eq. 22-3, replacing q in that equation with either λds or λdx . If the charge on the line is positive, then at P draw a vector $d\vec{E}$ that points directly away from dq . If the charge is negative, draw the vector pointing directly toward dq .

Step 4. Always look for any symmetry in the situation. If P is on an axis of symmetry of the charge distribution, resolve the field $d\vec{E}$ produced by dq into components that are perpendicular and parallel to the axis of symmetry. Then consider a second element dq' that is located symmetrically to dq about the line of symmetry. At P draw the vector $d\vec{E}'$ that this symmetrical element pro-

duces and resolve it into components. One of the components produced by dq is a *cancelling component*; it is canceled by the corresponding component produced by dq' and needs no further attention. The other component produced by dq is an *adding component*; it adds to the corresponding component produced by dq' . Add the adding components of all the elements via integration.

Step 5. Here are four general types of uniform charge distributions, with strategies for the integral of step 4.

Ring, with point P on (central) axis of symmetry, as in Fig. 22-11. In the expression for dE , replace r^2 with $z^2 + R^2$, as in Eq. 22-12. Express the adding component of $d\vec{E}$ in terms of θ . That introduces $\cos \theta$, but θ is identical for all elements and thus is not a variable. Replace $\cos \theta$ as in Eq. 22-13. Integrate over s , around the circumference of the ring.

Circular arc, with point P at the center of curvature, as in Fig. 22-13. Express the adding component of $d\vec{E}$ in terms of θ . That introduces either $\sin \theta$ or $\cos \theta$. Reduce the resulting two variables s and θ to one, θ , by replacing ds with $r d\theta$. Integrate over θ from one end of the arc to the other end.

Straight line, with point P on an extension of the line, as in Fig. 22-14a. In the expression for dE , replace r with x . Integrate over x , from end to end of the line of charge.

Straight line, with point P at perpendicular distance y from the line of charge, as in Fig. 22-14b. In the expression for dE , replace r with an expression involving x and y . If P is on the perpendicular bisector of the line of charge, find an expression for the adding component of $d\vec{E}$. That will introduce either $\sin \theta$ or $\cos \theta$. Reduce the resulting two variables x and θ to one, x , by replacing the trigonometric function with an expression (its definition) involving x and y . Integrate over x from end to end of the line of charge. If P is not on a line of symmetry, as in Fig. 22-14c, set up an integral to sum the components dE_x , and integrate over x to find E_x . Also set up an integral to sum the components dE_y , and integrate over x again to find E_y . Use the components E_x and E_y in the usual way to find the magnitude E and the orientation of \vec{E} .

Step 6. One arrangement of the integration limits gives a positive result. The reverse gives the same result with a mi-

nus sign; discard the minus sign. If the result is to be stated in terms of the total charge Q of the distribution, replace λ with Q/L , in which L is the length of the distribution.

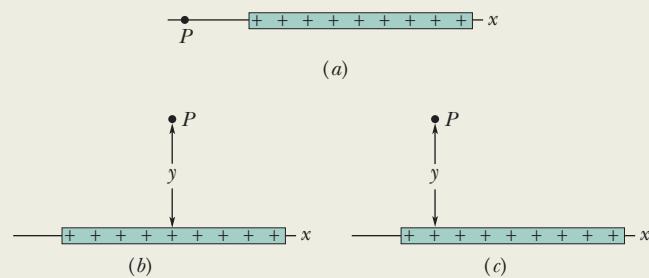


Figure 22-14 (a) Point P is on an extension of the line of charge. (b) P is on a line of symmetry of the line of charge, at perpendicular distance y from that line. (c) Same as (b) except that P is not on a line of symmetry.

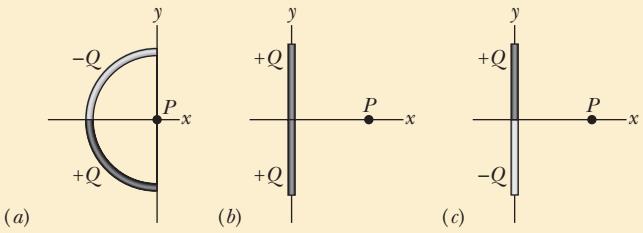


Additional examples, video, and practice available at WileyPLUS



Checkpoint 2

The figure here shows three nonconducting rods, one circular and two straight. Each has a uniform charge of magnitude Q along its top half and another along its bottom half. For each rod, what is the direction of the net electric field at point P ?



22-5 THE ELECTRIC FIELD DUE TO A CHARGED DISK

Learning Objectives

After reading this module, you should be able to . . .

22.19 Sketch a disk with uniform charge and indicate the direction of the electric field at a point on the central axis if the charge is positive and if it is negative.

22.20 Explain how the equation for the electric field on the central axis of a uniformly charged ring can be used to find

the equation for the electric field on the central axis of a uniformly charged disk.

22.21 For a point on the central axis of a uniformly charged disk, apply the relationship between the surface charge density σ , the disk radius R , and the distance z to that point.

Key Idea

On the central axis through a uniformly charged disk,

$$E = \frac{\sigma}{2\epsilon_0} \left(1 - \frac{z}{\sqrt{z^2 + R^2}} \right)$$

gives the electric field magnitude. Here z is the distance along the axis from the center of the disk, R is the radius of the disk, and σ is the surface charge density.

The Electric Field Due to a Charged Disk

Now we switch from a line of charge to a surface of charge by examining the electric field of a circular plastic disk, with a radius R and a uniform surface charge density σ (charge per unit area, Table 22-1) on its top surface. The disk sets up a

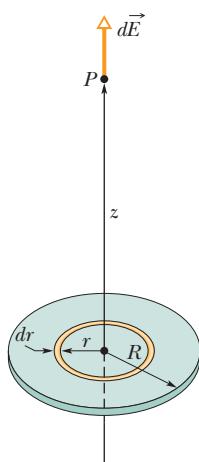


Figure 22-15 A disk of radius R and uniform positive charge. The ring shown has radius r and radial width dr . It sets up a differential electric field $d\vec{E}$ at point P on its central axis.

pattern of electric field lines around it, but here we restrict our attention to the electric field at an arbitrary point P on the central axis, at distance z from the center of the disk, as indicated in Fig. 22-15.

We could proceed as in the preceding module but set up a two-dimensional integral to include all of the field contributions from the two-dimensional distribution of charge on the top surface. However, we can save a lot of work with a neat shortcut using our earlier work with the field on the central axis of a thin ring.

We superimpose a ring on the disk as shown in Fig. 22-15, at an arbitrary radius $r \leq R$. The ring is so thin that we can treat the charge on it as a charge element dq . To find its small contribution dE to the electric field at point P , we rewrite Eq. 22-16 in terms of the ring's charge dq and radius r :

$$dE = \frac{dq}{4\pi\epsilon_0(z^2 + r^2)^{3/2}}. \quad (22-22)$$

The ring's field points in the positive direction of the z axis.

To find the total field at P , we are going to integrate Eq. 22-22 from the center of the disk at $r = 0$ out to the rim at $r = R$ so that we sum all the dE contributions (by sweeping our arbitrary ring over the entire disk surface). However, that means we want to integrate with respect to a variable radius r of the ring.

We get dr into the expression by substituting for dq in Eq. 22-22. Because the ring is so thin, call its thickness dr . Then its surface area dA is the product of its circumference $2\pi r$ and thickness dr . So, in terms of the surface charge density σ , we have

$$dq = \sigma dA = \sigma(2\pi r dr). \quad (22-23)$$

After substituting this into Eq. 22-22 and simplifying slightly, we can sum all the dE contributions with

$$E = \int dE = \frac{\sigma z}{4\epsilon_0} \int_0^R (z^2 + r^2)^{-3/2} (2r) dr, \quad (22-24)$$

where we have pulled the constants (including z) out of the integral. To solve this integral, we cast it in the form $\int X^m dX$ by setting $X = (z^2 + r^2)$, $m = -\frac{3}{2}$, and $dX = (2r) dr$. For the recast integral we have

$$\int X^m dX = \frac{X^{m+1}}{m+1},$$

and so Eq. 22-24 becomes

$$E = \frac{\sigma z}{4\epsilon_0} \left[\frac{(z^2 + r^2)^{-1/2}}{-\frac{1}{2}} \right]_0^R. \quad (22-25)$$

Taking the limits in Eq. 22-25 and rearranging, we find

$$E = \frac{\sigma}{2\epsilon_0} \left(1 - \frac{z}{\sqrt{z^2 + R^2}} \right) \quad (\text{charged disk}) \quad (22-26)$$

as the magnitude of the electric field produced by a flat, circular, charged disk at points on its central axis. (In carrying out the integration, we assumed that $z \geq 0$.)

If we let $R \rightarrow \infty$ while keeping z finite, the second term in the parentheses in Eq. 22-26 approaches zero, and this equation reduces to

$$E = \frac{\sigma}{2\epsilon_0} \quad (\text{infinite sheet}). \quad (22-27)$$

This is the electric field produced by an infinite sheet of uniform charge located on one side of a nonconductor such as plastic. The electric field lines for such a situation are shown in Fig. 22-4.

We also get Eq. 22-27 if we let $z \rightarrow 0$ in Eq. 22-26 while keeping R finite. This shows that at points very close to the disk, the electric field set up by the disk is the same as if the disk were infinite in extent.

22-6 A POINT CHARGE IN AN ELECTRIC FIELD

Learning Objectives

After reading this module, you should be able to . . .

22.22 For a charged particle placed in an external electric field (a field due to other charged objects), apply the relationship between the electric field \vec{E} at that point, the particle's charge q , and the electrostatic force \vec{F} that acts on the particle, and identify the relative directions of the force

and the field when the particle is positively charged and negatively charged.

22.23 Explain Millikan's procedure of measuring the elementary charge.

22.24 Explain the general mechanism of ink-jet printing.

Key Ideas

- If a particle with charge q is placed in an external electric field \vec{E} , an electrostatic force \vec{F} acts on the particle:

$$\vec{F} = q\vec{E}.$$

- If charge q is positive, the force vector is in the same direction as the field vector. If charge q is negative, the force vector is in the opposite direction (the minus sign in the equation reverses the force vector from the field vector).

A Point Charge in an Electric Field

In the preceding four modules we worked at the first of our two tasks: given a charge distribution, to find the electric field it produces in the surrounding space. Here we begin the second task: to determine what happens to a charged particle when it is in an electric field set up by other stationary or slowly moving charges.

What happens is that an electrostatic force acts on the particle, as given by

$$\vec{F} = q\vec{E}, \quad (22-28)$$

in which q is the charge of the particle (including its sign) and \vec{E} is the electric field that other charges have produced at the location of the particle. (The field is *not* the field set up by the particle itself; to distinguish the two fields, the field acting on the particle in Eq. 22-28 is often called the *external field*. A charged particle or object is not affected by its own electric field.) Equation 22-28 tells us



The electrostatic force \vec{F} acting on a charged particle located in an external electric field \vec{E} has the direction of \vec{E} if the charge q of the particle is positive and has the opposite direction if q is negative.

Measuring the Elementary Charge

Equation 22-28 played a role in the measurement of the elementary charge e by American physicist Robert A. Millikan in 1910–1913. Figure 22-16 is a representation of his apparatus. When tiny oil drops are sprayed into chamber A, some of them become charged, either positively or negatively, in the process. Consider a drop that drifts downward through the small hole in plate P_1 and into chamber C. Let us assume that this drop has a negative charge q .

If switch S in Fig. 22-16 is open as shown, battery B has no electrical effect on chamber C. If the switch is closed (the connection between chamber C and the positive terminal of the battery is then complete), the battery causes an excess positive charge on conducting plate P_1 and an excess negative charge on conducting plate P_2 . The charged plates set up a downward-directed electric field \vec{E} in chamber C. According to Eq. 22-28, this field exerts an electrostatic force on any charged drop that happens to be in the chamber and affects its motion. In particular, our negatively charged drop will tend to drift upward.

By timing the motion of oil drops with the switch opened and with it closed and thus determining the effect of the charge q , Millikan discovered that the

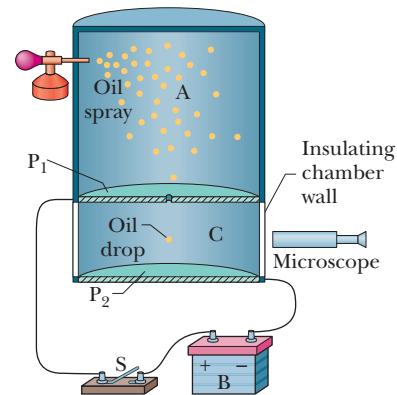


Figure 22-16 The Millikan oil-drop apparatus for measuring the elementary charge e . When a charged oil drop drifted into chamber C through the hole in plate P_1 , its motion could be controlled by closing and opening switch S and thereby setting up or eliminating an electric field in chamber C. The microscope was used to view the drop, to permit timing of its motion.

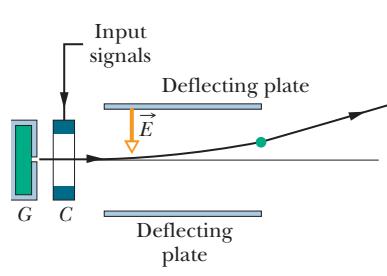


Figure 22-17 Ink-jet printer. Drops shot from generator G receive a charge in charging unit C . An input signal from a computer controls the charge and thus the effect of field \vec{E} on where the drop lands on the paper.

values of q were always given by

$$q = ne, \quad \text{for } n = 0, \pm 1, \pm 2, \pm 3, \dots, \quad (22-29)$$

in which e turned out to be the fundamental constant we call the *elementary charge*, $1.60 \times 10^{-19} \text{ C}$. Millikan's experiment is convincing proof that charge is quantized, and he earned the 1923 Nobel Prize in physics in part for this work. Modern measurements of the elementary charge rely on a variety of interlocking experiments, all more precise than the pioneering experiment of Millikan.

Ink-Jet Printing

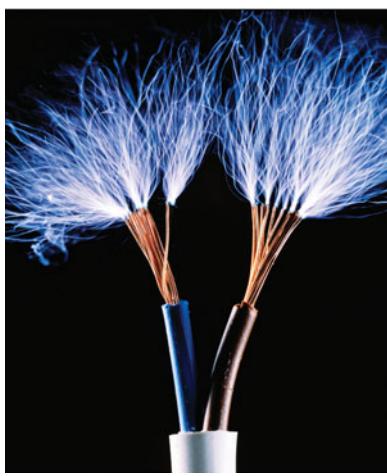
The need for high-quality, high-speed printing has caused a search for an alternative to impact printing, such as occurs in a standard typewriter. Building up letters by squirting tiny drops of ink at the paper is one such alternative.

Figure 22-17 shows a negatively charged drop moving between two conducting deflecting plates, between which a uniform, downward-directed electric field \vec{E} has been set up. The drop is deflected upward according to Eq. 22-28 and then strikes the paper at a position that is determined by the magnitudes of \vec{E} and the charge q of the drop.

In practice, E is held constant and the position of the drop is determined by the charge q delivered to the drop in the charging unit, through which the drop must pass before entering the deflecting system. The charging unit, in turn, is activated by electronic signals that encode the material to be printed.

Electrical Breakdown and Sparking

If the magnitude of an electric field in air exceeds a certain critical value E_c , the air undergoes *electrical breakdown*, a process whereby the field removes electrons from the atoms in the air. The air then begins to conduct electric current because the freed electrons are propelled into motion by the field. As they move, they collide with any atoms in their path, causing those atoms to emit light. We can see the paths, commonly called sparks, taken by the freed electrons because of that emitted light. Figure 22-18 shows sparks above charged metal wires where the electric fields due to the wires cause electrical breakdown of the air.



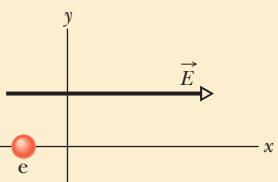
Adam Hart-Davis/Photo Researchers, Inc.

Figure 22-18 The metal wires are so charged that the electric fields they produce in the surrounding space cause the air there to undergo electrical breakdown.



Checkpoint 3

- (a) In the figure, what is the direction of the electrostatic force on the electron due to the external electric field shown? (b) In which direction will the electron accelerate if it is moving parallel to the y axis before it encounters the external field? (c) If, instead, the electron is initially moving rightward, will its speed increase, decrease, or remain constant?



Sample Problem 22.04**Motion of a charged particle in an electric field**

Figure 22-19 shows the deflecting plates of an ink-jet printer, with superimposed coordinate axes. An ink drop with a mass m of 1.3×10^{-10} kg and a negative charge of magnitude $Q = 1.5 \times 10^{-13}$ C enters the region between the plates, initially moving along the x axis with speed $v_x = 18$ m/s. The length L of each plate is 1.6 cm. The plates are charged and thus produce an electric field at all points between them. Assume that field \vec{E} is downward directed, is uniform, and has a magnitude of 1.4×10^6 N/C. What is the vertical deflection of the drop at the far edge of the plates? (The gravitational force on the drop is small relative to the electrostatic force acting on the drop and can be neglected.)

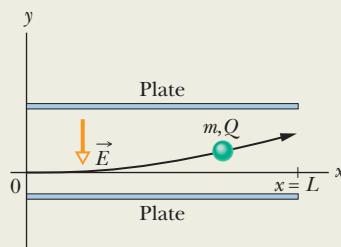


Figure 22-19 An ink drop of mass m and charge magnitude Q is deflected in the electric field of an ink-jet printer.

KEY IDEA

The drop is negatively charged and the electric field is directed *downward*. From Eq. 22-28, a constant electrostatic force of

magnitude QE acts *upward* on the charged drop. Thus, as the drop travels parallel to the x axis at constant speed v_x , it accelerates upward with some constant acceleration a_y .

Calculations: Applying Newton's second law ($F = ma$) for components along the y axis, we find that

$$a_y = \frac{F}{m} = \frac{QE}{m}. \quad (22-30)$$

Let t represent the time required for the drop to pass through the region between the plates. During t the vertical and horizontal displacements of the drop are

$$y = \frac{1}{2}a_y t^2 \quad \text{and} \quad L = v_x t, \quad (22-31)$$

respectively. Eliminating t between these two equations and substituting Eq. 22-30 for a_y , we find

$$\begin{aligned} y &= \frac{QEL^2}{2mv_x^2} \\ &= \frac{(1.5 \times 10^{-13} \text{ C})(1.4 \times 10^6 \text{ N/C})(1.6 \times 10^{-2} \text{ m})^2}{(2)(1.3 \times 10^{-10} \text{ kg})(18 \text{ m/s})^2} \\ &= 6.4 \times 10^{-4} \text{ m} \\ &= 0.64 \text{ mm.} \end{aligned} \quad (\text{Answer})$$



Additional examples, video, and practice available at WileyPLUS

22-7 A DIPOLE IN AN ELECTRIC FIELD

Learning Objectives

After reading this module, you should be able to . . .

22.25 On a sketch of an electric dipole in an external electric field, indicate the direction of the field, the direction of the dipole moment, the direction of the electrostatic forces on the two ends of the dipole, and the direction in which those forces tend to rotate the dipole, and identify the value of the net force on the dipole.

22.26 Calculate the torque on an electric dipole in an external electric field by evaluating a cross product of the dipole moment vector and the electric field vector, in magnitude-angle notation and unit-vector notation.

Key Ideas

- The torque on an electric dipole of dipole moment \vec{p} when placed in an external electric field \vec{E} is given by a cross product:

$$\vec{\tau} = \vec{p} \times \vec{E}.$$

- A potential energy U is associated with the orientation of the dipole moment in the field, as given by a dot product:

$$U = -\vec{p} \cdot \vec{E}.$$

22.27 For an electric dipole in an external electric field, relate the potential energy of the dipole to the work done by a torque as the dipole rotates in the electric field.

22.28 For an electric dipole in an external electric field, calculate the potential energy by taking a dot product of the dipole moment vector and the electric field vector, in magnitude-angle notation and unit-vector notation.

22.29 For an electric dipole in an external electric field, identify the angles for the minimum and maximum potential energies and the angles for the minimum and maximum torque magnitudes.

- If the dipole orientation changes, the work done by the electric field is

$$W = -\Delta U.$$

If the change in orientation is due to an external agent, the work done by the agent is $W_a = -W$.

A Dipole in an Electric Field

We have defined the electric dipole moment \vec{p} of an electric dipole to be a vector that points from the negative to the positive end of the dipole. As you will see, the behavior of a dipole in a uniform external electric field \vec{E} can be described completely in terms of the two vectors \vec{E} and \vec{p} , with no need of any details about the dipole's structure.

A molecule of water (H_2O) is an electric dipole; Fig. 22-20 shows why. There the black dots represent the oxygen nucleus (having eight protons) and the two hydrogen nuclei (having one proton each). The colored enclosed areas represent the regions in which electrons can be located around the nuclei.

In a water molecule, the two hydrogen atoms and the oxygen atom do not lie on a straight line but form an angle of about 105° , as shown in Fig. 22-20. As a result, the molecule has a definite "oxygen side" and "hydrogen side." Moreover, the 10 electrons of the molecule tend to remain closer to the oxygen nucleus than to the hydrogen nuclei. This makes the oxygen side of the molecule slightly more negative than the hydrogen side and creates an electric dipole moment \vec{p} that points along the symmetry axis of the molecule as shown. If the water molecule is placed in an external electric field, it behaves as would be expected of the more abstract electric dipole of Fig. 22-9.

To examine this behavior, we now consider such an abstract dipole in a uniform external electric field \vec{E} , as shown in Fig. 22-21a. We assume that the dipole is a rigid structure that consists of two centers of opposite charge, each of magnitude q , separated by a distance d . The dipole moment \vec{p} makes an angle θ with field \vec{E} .

Electrostatic forces act on the charged ends of the dipole. Because the electric field is uniform, those forces act in opposite directions (as shown in Fig. 22-21a) and with the same magnitude $F = qE$. Thus, *because the field is uniform*, the net force on the dipole from the field is zero and the center of mass of the dipole does not move. However, the forces on the charged ends do produce a net torque $\vec{\tau}$ on the dipole about its center of mass. The center of mass lies on the line connecting the charged ends, at some distance x from one end and thus a distance $d - x$ from the other end. From Eq. 10-39 ($\tau = rF \sin \phi$), we can write the magnitude of the net torque $\vec{\tau}$ as

$$\tau = Fx \sin \theta + F(d - x) \sin \theta = Fd \sin \theta. \quad (22-32)$$

We can also write the magnitude of $\vec{\tau}$ in terms of the magnitudes of the electric field E and the dipole moment $p = qd$. To do so, we substitute qE for F and p/q for d in Eq. 22-32, finding that the magnitude of $\vec{\tau}$ is

$$\tau = pE \sin \theta. \quad (22-33)$$

We can generalize this equation to vector form as

$$\vec{\tau} = \vec{p} \times \vec{E} \quad (\text{torque on a dipole}). \quad (22-34)$$

Vectors \vec{p} and \vec{E} are shown in Fig. 22-21b. The torque acting on a dipole tends to rotate \vec{p} (hence the dipole) into the direction of field \vec{E} , thereby reducing θ . In Fig. 22-21, such rotation is clockwise. As we discussed in Chapter 10, we can represent a torque that gives rise to a clockwise rotation by including a minus sign with the magnitude of the torque. With that notation, the torque of Fig. 22-21 is

$$\tau = -pE \sin \theta. \quad (22-35)$$

Potential Energy of an Electric Dipole

Potential energy can be associated with the orientation of an electric dipole in an electric field. The dipole has its least potential energy when it is in its equilibrium orientation, which is when its moment \vec{p} is lined up with the field \vec{E} (then $\vec{\tau} = \vec{p} \times \vec{E} = 0$). It has greater potential energy in all other orientations. Thus the dipole is like a pendulum, which has its least gravitational potential

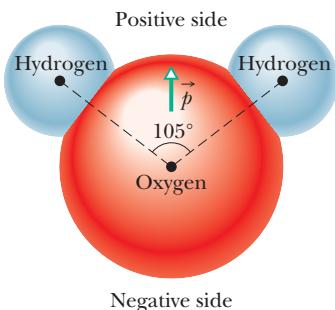


Figure 22-20 A molecule of H_2O , showing the three nuclei (represented by dots) and the regions in which the electrons can be located. The electric dipole moment \vec{p} points from the (negative) oxygen side to the (positive) hydrogen side of the molecule.

energy in *its* equilibrium orientation—at its lowest point. To rotate the dipole or the pendulum to any other orientation requires work by some external agent.

In any situation involving potential energy, we are free to define the zero-potential-energy configuration in an arbitrary way because only differences in potential energy have physical meaning. The expression for the potential energy of an electric dipole in an external electric field is simplest if we choose the potential energy to be zero when the angle θ in Fig. 22-21 is 90° . We then can find the potential energy U of the dipole at any other value of θ with Eq. 8-1 ($\Delta U = -W$) by calculating the work W done by the field on the dipole when the dipole is rotated to that value of θ from 90° . With the aid of Eq. 10-53 ($W = \int \tau d\theta$) and Eq. 22-35, we find that the potential energy U at any angle θ is

$$U = -W = - \int_{90^\circ}^{\theta} \tau d\theta = \int_{90^\circ}^{\theta} pE \sin \theta d\theta. \quad (22-36)$$

Evaluating the integral leads to

$$U = -pE \cos \theta. \quad (22-37)$$

We can generalize this equation to vector form as

$$U = -\vec{p} \cdot \vec{E} \quad (\text{potential energy of a dipole}). \quad (22-38)$$

Equations 22-37 and 22-38 show us that the potential energy of the dipole is least ($U = -pE$) when $\theta = 0$ (\vec{p} and \vec{E} are in the same direction); the potential energy is greatest ($U = pE$) when $\theta = 180^\circ$ (\vec{p} and \vec{E} are in opposite directions).

When a dipole rotates from an initial orientation θ_i to another orientation θ_f , the work W done on the dipole by the electric field is

$$W = -\Delta U = -(U_f - U_i), \quad (22-39)$$

where U_f and U_i are calculated with Eq. 22-38. If the change in orientation is caused by an applied torque (commonly said to be due to an external agent), then the work W_a done on the dipole by the applied torque is the negative of the work done on the dipole by the field; that is,

$$W_a = -W = (U_f - U_i). \quad (22-40)$$

Microwave Cooking

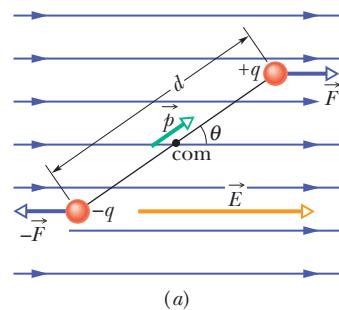
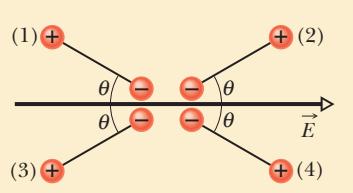
Food can be warmed and cooked in a microwave oven if the food contains water because water molecules are electric dipoles. When you turn on the oven, the microwave source sets up a rapidly oscillating electric field \vec{E} within the oven and thus also within the food. From Eq. 22-34, we see that any electric field \vec{E} produces a torque on an electric dipole moment \vec{p} to align \vec{p} with \vec{E} . Because the oven's \vec{E} oscillates, the water molecules continuously flip-flop in a frustrated attempt to align with \vec{E} .

Energy is transferred from the electric field to the thermal energy of the water (and thus of the food) where three water molecules happened to have bonded together to form a group. The flip-flop breaks some of the bonds. When the molecules reform the bonds, energy is transferred to the random motion of the group and then to the surrounding molecules. Soon, the thermal energy of the water is enough to cook the food.



Checkpoint 4

The figure shows four orientations of an electric dipole in an external electric field. Rank the orientations according to (a) the magnitude of the torque on the dipole and (b) the potential energy of the dipole, greatest first.



The dipole is being torqued into alignment.

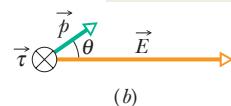


Figure 22-21 (a) An electric dipole in a uniform external electric field \vec{E} . Two centers of equal but opposite charge are separated by distance d . The line between them represents their rigid connection. (b) Field \vec{E} causes a torque $\vec{\tau}$ on the dipole. The direction of $\vec{\tau}$ is into the page, as represented by the symbol \otimes .

Sample Problem 22.05 Torque and energy of an electric dipole in an electric field

A neutral water molecule (H_2O) in its vapor state has an electric dipole moment of magnitude $6.2 \times 10^{-30} \text{ C}\cdot\text{m}$.

- (a) How far apart are the molecule's centers of positive and negative charge?

KEY IDEA

A molecule's dipole moment depends on the magnitude q of the molecule's positive or negative charge and the charge separation d .

Calculations: There are 10 electrons and 10 protons in a neutral water molecule; so the magnitude of its dipole moment is

$$p = qd = (10e)(d),$$

in which d is the separation we are seeking and e is the elementary charge. Thus,

$$\begin{aligned} d &= \frac{p}{10e} = \frac{6.2 \times 10^{-30} \text{ C}\cdot\text{m}}{(10)(1.60 \times 10^{-19} \text{ C})} \\ &= 3.9 \times 10^{-12} \text{ m} = 3.9 \text{ pm}. \end{aligned} \quad (\text{Answer})$$

This distance is not only small, but it is also actually smaller than the radius of a hydrogen atom.

- (b) If the molecule is placed in an electric field of $1.5 \times 10^4 \text{ N/C}$, what maximum torque can the field exert on it? (Such a field can easily be set up in the laboratory.)



Additional examples, video, and practice available at WileyPLUS

Review & Summary

Electric Field To explain the electrostatic force between two charges, we assume that each charge sets up an electric field in the space around it. The force acting on each charge is then due to the electric field set up at its location by the other charge.

Definition of Electric Field The *electric field* \vec{E} at any point is defined in terms of the electrostatic force \vec{F} that would be exerted on a positive test charge q_0 placed there:

$$\vec{E} = \frac{\vec{F}}{q_0}. \quad (22-1)$$

Electric Field Lines *Electric field lines* provide a means for visualizing the direction and magnitude of electric fields. The electric field vector at any point is tangent to a field line through that point. The density of field lines in any region is proportional to the magnitude of the electric field in that region. Field lines originate on positive charges and terminate on negative charges.

Field Due to a Point Charge The magnitude of the electric field \vec{E} set up by a point charge q at a distance r from the charge is

$$E = \frac{1}{4\pi\epsilon_0} \frac{|q|}{r^2}. \quad (22-3)$$

KEY IDEA

The torque on a dipole is maximum when the angle θ between \vec{p} and \vec{E} is 90° .

Calculation: Substituting $\theta = 90^\circ$ in Eq. 22-33 yields

$$\begin{aligned} \tau &= pE \sin \theta \\ &= (6.2 \times 10^{-30} \text{ C}\cdot\text{m})(1.5 \times 10^4 \text{ N/C})(\sin 90^\circ) \\ &= 9.3 \times 10^{-26} \text{ N}\cdot\text{m}. \end{aligned} \quad (\text{Answer})$$

- (c) How much work must an *external agent* do to rotate this molecule by 180° in this field, starting from its fully aligned position, for which $\theta = 0^\circ$?

KEY IDEA

The work done by an external agent (by means of a torque applied to the molecule) is equal to the change in the molecule's potential energy due to the change in orientation.

Calculation: From Eq. 22-40, we find

$$\begin{aligned} W_a &= U_{180^\circ} - U_0 \\ &= (-pE \cos 180^\circ) - (-pE \cos 0^\circ) \\ &= 2pE = (2)(6.2 \times 10^{-30} \text{ C}\cdot\text{m})(1.5 \times 10^4 \text{ N/C}) \\ &= 1.9 \times 10^{-25} \text{ J}. \end{aligned} \quad (\text{Answer})$$

The direction of \vec{E} is away from the point charge if the charge is positive and toward it if the charge is negative.

Field Due to an Electric Dipole An *electric dipole* consists of two particles with charges of equal magnitude q but opposite sign, separated by a small distance d . Their **electric dipole moment** \vec{p} has magnitude qd and points from the negative charge to the positive charge. The magnitude of the electric field set up by the dipole at a distant point on the dipole axis (which runs through both charges) is

$$E = \frac{1}{2\pi\epsilon_0} \frac{p}{z^3}, \quad (22-9)$$

where z is the distance between the point and the center of the dipole.

Field Due to a Continuous Charge Distribution The electric field due to a *continuous charge distribution* is found by treating charge elements as point charges and then summing, via integration, the electric field vectors produced by all the charge elements to find the net vector.

Field Due to a Charged Disk The electric field magnitude at a point on the central axis through a uniformly charged disk is given by

$$E = \frac{\sigma}{2\epsilon_0} \left(1 - \frac{z}{\sqrt{z^2 + R^2}} \right), \quad (22-26)$$

where z is the distance along the axis from the center of the disk, R is the radius of the disk, and σ is the surface charge density.

Force on a Point Charge in an Electric Field When a point charge q is placed in an external electric field \vec{E} , the electrostatic force \vec{F} that acts on the point charge is

$$\vec{F} = q\vec{E}. \quad (22-28)$$

Questions

- 1 Figure 22-22 shows three arrangements of electric field lines. In each arrangement, a proton is released from rest at point A and is then accelerated through point B by the electric field. Points A and B have equal separations in the three arrangements. Rank the arrangements according to the linear momentum of the proton at point B , greatest first.

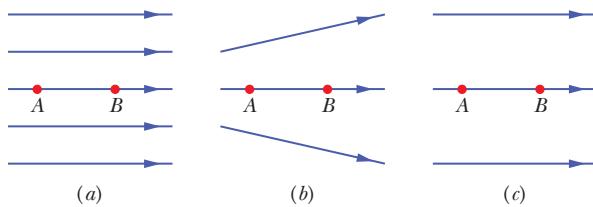


Figure 22-22 Question 1.

- 2 Figure 22-23 shows two square arrays of charged particles. The squares, which are centered on point P , are misaligned. The particles are separated by either d or $d/2$ along the perimeters of the squares. What are the magnitude and direction of the net electric field at P ?

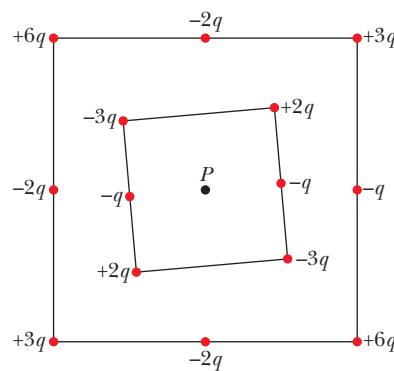


Figure 22-23 Question 2.

- 3 In Fig. 22-24, two particles of charge $-q$ are arranged symmetrically about the y axis; each produces an electric field at point P on that axis. (a) Are the magnitudes of the fields at P equal? (b) Is each electric field directed toward or away from the charge producing it? (c) Is the magnitude of the net electric field at P equal to the sum of the magnitudes E of the two field vectors (is it equal to $2E$)? (d) Do the x components of those two field vectors add or cancel? (e) Do their y components add or cancel? (f) Is the direction of the net field at P that of the canceling components or the adding components? (g) What is the direction of the net field?

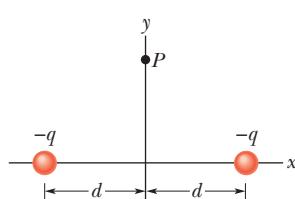


Figure 22-24 Question 3.

Force \vec{F} has the same direction as \vec{E} if q is positive and the opposite direction if q is negative.

Dipole in an Electric Field When an electric dipole of dipole moment \vec{p} is placed in an electric field \vec{E} , the field exerts a torque $\vec{\tau}$ on the dipole:

$$\vec{\tau} = \vec{p} \times \vec{E}. \quad (22-34)$$

The dipole has a potential energy U associated with its orientation in the field:

$$U = -\vec{p} \cdot \vec{E}. \quad (22-38)$$

This potential energy is defined to be zero when \vec{p} is perpendicular to \vec{E} ; it is least ($U = -pE$) when \vec{p} is aligned with \vec{E} and greatest ($U = pE$) when \vec{p} is directed opposite \vec{E} .

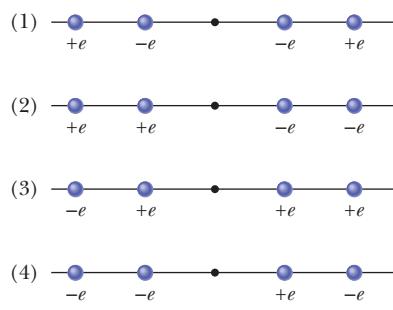


Figure 22-25 Question 4.



Figure 22-26 Question 5.

- 4 Figure 22-25 shows four situations in which four charged particles are evenly spaced to the left and right of a central point. The charge values are indicated. Rank the situations according to the magnitude of the net electric field at the central point, greatest first.
- 5 Figure 22-26 shows two charged particles fixed in place on an axis. (a) Where on the axis (other than at an infinite distance) is there a point at which their net electric field is zero: between the charges, to their left, or to their right? (b) Is there a point of zero net electric field anywhere off the axis (other than at an infinite distance)?

- 6 In Fig. 22-27, two identical circular nonconducting rings are centered on the same line with their planes perpendicular to the line. Each ring has charge that is uniformly distributed along its circumference. The rings each produce electric fields at points along the line. For three situations, the charges on rings A and B are, respectively, (1) q_0 and q_0 , (2) $-q_0$ and $-q_0$, and (3) $-q_0$ and q_0 . Rank the situations according to the magnitude of the net electric field at (a) point P_1 midway between the rings, (b) point P_2 at the center of ring B , and (c) point P_3 to the right of ring B , greatest first.

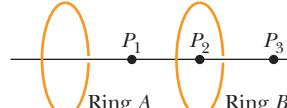


Figure 22-27 Question 6.

- 7 The potential energies associated with four orientations of an electric dipole in an electric field are (1) $-5U_0$, (2) $-7U_0$, (3) $3U_0$, and (4) $5U_0$, where U_0 is positive. Rank the orientations according to (a) the angle between the electric dipole moment \vec{p} and the electric field \vec{E} and (b) the magnitude of the torque on the electric dipole, greatest first.

- 8 (a) In Checkpoint 4, if the dipole rotates from orientation 1 to orientation 2, is the work done on the dipole by the field positive, negative, or zero? (b) If, instead, the dipole rotates from orientation 1 to orientation 4, is the work done by the field more than, less than, or the same as in (a)?

- 9** Figure 22-28 shows two disks and a flat ring, each with the same uniform charge Q . Rank the objects according to the magnitude of the electric field they create at points P (which are at the same vertical heights), greatest first.

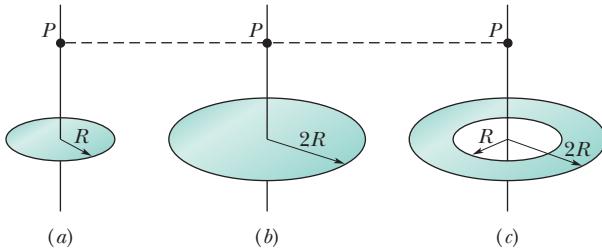


Figure 22-28 Question 9.

- 10** In Fig. 22-29, an electron e travels through a small hole in plate A and then toward plate B . A uniform electric field in the region between the plates then slows the electron without deflecting it. (a) What is the direction of the field? (b) Four other particles similarly travel through small holes in either plate A or plate B and then into the region between the plates. Three have charges $+q_1$, $+q_2$, and $-q_3$. The fourth (labeled n) is a neutron, which is electrically neutral. Does the speed of each of those four other particles increase, decrease, or remain the same in the region between the plates?

- 11** In Fig. 22-30a, a circular plastic rod with uniform charge $+Q$ produces an electric field of magnitude E at the center of

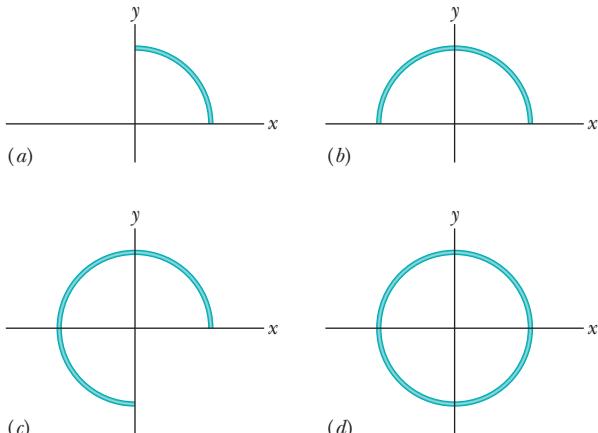


Figure 22-30 Question 11.

curvature (at the origin). In Figs. 22-30b, c, and d, more circular rods, each with identical uniform charges $+Q$, are added until the circle is complete. A fifth arrangement (which would be labeled e) is like that in d except the rod in the fourth quadrant has charge $-Q$. Rank the five arrangements according to the magnitude of the electric field at the center of curvature, greatest first.

- 12** When three electric dipoles are near each other, they each experience the electric field of the other two, and the three-dipole system has a certain potential energy. Figure 22-31 shows two arrangements in which three electric dipoles are side by side. Each dipole has the same magnitude of electric dipole moment, and the spacings between adjacent dipoles are identical. In which arrangement is the potential energy of the three-dipole system greater?

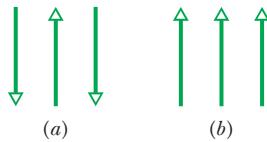


Figure 22-31 Question 12.

- 13** Figure 22-32 shows three rods, each with the same charge Q spread uniformly along its length. Rods a (of length L) and b (of length $L/2$) are straight, and points P are aligned with their midpoints. Rod c (of length $L/2$) forms a complete circle about point P . Rank the rods according to the magnitude of the electric field they create at points P , greatest first.

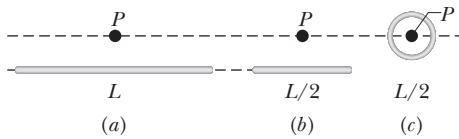


Figure 22-32 Question 13.

- 14** Figure 22-33 shows five protons that are launched in a uniform electric field \vec{E} ; the magnitude and direction of the launch velocities are indicated. Rank the protons according to the magnitude of their accelerations due to the field, greatest first.

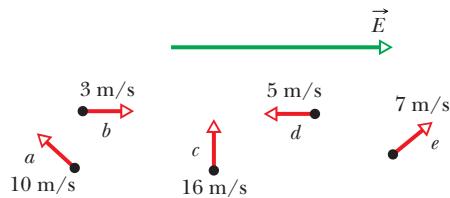


Figure 22-33 Question 14.

Problems



Tutoring problem available (at instructor's discretion) in WileyPLUS and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com

WWW Worked-out solution is at

ILW Interactive solution is at

<http://www.wiley.com/college/halliday>

Module 22-1 The Electric Field

- Sketch qualitatively the electric field lines both between and outside two concentric conducting spherical shells when a uniform

positive charge q_1 is on the inner shell and a uniform negative charge $-q_2$ is on the outer. Consider the cases $q_1 > q_2$, $q_1 = q_2$, and $q_1 < q_2$.

- 2 In Fig. 22-34 the electric field lines on the left have twice the separation of those on the right. (a) If the magnitude of the field at A is 40 N/C, what is the magnitude of the force on a proton at A ? (b) What is the magnitude of the field at B ?

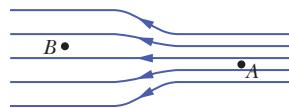


Figure 22-34 Problem 2.

Module 22-2 The Electric Field Due to a Charged Particle

- 3 **SSM** The nucleus of a plutonium-239 atom contains 94 protons. Assume that the nucleus is a sphere with radius 6.64 fm and with the charge of the protons uniformly spread through the sphere. At the surface of the nucleus, what are the (a) magnitude and (b) direction (radially inward or outward) of the electric field produced by the protons?

- 4 Two charged particles are attached to an x axis: Particle 1 of charge -2.00×10^{-7} C is at position $x = 6.00$ cm and particle 2 of charge $+2.00 \times 10^{-7}$ C is at position $x = 21.0$ cm. Midway between the particles, what is their net electric field in unit-vector notation?

- 5 **SSM** A charged particle produces an electric field with a magnitude of 2.0 N/C at a point that is 50 cm away from the particle. What is the magnitude of the particle's charge?

- 6 What is the magnitude of a point charge that would create an electric field of 1.00 N/C at points 1.00 m away?

- 7 **SSM ILW WWW** In Fig. 22-35, the four particles form a square of edge length $a = 5.00$ cm and have charges $q_1 = +10.0$ nC, $q_2 = -20.0$ nC, $q_3 = +20.0$ nC, and $q_4 = -10.0$ nC. In unit-vector notation, what net electric field do the particles produce at the square's center?

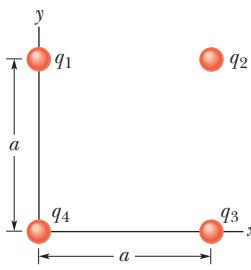


Figure 22-35 Problem 7.

- 8 **GO** In Fig. 22-36, the four particles are fixed in place and have charges $q_1 = q_2 = +5e$, $q_3 = +3e$, and $q_4 = -12e$. Distance $d = 5.0 \mu\text{m}$. What is the magnitude of the net electric field at point P due to the particles?

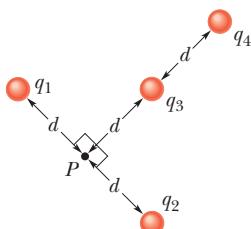


Figure 22-36 Problem 8.

- 9 **GO** Figure 22-37 shows two charged particles on an x axis: $-q = -3.20 \times 10^{-19}$ C at $x = -3.00$ m and $q = 3.20 \times 10^{-19}$ C at $x = +3.00$ m. What are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the net electric field produced at point P at $y = 4.00$ m?

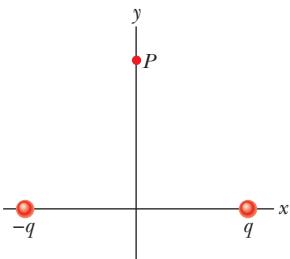


Figure 22-37 Problem 9.

- 10 **GO** Figure 22-38a shows two charged particles fixed in place on an x axis with separation L . The ratio q_1/q_2 of their charge magnitudes is 4.00. Figure 22-38b shows the x component $E_{\text{net},x}$ of their net electric field along the x axis just to the right of particle 2. The x axis scale is set by $x_s = 30.0$ cm. (a) At what value of $x > 0$ is $E_{\text{net},x}$ maximum? (b) If particle 2 has charge $-q_2 = -3e$, what is the value of that maximum?

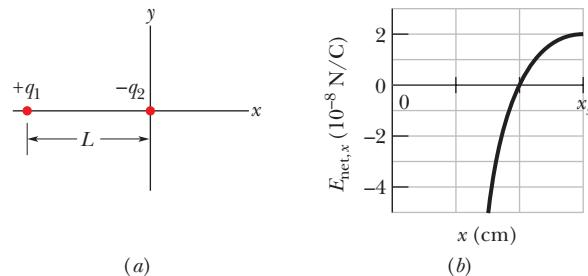


Figure 22-38 Problem 10.

- 11 **SSM** Two charged particles are fixed to an x axis: Particle 1 of charge $q_1 = 2.1 \times 10^{-8}$ C is at position $x = 20$ cm and particle 2 of charge $q_2 = -4.00q_1$ is at position $x = 70$ cm. At what coordinate on the axis (other than at infinity) is the net electric field produced by the two particles equal to zero?

- 12 **GO** Figure 22-39 shows an uneven arrangement of electrons (e) and protons (p) on a circular arc of radius $r = 2.00$ cm, with angles $\theta_1 = 30.0^\circ$, $\theta_2 = 50.0^\circ$, $\theta_3 = 30.0^\circ$, and $\theta_4 = 20.0^\circ$. What are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the net electric field produced at the center of the arc?

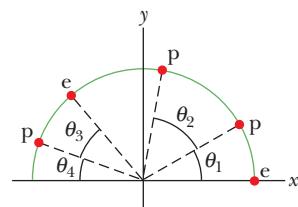


Figure 22-39 Problem 12.

- 13 **GO** Figure 22-40 shows a proton (p) on the central axis through a disk with a uniform charge density due to excess electrons. The disk is seen from an edge-on view. Three of those electrons are shown: electron e_c at the disk center and electrons e_s at opposite sides of the disk, at radius R from the center. The proton is initially at distance $z = R = 2.00$ cm from the disk. At that location, what are the magnitudes of (a) the electric field \vec{E}_c due to electron e_c and (b) the net electric field $\vec{E}_{s,\text{net}}$ due to electrons e_s ? The proton is then moved to $z = R/10.0$. What then are the magnitudes of (c) \vec{E}_c and (d) $\vec{E}_{s,\text{net}}$ at the proton's location? (e) From (a) and (c) we see that as the proton gets nearer to the disk, the magnitude of \vec{E}_c increases, as expected. Why does the magnitude of $\vec{E}_{s,\text{net}}$ from the two side electrons decrease, as we see from (b) and (d)?

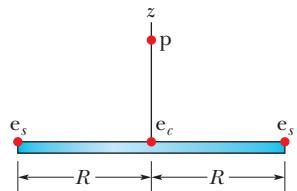


Figure 22-40 Problem 13.

- 14 In Fig. 22-41, particle 1 of charge $q_1 = -5.00q$ and particle 2 of charge $q_2 = +2.00q$ are fixed to an x axis. (a) As a multiple of distance L , at what coordinate on the axis is the net electric field of the particles zero? (b) Sketch the net electric field lines between and around the particles.

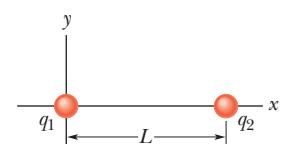


Figure 22-41 Problem 14.

••15 In Fig. 22-42, the three particles are fixed in place and have charges $q_1 = q_2 = +e$ and $q_3 = +2e$. Distance $a = 6.00 \mu\text{m}$. What are the (a) magnitude and (b) direction of the net electric field at point P due to the particles?

••16 Figure 22-43 shows a plastic ring of radius $R = 50.0 \text{ cm}$. Two small charged beads are on the ring; Bead 1 of charge $+2.00 \mu\text{C}$ is fixed in place at the left side; bead 2 of charge $+6.00 \mu\text{C}$ can be moved along the ring. The two beads produce a net electric field of magnitude E at the center of the ring. At what (a) positive and (b) negative value of angle θ should bead 2 be positioned such that $E = 2.00 \times 10^5 \text{ N/C}$?

••17 Two charged beads are on the plastic ring in Fig. 22-44a. Bead 2, which is not shown, is fixed in place on the ring, which has radius $R = 60.0 \text{ cm}$. Bead 1, which is not fixed in place, is initially on the x axis at angle $\theta = 0^\circ$. It is then moved to the opposite side, at angle $\theta = 180^\circ$, through the first and second quadrants of the xy coordinate system. Figure 22-44b gives the x component of the net electric field produced at the origin by the two beads as a function of θ , and Fig. 22-44c gives the y component of that net electric field. The vertical axis scales are set by $E_{xs} = 5.0 \times 10^4 \text{ N/C}$ and $E_{ys} = -9.0 \times 10^4 \text{ N/C}$. (a) At what angle θ is bead 2 located? What are the charges of (b) bead 1 and (c) bead 2?

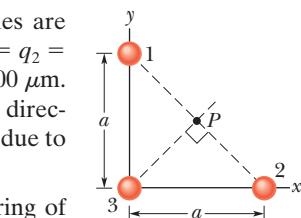


Figure 22-42
Problem 15.

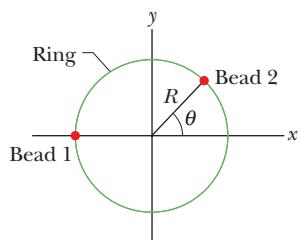


Figure 22-43 Problem 16.

••18 Two small charged beads are on a horizontal ring of radius $R = 60.0 \text{ cm}$. One bead is fixed in place at the left side of the ring. The other bead is free to move along the ring. The two beads produce a net electric field of magnitude E at the center of the ring. At what (a) positive and (b) negative value of angle θ should the free bead be positioned such that $E = 2.00 \times 10^5 \text{ N/C}$?

••19 Figure 22-45 shows an electric dipole. What are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the dipole's electric field at point P , located at distance $r \gg d$?

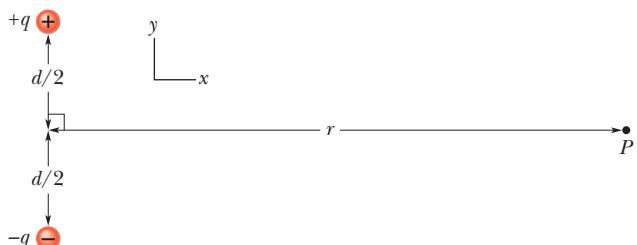


Figure 22-45 Problem 19.

••20 Equations 22-8 and 22-9 are approximations of the magnitude of the electric field of an electric dipole, at points along the dipole axis. Consider a point P on that axis at distance $z = 5.00d$ from the dipole center (d is the separation distance between the particles of the dipole). Let E_{appr} be the magnitude of the field at point P as approximated by Eqs. 22-8 and 22-9. Let E_{act} be the actual magnitude. What is the ratio $E_{\text{appr}}/E_{\text{act}}$?

••21 SSM *Electric quadrupole.* Figure 22-46 shows a generic electric quadrupole. It consists of two dipoles with dipole moments that are equal in magnitude but opposite in direction. Show that the value of E on the axis of the quadrupole for a point P a distance z from its center (assume $z \gg d$) is given by

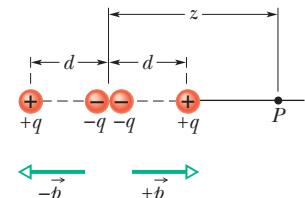


Figure 22-46 Problem 21.

$$E = \frac{3Q}{4\pi\epsilon_0 z^4},$$

in which $Q (= 2qd^2)$ is known as the *quadrupole moment* of the charge distribution.

Module 22-4 The Electric Field Due to a Line of Charge

•22 *Density, density, density.* (a) A charge $-300e$ is uniformly distributed along a circular arc of radius 4.00 cm , which subtends an angle of 40° . What is the linear charge density along the arc? (b) A charge $-300e$ is uniformly distributed over one face of a circular disk of radius 2.00 cm . What is the surface charge density over that face? (c) A charge $-300e$ is uniformly distributed over the surface of a sphere of radius 2.00 cm . What is the surface charge density over that surface? (d) A charge $-300e$ is uniformly spread through the volume of a sphere of radius 2.00 cm . What is the volume charge density in that sphere?

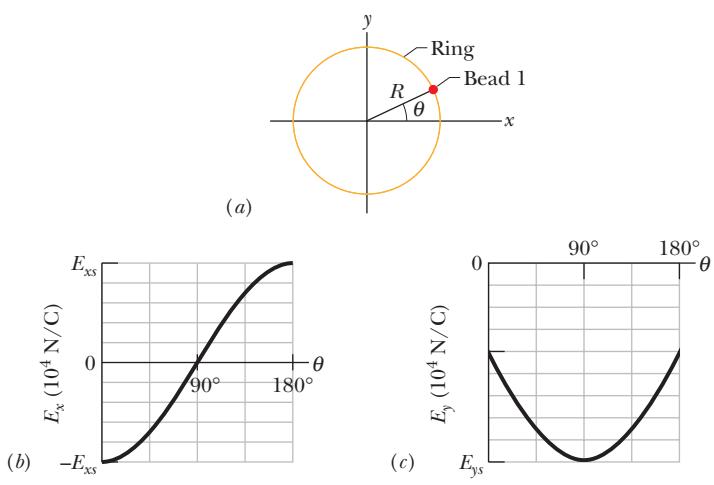


Figure 22-44 Problem 17.

Module 22-3 The Electric Field Due to a Dipole

••18 The electric field of an electric dipole along the dipole axis is approximated by Eqs. 22-8 and 22-9. If a binomial expansion is made of Eq. 22-7, what is the next term in the expression for the dipole's electric field along the dipole axis? That is, what is E_{next} in the expression

$$E = \frac{1}{2\pi\epsilon_0} \frac{qd}{z^3} + E_{\text{next}}?$$

•23 Figure 22-47 shows two parallel nonconducting rings with their central axes along a common line. Ring 1 has uniform charge q_1 and radius R ; ring 2 has uniform charge q_2 and the same radius R . The rings are separated by distance $d = 3.00R$. The net electric field at point P on the common line, at distance R from ring 1, is zero. What is the ratio q_1/q_2 ?

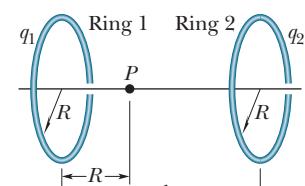


Figure 22-47 Problem 23.

••24 A thin nonconducting rod with a uniform distribution of positive charge Q is bent into a complete circle of radius R .

(Fig. 22-48). The central perpendicular axis through the ring is a z axis, with the origin at the center of the ring. What is the magnitude of the electric field due to the rod at (a) $z = 0$ and (b) $z = \infty$? (c) In terms of R , at what positive value of z is that magnitude maximum? (d) If $R = 2.00\text{ cm}$ and $Q = 4.00\text{ }\mu\text{C}$, what is the maximum magnitude?

••25 Figure 22-49 shows three circular arcs centered on the origin of a coordinate system. On each arc, the uniformly distributed charge is given in terms of $Q = 2.00\text{ }\mu\text{C}$. The radii are given in terms of $R = 10.0\text{ cm}$. What are the (a) magnitude and (b) direction (relative to the positive x direction) of the net electric field at the origin due to the arcs?

••26 **GO ILW** In Fig. 22-50, a thin glass rod forms a semicircle of radius $r = 5.00\text{ cm}$. Charge is uniformly distributed along the rod, with $+q = 4.50\text{ pC}$ in the upper half and $-q = -4.50\text{ pC}$ in the lower half. What are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the electric field \vec{E} at P , the center of the semicircle?

••27 **GO** In Fig. 22-51, two curved plastic rods, one of charge $+q$ and the other of charge $-q$, form a circle of radius $R = 8.50\text{ cm}$ in an xy plane. The x axis passes through both of the connecting points, and the charge is distributed uniformly on both rods. If $q = 15.0\text{ pC}$, what are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the electric field \vec{E} produced at P , the center of the circle?

••28 Charge is uniformly distributed around a ring of radius $R = 2.40\text{ cm}$, and the resulting electric field magnitude E is measured along the ring's central axis (perpendicular to the plane of the ring). At what distance from the ring's center is E maximum?

••29 **GO** Figure 22-52a shows a nonconducting rod with a uniformly distributed charge $+Q$. The rod forms a half-circle with radius R and produces an electric field of magnitude E_{arc} at its center of curvature P . If the arc is collapsed to a point at distance R from P (Fig. 22-52b), by what factor is the magnitude of the electric field at P multiplied?

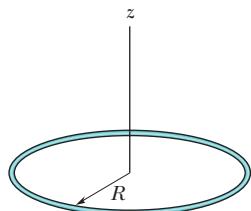


Figure 22-48 Problem 24.

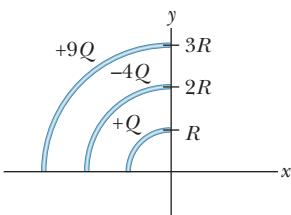


Figure 22-49 Problem 25.

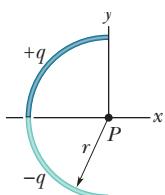


Figure 22-50
Problem 26.

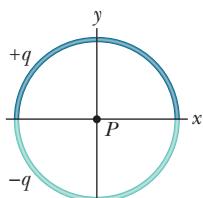


Figure 22-51
Problem 27.

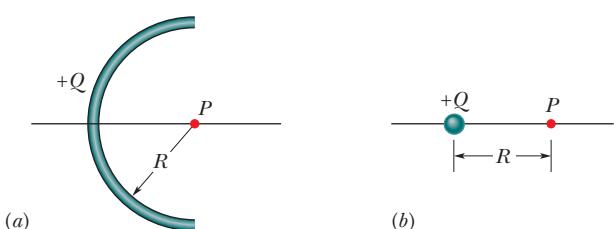


Figure 22-52 Problem 29.

••30 **GO** Figure 22-53 shows two concentric rings, of radii R and $R' = 3.00R$, that lie on the same plane. Point P lies on the central z axis, at distance $D = 2.00R$ from the center of the rings. The smaller ring has uniformly distributed charge $+Q$. In terms of Q , what is the uniformly distributed charge on the larger ring if the net electric field at P is zero?

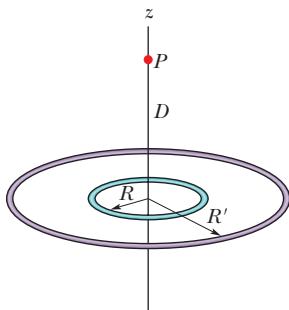


Figure 22-53 Problem 30.

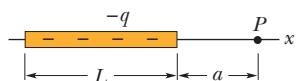


Figure 22-54 Problem 31.

••31 **SSM ILW WWW** In Fig. 22-54, a nonconducting rod of length $L = 8.15\text{ cm}$ has a charge $-q = -4.23\text{ fC}$ uniformly distributed along its length. (a) What is the linear charge density of the rod? What are the (b) magnitude and (c) direction (relative to the positive direction of the x axis) of the electric field produced at point P , at distance $a = 12.0\text{ cm}$ from the rod? What is the electric field magnitude produced at distance $a = 50\text{ m}$ by (d) the rod and (e) a particle of charge $-q = -4.23\text{ fC}$ that we use to replace the rod? (At that distance, the rod "looks" like a particle.)

••32 **GO** In Fig. 22-55, positive charge $q = 7.81\text{ pC}$ is spread uniformly along a thin nonconducting rod of length $L = 14.5\text{ cm}$. What are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the electric field produced at point P , at distance $R = 6.00\text{ cm}$ from the rod along its perpendicular bisector?

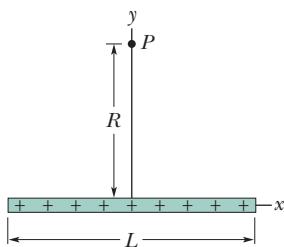


Figure 22-55 Problem 32.

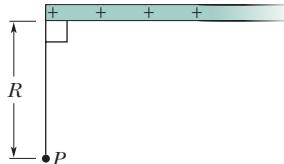


Figure 22-56 Problem 33.

••33 **GO** In Fig. 22-56, a "semi-infinite" nonconducting rod (that is, infinite in one direction only) has uniform linear charge density λ . Show that the electric field \vec{E}_P at point P makes an angle of 45° with the rod and that this result is independent of the distance R . (Hint: Separately find the component of \vec{E}_P parallel to the rod and the component perpendicular to the rod.)

Module 22-5 The Electric Field Due to a Charged Disk

••34 A disk of radius 2.5 cm has a surface charge density of $5.3\text{ }\mu\text{C/m}^2$ on its upper face. What is the magnitude of the electric field produced by the disk at a point on its central axis at distance $z = 12\text{ cm}$ from the disk?

••35 **SSM WWW** At what distance along the central perpendicular axis of a uniformly charged plastic disk of radius 0.600 m is the magnitude of the electric field equal to one-half the magnitude of the field at the center of the surface of the disk?

••36 A circular plastic disk with radius $R = 2.00\text{ cm}$ has a uniformly distributed charge $Q = +(2.00 \times 10^6)e$ on one face. A circular ring of width $30\text{ }\mu\text{m}$ is centered on that face, with the center of that width at radius $r = 0.50\text{ cm}$. In coulombs, what charge is contained within the width of the ring?

••37 Suppose you design an apparatus in which a uniformly charged disk of radius R is to produce an electric field. The field magnitude is most important along the central perpendicular axis of the disk, at a point P at distance $2.00R$ from the disk (Fig. 22-57a). Cost analysis suggests that you switch to a ring of the same outer radius R but with inner radius $R/2.00$ (Fig. 22-57b). Assume that the ring will have the same surface charge density as the original disk. If you switch to the ring, by what percentage will you decrease the electric field magnitude at P ?

••38 Figure 22-58a shows a circular disk that is uniformly charged. The central z axis is perpendicular to the disk face, with the origin at the disk. Figure 22-58b gives the magnitude of the electric field along that axis in terms of the maximum magnitude E_m at the disk surface. The z axis scale is set by $z_s = 8.0$ cm. What is the radius of the disk?

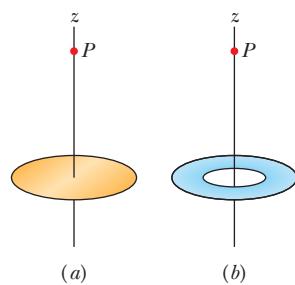


Figure 22-57 Problem 37.

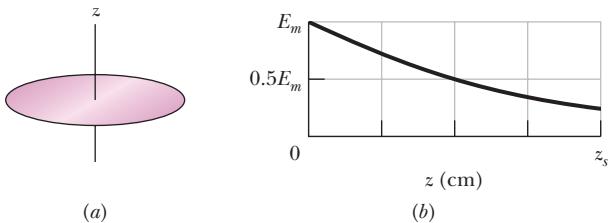


Figure 22-58 Problem 38.

Module 22-6 A Point Charge in an Electric Field

•39 In Millikan's experiment, an oil drop of radius $1.64 \mu\text{m}$ and density 0.851 g/cm^3 is suspended in chamber C (Fig. 22-16) when a downward electric field of $1.92 \times 10^5 \text{ N/C}$ is applied. Find the charge on the drop, in terms of e .

•40 **GO** An electron with a speed of $5.00 \times 10^8 \text{ cm/s}$ enters an electric field of magnitude $1.00 \times 10^3 \text{ N/C}$, traveling along a field line in the direction that retards its motion. (a) How far will the electron travel in the field before stopping momentarily, and (b) how much time will have elapsed? (c) If the region containing the electric field is 8.00 mm long (too short for the electron to stop within it), what fraction of the electron's initial kinetic energy will be lost in that region?

•41 **SSM** A charged cloud system produces an electric field in the air near Earth's surface. A particle of charge $-2.0 \times 10^{-9} \text{ C}$ is acted on by a downward electrostatic force of $3.0 \times 10^{-6} \text{ N}$ when placed in this field. (a) What is the magnitude of the electric field? What are the (b) magnitude and (c) direction of the electrostatic force \vec{F}_{el} on the proton placed in this field? (d) What is the magnitude of the gravitational force \vec{F}_g on the proton? (e) What is the ratio F_{el}/F_g in this case?

•42 Humid air breaks down (its molecules become ionized) in an electric field of $3.0 \times 10^6 \text{ N/C}$. In that field, what is the magnitude of the electrostatic force on (a) an electron and (b) an ion with a single electron missing?

•43 **SSM** An electron is released from rest in a uniform electric field of magnitude $2.00 \times 10^4 \text{ N/C}$. Calculate the acceleration of the electron. (Ignore gravitation.)

•44 An alpha particle (the nucleus of a helium atom) has a mass of $6.64 \times 10^{-27} \text{ kg}$ and a charge of $+2e$. What are the (a) magnitude and (b) direction of the electric field that will balance the gravitational force on the particle?

•45 **ILW** An electron on the axis of an electric dipole is 25 nm from the center of the dipole. What is the magnitude of the electrostatic force on the electron if the dipole moment is $3.6 \times 10^{-29} \text{ C}\cdot\text{m}$? Assume that 25 nm is much larger than the separation of the charged particles that form the dipole.

•46 An electron is accelerated eastward at $1.80 \times 10^9 \text{ m/s}^2$ by an electric field. Determine the field (a) magnitude and (b) direction.

•47 **SSM** Beams of high-speed protons can be produced in "guns" using electric fields to accelerate the protons. (a) What acceleration would a proton experience if the gun's electric field were $2.00 \times 10^4 \text{ N/C}$? (b) What speed would the proton attain if the field accelerated the proton through a distance of 1.00 cm?

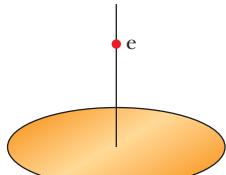


Figure 22-59
Problem 48.

•48 In Fig. 22-59, an electron (e) is to be released from rest on the central axis of a uniformly charged disk of radius R . The surface charge density on the disk is $+4.00 \mu\text{C/m}^2$. What is the magnitude of the electron's initial acceleration if it is released at a distance (a) R , (b) $R/100$, and (c) $R/1000$ from the center of the disk? (d) Why does the acceleration magnitude increase only slightly as the release point is moved closer to the disk?

•49 A 10.0 g block with a charge of $+8.00 \times 10^{-5} \text{ C}$ is placed in an electric field $\vec{E} = (3000\hat{i} - 600\hat{j}) \text{ N/C}$. What are the (a) magnitude and (b) direction (relative to the positive direction of the x axis) of the electrostatic force on the block? If the block is released from rest at the origin at time $t = 0$, what are its (c) x and (d) y coordinates at $t = 3.00 \text{ s}$?

•50 At some instant the velocity components of an electron moving between two charged parallel plates are $v_x = 1.5 \times 10^5 \text{ m/s}$ and $v_y = 3.0 \times 10^3 \text{ m/s}$. Suppose the electric field between the plates is uniform and given by $\vec{E} = (120 \text{ N/C})\hat{j}$. In unit-vector notation, what are (a) the electron's acceleration in that field and (b) the electron's velocity when its x coordinate has changed by 2.0 cm?

•51 **GO** Assume that a honeybee is a sphere of diameter 1.000 cm with a charge of $+45.0 \text{ pC}$ uniformly spread over its surface. Assume also that a spherical pollen grain of diameter $40.0 \mu\text{m}$ is electrically held on the surface of the bee because the bee's charge induces a charge of -1.00 pC on the near side of the grain and a charge of $+1.00 \text{ pC}$ on the far side. (a) What is the magnitude of the net electrostatic force on the grain due to the bee? Next, assume that the bee brings the grain to a distance of 1.000 mm from the tip of a flower's stigma and that the tip is a particle of charge -45.0 pC . (b) What is the magnitude of the net electrostatic force on the grain due to the stigma? (c) Does the grain remain on the bee or does it move to the stigma?

•52 An electron enters a region of uniform electric field with an initial velocity of 40 km/s in the same direction as the electric field, which has magnitude $E = 50 \text{ N/C}$. (a) What is the speed of the electron 1.5 ns after entering this region? (b) How far does the electron travel during the 1.5 ns interval?

- 53 GO** Two large parallel copper plates are 5.0 cm apart and have a uniform electric field between them as depicted in Fig. 22-60. An electron is released from the negative plate at the same time that a proton is released from the positive plate. Neglect the force of the particles on each other and find their distance from the positive plate when they pass each other. (Does it surprise you that you need not know the electric field to solve this problem?)

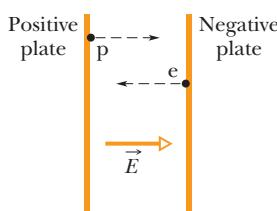


Figure 22-60 Problem 53.

- 54 GO** In Fig. 22-61, an electron is shot at an initial speed of $v_0 = 2.00 \times 10^6 \text{ m/s}$, at angle $\theta_0 = 40.0^\circ$ from an x axis. It moves through a uniform electric field $\vec{E} = (5.00 \text{ N/C})\hat{j}$. A screen for detecting electrons is positioned parallel to the y axis, at distance $x = 3.00 \text{ m}$. In unit-vector notation, what is the velocity of the electron when it hits the screen?

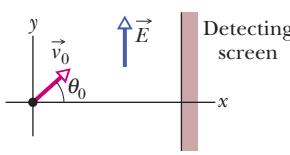


Figure 22-61 Problem 54.

- 55 ILW** A uniform electric field exists in a region between two oppositely charged plates. An electron is released from rest at the surface of the negatively charged plate and strikes the surface of the opposite plate, 2.0 cm away, in a time $1.5 \times 10^{-8} \text{ s}$. (a) What is the speed of the electron as it strikes the second plate? (b) What is the magnitude of the electric field \vec{E} ?

Module 22-7 A Dipole in an Electric Field

- 56** An electric dipole consists of charges $+2e$ and $-2e$ separated by 0.78 nm. It is in an electric field of strength $3.4 \times 10^6 \text{ N/C}$. Calculate the magnitude of the torque on the dipole when the dipole moment is (a) parallel to, (b) perpendicular to, and (c) antiparallel to the electric field.

- 57 SSM** An electric dipole consisting of charges of magnitude 1.50 nC separated by $6.20 \mu\text{m}$ is in an electric field of strength 1100 N/C. What are (a) the magnitude of the electric dipole moment and (b) the difference between the potential energies for dipole orientations parallel and antiparallel to \vec{E} ?

- 58** A certain electric dipole is placed in a uniform electric field \vec{E} of magnitude 20 N/C. Figure 22-62 gives the potential energy U of the dipole versus the angle θ between \vec{E} and the dipole moment \vec{p} . The vertical axis scale is set by $U_s = 100 \times 10^{-28} \text{ J}$. What is the magnitude of \vec{p} ?

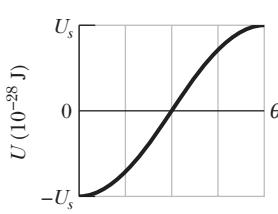


Figure 22-62 Problem 58.

- 59** How much work is required to turn an electric dipole 180° in a uniform electric field of magnitude $E = 46.0 \text{ N/C}$ if the dipole moment has a magnitude of $p = 3.02 \times 10^{-25} \text{ C} \cdot \text{m}$ and the initial angle is 64° ?

- 60** A certain electric dipole is placed in a uniform electric field \vec{E} of magnitude 40 N/C. Figure 22-63 gives the magnitude τ of the torque on the dipole versus the angle θ between field \vec{E} and the dipole moment \vec{p} . The vertical axis scale is set by $\tau_s = 100 \times 10^{-28} \text{ N} \cdot \text{m}$. What is the magnitude of \vec{p} ?

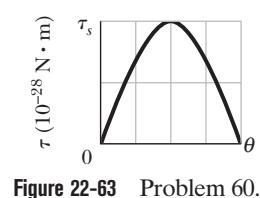


Figure 22-63 Problem 60.

- 61** Find an expression for the oscillation frequency of an electric dipole of dipole moment \vec{p} and rotational inertia I for small amplitudes of oscillation about its equilibrium position in a uniform electric field of magnitude E .

Additional Problems

- 62** (a) What is the magnitude of an electron's acceleration in a uniform electric field of magnitude $1.40 \times 10^6 \text{ N/C}$? (b) How long would the electron take, starting from rest, to attain one-tenth the speed of light? (c) How far would it travel in that time?

- 63** A spherical water drop $1.20 \mu\text{m}$ in diameter is suspended in calm air due to a downward-directed atmospheric electric field of magnitude $E = 462 \text{ N/C}$. (a) What is the magnitude of the gravitational force on the drop? (b) How many excess electrons does it have?

- 64** Three particles, each with positive charge Q , form an equilateral triangle, with each side of length d . What is the magnitude of the electric field produced by the particles at the midpoint of any side?

- 65** In Fig. 22-64a, a particle of charge $+Q$ produces an electric field of magnitude E_{part} at point P , at distance R from the particle. In Fig. 22-64b, that same amount of charge is spread uniformly along a circular arc that has radius R and subtends an angle θ . The charge on the arc produces an electric field of magnitude E_{arc} at its center of curvature P . For what value of θ does $E_{\text{arc}} = 0.500E_{\text{part}}$? (Hint: You will probably resort to a graphical solution.)

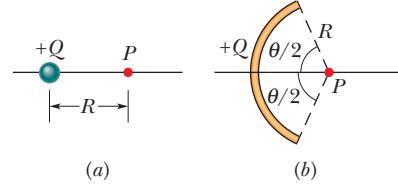
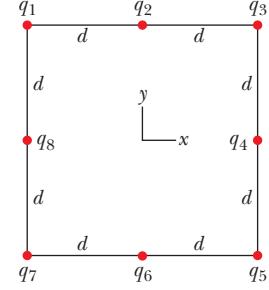


Figure 22-64 Problem 65.

- 66** A proton and an electron form two corners of an equilateral triangle of side length $2.0 \times 10^{-6} \text{ m}$. What is the magnitude of the net electric field these two particles produce at the third corner?

- 67** A charge (uniform linear density = 9.0 nC/m) lies on a string that is stretched along an x axis from $x = 0$ to $x = 3.0 \text{ m}$. Determine the magnitude of the electric field at $x = 4.0 \text{ m}$ on the x axis.

- 68** In Fig. 22-65, eight particles form a square in which distance $d = 2.0 \text{ cm}$. The charges are $q_1 = +3e$, $q_2 = +e$, $q_3 = -5e$, $q_4 = -2e$, $q_5 = +3e$, $q_6 = +e$, $q_7 = -5e$, and $q_8 = +e$. In unit-vector notation, what is the net electric field at the square's center?

Figure 22-65
Problem 68.

- 69** Two particles, each with a charge of magnitude 12 nC, are at two of the vertices of an equilateral triangle with edge length 2.0 m. What is the magnitude of the electric field at the third vertex if (a) both charges are positive and (b) one charge is positive and the other is negative?

- 70** The following table gives the charge seen by Millikan at different times on a single drop in his experiment. From the data, calculate the elementary charge e .

$6.563 \times 10^{-19} \text{ C}$	$13.13 \times 10^{-19} \text{ C}$	$19.71 \times 10^{-19} \text{ C}$
$8.204 \times 10^{-19} \text{ C}$	$16.48 \times 10^{-19} \text{ C}$	$22.89 \times 10^{-19} \text{ C}$
$11.50 \times 10^{-19} \text{ C}$	$18.08 \times 10^{-19} \text{ C}$	$26.13 \times 10^{-19} \text{ C}$

71 A charge of 20 nC is uniformly distributed along a straight rod of length 4.0 m that is bent into a circular arc with a radius of 2.0 m . What is the magnitude of the electric field at the center of curvature of the arc?

72 An electron is constrained to the central axis of the ring of charge of radius R in Fig. 22-11, with $z \ll R$. Show that the electrostatic force on the electron can cause it to oscillate through the ring center with an angular frequency

$$\omega = \sqrt{\frac{eq}{4\pi\epsilon_0 m R^3}},$$

where q is the ring's charge and m is the electron's mass.

73 SSM The electric field in an xy plane produced by a positively charged particle is $7.2(4.0\hat{i} + 3.0\hat{j}) \text{ N/C}$ at the point $(3.0, 3.0) \text{ cm}$ and $100\hat{i} \text{ N/C}$ at the point $(2.0, 0) \text{ cm}$. What are the (a) x and (b) y coordinates of the particle? (c) What is the charge of the particle?

74 (a) What total (excess) charge q must the disk in Fig. 22-15 have for the electric field on the surface of the disk at its center to have magnitude $3.0 \times 10^6 \text{ N/C}$, the E value at which air breaks down electrically, producing sparks? Take the disk radius as 2.5 cm . (b) Suppose each surface atom has an effective cross-sectional area of 0.015 nm^2 . How many atoms are needed to make up the disk surface? (c) The charge calculated in (a) results from some of the surface atoms having one excess electron. What fraction of these atoms must be so charged?

75 In Fig. 22-66, particle 1 (of charge $+1.00 \mu\text{C}$), particle 2 (of charge $+1.00 \mu\text{C}$), and particle 3 (of charge Q) form an equilateral triangle of edge length a . For what value of Q (both sign and magnitude) does the net electric field produced by the particles at the center of the triangle vanish?

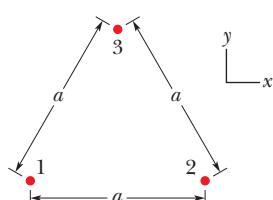


Figure 22-66 Problems 75 and 86.

76 In Fig. 22-67, an electric dipole swings from an initial orientation i ($\theta_i = 20.0^\circ$) to a final orientation f ($\theta_f = 20.0^\circ$) in a uniform external electric field \vec{E} . The electric dipole moment is $1.60 \times 10^{-27} \text{ C}\cdot\text{m}$; the field magnitude is $3.00 \times 10^6 \text{ N/C}$. What is the change in the dipole's potential energy?

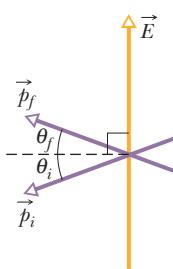


Figure 22-67 Problem 76.

77 A particle of charge $-q_1$ is at the origin of an x axis. (a) At what location on the axis should a particle of charge $-4q_1$ be placed so that the net electric field is zero at $x = 2.0 \text{ mm}$ on the axis? (b) If, instead, a particle of charge $+4q_1$ is placed at that location, what is the direction (relative to the positive direction of the x axis) of the net electric field at $x = 2.0 \text{ mm}$?

78 Two particles, each of positive charge q , are fixed in place on a y axis, one at $y = d$ and the other at $y = -d$. (a) Write an expression that gives the magnitude E of the net electric field at points on the x axis given by $x = \alpha d$. (b) Graph E versus α for the range $0 < \alpha < 4$. From the graph, determine the values of α that give (c) the maximum value of E and (d) half the maximum value of E .

79 A clock face has negative point charges $-q, -2q, -3q, \dots, -12q$ fixed at the positions of the corresponding numerals. The clock hands do not perturb the net field due to the point charges. At

what time does the hour hand point in the same direction as the electric field vector at the center of the dial? (Hint: Use symmetry.)

80 Calculate the electric dipole moment of an electron and a proton 4.30 nm apart.

81 An electric field \vec{E} with an average magnitude of about 150 N/C points downward in the atmosphere near Earth's surface. We wish to "float" a sulfur sphere weighing 4.4 N in this field by charging the sphere. (a) What charge (both sign and magnitude) must be used? (b) Why is the experiment impractical?

82 A circular rod has a radius of curvature $R = 9.00 \text{ cm}$ and a uniformly distributed positive charge $Q = 6.25 \text{ pC}$ and subtends an angle $\theta = 2.40 \text{ rad}$. What is the magnitude of the electric field that Q produces at the center of curvature?

83 SSM An electric dipole with dipole moment

$$\vec{p} = (3.00\hat{i} + 4.00\hat{j})(1.24 \times 10^{-30} \text{ C}\cdot\text{m})$$

is in an electric field $\vec{E} = (4000 \text{ N/C})\hat{i}$. (a) What is the potential energy of the electric dipole? (b) What is the torque acting on it? (c) If an external agent turns the dipole until its electric dipole moment is

$$\vec{p} = (-4.00\hat{i} + 3.00\hat{j})(1.24 \times 10^{-30} \text{ C}\cdot\text{m}),$$

how much work is done by the agent?

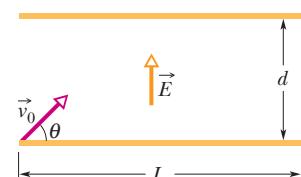


Figure 22-68 Problem 84.

84 In Fig. 22-68, a uniform, upward electric field \vec{E} of magnitude $2.00 \times 10^3 \text{ N/C}$ has been set up between two horizontal plates by charging the lower plate positively and the upper plate negatively. The plates have length $L = 10.0 \text{ cm}$ and separation $d = 2.00 \text{ cm}$. An electron is then shot between the plates from the left edge of the lower plate. The initial velocity v_0 of the electron makes an angle $\theta = 45.0^\circ$ with the lower plate and has a magnitude of $6.00 \times 10^6 \text{ m/s}$. (a) Will the electron strike one of the plates? (b) If so, which plate and how far horizontally from the left edge will the electron strike?

85 For the data of Problem 70, assume that the charge q on the drop is given by $q = ne$, where n is an integer and e is the elementary charge. (a) Find n for each given value of q . (b) Do a linear regression fit of the values of q versus the values of n and then use that fit to find e .

86 In Fig. 22-66, particle 1 (of charge $+2.00 \text{ pC}$), particle 2 (of charge -2.00 pC), and particle 3 (of charge $+5.00 \text{ pC}$) form an equilateral triangle of edge length $a = 9.50 \text{ cm}$. (a) Relative to the positive direction of the x axis, determine the direction of the force \vec{F}_3 on particle 3 due to the other particles by sketching electric field lines of the other particles. (b) Calculate the magnitude of \vec{F}_3 .

87 In Fig. 22-69, particle 1 of charge $q_1 = 1.00 \text{ pC}$ and particle 2 of charge $q_2 = -2.00 \text{ pC}$ are fixed at a distance $d = 5.00 \text{ cm}$ apart. In unit-vector notation, what is the net electric field at points (a) A , (b) B , and (c) C ? (d) Sketch the electric field lines.

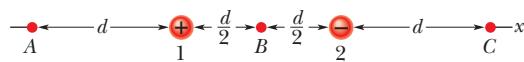


Figure 22-69 Problem 87.

Gauss' Law

23-1 ELECTRIC FLUX

Learning Objectives

After reading this module, you should be able to . . .

- 23.01** Identify that Gauss' law relates the electric field at points on a closed surface (real or imaginary, said to be a Gaussian surface) to the net charge enclosed by that surface.
- 23.02** Identify that the amount of electric field piercing a surface (not skimming along the surface) is the electric flux Φ through the surface.
- 23.03** Identify that an area vector for a flat surface is a vector that is perpendicular to the surface and that has a magnitude equal to the area of the surface.
- 23.04** Identify that any surface can be divided into area elements (patch elements) that are each small enough and flat enough for an area vector $d\vec{A}$ to be assigned to it, with the vector perpendicular to the element and having a magnitude equal to the area of the element.

Key Ideas

- The electric flux Φ through a surface is the amount of electric field that pierces the surface.
- The area vector $d\vec{A}$ for an area element (patch element) on a surface is a vector that is perpendicular to the element and has a magnitude equal to the area dA of the element.
- The electric flux $d\Phi$ through a patch element with area vector $d\vec{A}$ is given by a dot product:

$$d\Phi = \vec{E} \cdot d\vec{A}.$$

- The total flux through a surface is given by

- 23.05** Calculate the flux Φ through a surface by integrating the dot product of the electric field vector \vec{E} and the area vector $d\vec{A}$ (for patch elements) over the surface, in magnitude-angle notation and unit-vector notation.
- 23.06** For a closed surface, explain the algebraic signs associated with inward flux and outward flux.
- 23.07** Calculate the *net* flux Φ through a *closed* surface, algebraic sign included, by integrating the dot product of the electric field vector \vec{E} and the area vector $d\vec{A}$ (for patch elements) over the full surface.
- 23.08** Determine whether a closed surface can be broken up into parts (such as the sides of a cube) to simplify the integration that yields the net flux through the surface.

$$\Phi = \int \vec{E} \cdot d\vec{A} \quad (\text{total flux}),$$

where the integration is carried out over the surface.

- The net flux through a closed surface (which is used in Gauss' law) is given by

$$\Phi = \oint \vec{E} \cdot d\vec{A} \quad (\text{net flux}),$$

where the integration is carried out over the entire surface.

What Is Physics?

In the preceding chapter we found the electric field at points near extended charged objects, such as rods. Our technique was labor-intensive: We split the charge distribution up into charge elements dq , found the field $d\vec{E}$ due to an element, and resolved the vector into components. Then we determined whether the components from all the elements would end up canceling or adding. Finally we summed the adding components by integrating over all the elements, with several changes in notation along the way.

One of the primary goals of physics is to find simple ways of solving such labor-intensive problems. One of the main tools in reaching this goal is the use of symmetry. In this chapter we discuss a beautiful relationship between charge and

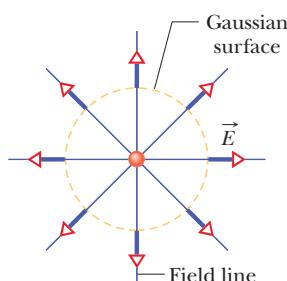


Figure 23-1 Electric field vectors and field lines pierce an imaginary, spherical Gaussian surface that encloses a particle with charge $+Q$.

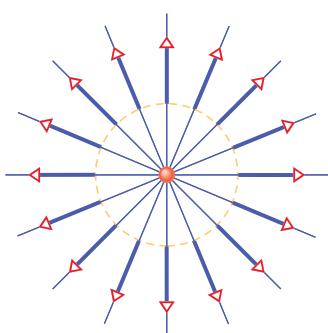


Figure 23-2 Now the enclosed particle has charge $+2Q$.

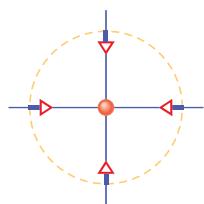


Figure 23-3 Can you tell what the enclosed charge is now?

electric field that allows us, in certain symmetric situations, to find the electric field of an extended charged object with a few lines of algebra. The relationship is called **Gauss' law**, which was developed by German mathematician and physicist Carl Friedrich Gauss (1777–1855).

Let's first take a quick look at some simple examples that give the spirit of Gauss' law. Figure 23-1 shows a particle with charge $+Q$ that is surrounded by an imaginary concentric sphere. At points on the sphere (said to be a *Gaussian surface*), the electric field vectors have a moderate magnitude (given by $E = kQ/r^2$) and point radially away from the particle (because it is positively charged). The electric field lines are also outward and have a moderate density (which, recall, is related to the field magnitude). We say that the field vectors and the field lines *pierce* the surface.

Figure 23-2 is similar except that the enclosed particle has charge $+2Q$. Because the enclosed charge is now twice as much, the magnitude of the field vectors piercing outward through the (same) Gaussian surface is twice as much as in Fig. 23-1, and the density of the field lines is also twice as much. That sentence, in a nutshell, is Gauss' law.



Gauss' law relates the electric field at points on a (closed) Gaussian surface to the net charge enclosed by that surface.

Let's check this with a third example with a particle that is also enclosed by the same spherical Gaussian surface (a *Gaussian sphere*, if you like, or even the catchy *G-sphere*) as shown in Fig. 23-3. What is the amount and sign of the enclosed charge? Well, from the inward piercing we see immediately that the charge must be negative. From the fact that the density of field lines is half that of Fig. 23-1, we also see that the charge must be $-0.5Q$. (Using Gauss' law is like being able to tell what is inside a gift box by looking at the wrapping paper on the box.)

The problems in this chapter are of two types. Sometimes we know the charge and we use Gauss' law to find the field at some point. Sometimes we know the field on a Gaussian surface and we use Gauss' law to find the charge enclosed by the surface. However, we cannot do all this by simply comparing the density of field lines in a drawing as we just did. We need a quantitative way of determining how much electric field pierces a surface. That measure is called the electric flux.

Electric Flux

Flat Surface, Uniform Field. We begin with a flat surface with area A in a uniform electric field \vec{E} . Figure 23-4a shows one of the electric field vectors \vec{E} piercing a small square patch with area ΔA (where Δ indicates “small”). Actually, only the x component (with magnitude $E_x = E \cos \theta$ in Fig. 23-4b) pierces the patch. The y component merely skims along the surface (no piercing in that) and does not come into play in Gauss' law. The *amount* of electric field piercing the patch is defined to be the **electric flux** $\Delta\Phi$ through it:

$$\Delta\Phi = (E \cos \theta) \Delta A.$$

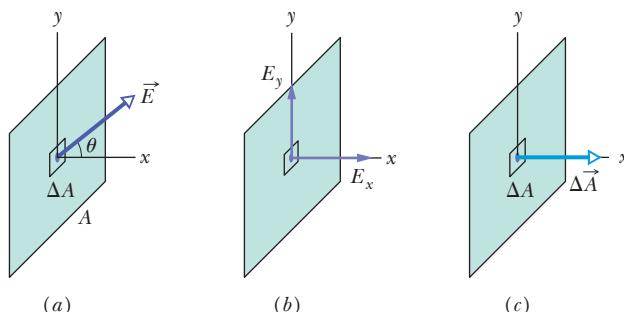


Figure 23-4 (a) An electric field vector pierces a small square patch on a flat surface. (b) Only the x component actually pierces the patch; the y component skims across it. (c) The area vector of the patch is perpendicular to the patch, with a magnitude equal to the patch's area.

There is another way to write the right side of this statement so that we have only the piercing component of \vec{E} . We define an area vector $\Delta\vec{A}$ that is perpendicular to the patch and that has a magnitude equal to the area ΔA of the patch (Fig. 23-4c). Then we can write

$$\Delta\Phi = \vec{E} \cdot \Delta\vec{A},$$

and the dot product automatically gives us the component of \vec{E} that is parallel to $\Delta\vec{A}$ and thus piercing the patch.

To find the total flux Φ through the surface in Fig. 23-4, we sum the flux through every patch on the surface:

$$\Phi = \sum \vec{E} \cdot \Delta\vec{A}. \quad (23-1)$$

However, because we do not want to sum hundreds (or more) flux values, we transform the summation into an integral by shrinking the patches from small squares with area ΔA to *patch elements* (or *area elements*) with area dA . The total flux is then

$$\Phi = \int \vec{E} \cdot d\vec{A} \quad (\text{total flux}). \quad (23-2)$$

Now we can find the total flux by integrating the dot product over the full surface.

Dot Product. We can evaluate the dot product inside the integral by writing the two vectors in unit-vector notation. For example, in Fig. 23-4, $d\vec{A} = dA\hat{i}$ and \vec{E} might be, say, $(4\hat{i} + 4\hat{j})$ N/C. Instead, we can evaluate the dot product in magnitude-angle notation: $E \cos \theta dA$. When the electric field is uniform and the surface is flat, the product $E \cos \theta$ is a constant and comes outside the integral. The remaining $\int dA$ is just an instruction to sum the areas of all the patch elements to get the total area, but we already know that the total area is A . So the total flux in this simple situation is

$$\Phi = (E \cos \theta)A \quad (\text{uniform field, flat surface}). \quad (23-3)$$

Closed Surface. To use Gauss' law to relate flux and charge, we need a closed surface. Let's use the closed surface in Fig. 23-5 that sits in a nonuniform electric field. (Don't worry. The homework problems involve less complex surfaces.) As before, we first consider the flux through small square patches. However, now we are interested in not only the piercing components of the field but also on whether the piercing is inward or outward (just as we did with Figs. 23-1 through 23-3).

Directions. To keep track of the piercing direction, we again use an area vector $\Delta\vec{A}$ that is perpendicular to a patch, but now we always draw it pointing outward from the surface (*away from the interior*). Then if a field vector pierces outward, it and the area vector are in the same direction, the angle is $\theta = 0$, and $\cos \theta = 1$. Thus, the dot product $\vec{E} \cdot \Delta\vec{A}$ is positive and so is the flux. Conversely, if a field vector pierces inward, the angle is $\theta = 180^\circ$ and $\cos \theta = -1$. Thus, the dot product is negative and so is the flux. If a field vector skims the surface (no piercing), the dot product is zero (because $\cos 90^\circ = 0$) and so is the flux. Figure 23-5 gives some general examples and here is a summary:



An inward piercing field is negative flux. An outward piercing field is positive flux. A skimming field is zero flux.

Net Flux. In principle, to find the **net flux** through the surface in Fig. 23-5, we find the flux at every patch and then sum the results (with the algebraic signs included). However, we are not about to do that much work. Instead, we shrink the squares to patch elements with area vectors $d\vec{A}$ and then integrate:

$$\Phi = \oint \vec{E} \cdot d\vec{A} \quad (\text{net flux}). \quad (23-4)$$

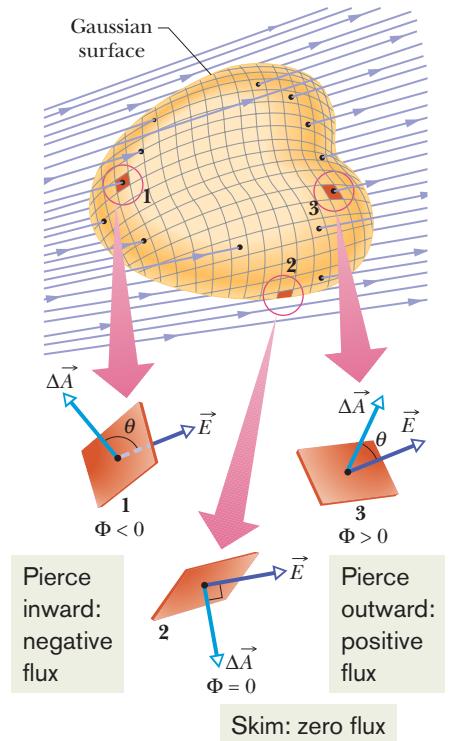


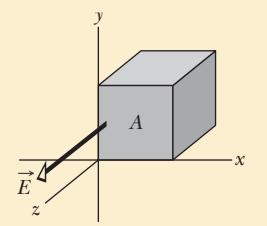
Figure 23-5 A Gaussian surface of arbitrary shape immersed in an electric field. The surface is divided into small squares of area ΔA . The electric field vectors \vec{E} and the area vectors $\Delta\vec{A}$ for three representative squares, marked 1, 2, and 3, are shown.

The loop on the integral sign indicates that we must integrate over the entire closed surface, to get the *net* flux through the surface (as in Fig. 23-5, flux might enter on one side and leave on another side). Keep in mind that we want to determine the net flux through a surface because that is what Gauss' law relates to the charge enclosed by the surface. (The law is coming up next.) Note that flux is a scalar (yes, we talk about field vectors but flux is the *amount* of piercing field, not a vector itself). The SI unit of flux is the newton-square-meter per coulomb ($N \cdot m^2/C$).



Checkpoint 1

The figure here shows a Gaussian cube of face area A immersed in a uniform electric field \vec{E} that has the positive direction of the z axis. In terms of E and A , what is the flux through (a) the front face (which is in the xy plane), (b) the rear face, (c) the top face, and (d) the whole cube?



Sample Problem 23.01 Flux through a closed cylinder, uniform field

Figure 23-6 shows a Gaussian surface in the form of a closed cylinder (a Gaussian cylinder or G-cylinder) of radius R . It lies in a uniform electric field \vec{E} with the cylinder's central axis (along the length of the cylinder) parallel to the field. What is the net flux Φ of the electric field through the cylinder?

KEY IDEAS

We can find the net flux Φ with Eq. 23-4 by integrating the dot product $\vec{E} \cdot d\vec{A}$ over the cylinder's surface. However, we cannot write out functions so that we can do that with one integral. Instead, we need to be a bit clever: We break up the surface into sections with which we can actually evaluate an integral.

Calculations: We break the integral of Eq. 23-4 into three terms: integrals over the left cylinder cap a , the curved cylindrical surface b , and the right cap c :

$$\begin{aligned}\Phi &= \oint \vec{E} \cdot d\vec{A} \\ &= \int_a \vec{E} \cdot d\vec{A} + \int_b \vec{E} \cdot d\vec{A} + \int_c \vec{E} \cdot d\vec{A}. \quad (23-5)\end{aligned}$$

Pick a patch element on the left cap. Its area vector $d\vec{A}$ must be perpendicular to the patch and pointing away from the interior of the cylinder. In Fig. 23-6, that means the angle between it and the field piercing the patch is 180° . Also, note that the electric field through the end cap is uniform and thus E can be pulled out of the integration. So, we can write the flux through the left cap as

$$\int_a \vec{E} \cdot d\vec{A} = \int E(\cos 180^\circ) dA = -E \int dA = -EA,$$

where $\int dA$ gives the cap's area A ($= \pi R^2$). Similarly, for the right cap, where $\theta = 0$ for all points,

$$\int_c \vec{E} \cdot d\vec{A} = \int E(\cos 0) dA = EA.$$

Finally, for the cylindrical surface, where the angle θ is 90° at all points,

$$\int_b \vec{E} \cdot d\vec{A} = \int E(\cos 90^\circ) dA = 0.$$

Substituting these results into Eq. 23-5 leads us to

$$\Phi = -EA + 0 + EA = 0. \quad (\text{Answer})$$

The net flux is zero because the field lines that represent the electric field all pass entirely through the Gaussian surface, from the left to the right.

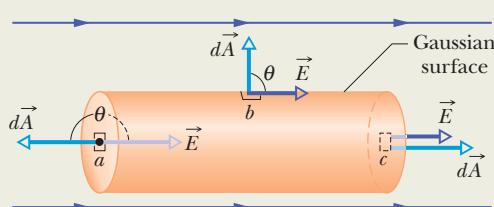


Figure 23-6 A cylindrical Gaussian surface, closed by end caps, is immersed in a uniform electric field. The cylinder axis is parallel to the field direction.



Additional examples, video, and practice available at WileyPLUS

Sample Problem 23.02 Flux through a closed cube, nonuniform field

A *nonuniform* electric field given by $\vec{E} = 3.0x\hat{i} + 4.0\hat{j}$ pierces the Gaussian cube shown in Fig. 23-7a. (E is in newtons per coulomb and x is in meters.) What is the electric flux through the right face, the left face, and the top face? (We consider the other faces in another sample problem.)

KEY IDEA

We can find the flux Φ through the surface by integrating the scalar product $\vec{E} \cdot d\vec{A}$ over each face.

Right face: An area vector \vec{A} is always perpendicular to its surface and always points away from the interior of a Gaussian surface. Thus, the vector $d\vec{A}$ for any patch element (small section) on the right face of the cube must point in the positive direction of the x axis. An example of such an element is shown in Figs. 23-7b and c, but we would have an identical vector for any other choice of a patch element on that face. The most convenient way to express the vector is in unit-vector notation,

$$d\vec{A} = dA\hat{i}.$$

From Eq. 23-4, the flux Φ_r through the right face is then

$$\begin{aligned}\Phi_r &= \int \vec{E} \cdot d\vec{A} = \int (3.0x\hat{i} + 4.0\hat{j}) \cdot (dA\hat{i}) \\ &= \int [(3.0x)(dA)\hat{i} \cdot \hat{i} + (4.0)(dA)\hat{j} \cdot \hat{i}] \\ &= \int (3.0x dA + 0) = 3.0 \int x dA.\end{aligned}$$

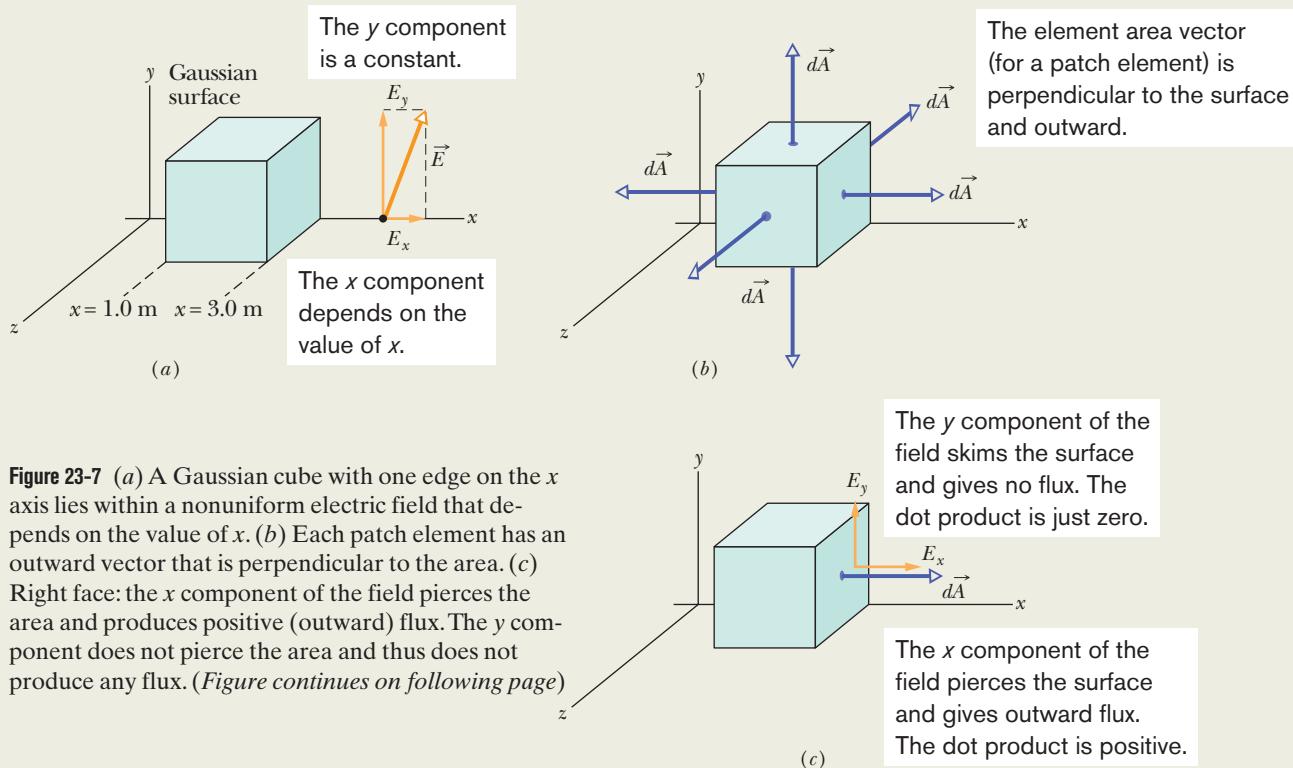
We are about to integrate over the right face, but we note that x has the same value everywhere on that face—namely, $x = 3.0$ m. This means we can substitute that constant value for x . This can be a confusing argument. Although x is certainly a variable as we move left to right across the figure, because the right face is perpendicular to the x axis, every point on the face has the same x coordinate. (The y and z coordinates do not matter in our integral.) Thus, we have

$$\Phi_r = 3.0 \int (3.0) dA = 9.0 \int dA.$$

The integral $\int dA$ merely gives us the area $A = 4.0$ m² of the right face, so

$$\Phi_r = (9.0 \text{ N/C})(4.0 \text{ m}^2) = 36 \text{ N} \cdot \text{m}^2/\text{C}. \quad (\text{Answer})$$

Left face: We repeat this procedure for the left face. However,



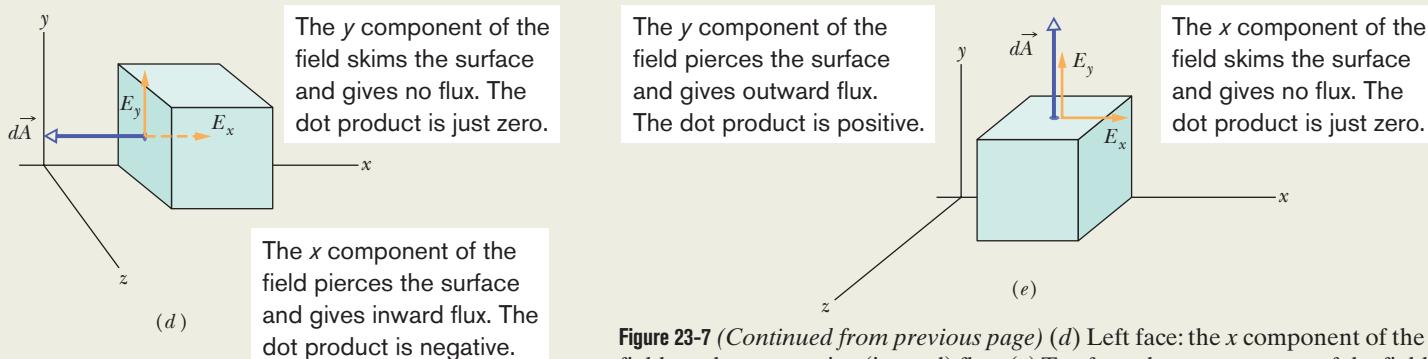


Figure 23-7 (Continued from previous page) (d) Left face: the x component of the field produces negative (inward) flux. (e) Top face: the y component of the field produces positive (outward) flux.

two factors change. (1) The element area vector $d\vec{A}$ points in the negative direction of the x axis, and thus $d\vec{A} = -dA\hat{i}$ (Fig. 23-7d). (2) On the left face, $x = 1.0 \text{ m}$. With these changes, we find that the flux Φ_l through the left face is

$$\Phi_l = -12 \text{ N} \cdot \text{m}^2/\text{C}. \quad (\text{Answer})$$

Top face: Now $d\vec{A}$ points in the positive direction of the y axis, and thus $d\vec{A} = dA\hat{j}$ (Fig. 23-7e). The flux Φ_t is

$$\begin{aligned}\Phi_t &= \int (3.0x\hat{i} + 4.0\hat{j}) \cdot (dA\hat{j}) \\ &= \int [(3.0x)(dA)\hat{i} \cdot \hat{j} + (4.0)(dA)\hat{j} \cdot \hat{j}] \\ &= \int (0 + 4.0 dA) = 4.0 \int dA \\ &= 16 \text{ N} \cdot \text{m}^2/\text{C}. \quad (\text{Answer})\end{aligned}$$



Additional examples, video, and practice available at WileyPLUS

23-2 GAUSS' LAW

Learning Objectives

After reading this module, you should be able to ...

- 23.09** Apply Gauss' law to relate the net flux Φ through a closed surface to the net enclosed charge q_{enc} .
- 23.10** Identify how the algebraic sign of the net enclosed charge corresponds to the direction (inward or outward) of the net flux through a Gaussian surface.
- 23.11** Identify that charge outside a Gaussian surface makes

Key Ideas

- Gauss' law relates the net flux Φ penetrating a closed surface to the net charge q_{enc} enclosed by the surface:

$$\epsilon_0\Phi = q_{\text{enc}} \quad (\text{Gauss' law}).$$

no contribution to the *net* flux through the closed surface.

- 23.12** Derive the expression for the magnitude of the electric field of a charged particle by using Gauss' law.
- 23.13** Identify that for a charged particle or uniformly charged sphere, Gauss' law is applied with a Gaussian surface that is a concentric sphere.

- Gauss' law can also be written in terms of the electric field piercing the enclosing Gaussian surface:

$$\epsilon_0 \oint \vec{E} \cdot d\vec{A} = q_{\text{enc}} \quad (\text{Gauss' law}).$$

Gauss' Law

Gauss' law relates the net flux Φ of an electric field through a closed surface (a Gaussian surface) to the *net* charge q_{enc} that is *enclosed* by that surface. It tells us that

$$\epsilon_0\Phi = q_{\text{enc}} \quad (\text{Gauss' law}).$$

(23-6)

By substituting Eq. 23-4, the definition of flux, we can also write Gauss' law as

$$\epsilon_0 \oint \vec{E} \cdot d\vec{A} = q_{\text{enc}} \quad (\text{Gauss' law}). \quad (23-7)$$

Equations 23-6 and 23-7 hold only when the net charge is located in a vacuum or (what is the same for most practical purposes) in air. In Chapter 25, we modify Gauss' law to include situations in which a material such as mica, oil, or glass is present.

In Eqs. 23-6 and 23-7, the net charge q_{enc} is the algebraic sum of all the *enclosed* positive and negative charges, and it can be positive, negative, or zero. We include the sign, rather than just use the magnitude of the enclosed charge, because the sign tells us something about the net flux through the Gaussian surface: If q_{enc} is positive, the net flux is *outward*; if q_{enc} is negative, the net flux is *inward*.

Charge outside the surface, no matter how large or how close it may be, is not included in the term q_{enc} in Gauss' law. The exact form and location of the charges inside the Gaussian surface are also of no concern; the only things that matter on the right side of Eqs. 23-6 and 23-7 are the magnitude and sign of the net enclosed charge. The quantity \vec{E} on the left side of Eq. 23-7, however, is the electric field resulting from *all* charges, both those inside and those outside the Gaussian surface. This statement may seem to be inconsistent, but keep this in mind: The electric field due to a charge outside the Gaussian surface contributes zero net flux *through* the surface, because as many field lines due to that charge enter the surface as leave it.

Let us apply these ideas to Fig. 23-8, which shows two particles, with charges equal in magnitude but opposite in sign, and the field lines describing their electric fields. Four Gaussian surfaces are also shown in cross section. Let us consider each in turn.

Surface S_1 . The electric field is outward for all points on this surface. Thus, the flux of the electric field through this surface is positive, and so is the net charge within the surface, as Gauss' law requires. (That is, in Eq. 23-6, if Φ is positive, q_{enc} must be also.)

Surface S_2 . The electric field is inward for all points on this surface. Thus, the flux of the electric field through this surface is negative and so is the enclosed charge, as Gauss' law requires.

Surface S_3 . This surface encloses no charge, and thus $q_{\text{enc}} = 0$. Gauss' law (Eq. 23-6) requires that the net flux of the electric field through this surface be zero. That is reasonable because all the field lines pass entirely through the surface, entering it at the top and leaving at the bottom.

Surface S_4 . This surface encloses no *net* charge, because the enclosed positive and negative charges have equal magnitudes. Gauss' law requires that the net flux of the electric field through this surface be zero. That is reasonable because there are as many field lines leaving surface S_4 as entering it.

What would happen if we were to bring an enormous charge Q up close to surface S_4 in Fig. 23-8? The pattern of the field lines would certainly change, but the net flux for each of the four Gaussian surfaces would not change. Thus, the value of Q would not enter Gauss' law in any way, because Q lies outside all four of the Gaussian surfaces that we are considering.



Checkpoint 2

The figure shows three situations in which a Gaussian cube sits in an electric field. The arrows and the values indicate the directions of the field lines and the magnitudes (in $\text{N} \cdot \text{m}^2/\text{C}$) of the flux through the six sides of each cube. (The lighter arrows are for the hidden faces.) In which situation does the cube enclose (a) a positive net charge, (b) a negative net charge, and (c) zero net charge?

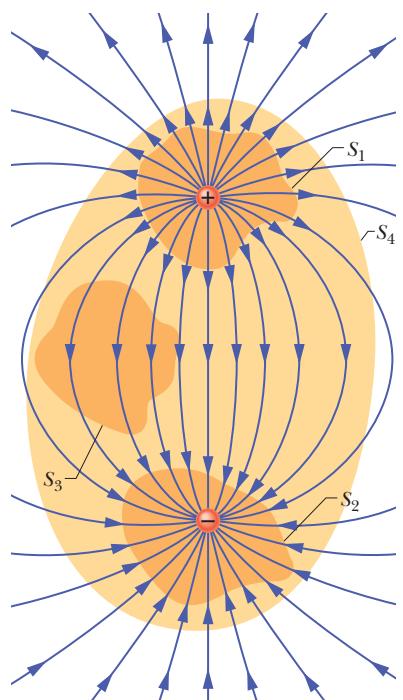
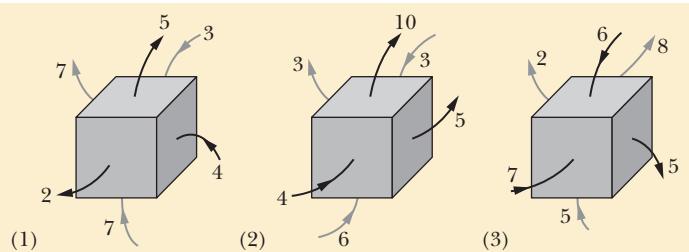


Figure 23-8 Two charges, equal in magnitude but opposite in sign, and the field lines that represent their net electric field. Four Gaussian surfaces are shown in cross section. Surface S_1 encloses the positive charge. Surface S_2 encloses the negative charge. Surface S_3 encloses no charge. Surface S_4 encloses both charges and thus no net charge.



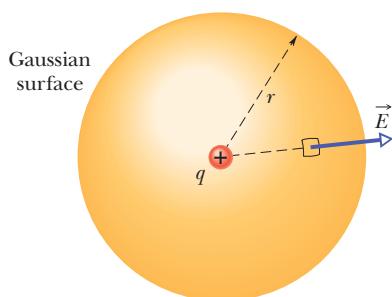


Figure 23-9 A spherical Gaussian surface centered on a particle with charge q .

Gauss' Law and Coulomb's Law

One of the situations in which we can apply Gauss' law is in finding the electric field of a charged particle. That field has spherical symmetry (the field depends on the distance r from the particle but not the direction). So, to make use of that symmetry, we enclose the particle in a Gaussian sphere that is centered on the particle, as shown in Fig. 23-9 for a particle with positive charge q . Then the electric field has the same magnitude E at any point on the sphere (all points are at the same distance r). That feature will simplify the integration.

The drill here is the same as previously. Pick a patch element on the surface and draw its area vector $d\vec{A}$ perpendicular to the patch and directed outward. From the symmetry of the situation, we know that the electric field \vec{E} at the patch is also radially outward and thus at angle $\theta = 0$ with $d\vec{A}$. So, we rewrite Gauss' law as

$$\epsilon_0 \oint \vec{E} \cdot d\vec{A} = \epsilon_0 \oint E dA = q_{\text{enc}}. \quad (23-8)$$

Here $q_{\text{enc}} = q$. Because the field magnitude E is the same at every patch element, E can be pulled outside the integral:

$$\epsilon_0 E \oint dA = q. \quad (23-9)$$

The remaining integral is just an instruction to sum all the areas of the patch elements on the sphere, but we already know that the total area is $4\pi r^2$. Substituting this, we have

$$\begin{aligned} \epsilon_0 E (4\pi r^2) &= q \\ \text{or} \quad E &= \frac{1}{4\pi\epsilon_0} \frac{q}{r^2}. \end{aligned} \quad (23-10)$$

This is exactly Eq. 22-3, which we found using Coulomb's law.



Checkpoint 3

There is a certain net flux Φ_i through a Gaussian sphere of radius r enclosing an isolated charged particle. Suppose the enclosing Gaussian surface is changed to (a) a larger Gaussian sphere, (b) a Gaussian cube with edge length equal to r , and (c) a Gaussian cube with edge length equal to $2r$. In each case, is the net flux through the new Gaussian surface greater than, less than, or equal to Φ_i ?

Sample Problem 23.03 Using Gauss' law to find the electric field

Figure 23-10a shows, in cross section, a plastic, spherical shell with uniform charge $Q = -16e$ and radius $R = 10$ cm. A particle with charge $q = +5e$ is at the center. What is the electric field (magnitude and direction) at (a) point P_1 at radial distance $r_1 = 6.00$ cm and (b) point P_2 at radial distance $r_2 = 12.0$ cm?

KEY IDEAS

- (1) Because the situation in Fig. 23-10a has spherical symmetry, we can apply Gauss' law (Eq. 23-7) to find the electric field at a point if we use a Gaussian surface in the form of a sphere concentric with the particle and shell.
- (2) To find the electric field at a point, we put that point on a Gaussian surface (so that the \vec{E} we want is the \vec{E} in the dot product inside the integral in Gauss' law).
- (3) Gauss' law relates the net electric flux through a closed surface to the net enclosed charge. Any external charge is not included.

Calculations: To find the field at point P_1 , we construct a Gaussian sphere with P_1 on its surface and thus with a radius of r_1 . Because the charge enclosed by the Gaussian sphere is positive, the electric flux through the surface must be positive and thus outward. So, the electric field \vec{E} pierces the surface outward and, because of the spherical symmetry, must be *radially* outward, as drawn in Fig. 23-10b. That figure does not include the plastic shell because the shell is not enclosed by the Gaussian sphere.

Consider a patch element on the sphere at P_1 . Its area vector $d\vec{A}$ is radially outward (it must always be outward from a Gaussian surface). Thus the angle θ between \vec{E} and $d\vec{A}$ is zero. We can now rewrite the left side of Eq. 23-7 (Gauss' law) as

$$\epsilon_0 \oint \vec{E} \cdot d\vec{A} = \epsilon_0 \oint E \cos 0 dA = \epsilon_0 \oint E dA = \epsilon_0 E \oint dA,$$

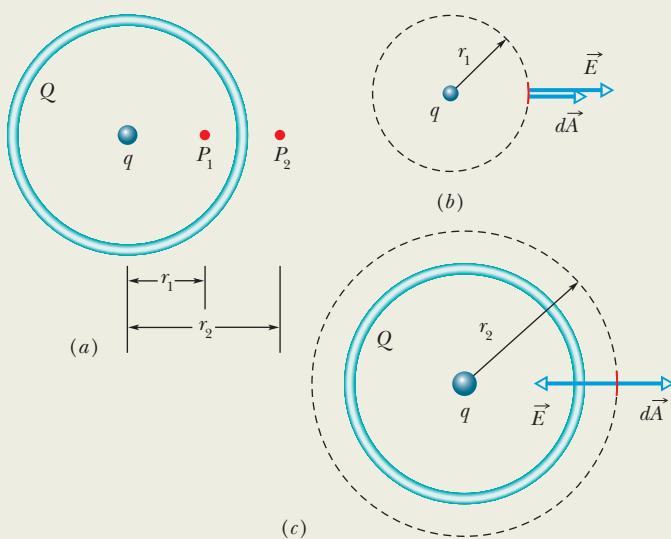


Figure 23-10 (a) A charged plastic spherical shell encloses a charged particle. (b) To find the electric field at P_1 , arrange for the point to be on a Gaussian sphere. The electric field pierces outward. The area vector for the patch element is outward. (c) P_2 is on a Gaussian sphere, \vec{E} is inward, and $d\vec{A}$ is still outward.

where in the last step we pull the field magnitude E out of the integral because it is the same at all points on the Gaussian sphere and thus is a constant. The remaining integral is simply an instruction for us to sum the areas of all the patch elements on the sphere, but we already know that the surface area of a sphere is $4\pi r^2$. Substituting these results, Eq. 23-7 for Gauss' law gives us

$$\epsilon_0 E 4\pi r^2 = q_{\text{enc}}$$

Sample Problem 23.04 Using Gauss' law to find the enclosed charge

What is the net charge enclosed by the Gaussian cube of Sample Problem 23.02?

KEY IDEA

The net charge enclosed by a (real or mathematical) closed surface is related to the total electric flux through the surface by Gauss' law as given by Eq. 23-6 ($\epsilon_0 \Phi = q_{\text{enc}}$).

Flux: To use Eq. 23-6, we need to know the flux through all six faces of the cube. We already know the flux through the right face ($\Phi_r = 36 \text{ N} \cdot \text{m}^2/\text{C}$), the left face ($\Phi_l = -12 \text{ N} \cdot \text{m}^2/\text{C}$), and the top face ($\Phi_t = 16 \text{ N} \cdot \text{m}^2/\text{C}$).

For the bottom face, our calculation is just like that for the top face *except* that the element area vector $d\vec{A}$ is now directed downward along the y axis (recall, it must be *outward* from the Gaussian enclosure). Thus, we have

The only charge enclosed by the Gaussian surface through P_1 is that of the particle. Solving for E and substituting $q_{\text{enc}} = 5e$ and $r = r_1 = 6.00 \times 10^{-2} \text{ m}$, we find that the magnitude of the electric field at P_1 is

$$\begin{aligned} E &= \frac{q_{\text{enc}}}{4\pi\epsilon_0 r^2} \\ &= \frac{5(1.60 \times 10^{-19} \text{ C})}{4\pi(8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2)(0.0600 \text{ m})^2} \\ &= 2.00 \times 10^{-6} \text{ N/C.} \end{aligned} \quad (\text{Answer})$$

To find the electric field at P_2 , we follow the same procedure by constructing a Gaussian sphere with P_2 on its surface. This time, however, the net charge enclosed by the sphere is $q_{\text{enc}} = q + Q = 5e + (-16e) = -11e$. Because the net charge is negative, the electric field vectors on the sphere's surface pierce inward (Fig. 23-10c), the angle θ between \vec{E} and $d\vec{A}$ is 180° , and the dot product is $E(\cos 180^\circ) dA = -E dA$. Now solving Gauss' law for E and substituting $r = r_2 = 12.00 \times 10^{-2} \text{ m}$ and the new q_{enc} , we find

$$\begin{aligned} E &= \frac{-q_{\text{enc}}}{4\pi\epsilon_0 r^2} \\ &= \frac{-[-11(1.60 \times 10^{-19} \text{ C})]}{4\pi(8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2)(0.120 \text{ m})^2} \\ &= 1.10 \times 10^{-6} \text{ N/C.} \end{aligned} \quad (\text{Answer})$$

Note how different the calculations would have been if we had put P_1 or P_2 on the surface of a Gaussian cube instead of mimicking the spherical symmetry with a Gaussian sphere. Then angle θ and magnitude E would have varied considerably over the surface of the cube and evaluation of the integral in Gauss' law would have been difficult.

$$d\vec{A} = -dA\hat{j}, \text{ and we find}$$

$$\Phi_b = -16 \text{ N} \cdot \text{m}^2/\text{C}.$$

For the front face we have $d\vec{A} = dA\hat{k}$, and for the back face, $d\vec{A} = -dA\hat{k}$. When we take the dot product of the given electric field $\vec{E} = 3.0\hat{x} + 4.0\hat{j}$ with either of these expressions for $d\vec{A}$, we get 0 and thus there is no flux through those faces. We can now find the total flux through the six sides of the cube:

$$\begin{aligned} \Phi &= (36 - 12 + 16 - 16 + 0 + 0) \text{ N} \cdot \text{m}^2/\text{C} \\ &= 24 \text{ N} \cdot \text{m}^2/\text{C}. \end{aligned}$$

Enclosed charge: Next, we use Gauss' law to find the charge q_{enc} enclosed by the cube:

$$\begin{aligned} q_{\text{enc}} &= \epsilon_0 \Phi = (8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2)(24 \text{ N} \cdot \text{m}^2/\text{C}) \\ &= 2.1 \times 10^{-10} \text{ C.} \end{aligned} \quad (\text{Answer})$$

Thus, the cube encloses a *net* positive charge.

23-3 A CHARGED ISOLATED CONDUCTOR

Learning Objectives

After reading this module, you should be able to ...

23.14 Apply the relationship between surface charge density σ and the area over which the charge is uniformly spread.

23.15 Identify that if excess charge (positive or negative) is placed on an isolated conductor, that charge moves to the surface and none is in the interior.

23.16 Identify the value of the electric field inside an isolated conductor.

23.17 For a conductor with a cavity that contains a charged

object, determine the charge on the cavity wall and on the external surface.

23.18 Explain how Gauss' law is used to find the electric field magnitude E near an isolated conducting surface with a uniform surface charge density σ .

23.19 For a uniformly charged conducting surface, apply the relationship between the charge density σ and the electric field magnitude E at points near the conductor, and identify the direction of the field vectors.

Key Ideas

- An excess charge on an isolated conductor is located entirely on the outer surface of the conductor.

- The internal electric field of a charged, isolated conductor is zero, and the external field (at nearby points) is perpendicular to the surface and has a magnitude that depends on the surface charge density σ :

$$E = \frac{\sigma}{\epsilon_0}$$

A Charged Isolated Conductor

Gauss' law permits us to prove an important theorem about conductors:



If an excess charge is placed on an isolated conductor, that amount of charge will move entirely to the surface of the conductor. None of the excess charge will be found within the body of the conductor.

This might seem reasonable, considering that charges with the same sign repel one another. You might imagine that, by moving to the surface, the added charges are getting as far away from one another as they can. We turn to Gauss' law for verification of this speculation.

Figure 23-11a shows, in cross section, an isolated lump of copper hanging from an insulating thread and having an excess charge q . We place a Gaussian surface just inside the actual surface of the conductor.

The electric field inside this conductor must be zero. If this were not so, the field would exert forces on the conduction (free) electrons, which are always present in a conductor, and thus current would always exist within a conductor. (That is, charge would flow from place to place within the conductor.) Of course, there is no such perpetual current in an isolated conductor, and so the internal electric field is zero.

(An internal electric field *does* appear as a conductor is being charged. However, the added charge quickly distributes itself in such a way that the net internal electric field—the vector sum of the electric fields due to all the charges, both inside and outside—is zero. The movement of charge then ceases, because the net force on each charge is zero; the charges are then in *electrostatic equilibrium*.)

If \vec{E} is zero everywhere inside our copper conductor, it must be zero for all points on the Gaussian surface because that surface, though close to the surface of the conductor, is definitely inside the conductor. This means that the flux through the Gaussian surface must be zero. Gauss' law then tells us that the net charge inside the Gaussian surface must also be zero. Then because the excess charge is not inside the Gaussian surface, it must be outside that surface, which means it must lie on the actual surface of the conductor.

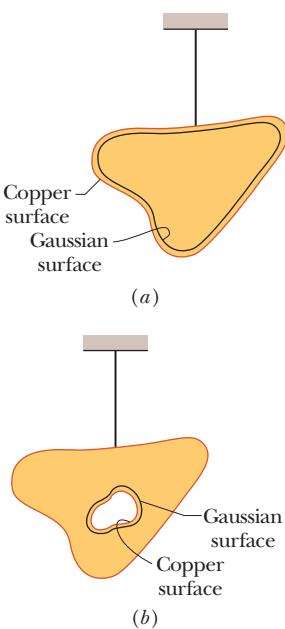


Figure 23-11 (a) A lump of copper with a charge q hangs from an insulating thread. A Gaussian surface is placed within the metal, just inside the actual surface. (b) The lump of copper now has a cavity within it. A Gaussian surface lies within the metal, close to the cavity surface.

An Isolated Conductor with a Cavity

Figure 23-11b shows the same hanging conductor, but now with a cavity that is totally within the conductor. It is perhaps reasonable to suppose that when we scoop out the electrically neutral material to form the cavity, we do not change the distribution of charge or the pattern of the electric field that exists in Fig. 23-11a. Again, we must turn to Gauss' law for a quantitative proof.

We draw a Gaussian surface surrounding the cavity, close to its surface but inside the conducting body. Because $\vec{E} = 0$ inside the conductor, there can be no flux through this new Gaussian surface. Therefore, from Gauss' law, that surface can enclose no net charge. We conclude that there is no net charge on the cavity walls; all the excess charge remains on the outer surface of the conductor, as in Fig. 23-11a.

The Conductor Removed

Suppose that, by some magic, the excess charges could be "frozen" into position on the conductor's surface, perhaps by embedding them in a thin plastic coating, and suppose that then the conductor could be removed completely. This is equivalent to enlarging the cavity of Fig. 23-11b until it consumes the entire conductor, leaving only the charges. The electric field would not change at all; it would remain zero inside the thin shell of charge and would remain unchanged for all external points. This shows us that the electric field is set up by the charges and not by the conductor. The conductor simply provides an initial pathway for the charges to take up their positions.

The External Electric Field

You have seen that the excess charge on an isolated conductor moves entirely to the conductor's surface. However, unless the conductor is spherical, the charge does not distribute itself uniformly. Put another way, the surface charge density σ (charge per unit area) varies over the surface of any nonspherical conductor. Generally, this variation makes the determination of the electric field set up by the surface charges very difficult.

However, the electric field just outside the surface of a conductor is easy to determine using Gauss' law. To do this, we consider a section of the surface that is small enough to permit us to neglect any curvature and thus to take the section to be flat. We then imagine a tiny cylindrical Gaussian surface to be partially embedded in the section as shown in Fig. 23-12: One end cap is fully inside the conductor, the other is fully outside, and the cylinder is perpendicular to the conductor's surface.

The electric field \vec{E} at and just outside the conductor's surface must also be perpendicular to that surface. If it were not, then it would have a component along the conductor's surface that would exert forces on the surface charges, causing them to move. However, such motion would violate our implicit assumption that we are dealing with electrostatic equilibrium. Therefore, \vec{E} is perpendicular to the conductor's surface.

We now sum the flux through the Gaussian surface. There is no flux through the internal end cap, because the electric field within the conductor is zero. There is no flux through the curved surface of the cylinder, because internally (in the conductor) there is no electric field and externally the electric field is parallel to the curved portion of the Gaussian surface. The only flux through the Gaussian surface is that through the external end cap, where \vec{E} is perpendicular to the plane of the cap. We assume that the cap area A is small enough that the field magnitude E is constant over the cap. Then the flux through the cap is EA , and that is the net flux Φ through the Gaussian surface.

The charge q_{enc} enclosed by the Gaussian surface lies on the conductor's surface in an area A . (Think of the cylinder as a cookie cutter.) If σ is the charge per unit area, then q_{enc} is equal to σA . When we substitute σA for q_{enc} and EA for Φ ,

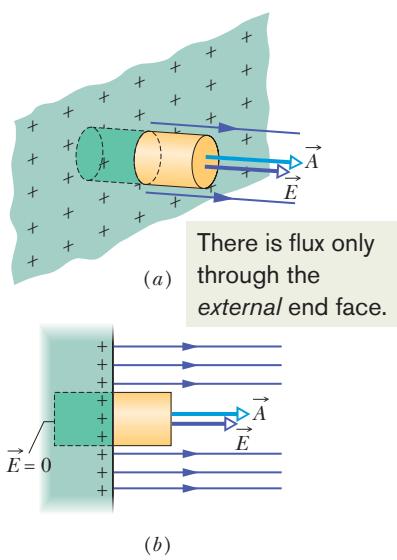


Figure 23-12 (a) Perspective view and (b) side view of a tiny portion of a large, isolated conductor with excess positive charge on its surface. A (closed) cylindrical Gaussian surface, embedded perpendicularly in the conductor, encloses some of the charge. Electric field lines pierce the external end cap of the cylinder, but not the internal end cap. The external end cap has area A and area vector \vec{A} .

Gauss' law (Eq. 23-6) becomes

$$\epsilon_0 EA = \sigma A,$$

from which we find

$$E = \frac{\sigma}{\epsilon_0} \quad (\text{conducting surface}). \quad (23-11)$$

Thus, the magnitude of the electric field just outside a conductor is proportional to the surface charge density on the conductor. The sign of the charge gives us the direction of the field. If the charge on the conductor is positive, the electric field is directed away from the conductor as in Fig. 23-12. It is directed toward the conductor if the charge is negative.

The field lines in Fig. 23-12 must terminate on negative charges somewhere in the environment. If we bring those charges near the conductor, the charge density at any given location on the conductor's surface changes, and so does the magnitude of the electric field. However, the relation between σ and E is still given by Eq. 23-11.

Sample Problem 23.05 Spherical metal shell, electric field and enclosed charge

Figure 23-13a shows a cross section of a spherical metal shell of inner radius R . A particle with a charge of $-5.0 \mu\text{C}$ is located at a distance $R/2$ from the center of the shell. If the shell is electrically neutral, what are the (induced) charges on its inner and outer surfaces? Are those charges uniformly distributed? What is the field pattern inside and outside the shell?

KEY IDEAS

Figure 23-13b shows a cross section of a spherical Gaussian surface within the metal, just outside the inner wall of the shell. The electric field must be zero inside the metal (and thus on the Gaussian surface inside the metal). This means that the electric flux through the Gaussian surface must also be zero. Gauss' law then tells us that the *net* charge enclosed by the Gaussian surface must be zero.

Reasoning: With a particle of charge $-5.0 \mu\text{C}$ within the shell, a charge of $+5.0 \mu\text{C}$ must lie on the inner wall of the shell in order that the net enclosed charge be zero. If the particle were centered, this positive charge would be uniformly distributed along the inner wall. However, since the particle is off-center, the distribution of positive charge is skewed, as suggested by Fig. 23-13b, because the positive charge tends to collect on the section of the inner wall nearest the (negative) particle.

Because the shell is electrically neutral, its inner wall can have a charge of $+5.0 \mu\text{C}$ only if electrons, with a total charge of $-5.0 \mu\text{C}$, leave the inner wall and move to the outer wall. There they spread out uniformly, as is also suggested by Fig. 23-13b. This distribution of negative charge is

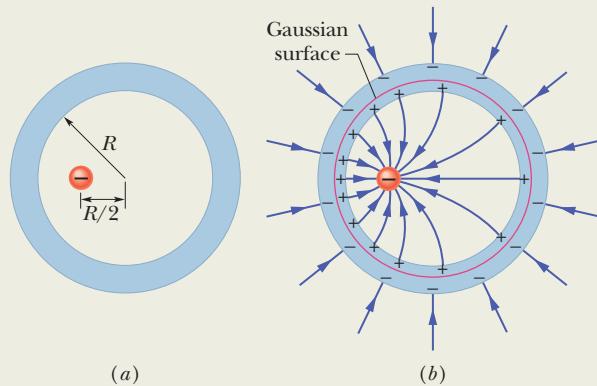


Figure 23-13 (a) A negatively charged particle is located within a spherical metal shell that is electrically neutral. (b) As a result, positive charge is nonuniformly distributed on the inner wall of the shell, and an equal amount of negative charge is uniformly distributed on the outer wall.

uniform because the shell is spherical and because the skewed distribution of positive charge on the inner wall cannot produce an electric field in the shell to affect the distribution of charge on the outer wall. Furthermore, these negative charges repel one another.

The field lines inside and outside the shell are shown approximately in Fig. 23-13b. All the field lines intersect the shell and the particle perpendicularly. Inside the shell the pattern of field lines is skewed because of the skew of the positive charge distribution. Outside the shell the pattern is the same as if the particle were centered and the shell were missing. In fact, this would be true no matter where inside the shell the particle happened to be located.

23-4 APPLYING GAUSS' LAW: CYLINDRICAL SYMMETRY

Learning Objectives

After reading this module, you should be able to . . .

23.20 Explain how Gauss' law is used to derive the electric field magnitude outside a line of charge or a cylindrical surface (such as a plastic rod) with a uniform linear charge density λ .

23.21 Apply the relationship between linear charge density λ

on a cylindrical surface and the electric field magnitude E at radial distance r from the central axis.

23.22 Explain how Gauss' law can be used to find the electric field magnitude *inside* a cylindrical nonconducting surface (such as a plastic rod) with a uniform volume charge density ρ .

Key Idea

- The electric field at a point near an infinite line of charge (or charged rod) with uniform linear charge density λ is perpendicular to the line and has magnitude

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \quad (\text{line of charge}),$$

where r is the perpendicular distance from the line to the point.

Applying Gauss' Law: Cylindrical Symmetry

Figure 23-14 shows a section of an infinitely long cylindrical plastic rod with a uniform charge density λ . We want to find an expression for the electric field magnitude E at radius r from the central axis of the rod, outside the rod. We could do that using the approach of Chapter 22 (charge element dq , field vector $d\vec{E}$, etc.). However, Gauss' law gives a much faster and easier (and prettier) approach.

The charge distribution and the field have cylindrical symmetry. To find the field at radius r , we enclose a section of the rod with a concentric Gaussian cylinder of radius r and height h . (If you want the field at a certain point, put a Gaussian surface through that point.) We can now apply Gauss' law to relate the charge enclosed by the cylinder and the net flux through the cylinder's surface.

First note that because of the symmetry, the electric field at any point must be radially outward (the charge is positive). That means that at any point on the end caps, the field only skims the surface and does not pierce it. So, the flux through each end cap is zero.

To find the flux through the cylinder's curved surface, first note that for any patch element on the surface, the area vector $d\vec{A}$ is radially outward (away from the interior of the Gaussian surface) and thus in the same direction as the field piercing the patch. The dot product in Gauss' law is then simply $E dA \cos 0 = E dA$, and we can pull E out of the integral. The remaining integral is just the instruction to sum the areas of all patch elements on the cylinder's curved surface, but we already know that the total area is the product of the cylinder's height h and circumference $2\pi r$. The net flux through the cylinder is then

$$\Phi = EA \cos \theta = E(2\pi rh)\cos 0 = E(2\pi rh).$$

On the other side of Gauss's law we have the charge q_{enc} enclosed by the cylinder. Because the linear charge density (charge per unit length, remember) is uniform, the enclosed charge is λh . Thus, Gauss' law,

$$\epsilon_0 \Phi = q_{\text{enc}},$$

reduces to

$$\epsilon_0 E(2\pi rh) = \lambda h,$$

yielding

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \quad (\text{line of charge}). \quad (23-12)$$

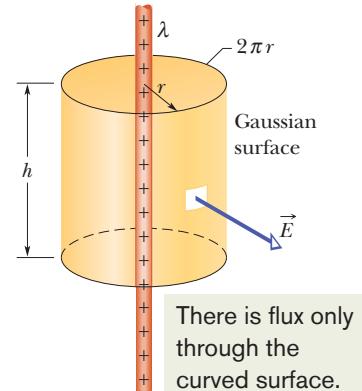


Figure 23-14 A Gaussian surface in the form of a closed cylinder surrounds a section of a very long, uniformly charged, cylindrical plastic rod.

This is the electric field due to an infinitely long, straight line of charge, at a point that is a radial distance r from the line. The direction of \vec{E} is radially outward

from the line of charge if the charge is positive, and radially inward if it is negative. Equation 23-12 also approximates the field of a *finite* line of charge at points that are not too near the ends (compared with the distance from the line).

If the rod has a uniform volume charge density ρ , we could use a similar procedure to find the electric field magnitude *inside* the rod. We would just shrink the Gaussian cylinder shown in Fig. 23-14 until it is inside the rod. The charge q_{enc} enclosed by the cylinder would then be proportional to the volume of the rod enclosed by the cylinder because the charge density is uniform.

Sample Problem 23.06

Gauss' law and an upward streamer in a lightning storm

Upward streamer in a lightning storm. The woman in Fig. 23-15 was standing on a lookout platform high in the Sequoia National Park when a large storm cloud moved overhead. Some of the conduction electrons in her body were driven into the ground by the cloud's negatively charged base (Fig. 23-16a), leaving her positively charged. You can tell she was highly charged because her hair strands repelled one another and extended away from her along the electric field lines produced by the charge on her.

Lightning did not strike the woman, but she was in extreme danger because that electric field was on the verge of causing electrical breakdown in the surrounding air. Such a breakdown would have occurred along a path extending away from her in what is called an *upward streamer*. An upward streamer is dangerous because the resulting ionization of molecules in the air suddenly frees a tremendous number of electrons from those molecules. Had the woman in Fig. 23-15 developed an upward streamer, the free electrons in the air would have moved to neutralize her (Fig. 23-16b), producing a large, perhaps fatal, charge flow through her body. That charge flow is dangerous because it could have interfered with or even stopped her breathing (which is obviously necessary for oxygen) and the steady beat of her heart (which is obviously necessary for the blood flow that carries the oxygen). The charge flow could also have caused burns.

Let's model her body as a narrow vertical cylinder of height $L = 1.8 \text{ m}$ and radius $R = 0.10 \text{ m}$ (Fig. 23-16c). Assume that charge Q was uniformly distributed along the cylinder and that electrical breakdown would have occurred if the electric



Courtesy NOAA

Figure 23-15 This woman has become positively charged by an overhead storm cloud.

field magnitude along her body had exceeded the critical value $E_c = 2.4 \text{ MN/C}$. What value of Q would have put the air along her body on the verge of breakdown?

KEY IDEA

Because $R \ll L$, we can approximate the charge distribution as a long line of charge. Further, because we assume that the charge is uniformly distributed along this line, we can approximate the magnitude of the electric field along the side of her body with Eq. 23-12 ($E = \lambda/2\pi\epsilon_0 r$).

Calculations: Substituting the critical value E_c for E , the cylinder radius R for radial distance r , and the ratio Q/L for linear charge density λ , we have

$$E_c = \frac{Q/L}{2\pi\epsilon_0 R},$$

or

$$Q = 2\pi\epsilon_0 RLE_c.$$

Substituting given data then gives us

$$\begin{aligned} Q &= (2\pi)(8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2)(0.10 \text{ m}) \\ &\quad \times (1.8 \text{ m})(2.4 \times 10^6 \text{ N/C}) \\ &= 2.402 \times 10^{-5} \text{ C} \approx 24 \mu\text{C}. \end{aligned} \quad (\text{Answer})$$

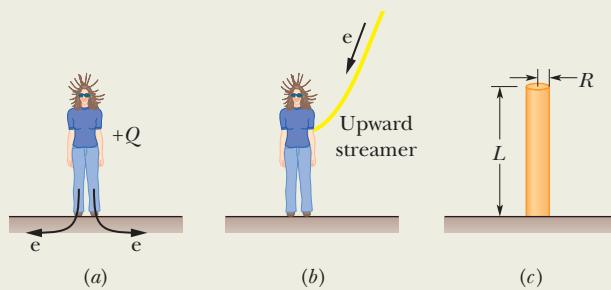


Figure 23-16 (a) Some of the conduction electrons in the woman's body are driven into the ground, leaving her positively charged. (b) An upward streamer develops if the air undergoes electrical breakdown, which provides a path for electrons freed from molecules in the air to move to the woman. (c) A cylinder represents the woman.

23-5 APPLYING GAUSS' LAW: PLANAR SYMMETRY

Learning Objectives

After reading this module, you should be able to . . .

- 23.23** Apply Gauss' law to derive the electric field magnitude E near a large, flat, nonconducting surface with a uniform surface charge density σ .
- 23.24** For points near a large, flat *nonconducting* surface with a uniform charge density σ , apply the relationship be-

tween the charge density and the electric field magnitude E and also specify the direction of the field.

- 23.25** For points near two large, flat, parallel, *conducting* surfaces with a uniform charge density σ , apply the relationship between the charge density and the electric field magnitude E and also specify the direction of the field.

Key Ideas

- The electric field due to an infinite nonconducting sheet with uniform surface charge density σ is perpendicular to the plane of the sheet and has magnitude

$$E = \frac{\sigma}{2\epsilon_0} \quad (\text{nonconducting sheet of charge}).$$

- The external electric field just outside the surface of an isolated charged conductor with surface charge density σ is perpendicular to the surface and has magnitude

$$E = \frac{\sigma}{\epsilon_0} \quad (\text{external, charged conductor}).$$

Inside the conductor, the electric field is zero.

Applying Gauss' Law: Planar Symmetry

Nonconducting Sheet

Figure 23-17 shows a portion of a thin, infinite, nonconducting sheet with a uniform (positive) surface charge density σ . A sheet of thin plastic wrap, uniformly charged on one side, can serve as a simple model. Let us find the electric field \vec{E} a distance r in front of the sheet.

A useful Gaussian surface is a closed cylinder with end caps of area A , arranged to pierce the sheet perpendicularly as shown. From symmetry, \vec{E} must be perpendicular to the sheet and hence to the end caps. Furthermore, since the charge is positive, \vec{E} is directed *away* from the sheet, and thus the electric field lines pierce the two Gaussian end caps in an outward direction. Because the field lines do not pierce the curved surface, there is no flux through this portion of the Gaussian surface. Thus $\vec{E} \cdot d\vec{A}$ is simply $E dA$; then Gauss' law,

$$\epsilon_0 \oint \vec{E} \cdot d\vec{A} = q_{\text{enc}},$$

becomes

$$\epsilon_0(EA + EA) = \sigma A,$$

where σA is the charge enclosed by the Gaussian surface. This gives

$$E = \frac{\sigma}{2\epsilon_0} \quad (\text{sheet of charge}). \quad (23-13)$$

Since we are considering an infinite sheet with uniform charge density, this result holds for any point at a finite distance from the sheet. Equation 23-13 agrees with Eq. 22-27, which we found by integration of electric field components.

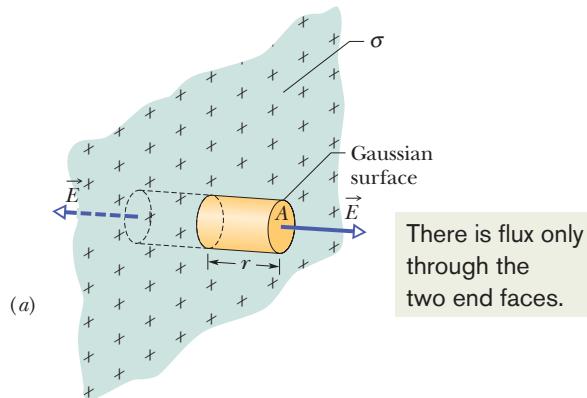
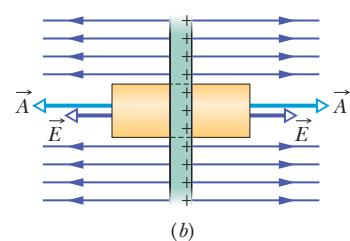


Figure 23-17 (a) Perspective view and (b) side view of a portion of a very large, thin plastic sheet, uniformly charged on one side to surface charge density σ . A closed cylindrical Gaussian surface passes through the sheet and is perpendicular to it.



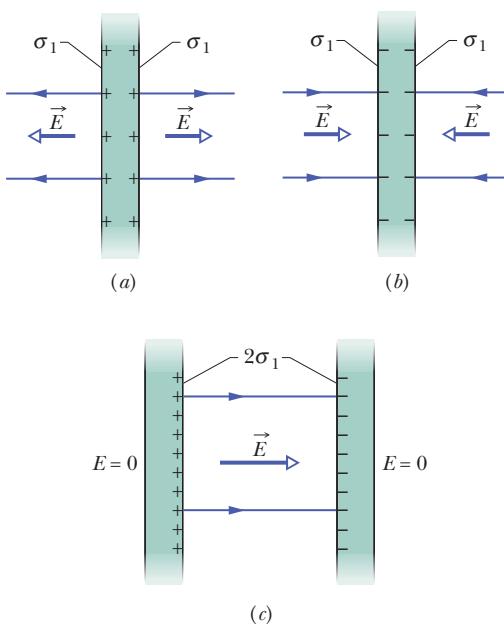


Figure 23-18 (a) A thin, very large conducting plate with excess positive charge. (b) An identical plate with excess negative charge. (c) The two plates arranged so they are parallel and close.

Two Conducting Plates

Figure 23-18a shows a cross section of a thin, infinite conducting plate with excess positive charge. From Module 23-3 we know that this excess charge lies on the surface of the plate. Since the plate is thin and very large, we can assume that essentially all the excess charge is on the two large faces of the plate.

If there is no external electric field to force the positive charge into some particular distribution, it will spread out on the two faces with a uniform surface charge density of magnitude σ_1 . From Eq. 23-11 we know that just outside the plate this charge sets up an electric field of magnitude $E = \sigma_1/\epsilon_0$. Because the excess charge is positive, the field is directed away from the plate.

Figure 23-18b shows an identical plate with excess negative charge having the same magnitude of surface charge density σ_1 . The only difference is that now the electric field is directed toward the plate.

Suppose we arrange for the plates of Figs. 23-18a and b to be close to each other and parallel (Fig. 23-18c). Since the plates are conductors, when we bring them into this arrangement, the excess charge on one plate attracts the excess charge on the other plate, and all the excess charge moves onto the inner faces of the plates as in Fig. 23-18c. With twice as much charge now on each inner face, the new surface charge density (call it σ) on each inner face is twice σ_1 . Thus, the electric field at any point between the plates has the magnitude

$$E = \frac{2\sigma_1}{\epsilon_0} = \frac{\sigma}{\epsilon_0}. \quad (23-14)$$

This field is directed away from the positively charged plate and toward the negatively charged plate. Since no excess charge is left on the outer faces, the electric field to the left and right of the plates is zero.

Because the charges moved when we brought the plates close to each other, the charge distribution of the two-plate system is not merely the sum of the charge distributions of the individual plates.

One reason why we discuss seemingly unrealistic situations, such as the field set up by an infinite sheet of charge, is that analyses for “infinite” situations yield good approximations to many real-world problems. Thus, Eq. 23-13 holds well for a finite nonconducting sheet as long as we are dealing with points close to the sheet and not too near its edges. Equation 23-14 holds well for a pair of finite conducting plates as long as we consider points that are not too close to their edges. The trouble with the edges is that near an edge we can no longer use planar symmetry to find expressions for the fields. In fact, the field lines there are curved (said to be an *edge effect* or *fringing*), and the fields can be very difficult to express algebraically.

Sample Problem 23.07 Electric field near two parallel nonconducting sheets with charge

Figure 23-19a shows portions of two large, parallel, nonconducting sheets, each with a fixed uniform charge on one side. The magnitudes of the surface charge densities are $\sigma_{(+)} = 6.8 \mu\text{C}/\text{m}^2$ for the positively charged sheet and $\sigma_{(-)} = 4.3 \mu\text{C}/\text{m}^2$ for the negatively charged sheet.

Find the electric field \vec{E} (a) to the left of the sheets, (b) between the sheets, and (c) to the right of the sheets.

KEY IDEA

With the charges fixed in place (they are on nonconductors), we can find the electric field of the sheets in Fig. 23-19a by (1) finding the field of each sheet as if that sheet were isolated and (2) algebraically adding the fields of the isolated sheets

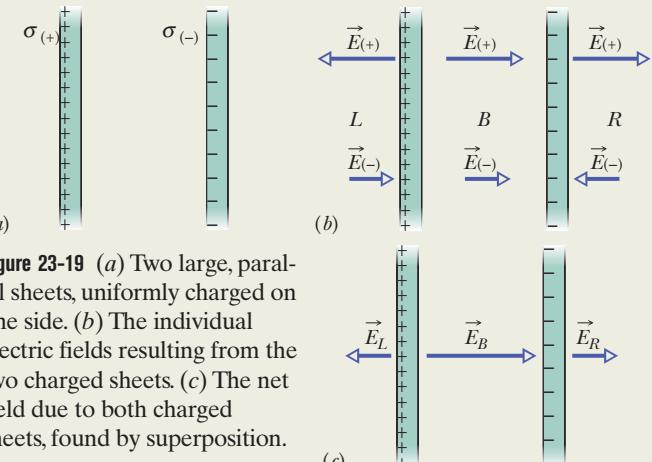


Figure 23-19 (a) Two large, parallel sheets, uniformly charged on one side. (b) The individual electric fields resulting from the two charged sheets. (c) The net field due to both charged sheets, found by superposition.

via the superposition principle. (We can add the fields algebraically because they are parallel to each other.)

Calculations: At any point, the electric field $\vec{E}_{(+)}$ due to the positive sheet is directed *away* from the sheet and, from Eq. 23-13, has the magnitude

$$\begin{aligned} E_{(+)} &= \frac{\sigma_{(+)}}{2\epsilon_0} = \frac{6.8 \times 10^{-6} \text{ C/m}^2}{(2)(8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2)} \\ &= 3.84 \times 10^5 \text{ N/C.} \end{aligned}$$

Similarly, at any point, the electric field $\vec{E}_{(-)}$ due to the negative sheet is directed *toward* that sheet and has the magnitude

$$\begin{aligned} E_{(-)} &= \frac{\sigma_{(-)}}{2\epsilon_0} = \frac{4.3 \times 10^{-6} \text{ C/m}^2}{(2)(8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2)} \\ &= 2.43 \times 10^5 \text{ N/C.} \end{aligned}$$



Additional examples, video, and practice available at WileyPLUS

Figure 23-19b shows the fields set up by the sheets to the left of the sheets (L), between them (B), and to their right (R).

The resultant fields in these three regions follow from the superposition principle. To the left, the field magnitude is

$$\begin{aligned} E_L &= E_{(+)} - E_{(-)} \\ &= 3.84 \times 10^5 \text{ N/C} - 2.43 \times 10^5 \text{ N/C} \\ &= 1.4 \times 10^5 \text{ N/C.} \end{aligned} \quad (\text{Answer})$$

Because $E_{(+)}$ is larger than $E_{(-)}$, the net electric field \vec{E}_L in this region is directed to the left, as Fig. 23-19c shows. To the right of the sheets, the net electric field has the same magnitude but is directed to the right, as Fig. 23-19c shows.

Between the sheets, the two fields add and we have

$$\begin{aligned} E_B &= E_{(+)} + E_{(-)} \\ &= 3.84 \times 10^5 \text{ N/C} + 2.43 \times 10^5 \text{ N/C} \\ &= 6.3 \times 10^5 \text{ N/C.} \end{aligned} \quad (\text{Answer})$$

The electric field \vec{E}_B is directed to the right.

23-6 APPLYING GAUSS' LAW: SPHERICAL SYMMETRY

Learning Objectives

After reading this module, you should be able to . . .

- 23.26 Identify that a shell of uniform charge attracts or repels a charged particle that is outside the shell as if all the shell's charge is concentrated at the center of the shell.
- 23.27 Identify that if a charged particle is enclosed by a shell of uniform charge, there is no electrostatic force on the particle from the shell.
- 23.28 For a point outside a spherical shell with uniform

charge, apply the relationship between the electric field magnitude E , the charge q on the shell, and the distance r from the shell's center.

- 23.29 Identify the magnitude of the electric field for points enclosed by a spherical shell with uniform charge.
- 23.30 For a uniform spherical charge distribution (a uniform ball of charge), determine the magnitude and direction of the electric field at interior and exterior points.

Key Ideas

- Outside a spherical shell of uniform charge q , the electric field due to the shell is radial (inward or outward, depending on the sign of the charge) and has the magnitude

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \quad (\text{outside spherical shell}),$$

where r is the distance to the point of measurement from the center of the shell. The field is the same as though all of the charge is concentrated as a particle at the center of the shell.

- Inside the shell, the field due to the shell is zero.
- Inside a sphere with a uniform volume charge density, the field is radial and has the magnitude

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{R^3} r \quad (\text{inside sphere of charge}),$$

where q is the total charge, R is the sphere's radius, and r is the radial distance from the center of the sphere to the point of measurement.

Applying Gauss' Law: Spherical Symmetry

Here we use Gauss' law to prove the two shell theorems presented without proof in Module 21-1:



A shell of uniform charge attracts or repels a charged particle that is outside the shell as if all the shell's charge were concentrated at the center of the shell.

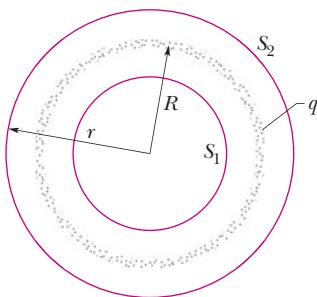
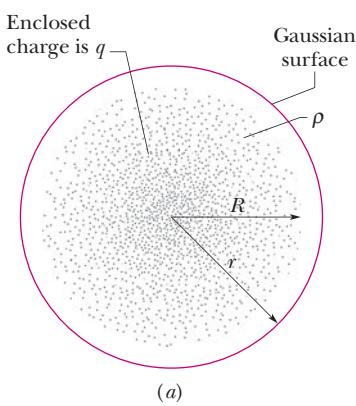
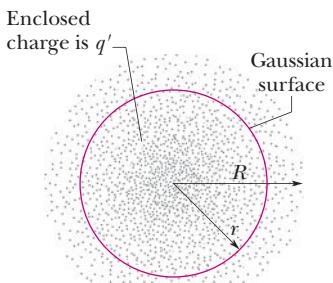


Figure 23-20 A thin, uniformly charged, spherical shell with total charge q , in cross section. Two Gaussian surfaces S_1 and S_2 are also shown in cross section. Surface S_2 encloses the shell, and S_1 encloses only the empty interior of the shell.



(a)



(b)

The flux through the surface depends on only the *enclosed* charge.

Figure 23-21 The dots represent a spherically symmetric distribution of charge of radius R , whose volume charge density ρ is a function only of distance from the center. The charged object is not a conductor, and therefore the charge is assumed to be fixed in position. A concentric spherical Gaussian surface with $r > R$ is shown in (a). A similar Gaussian surface with $r < R$ is shown in (b).

Figure 23-20 shows a charged spherical shell of total charge q and radius R and two concentric spherical Gaussian surfaces, S_1 and S_2 . If we followed the procedure of Module 23-2 as we applied Gauss' law to surface S_2 , for which $r \geq R$, we would find that

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \quad (\text{spherical shell, field at } r \geq R). \quad (23-15)$$

This field is the same as one set up by a particle with charge q at the center of the shell of charge. Thus, the force produced by a shell of charge q on a charged particle placed outside the shell is the same as if all the shell's charge is concentrated as a particle at the shell's center. This proves the first shell theorem.

Applying Gauss' law to surface S_1 , for which $r < R$, leads directly to

$$E = 0 \quad (\text{spherical shell, field at } r < R), \quad (23-16)$$

because this Gaussian surface encloses no charge. Thus, if a charged particle were enclosed by the shell, the shell would exert no net electrostatic force on the particle. This proves the second shell theorem.



If a charged particle is located inside a shell of uniform charge, there is no electrostatic force on the particle from the shell.

Any spherically symmetric charge distribution, such as that of Fig. 23-21, can be constructed with a nest of concentric spherical shells. For purposes of applying the two shell theorems, the volume charge density ρ should have a single value for each shell but need not be the same from shell to shell. Thus, for the charge distribution as a whole, ρ can vary, but only with r , the radial distance from the center. We can then examine the effect of the charge distribution "shell by shell."

In Fig. 23-21a, the entire charge lies within a Gaussian surface with $r > R$. The charge produces an electric field on the Gaussian surface as if the charge were that of a particle located at the center, and Eq. 23-15 holds.

Figure 23-21b shows a Gaussian surface with $r < R$. To find the electric field at points on this Gaussian surface, we separately consider the charge inside it and the charge outside it. From Eq. 23-16, the outside charge does not set up a field on the Gaussian surface. From Eq. 23-15, the inside charge sets up a field as though it is concentrated at the center. Letting q' represent that enclosed charge, we can then rewrite Eq. 23-15 as

$$E = \frac{1}{4\pi\epsilon_0} \frac{q'}{r^2} \quad (\text{spherical distribution, field at } r \leq R). \quad (23-17)$$

If the full charge q enclosed within radius R is uniform, then q' enclosed within radius r in Fig. 23-21b is proportional to q :

$$\frac{\left(\text{charge enclosed by sphere of radius } r\right)}{\left(\text{volume enclosed by sphere of radius } r\right)} = \frac{\text{full charge}}{\text{full volume}}$$

or

$$\frac{q'}{\frac{4}{3}\pi r^3} = \frac{q}{\frac{4}{3}\pi R^3}. \quad (23-18)$$

This gives us

$$q' = q \frac{r^3}{R^3}. \quad (23-19)$$

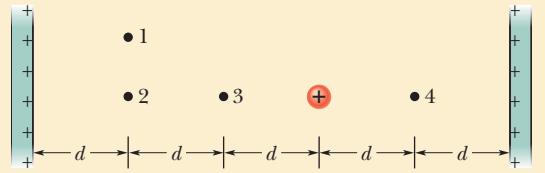
Substituting this into Eq. 23-17 yields

$$E = \left(\frac{q}{4\pi\epsilon_0 R^3} \right) r \quad (\text{uniform charge, field at } r \leq R). \quad (23-20)$$



Checkpoint 4

The figure shows two large, parallel, nonconducting sheets with identical (positive) uniform surface charge densities, and a sphere with a uniform (positive) volume charge density. Rank the four numbered points according to the magnitude of the net electric field there, greatest first.



Review & Summary

Gauss' Law Gauss' law and Coulomb's law are different ways of describing the relation between charge and electric field in static situations. Gauss' law is

$$q_0\Phi = q_{\text{enc}} \quad (\text{Gauss' law}), \quad (23-6)$$

in which q_{enc} is the net charge inside an imaginary closed surface (a Gaussian surface) and Φ is the net flux of the electric field through the surface:

$$\Phi = \oint \vec{E} \cdot d\vec{A} \quad (\text{electric flux through a Gaussian surface}). \quad (23-4)$$

Coulomb's law can be derived from Gauss' law.

Applications of Gauss' Law Using Gauss' law and, in some cases, symmetry arguments, we can derive several important results in electrostatic situations. Among these are:

1. An excess charge on an isolated conductor is located entirely on the outer surface of the conductor.
2. The external electric field near the surface of a charged conductor is perpendicular to the surface and has a magnitude that depends on the surface charge density σ :

$$E = \frac{\sigma}{\epsilon_0} \quad (\text{conducting surface}). \quad (23-11)$$

Within the conductor, $E = 0$.

3. The electric field at any point due to an infinite line of charge

with uniform linear charge density λ is perpendicular to the line of charge and has magnitude

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \quad (\text{line of charge}), \quad (23-12)$$

where r is the perpendicular distance from the line of charge to the point.

4. The electric field due to an infinite nonconducting sheet with uniform surface charge density σ is perpendicular to the plane of the sheet and has magnitude

$$E = \frac{\sigma}{2\epsilon_0} \quad (\text{sheet of charge}). \quad (23-13)$$

5. The electric field outside a spherical shell of charge with radius R and total charge q is directed radially and has magnitude

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \quad (\text{spherical shell, for } r \geq R). \quad (23-15)$$

Here r is the distance from the center of the shell to the point at which E is measured. (The charge behaves, for external points, as if it were all located at the center of the sphere.) The field inside a uniform spherical shell of charge is exactly zero:

$$E = 0 \quad (\text{spherical shell, for } r < R). \quad (23-16)$$

6. The electric field inside a uniform sphere of charge is directed radially and has magnitude

$$E = \left(\frac{q}{4\pi\epsilon_0 R^3} \right) r. \quad (23-20)$$

Questions

- 1 A surface has the area vector $\vec{A} = (2\hat{i} + 3\hat{j}) \text{ m}^2$. What is the flux of a uniform electric field through the area if the field is (a) $\vec{E} = 4\hat{i} \text{ N/C}$ and (b) $\vec{E} = 4\hat{k} \text{ N/C}$?

- 2 Figure 23-22 shows, in cross section, three solid cylinders, each of length L and uniform charge Q . Concentric with each cylinder is a cylindrical Gaussian surface, with all three surfaces having the same radius. Rank the Gaussian surfaces according to the electric field at any point on the surface, greatest first.

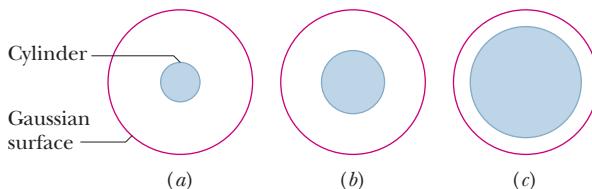


Figure 23-22 Question 2.

- 3 Figure 23-23 shows, in cross section, a central metal ball, two spherical metal shells, and three spherical Gaussian surfaces of radii R , $2R$, and $3R$, all with the same center. The uniform charges on the three objects are: ball, Q ; smaller shell, $3Q$; larger shell, $5Q$. Rank the Gaussian surfaces according to the magnitude of the electric field at any point on the surface, greatest first.

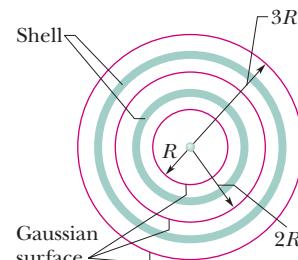


Figure 23-23 Question 3.

- 4** Figure 23-24 shows, in cross section, two Gaussian spheres and two Gaussian cubes that are centered on a positively charged particle. (a) Rank the net flux through the four Gaussian surfaces, greatest first. (b) Rank the magnitudes of the electric fields on the surfaces, greatest first, and indicate whether the magnitudes are uniform or variable along each surface.

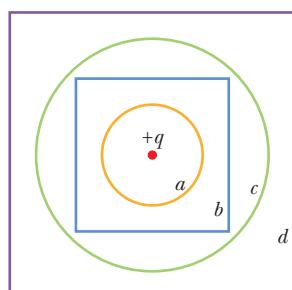


Figure 23-24 Question 4.

- 5** In Fig. 23-25, an electron is released between two infinite nonconducting sheets that are horizontal and have uniform surface charge densities $\sigma_{(+)}$ and $\sigma_{(-)}$, as indicated. The electron is subjected to the following three situations involving surface charge densities and sheet separations. Rank the magnitudes of the electron's acceleration, greatest first.

Situation	$\sigma_{(+)}$	$\sigma_{(-)}$	Separation
1	$+4\sigma$	-4σ	d
2	$+7\sigma$	$-\sigma$	$4d$
3	$+3\sigma$	-5σ	$9d$

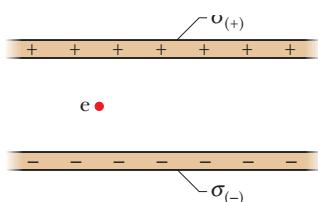


Figure 23-25 Question 5.

- 6** Three infinite nonconducting sheets, with uniform positive surface charge densities σ , 2σ , and 3σ , are arranged to be parallel like the two sheets in Fig. 23-19a. What is their order, from left to right, if the electric field \vec{E} produced by the arrangement has magnitude $E = 0$ in one region and $E = 2\sigma/\epsilon_0$ in another region?

- 7** Figure 23-26 shows four situations in which four very long rods extend into and out of the page (we see only their cross sections). The value below each cross section gives that particular rod's uniform charge density in microcoulombs per meter. The rods are separated by either d or $2d$ as drawn, and a central point is shown midway between the inner rods. Rank the situations according to the magnitude of the net electric field at that central point, greatest first.

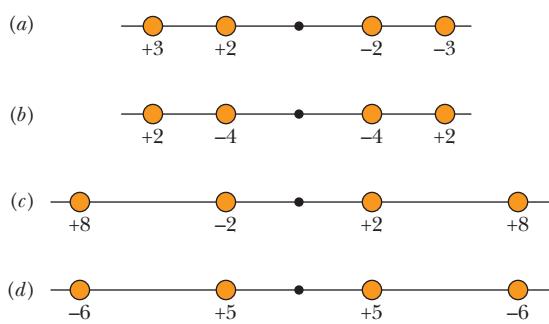


Figure 23-26 Question 7.

- 8** Figure 23-27 shows four solid spheres, each with charge Q uniformly distributed through its volume. (a) Rank the spheres according to their volume charge density, greatest first. The figure also shows a point P for each sphere, all at the same distance from the center of the sphere. (b) Rank the spheres according to the magnitude of the electric field they produce at point P , greatest first.

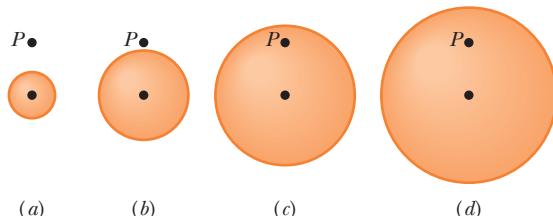


Figure 23-27 Question 8.

- 9** A small charged ball lies within the hollow of a metallic spherical shell of radius R . For three situations, the net charges on the ball and shell, respectively, are (1) $+4q$, 0; (2) $-6q$, $+10q$; (3) $+16q$, $-12q$. Rank the situations according to the charge on (a) the inner surface of the shell and (b) the outer surface, most positive first.

- 10** Rank the situations of Question 9 according to the magnitude of the electric field (a) halfway through the shell and (b) at a point $2R$ from the center of the shell, greatest first.

- 11** Figure 23-28 shows a section of three long charged cylinders centered on the same axis. Central cylinder A has a uniform charge $q_A = +3q_0$. What uniform charges q_B and q_C should be on cylinders B and C so that (if possible) the net electric field is zero at (a) point 1, (b) point 2, and (c) point 3?

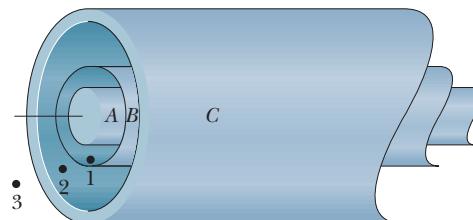


Figure 23-28 Question 11.

- 12** Figure 23-29 shows four Gaussian surfaces consisting of identical cylindrical midsections but different end caps. The surfaces are in a uniform electric field \vec{E} that is directed parallel to the central axis of each cylindrical midsection. The end caps have these shapes: S_1 , convex hemispheres; S_2 , concave hemispheres; S_3 , cones; S_4 , flat disks. Rank the surfaces according to (a) the net electric flux through them and (b) the electric flux through the top end caps, greatest first.

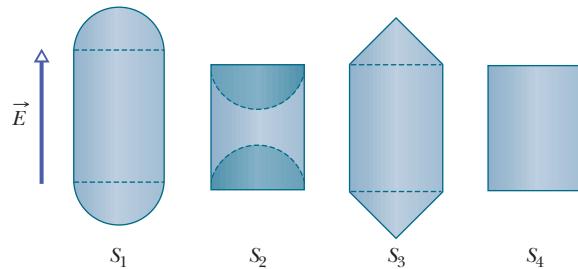


Figure 23-29 Question 12.

Problems



Tutoring problem available (at instructor's discretion) in WileyPLUS and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com

WWW Worked-out solution is at

ILW Interactive solution is at

<http://www.wiley.com/college/halliday>

Module 23-1 Electric Flux

- 1 **SSM** The square surface shown in Fig. 23-30 measures 3.2 mm on each side. It is immersed in a uniform electric field with magnitude $E = 1800 \text{ N/C}$ and with field lines at an angle of $\theta = 35^\circ$ with a normal to the surface, as shown. Take that normal to be directed "outward," as though the surface were one face of a box. Calculate the electric flux through the surface.

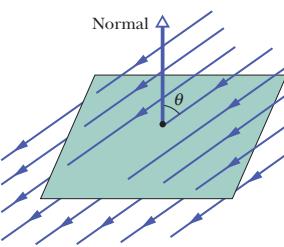


Figure 23-30 Problem 1.

- 2 An electric field given by $\vec{E} = 4.0\hat{i} - 3.0(y^2 + 2.0)\hat{j}$ pierces a Gaussian cube of edge length 2.0 m and positioned as shown in Fig. 23-7. (The magnitude E is in newtons per coulomb and the position x is in meters.) What is the electric flux through the (a) top face, (b) bottom face, (c) left face, and (d) back face? (e) What is the net electric flux through the cube?

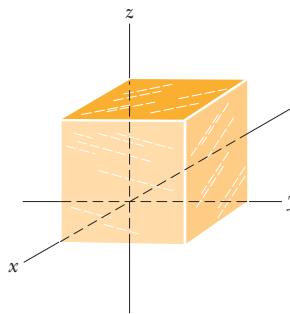


Figure 23-31 Problems 3, 6, and 9.

- 3 The cube in Fig. 23-31 has edge length 1.40 m and is oriented as shown in a region of uniform electric field. Find the electric flux through the right face if the electric field, in newtons per coulomb, is given by (a) $6.00\hat{i}$, (b) $-2.00\hat{j}$, and (c) $-3.00\hat{i} + 4.00\hat{k}$. (d) What is the total flux through the cube for each field?

Module 23-2 Gauss' Law

- 4 In Fig. 23-32, a butterfly net is in a uniform electric field of magnitude $E = 3.0 \text{ mN/C}$. The rim, a circle of radius $a = 11 \text{ cm}$, is aligned perpendicular to the field. The net contains no net charge. Find the electric flux through the netting.

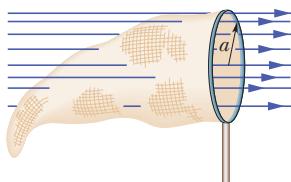


Figure 23-32 Problem 4.

- 5 In Fig. 23-33, a proton is a distance $d/2$ directly above the center of a square of side d . What is the magnitude of the electric flux through the square? (Hint: Think of the square as one face of a cube with edge d .)

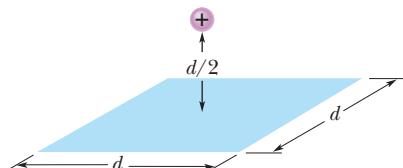


Figure 23-33 Problem 5.

- 6 At each point on the surface of the cube shown in Fig. 23-31, the electric field is parallel to the z axis. The length of each edge of the cube is 3.0 m. On the top face of the cube the field is

$\vec{E} = -34\hat{k} \text{ N/C}$, and on the bottom face it is $\vec{E} = +20\hat{k} \text{ N/C}$. Determine the net charge contained within the cube.

- 7 A particle of charge $1.8 \mu\text{C}$ is at the center of a Gaussian cube 55 cm on edge. What is the net electric flux through the surface?

- 8 When a shower is turned on in a closed bathroom, the splashing of the water on the bare tub can fill the room's air with negatively charged ions and produce an electric field in the air as great as 1000 N/C . Consider a bathroom with dimensions $2.5 \text{ m} \times 3.0 \text{ m} \times 2.0 \text{ m}$. Along the ceiling, floor, and four walls, approximate the electric field in the air as being directed perpendicular to the surface and as having a uniform magnitude of 600 N/C . Also, treat those surfaces as forming a closed Gaussian surface around the room's air. What are (a) the volume charge density ρ and (b) the number of excess elementary charges e per cubic meter in the room's air?

- 9 **ILW** Fig. 23-31 shows a Gaussian surface in the shape of a cube with edge length 1.40 m. What are (a) the net flux Φ through the surface and (b) the net charge q_{enc} enclosed by the surface if $\vec{E} = (3.00y\hat{j}) \text{ N/C}$, with y in meters? What are (c) Φ and (d) q_{enc} if $\vec{E} = [-4.00\hat{i} + (6.00 + 3.00y)\hat{j}] \text{ N/C}$?

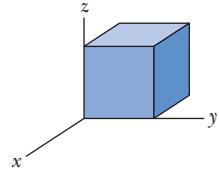


Figure 23-34
Problem 10.

- 10 Figure 23-34 shows a closed Gaussian surface in the shape of a cube of edge length 2.00 m. It lies in a region where the nonuniform electric field is given by $\vec{E} = (3.00x + 4.00)\hat{i} + 6.00\hat{j} + 7.00\hat{k} \text{ N/C}$, with x in meters. What is the net charge contained by the cube?

- 11 **GO** Figure 23-35 shows a closed Gaussian surface in the shape of a cube of edge length 2.00 m, with one corner at $x_1 = 5.00 \text{ m}$, $y_1 = 4.00 \text{ m}$. The cube lies in a region where the electric field vector is given by $\vec{E} = -3.00\hat{i} - 4.00y^2\hat{j} + 3.00\hat{k} \text{ N/C}$, with y in meters. What is the net charge contained by the cube?

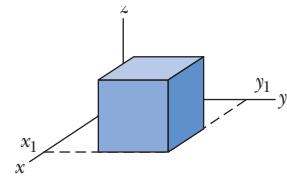


Figure 23-35 Problem 11.

- 12 Figure 23-36 shows two nonconducting spherical shells fixed in place. Shell 1 has uniform surface charge density $+6.0 \mu\text{C/m}^2$ on its outer surface and radius 3.0 cm; shell 2 has uniform surface charge density $+4.0 \mu\text{C/m}^2$ on its outer surface and radius 2.0 cm; the shell centers are separated by $L = 10 \text{ cm}$. In unit-vector notation, what is the net electric field at $x = 2.0 \text{ cm}$?

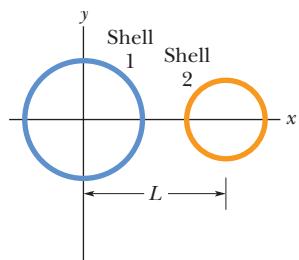


Figure 23-36 Problem 12.

••13 SSM The electric field in a certain region of Earth's atmosphere is directed vertically down. At an altitude of 300 m the field has magnitude 60.0 N/C; at an altitude of 200 m, the magnitude is 100 N/C. Find the net amount of charge contained in a cube 100 m on edge, with horizontal faces at altitudes of 200 and 300 m.

••14 GO Flux and nonconducting shells. A charged particle is suspended at the center of two concentric spherical shells that are very thin and made of nonconducting material. Figure 23-37a shows a cross section. Figure 23-37b gives the net flux Φ through a Gaussian sphere centered on the particle, as a function of the radius r of the sphere. The scale of the vertical axis is set by $\Phi_s = 5.0 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}$. (a) What is the charge of the central particle? What are the net charges of (b) shell A and (c) shell B?

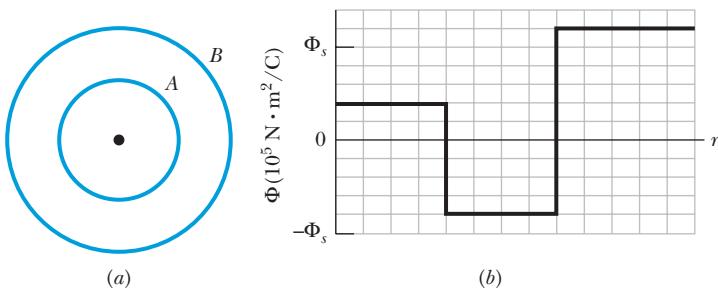


Figure 23-37 Problem 14.

••15 A particle of charge $+q$ is placed at one corner of a Gaussian cube. What multiple of q/ϵ_0 gives the flux through (a) each cube face forming that corner and (b) each of the other cube faces?

••16 GO The box-like Gaussian surface shown in Fig. 23-38 encloses a net charge of $+24.0\epsilon_0 \text{ C}$ and lies in an electric field given by $\vec{E} = [(10.0 + 2.00x)\hat{i} - 3.00\hat{j} + bz\hat{k}] \text{ N/C}$, with x and z in meters and b a constant. The bottom face is in the xz plane; the top face is in the horizontal plane passing through $y_2 = 1.00 \text{ m}$. For $x_1 = 1.00 \text{ m}$, $x_2 = 4.00 \text{ m}$, $z_1 = 1.00 \text{ m}$, and $z_2 = 3.00 \text{ m}$, what is b ?

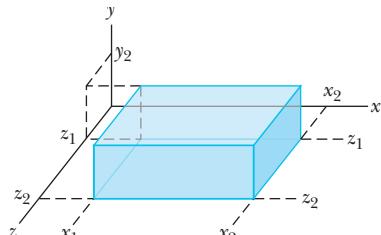


Figure 23-38 Problem 16.

Module 23-3 A Charged Isolated Conductor

•17 SSM A uniformly charged conducting sphere of 1.2 m diameter has surface charge density $8.1 \mu\text{C}/\text{m}^2$. Find (a) the net charge on the sphere and (b) the total electric flux leaving the surface.

•18 The electric field just above the surface of the charged conducting drum of a photocopying machine has a magnitude E of $2.3 \times 10^5 \text{ N/C}$. What is the surface charge density on the drum?

•19 Space vehicles traveling through Earth's radiation belts can intercept a significant number of electrons. The resulting charge buildup can damage electronic components and disrupt operations. Suppose a spherical metal satellite 1.3 m in diameter accumulates $2.4 \mu\text{C}$ of charge in one orbital revolution. (a) Find the resulting surface charge density. (b) Calculate the magnitude of the electric field just outside the surface of the satellite, due to the surface charge.

•20 GO Flux and conducting shells. A charged particle is held at the center of two concentric conducting spherical shells. Figure 23-39a shows a cross section. Figure 23-39b gives the net flux Φ through a Gaussian sphere centered on the particle, as a function of the radius r of the sphere. The scale of the vertical axis is set by $\Phi_s = 5.0 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}$. What are (a) the charge of the central particle and the net charges of (b) shell A and (c) shell B?

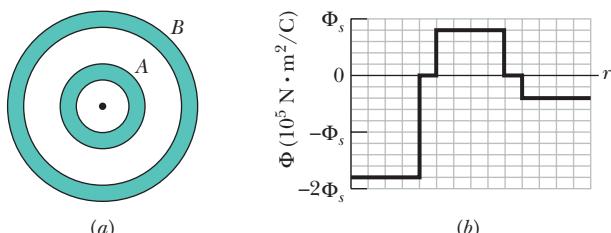


Figure 23-39 Problem 20.

•21 An isolated conductor has net charge $+10 \times 10^{-6} \text{ C}$ and a cavity with a particle of charge $q = +3.0 \times 10^{-6} \text{ C}$. What is the charge on (a) the cavity wall and (b) the outer surface?

Module 23-4 Applying Gauss' Law: Cylindrical Symmetry

•22 An electron is released 9.0 cm from a very long nonconducting rod with a uniform $6.0 \mu\text{C/m}$. What is the magnitude of the electron's initial acceleration?

•23 (a) The drum of a photocopying machine has a length of 42 cm and a diameter of 12 cm . The electric field just above the drum's surface is $2.3 \times 10^5 \text{ N/C}$. What is the total charge on the drum? (b) The manufacturer wishes to produce a desktop version of the machine. This requires reducing the drum length to 28 cm and the diameter to 8.0 cm . The electric field at the drum surface must not change. What must be the charge on this new drum?

•24 Figure 23-40 shows a section of a long, thin-walled metal tube of radius $R = 3.00 \text{ cm}$, with a charge per unit length of $\lambda = 2.00 \times 10^{-8} \text{ C/m}$. What is the magnitude E of the electric field at radial distance (a) $r = R/2.00$ and (b) $r = 2.00R$? (c) Graph E versus r for the range $r = 0$ to $2.00R$.

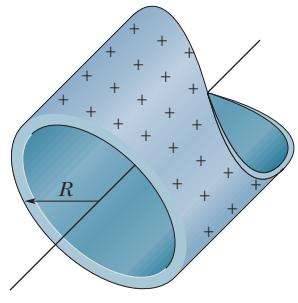


Figure 23-40 Problem 24.

•25 SSM An infinite line of charge produces a field of magnitude $4.5 \times 10^4 \text{ N/C}$ at distance 2.0 m . Find the linear charge density.

•26 Figure 23-41a shows a narrow charged solid cylinder that is coaxial with a larger charged cylindrical shell. Both are noncon-

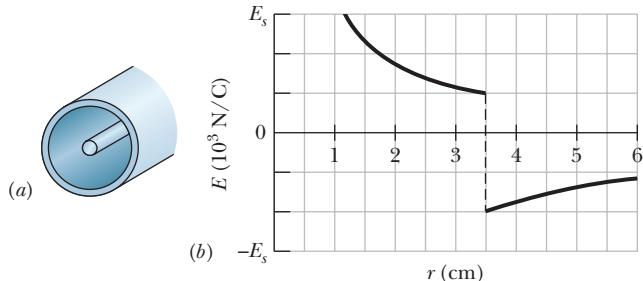


Figure 23-41 Problem 26.

ducting and thin and have uniform surface charge densities on their outer surfaces. Figure 23-41b gives the radial component E of the electric field versus radial distance r from the common axis, and $E_s = 3.0 \times 10^3 \text{ N/C}$. What is the shell's linear charge density?

•27 GO A long, straight wire has fixed negative charge with a linear charge density of magnitude 3.6 nC/m . The wire is to be enclosed by a coaxial, thin-walled nonconducting cylindrical shell of radius 1.5 cm . The shell is to have positive charge on its outside surface with a surface charge density σ that makes the net external electric field zero. Calculate σ .

•28 GO A charge of uniform linear density 2.0 nC/m is distributed along a long, thin, nonconducting rod. The rod is coaxial with a long conducting cylindrical shell (inner radius = 5.0 cm , outer radius = 10 cm). The net charge on the shell is zero. (a) What is the magnitude of the electric field 15 cm from the axis of the shell? What is the surface charge density on the (b) inner and (c) outer surface of the shell?

•29 SSM WWW Figure 23-42 is a section of a conducting rod of radius $R_1 = 1.30 \text{ mm}$ and length $L = 11.00 \text{ m}$ inside a thin-walled coaxial conducting cylindrical shell of radius $R_2 = 10.0R_1$ and the (same) length L . The net charge on the rod is $Q_1 = +3.40 \times 10^{-12} \text{ C}$; that on the shell is $Q_2 = -2.00Q_1$. What are the (a) magnitude E and (b) direction (radially inward or outward) of the electric field at radial distance $r = 2.00R_2$? What are (c) E and (d) the direction at $r = 5.00R_1$? What is the charge on the (e) interior and (f) exterior surface of the shell?

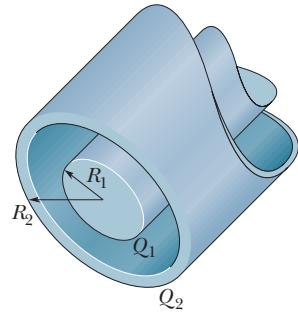


Figure 23-42 Problem 29.

•30 In Fig. 23-43, short sections of two very long parallel lines of charge are shown, fixed in place, separated by $L = 8.0 \text{ cm}$. The uniform linear charge densities are $+6.0 \mu\text{C/m}$ for line 1 and $-2.0 \mu\text{C/m}$ for line 2. Where along the x axis shown is the net electric field from the two lines zero?

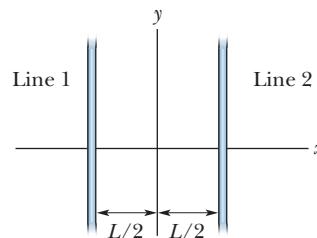
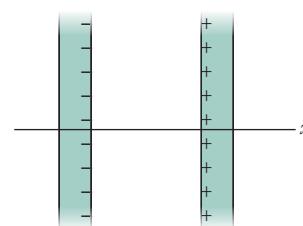


Figure 23-43 Problem 30.

•31 ILW Two long, charged, thin-walled, concentric cylindrical shells have radii of 3.0 and 6.0 cm . The charge per unit length is $5.0 \times 10^{-6} \text{ C/m}$ on the inner shell and $-7.0 \times 10^{-6} \text{ C/m}$ on the outer shell. What are the (a) magnitude E and (b) direction (radially inward or outward) of the electric field at radial distance $r = 4.0 \text{ cm}$? What are (c) E and (d) the direction at $r = 8.0 \text{ cm}$?

•32 GO A long, nonconducting, solid cylinder of radius 4.0 cm has a nonuniform volume charge density ρ that is a function of radial distance r from the cylinder axis: $\rho = Ar^2$. For $A = 2.5 \mu\text{C/m}^5$, what is the magnitude of the electric field at (a) $r = 3.0 \text{ cm}$ and (b) $r = 5.0 \text{ cm}$?



Module 23-5 Applying Gauss' Law: Planar Symmetry

•33 In Fig. 23-44, two large, thin metal plates are parallel and close to each other. On their inner faces,

the plates have excess surface charge densities of opposite signs and magnitude $7.00 \times 10^{-22} \text{ C/m}^2$. In unit-vector notation, what is the electric field at points (a) to the left of the plates, (b) to the right of them, and (c) between them?

•34 In Fig. 23-45, a small circular hole of radius $R = 1.80 \text{ cm}$ has been cut in the middle of an infinite, flat, nonconducting surface that has uniform charge density $\sigma = 4.50 \text{ pC/m}^2$. A z axis, with its origin at the hole's center, is perpendicular to the surface. In unit-vector notation, what is the electric field at point P at $z = 2.56 \text{ cm}$? (Hint: See Eq. 22-26 and use superposition.)

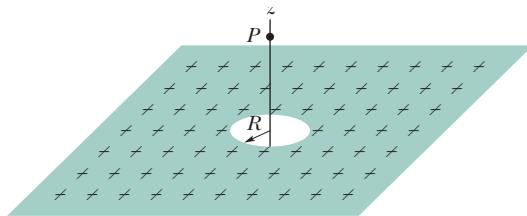


Figure 23-45 Problem 34.

•35 GO Figure 23-46a shows three plastic sheets that are large, parallel, and uniformly charged. Figure 23-46b gives the component of the net electric field along an x axis through the sheets. The scale of the vertical axis is set by $E_s = 6.0 \times 10^5 \text{ N/C}$. What is the ratio of the charge density on sheet 3 to that on sheet 2?

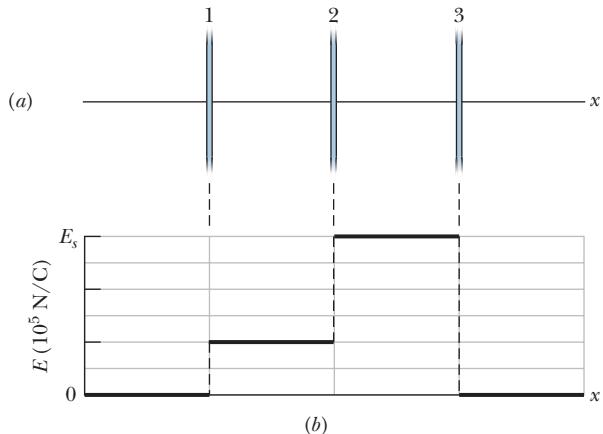


Figure 23-46 Problem 35.

•36 Figure 23-47 shows cross sections through two large, parallel, nonconducting sheets with identical distributions of positive charge with surface charge density $\sigma = 1.77 \times 10^{-22} \text{ C/m}^2$. In unit-vector notation, what is \vec{E} at points (a) above the sheets, (b) between them, and (c) below them?

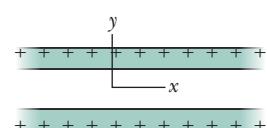


Figure 23-47
Problem 36.

•37 SSM WWW A square metal plate of edge length 8.0 cm and negligible thickness has a total charge of $6.0 \times 10^{-6} \text{ C}$. (a) Estimate the magnitude E of the electric field just off the center of the plate (at, say, a distance of 0.50 mm from the center) by assuming that the charge is spread uniformly over the two faces of the plate. (b) Estimate E at a distance of 30 m (large relative to the plate size) by assuming that the plate is a charged particle.

- 38 GO** In Fig. 23-48a, an electron is shot directly away from a uniformly charged plastic sheet, at speed $v_s = 2.0 \times 10^5 \text{ m/s}$. The sheet is nonconducting, flat, and very large. Figure 23-48b gives the electron's vertical velocity component v versus time t until the return to the launch point. What is the sheet's surface charge density?

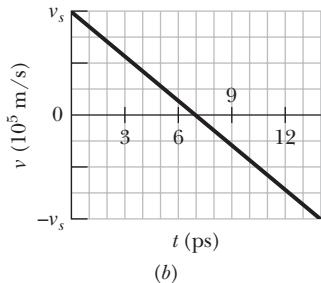
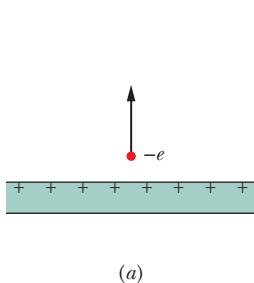


Figure 23-48 Problem 38.

- 39 SSM** In Fig. 23-49, a small, nonconducting ball of mass $m = 1.0 \text{ mg}$ and charge $q = 2.0 \times 10^{-8} \text{ C}$ (distributed uniformly through its volume) hangs from an insulating thread that makes an angle $\theta = 30^\circ$ with a vertical, uniformly charged nonconducting sheet (shown in cross section). Considering the gravitational force on the ball and assuming the sheet extends far vertically and into and out of the page, calculate the surface charge density σ of the sheet.

- 40** Figure 23-50 shows a very large nonconducting sheet that has a uniform surface charge density of $\sigma = -2.00 \mu\text{C/m}^2$; it also shows a particle of charge $Q = 6.00 \mu\text{C}$, at distance d from the sheet.

Both are fixed in place. If $d = 0.200 \text{ m}$, at what (a) positive and (b) negative coordinate on the x axis (other than infinity) is the net electric field \vec{E}_{net} of the sheet and particle zero? (c) If $d = 0.800 \text{ m}$, at what coordinate on the x axis is $\vec{E}_{\text{net}} = 0$?

- 41 GO** An electron is shot directly toward the center of a large metal plate that has surface charge density $-2.0 \times 10^{-6} \text{ C/m}^2$. If the initial kinetic energy of the electron is $1.60 \times 10^{-17} \text{ J}$ and if the electron is to stop (due to electrostatic repulsion from the plate) just as it reaches the plate, how far from the plate must the launch point be?

- 42** Two large metal plates of area 1.0 m^2 face each other, 5.0 cm apart, with equal charge magnitudes $|q|$ but opposite signs. The field magnitude E between them (neglect fringing) is 55 N/C . Find $|q|$.

- 43 GO** Figure 23-51 shows a cross section through a very large nonconducting slab of thickness $d = 9.40 \text{ mm}$ and uniform volume charge density $\rho = 5.80 \text{ fC/m}^3$. The origin of an x axis is at the slab's center. What is the magnitude of the slab's electric field at an x coordinate of (a) 0, (b) 2.00 mm , (c) 4.70 mm , and (d) 26.0 mm ?

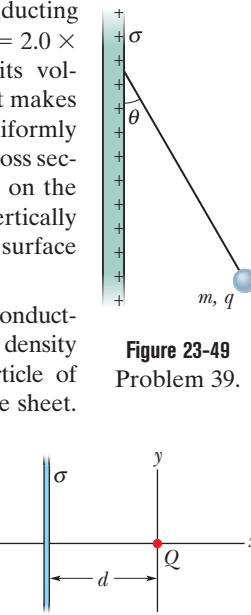


Figure 23-49 Problem 39.

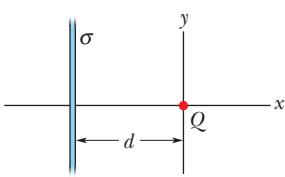


Figure 23-50 Problem 40.

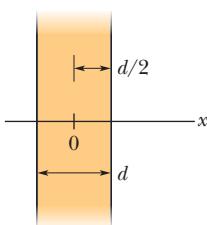


Figure 23-51 Problem 43.

Module 23-6 Applying Gauss' Law: Spherical Symmetry

- 44** Figure 23-52 gives the magnitude of the electric field inside and outside a sphere with a positive charge distributed uniformly throughout its volume. The scale of the vertical axis is set by $E_s = 5.0 \times 10^7 \text{ N/C}$. What is the charge on the sphere?

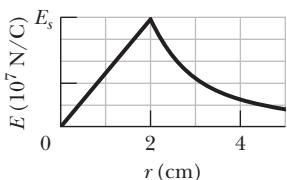


Figure 23-52 Problem 44.

- 45** Two charged concentric spherical shells have radii 10.0 cm and 15.0 cm . The charge on the inner shell is $4.00 \times 10^{-8} \text{ C}$, and that on the outer shell is $2.00 \times 10^{-8} \text{ C}$. Find the electric field (a) at $r = 12.0 \text{ cm}$ and (b) at $r = 20.0 \text{ cm}$.

- 46** Assume that a ball of charged particles has a uniformly distributed negative charge density except for a narrow radial tunnel through its center, from the surface on one side to the surface on the opposite side. Also assume that we can position a proton anywhere along the tunnel or outside the ball. Let F_R be the magnitude of the electrostatic force on the proton when it is located at the ball's surface, at radius R . As a multiple of R , how far from the surface is there a point where the force magnitude is $0.50F_R$ if we move the proton (a) away from the ball and (b) into the tunnel?

- 47 SSM** An unknown charge sits on a conducting solid sphere of radius 10 cm . If the electric field 15 cm from the center of the sphere has the magnitude $3.0 \times 10^3 \text{ N/C}$ and is directed radially inward, what is the net charge on the sphere?

- 48 GO** A charged particle is held at the center of a spherical shell. Figure 23-53 gives the magnitude E of the electric field versus radial distance r . The scale of the vertical axis is set by $E_s = 10.0 \times 10^7 \text{ N/C}$. Approximately, what is the net charge on the shell?

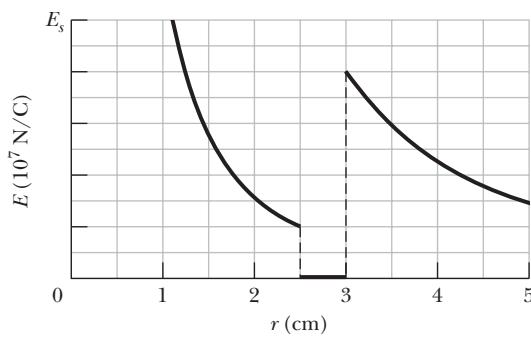


Figure 23-53 Problem 48.

- 49** In Fig. 23-54, a solid sphere of radius $a = 2.00 \text{ cm}$ is concentric with a spherical conducting shell of inner radius $b = 2.00a$ and outer radius $c = 2.40a$. The sphere has a net uniform charge $q_1 = +5.00 \text{ fC}$; the shell has a net charge $q_2 = -q_1$. What is the magnitude of the electric field at radial distances (a) $r = 0$, (b) $r = a/2.00$, (c) $r = a$, (d) $r = 1.50a$, (e) $r = 2.30a$, and (f) $r = 3.50a$? What is the net charge on the (g) inner and (h) outer surface of the shell?

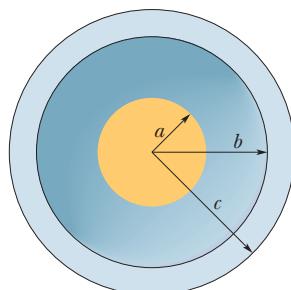


Figure 23-54 Problem 49.

- 50 GO** Figure 23-55 shows two nonconducting spherical shells fixed in place on an x axis. Shell 1 has uniform surface charge density $+4.0 \mu\text{C}/\text{m}^2$ on its outer surface and radius 0.50 cm, and shell 2 has uniform surface charge density $-2.0 \mu\text{C}/\text{m}^2$ on its outer surface and radius 2.0 cm; the centers are separated by $L = 6.0 \text{ cm}$. Other than at $x = \infty$, where on the x axis is the net electric field equal to zero?

- 51 SSM WWW** In Fig. 23-56, a nonconducting spherical shell of inner radius $a = 2.00 \text{ cm}$ and outer radius $b = 2.40 \text{ cm}$ has (within its thickness) a positive volume charge density $\rho = A/r$, where A is a constant and r is the distance from the center of the shell. In addition, a small ball of charge $q = 45.0 \text{ fC}$ is located at that center. What value should A have if the electric field in the shell ($a \leq r \leq b$) is to be uniform?

- 52 GO** Figure 23-57 shows a spherical shell with uniform volume charge density $\rho = 1.84 \text{ nC}/\text{m}^3$, inner radius $a = 10.0 \text{ cm}$, and outer radius $b = 2.00a$. What is the magnitude of the electric field at radial distances (a) $r = 0$; (b) $r = a/2.00$, (c) $r = a$, (d) $r = 1.50a$, (e) $r = b$, and (f) $r = 3.00b$?

- 53 ILW** The volume charge density of a solid nonconducting sphere of radius $R = 5.60 \text{ cm}$ varies with radial distance r as given by $\rho = (14.1 \text{ pC}/\text{m}^3)r/R$. (a) What is the sphere's total charge? What is the field magnitude E at (b) $r = 0$, (c) $r = R/2.00$, and (d) $r = R$? (e) Graph E versus r .

- 54** Figure 23-58 shows, in cross section, two solid spheres with uniformly distributed charge throughout their volumes. Each has radius R . Point P lies on a line connecting the centers of the spheres, at radial distance $R/2.00$ from the center of sphere 1. If the net electric field at point P is zero, what is the ratio q_2/q_1 of the total charges?

- 55** A charge distribution that is spherically symmetric but not uniform radially produces an electric field of magnitude $E = Kr^4$, directed radially outward from the center of the sphere. Here r is the radial distance from that center, and K is a constant. What is the volume density ρ of the charge distribution?

Additional Problems

- 56** The electric field in a particular space is $\vec{E} = (x + 2)\hat{i} \text{ N/C}$, with x in meters. Consider a cylindrical Gaussian surface of radius 20 cm that is coaxial with the x axis. One end of the cylinder is at $x = 0$. (a) What is the magnitude of the electric flux through the other end of the cylinder at $x = 2.0 \text{ m}$? (b) What net charge is enclosed within the cylinder?

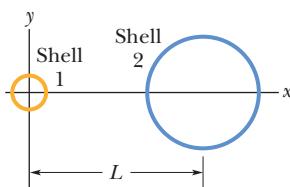


Figure 23-55 Problem 50.

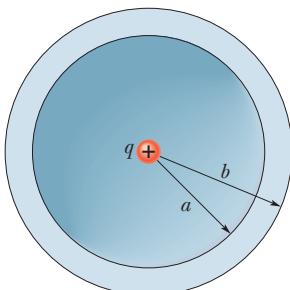


Figure 23-56 Problem 51.

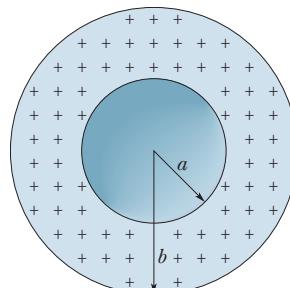


Figure 23-57 Problem 52.

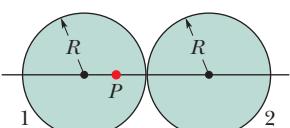


Figure 23-58 Problem 54.

- 57** A thin-walled metal spherical shell has radius 25.0 cm and charge $2.00 \times 10^{-7} \text{ C}$. Find E for a point (a) inside the shell, (b) just outside it, and (c) 3.00 m from the center.

- 58** A uniform surface charge of density $8.0 \text{ nC}/\text{m}^2$ is distributed over the entire xy plane. What is the electric flux through a spherical Gaussian surface centered on the origin and having a radius of 5.0 cm?

- 59** Charge of uniform volume density $\rho = 1.2 \text{ nC}/\text{m}^3$ fills an infinite slab between $x = -5.0 \text{ cm}$ and $x = +5.0 \text{ cm}$. What is the magnitude of the electric field at any point with the coordinate (a) $x = 4.0 \text{ cm}$ and (b) $x = 6.0 \text{ cm}$?

- 60** *The chocolate crumb mystery.* Explosions ignited by electrostatic discharges (sparks) constitute a serious danger in facilities handling grain or powder. Such an explosion occurred in chocolate crumb powder at a biscuit factory in the 1970s. Workers usually emptied newly delivered sacks of the powder into a loading bin, from which it was blown through electrically grounded plastic pipes to a silo for storage. Somewhere along this route, two conditions for an explosion were met: (1) The magnitude of an electric field became $3.0 \times 10^6 \text{ N/C}$ or greater, so that electrical breakdown and thus sparking could occur. (2) The energy of a spark was 150 mJ or greater so that it could ignite the powder explosively. Let us check for the first condition in the powder flow through the plastic pipes.

Suppose a stream of *negatively* charged powder was blown through a cylindrical pipe of radius $R = 5.0 \text{ cm}$. Assume that the powder and its charge were spread uniformly through the pipe with a volume charge density ρ . (a) Using Gauss' law, find an expression for the magnitude of the electric field \vec{E} in the pipe as a function of radial distance r from the pipe center. (b) Does E increase or decrease with increasing r ? (c) Is \vec{E} directed radially inward or outward? (d) For $\rho = 1.1 \times 10^{-3} \text{ C/m}^3$ (a typical value at the factory), find the maximum E and determine where that maximum field occurs. (e) Could sparking occur, and if so, where? (The story continues with Problem 70 in Chapter 24.)

- 61 SSM** A thin-walled metal spherical shell of radius a has a charge q_a . Concentric with it is a thin-walled metal spherical shell of radius $b > a$ and charge q_b . Find the electric field at points a distance r from the common center, where (a) $r < a$, (b) $a < r < b$, and (c) $r > b$. (d) Discuss the criterion you would use to determine how the charges are distributed on the inner and outer surfaces of the shells.

- 62** A particle of charge $q = 1.0 \times 10^{-7} \text{ C}$ is at the center of a spherical cavity of radius 3.0 cm in a chunk of metal. Find the electric field (a) 1.5 cm from the cavity center and (b) anywhere in the metal.

- 63** A proton at speed $v = 3.00 \times 10^5 \text{ m/s}$ orbits at radius $r = 1.00 \text{ cm}$ outside a charged sphere. Find the sphere's charge.

- 64** Equation 23-11 ($E = \sigma/\epsilon_0$) gives the electric field at points near a charged conducting surface. Apply this equation to a conducting sphere of radius r and charge q , and show that the electric field outside the sphere is the same as the field of a charged particle located at the center of the sphere.

- 65** Charge Q is uniformly distributed in a sphere of radius R . (a) What fraction of the charge is contained within the radius $r = R/2.00$? (b) What is the ratio of the electric field magnitude at $r = R/2.00$ to that on the surface of the sphere?

- 66** A charged particle causes an electric flux of $-750 \text{ N} \cdot \text{m}^2/\text{C}$ to pass through a spherical Gaussian surface of 10.0 cm radius centered on the charge. (a) If the radius of the Gaussian surface were

doubled, how much flux would pass through the surface? (b) What is the charge of the particle?

67 SSM The electric field at point P just outside the outer surface of a hollow spherical conductor of inner radius 10 cm and outer radius 20 cm has magnitude 450 N/C and is directed outward. When a particle of unknown charge Q is introduced into the center of the sphere, the electric field at P is still directed outward but is now 180 N/C. (a) What was the net charge enclosed by the outer surface before Q was introduced? (b) What is charge Q ? After Q is introduced, what is the charge on the (c) inner and (d) outer surface of the conductor?

68 The net electric flux through each face of a die (singular of dice) has a magnitude in units of $10^3 \text{ N} \cdot \text{m}^2/\text{C}$ that is exactly equal to the number of spots N on the face (1 through 6). The flux is inward for N odd and outward for N even. What is the net charge inside the die?

69 Figure 23-59 shows, in cross section, three infinitely large nonconducting sheets on which charge is uniformly spread. The surface charge densities are $\sigma_1 = +2.00 \mu\text{C}/\text{m}^2$, $\sigma_2 = +4.00 \mu\text{C}/\text{m}^2$, and $\sigma_3 = -5.00 \mu\text{C}/\text{m}^2$, and distance $L = 1.50 \text{ cm}$. In unit-vector notation, what is the net electric field at point P ?

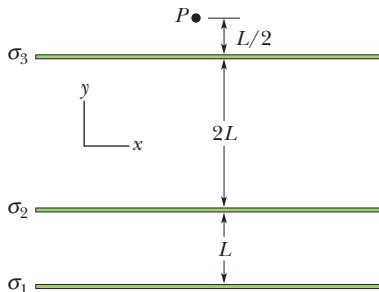


Figure 23-59 Problem 69.

70 Charge of uniform volume density $\rho = 3.2 \mu\text{C}/\text{m}^3$ fills a nonconducting solid sphere of radius 5.0 cm. What is the magnitude of the electric field (a) 3.5 cm and (b) 8.0 cm from the sphere's center?

71 A Gaussian surface in the form of a hemisphere of radius $R = 5.68 \text{ cm}$ lies in a uniform electric field of magnitude $E = 2.50 \text{ N/C}$. The surface encloses no net charge. At the (flat) base of the surface, the field is perpendicular to the surface and directed into the surface. What is the flux through (a) the base and (b) the curved portion of the surface?

72 What net charge is enclosed by the Gaussian cube of Problem 2?

73 A nonconducting solid sphere has a uniform volume charge density ρ . Let \vec{r} be the vector from the center of the sphere to a general point P within the sphere. (a) Show that the electric field at P is given by $\vec{E} = \rho \vec{r} / 3\epsilon_0$. (Note that the result is independent of the radius of the sphere.) (b) A spherical cavity is hollowed out of the sphere, as shown in Fig. 23-60. Using superposition concepts, show that the electric field at all points within the cavity is uniform and equal to $\vec{E} = \rho \vec{a} / 3\epsilon_0$, where \vec{a} is the position vector from the center of the sphere to the center of the cavity.



Figure 23-60
Problem 73.

74 A uniform charge density of 500 nC/m^3 is distributed throughout a spherical volume of radius 6.00 cm. Consider a cubical Gaussian surface with its center at the center of the sphere. What is the electric flux through this cubical surface if its edge length is (a) 4.00 cm and (b) 14.0 cm?

75 Figure 23-61 shows a Geiger counter, a device used to detect ionizing radiation, which causes ionization of atoms. A thin, posi-

tively charged central wire is surrounded by a concentric, circular, conducting cylindrical shell with an equal negative charge, creating a strong radial electric field. The shell contains a low-pressure inert gas. A particle of radiation entering the device through the shell wall ionizes a few of the gas atoms. The resulting free electrons (e^-) are drawn to the positive wire. However, the electric field is so intense that, between collisions with gas atoms, the free electrons gain energy sufficient to ionize these atoms also. More free electrons are thereby created, and the process is repeated until the electrons reach the wire. The resulting "avalanche" of electrons is collected by the wire, generating a signal that is used to record the passage of the original particle of radiation. Suppose that the radius of the central wire is $25 \mu\text{m}$, the inner radius of the shell 1.4 cm , and the length of the shell 16 cm . If the electric field at the shell's inner wall is $2.9 \times 10^4 \text{ N/C}$, what is the total positive charge on the central wire?

76 Charge is distributed uniformly throughout the volume of an infinitely long solid cylinder of radius R . (a) Show that, at a distance $r < R$ from the cylinder axis,

$$E = \frac{\rho r}{2\epsilon_0},$$

where ρ is the volume charge density. (b) Write an expression for E when $r > R$.

77 SSM A spherical conducting shell has a charge of $-14 \mu\text{C}$ on its outer surface and a charged particle in its hollow. If the net charge on the shell is $-10 \mu\text{C}$, what is the charge (a) on the inner surface of the shell and (b) of the particle?

78 A charge of 6.00 pC is spread uniformly throughout the volume of a sphere of radius $r = 4.00 \text{ cm}$. What is the magnitude of the electric field at a radial distance of (a) 6.00 cm and (b) 3.00 cm ?

79 Water in an irrigation ditch of width $w = 3.22 \text{ m}$ and depth $d = 1.04 \text{ m}$ flows with a speed of 0.207 m/s . The *mass flux* of the flowing water through an imaginary surface is the product of the water's density (1000 kg/m^3) and its volume flux through that surface. Find the mass flux through the following imaginary surfaces: (a) a surface of area wd , entirely in the water, perpendicular to the flow; (b) a surface with area $3wd/2$, of which wd is in the water, perpendicular to the flow; (c) a surface of area $wd/2$, entirely in the water, perpendicular to the flow; (d) a surface of area wd , half in the water and half out, perpendicular to the flow; (e) a surface of area wd , entirely in the water, with its normal 34.0° from the direction of flow.

80 Charge of uniform surface density 8.00 nC/m^2 is distributed over an entire xy plane; charge of uniform surface density 3.00 nC/m^2 is distributed over the parallel plane defined by $z = 2.00 \text{ m}$. Determine the magnitude of the electric field at any point having a z coordinate of (a) 1.00 m and (b) 3.00 m .

81 A spherical ball of charged particles has a uniform charge density. In terms of the ball's radius R , at what radial distances (a) inside and (b) outside the ball is the magnitude of the ball's electric field equal to $\frac{1}{4}$ of the maximum magnitude of that field?

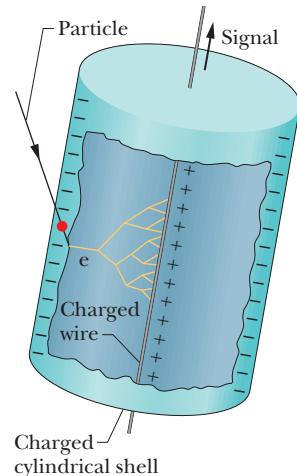


Figure 23-61 Problem 75.

Electric Potential

24-1 ELECTRIC POTENTIAL

Learning Objectives

After reading this module, you should be able to . . .

- 24.01** Identify that the electric force is conservative and thus has an associated potential energy.
- 24.02** Identify that at every point in a charged object's electric field, the object sets up an electric potential V , which is a scalar quantity that can be positive or negative depending on the sign of the object's charge.
- 24.03** For a charged particle placed at a point in an object's electric field, apply the relationship between the object's electric potential V at that point, the particle's charge q , and the potential energy U of the particle-object system.
- 24.04** Convert energies between units of joules and electron-volts.
- 24.05** If a charged particle moves from an initial point to a final point in an electric field, apply the relationships

Key Ideas

- The electric potential V at a point P in the electric field of a charged object is

$$V = \frac{-W_{\infty}}{q_0} = \frac{U}{q_0},$$

where W_{∞} is the work that would be done by the electric force on a positive test charge q_0 were it brought from an infinite distance to P , and U is the electric potential energy that would then be stored in the test charge-object system.

- If a particle with charge q is placed at a point where the electric potential of a charged object is V , the electric potential energy U of the particle-object system is

$$U = qV.$$

- If the particle moves through a potential difference ΔV , the change in the electric potential energy is

between the change ΔV in the potential, the particle's charge q , the change ΔU in the potential energy, and the work W done by the electric force.

- 24.06** If a charged particle moves between two given points in the electric field of a charged object, identify that the amount of work done by the electric force is path independent.
- 24.07** If a charged particle moves through a change ΔV in electric potential without an applied force acting on it, relate ΔV and the change ΔK in the particle's kinetic energy.
- 24.08** If a charged particle moves through a change ΔV in electric potential while an applied force acts on it, relate ΔV , the change ΔK in the particle's kinetic energy, and the work W_{app} done by the applied force.

$$\Delta U = q \Delta V = q(V_f - V_i).$$

- If a particle moves through a change ΔV in electric potential without an applied force acting on it, applying the conservation of mechanical energy gives the change in kinetic energy as

$$\Delta K = -q \Delta V.$$

- If, instead, an applied force acts on the particle, doing work W_{app} , the change in kinetic energy is

$$\Delta K = -q \Delta V + W_{\text{app}}.$$

- In the special case when $\Delta K = 0$, the work of an applied force involves only the motion of the particle through a potential difference:

$$W_{\text{app}} = q \Delta V.$$

What Is Physics?

One goal of physics is to identify basic forces in our world, such as the electric force we discussed in Chapter 21. A related goal is to determine whether a force is conservative—that is, whether a potential energy can be associated with it. The motivation for associating a potential energy with a force is that we can then



Figure 24-1 Particle 1 is located at point P in the electric field of particle 2.

apply the principle of the conservation of mechanical energy to closed systems involving the force. This extremely powerful principle allows us to calculate the results of experiments for which force calculations alone would be very difficult. Experimentally, physicists and engineers discovered that the electric force is conservative and thus has an associated electric potential energy. In this chapter we first define this type of potential energy and then put it to use.

For a quick taste, let's return to the situation we considered in Chapter 22: In Figure 24-1, particle 1 with positive charge q_1 is located at point P near particle 2 with positive charge q_2 . In Chapter 22 we explained how particle 2 is able to push on particle 1 without any contact. To account for the force \vec{F} (which is a vector quantity), we defined an electric field \vec{E} (also a vector quantity) that is set up at P by particle 2. That field exists regardless of whether particle 1 is at P . If we choose to place particle 1 there, the push on it is due to charge q_1 and that pre-existing field \vec{E} .

Here is a related problem. If we release particle 1 at P , it begins to move and thus has kinetic energy. Energy cannot appear by magic, so from where does it come? It comes from the electric potential energy U associated with the force between the two particles in the arrangement of Fig. 24-1. To account for the potential energy U (which is a scalar quantity), we define an **electric potential** V (also a scalar quantity) that is set up at P by particle 2. The electric potential exists regardless of whether particle 1 is at P . If we choose to place particle 1 there, the potential energy of the two-particle system is then due to charge q_1 and that pre-existing electric potential V .

Our goals in this chapter are to (1) define electric potential, (2) discuss how to calculate it for various arrangements of charged particles and objects, and (3) discuss how electric potential V is related to electric potential energy U .

Electric Potential and Electric Potential Energy

We are going to define the electric potential (or *potential* for short) in terms of electric potential energy, so our first job is to figure out how to measure that potential energy. Back in Chapter 8, we measured gravitational potential energy U of an object by (1) assigning $U = 0$ for a reference configuration (such as the object at table level) and (2) then calculating the work W the gravitational force does if the object is moved up or down from that level. We then defined the potential energy as being

$$U = -W \quad (\text{potential energy}). \quad (24-1)$$

Let's follow the same procedure with our new conservative force, the electric force. In Fig. 24-2a, we want to find the potential energy U associated with a positive test charge q_0 located at point P in the electric field of a charged rod. First, we need a reference configuration for which $U = 0$. A reasonable choice is for the test charge to be infinitely far from the rod, because then there is no interaction with the rod. Next, we bring the test charge in from infinity to point P to form the configuration of Fig. 24-2a. Along the way, we calculate the work done by the electric force on the test charge. The potential energy of the final configuration is then given by Eq. 24-1, where W is now the work done by the electric force. Let's use the notation W_∞ to emphasize that the test charge is brought in from infinity. The work and thus the potential energy can be positive or negative depending on the sign of the rod's charge.

Next, we define the electric potential V at P in terms of the work done by the electric force and the resulting potential energy:

$$V = \frac{-W_\infty}{q_0} = \frac{U}{q_0} \quad (\text{electric potential}). \quad (24-2)$$

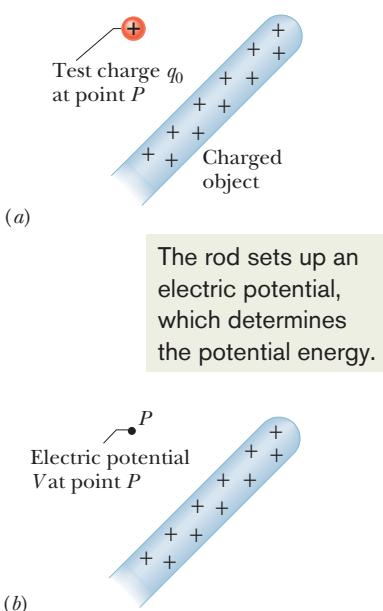


Figure 24-2 (a) A test charge has been brought in from infinity to point P in the electric field of the rod. (b) We define an electric potential V at P based on the potential energy of the configuration in (a).

That is, the electric potential is the amount of electric potential energy per unit charge when a positive test charge is brought in from infinity. The rod sets up this potential V at P regardless of whether the test charge (or anything else) happens to be there (Fig. 24-2b). From Eq. 24-2 we see that V is a scalar quantity (because there is no direction associated with potential energy or charge) and can be positive or negative (because potential energy and charge have signs).

Repeating this procedure we find that an electric potential is set up at every point in the rod's electric field. In fact, every charged object sets up electric potential V at points throughout its electric field. If we happen to place a particle with, say, charge q at a point where we know the pre-existing V , we can immediately find the potential energy of the configuration:

$$(\text{electric potential energy}) = (\text{particle's charge}) \left(\frac{\text{electric potential energy}}{\text{unit charge}} \right),$$

or

$$U = qV, \quad (24-3)$$

where q can be positive or negative.

Two Cautions. (1) The (now very old) decision to call V a *potential* was unfortunate because the term is easily confused with *potential energy*. Yes, the two quantities are related (that is the point here) but they are very different and not interchangeable. (2) Electric potential is a scalar, not a vector. (When you come to the homework problems, you will rejoice on this point.)

Language. A potential energy is a property of a system (or configuration) of objects, but sometimes we can get away with assigning it to a single object. For example, the gravitational potential energy of a baseball hit to outfield is actually a potential energy of the baseball–Earth system (because it is associated with the force between the baseball and Earth). However, because only the baseball noticeably moves (its motion does not noticeably affect Earth), we might assign the gravitational potential energy to it alone. In a similar way, if a charged particle is placed in an electric field and has no noticeable effect on the field (or the charged object that sets up the field), we usually assign the electric potential energy to the particle alone.

Units. The SI unit for potential that follows from Eq. 24-2 is the joule per coulomb. This combination occurs so often that a special unit, the *volt* (abbreviated V), is used to represent it. Thus,

$$1 \text{ volt} = 1 \text{ joule per coulomb.}$$

With two unit conversions, we can now switch the unit for electric field from newtons per coulomb to a more conventional unit:

$$\begin{aligned} 1 \text{ N/C} &= \left(1 \frac{\text{N}}{\text{C}}\right) \left(\frac{1 \text{ V}}{1 \text{ J/C}}\right) \left(\frac{1 \text{ J}}{1 \text{ N}\cdot\text{m}}\right) \\ &= 1 \text{ V/m.} \end{aligned}$$

The conversion factor in the second set of parentheses comes from our definition of volt given above; that in the third set of parentheses is derived from the definition of the joule. From now on, we shall express values of the electric field in volts per meter rather than in newtons per coulomb.

Motion Through an Electric Field

Change in Electric Potential. If we move from an initial point i to a second point f in the electric field of a charged object, the electric potential changes by

$$\Delta V = V_f - V_i$$

If we move a particle with charge q from i to f , then, from Eq. 24-3, the potential energy of the system changes by

$$\Delta U = q \Delta V = q(V_f - V_i). \quad (24-4)$$

The change can be positive or negative, depending on the signs of q and ΔV . It can also be zero, if there is no change in potential from i to f (the points have the same value of potential). Because the electric force is conservative, the change in potential energy ΔU between i and f is the same for all paths between those points (it is *path independent*).

Work by the Field. We can relate the potential energy change ΔU to the work W done by the electric force as the particle moves from i to f by applying the general relation for a conservative force (Eq. 8-1):

$$W = -\Delta U \quad (\text{work, conservative force}). \quad (24-5)$$

Next, we can relate that work to the change in the potential by substituting from Eq. 24-4:

$$W = -\Delta U = -q \Delta V = -q(V_f - V_i). \quad (24-6)$$

Up until now, we have always attributed work to a force but here can also say that W is the work done on the particle by the electric field (because it, of course, produces the force). The work can be positive, negative, or zero. Because ΔU between any two points is path independent, so is the work W done by the field. (If you need to calculate work for a difficult path, switch to an easier path—you get the same result.)

Conservation of Energy. If a charged particle moves through an electric field with no force acting on it other than the electric force due to the field, then the mechanical energy is conserved. Let's assume that we can assign the electric potential energy to the particle alone. Then we can write the conservation of mechanical energy of the particle that moves from point i to point f as

$$U_i + K_i = U_f + K_f, \quad (24-7)$$

$$\text{or} \quad \Delta K = -\Delta U. \quad (24-8)$$

Substituting Eq. 24-4, we find a very useful equation for the change in the particle's kinetic energy as a result of the particle moving through a potential difference:

$$\Delta K = -q \Delta V = -q(V_f - V_i). \quad (24-9)$$

Work by an Applied Force. If some force in addition to the electric force acts on the particle, we say that the additional force is an *applied force* or *external force*, which is often attributed to an *external agent*. Such an applied force can do work on the particle, but the force may not be conservative and thus, in general, we cannot associate a potential energy with it. We account for that work W_{app} by modifying Eq. 24-7:

$$(\text{initial energy}) + (\text{work by applied force}) = (\text{final energy})$$

$$\text{or} \quad U_i + K_i + W_{\text{app}} = U_f + K_f. \quad (24-10)$$

Rearranging and substituting from Eq. 24-4, we can also write this as

$$\Delta K = -\Delta U + W_{\text{app}} = -q \Delta V + W_{\text{app}}. \quad (24-11)$$

The work by the applied force can be positive, negative, or zero, and thus the energy of the system can increase, decrease, or remain the same.

In the special case where the particle is stationary before and after the move, the kinetic energy terms in Eqs. 24-10 and 24-11 are zero and we have

$$W_{\text{app}} = q \Delta V \quad (\text{for } K_i = K_f). \quad (24-12)$$

In this special case, the work W_{app} involves the motion of the particle through the potential difference ΔV and not a change in the particle's kinetic energy.

By comparing Eqs. 24-6 and 24-12, we see that in this special case, the work by the applied force is the negative of the work by the field:

$$W_{\text{app}} = -W \quad (\text{for } K_i = K_f). \quad (24-13)$$

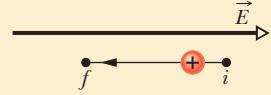
Electron-volts. In atomic and subatomic physics, energy measures in the SI unit of joules often require awkward powers of ten. A more convenient (but non-SI unit) is the *electron-volt* (eV), which is defined to be equal to the work required to move a single elementary charge e (such as that of an electron or proton) through a potential difference ΔV of exactly one volt. From Eq. 24-6, we see that the magnitude of this work is $q \Delta V$. Thus,

$$\begin{aligned} 1 \text{ eV} &= e(1 \text{ V}) \\ &= (1.602 \times 10^{-19} \text{ C})(1 \text{ J/C}) = 1.602 \times 10^{-19} \text{ J}. \end{aligned} \quad (24-14)$$



Checkpoint 1

In the figure, we move a proton from point i to point f in a uniform electric field. Is positive or negative work done by (a) the electric field and (b) our force? (c) Does the electric potential energy increase or decrease? (d) Does the proton move to a point of higher or lower electric potential?



Sample Problem 24.01 Work and potential energy in an electric field

Electrons are continually being knocked out of air molecules in the atmosphere by cosmic-ray particles coming in from space. Once released, each electron experiences an electric force \vec{F} due to the electric field \vec{E} that is produced in the atmosphere by charged particles already on Earth. Near Earth's surface the electric field has the magnitude $E = 150 \text{ N/C}$ and is directed downward. What is the change ΔU in the electric potential energy of a released electron when the electric force causes it to move vertically upward through a distance $d = 520 \text{ m}$ (Fig. 24-3)? Through what potential change does the electron move?

KEY IDEAS

- (1) The change ΔU in the electric potential energy of the electron is related to the work W done on the electron by the electric field. Equation 24-5 ($W = -\Delta U$) gives the relation.
- (2) The work done by a constant force \vec{F} on a particle undergoing a displacement \vec{d} is

$$W = \vec{F} \cdot \vec{d}.$$

- (3) The electric force and the electric field are related by the force equation $\vec{F} = q\vec{E}$, where here q is the charge of an electron ($= -1.6 \times 10^{-19} \text{ C}$).

Calculations: Substituting the force equation into the work equation and taking the dot product yield

$$W = q\vec{E} \cdot \vec{d} = qEd \cos \theta,$$

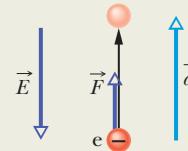


Figure 24-3 An electron in the atmosphere is moved upward through displacement \vec{d} by an electric force \vec{F} due to an electric field \vec{E} .

where θ is the angle between the directions of \vec{E} and \vec{d} . The field \vec{E} is directed downward and the displacement \vec{d} is directed upward; so $\theta = 180^\circ$. We can now evaluate the work as

$$\begin{aligned} W &= (-1.6 \times 10^{-19} \text{ C})(150 \text{ N/C})(520 \text{ m}) \cos 180^\circ \\ &= 1.2 \times 10^{-14} \text{ J}. \end{aligned}$$

Equation 24-5 then yields

$$\Delta U = -W = -1.2 \times 10^{-14} \text{ J}. \quad (\text{Answer})$$

This result tells us that during the 520 m ascent, the electric potential energy of the electron *decreases* by $1.2 \times 10^{-14} \text{ J}$. To find the change in electric potential, we apply Eq. 24-4:

$$\begin{aligned} \Delta V &= \frac{\Delta U}{-q} = \frac{-1.2 \times 10^{-14} \text{ J}}{-1.6 \times 10^{-19} \text{ C}} \\ &= 4.5 \times 10^4 \text{ V} = 45 \text{ kV}. \quad (\text{Answer}) \end{aligned}$$

This tells us that the electric force does work to move the electron to a *higher* potential.



Additional examples, video, and practice available at WileyPLUS

24-2 EQUIPOTENTIAL SURFACES AND THE ELECTRIC FIELD

Learning Objectives

After reading this module, you should be able to . . .

24.09 Identify an equipotential surface and describe how it is related to the direction of the associated electric field.

24.10 Given an electric field as a function of position, calculate the change in potential ΔV from an initial point to a final point by choosing a path between the points and integrating the dot product of the field \vec{E} and a length element $d\vec{s}$ along the path.

24.11 For a uniform electric field, relate the field magnitude E and the separation Δx and potential difference ΔV between adjacent equipotential lines.

24.12 Given a graph of electric field E versus position along an axis, calculate the change in potential ΔV from an initial point to a final point by graphical integration.

24.13 Explain the use of a zero-potential location.

Key Ideas

- The points on an equipotential surface all have the same electric potential. The work done on a test charge in moving it from one such surface to another is independent of the locations of the initial and final points on these surfaces and of the path that joins the points. The electric field \vec{E} is always directed perpendicularly to corresponding equipotential surfaces.
- The electric potential difference between two points i and f is

$$V_f - V_i = - \int_i^f \vec{E} \cdot d\vec{s},$$

where the integral is taken over any path connecting the points. If the integration is difficult along any particular path,

we can choose a different path along which the integration might be easier.

- If we choose $V_i = 0$, we have, for the potential at a particular point,

$$V = - \int_i^f \vec{E} \cdot d\vec{s}.$$

- In a uniform field of magnitude E , the change in potential from a higher equipotential surface to a lower one, separated by distance Δx , is

$$\Delta V = -E \Delta x.$$

Equipotential Surfaces

Adjacent points that have the same electric potential form an **equipotential surface**, which can be either an imaginary surface or a real, physical surface. No net work W is done on a charged particle by an electric field when the particle moves between two points i and f on the same equipotential surface. This follows from Eq. 24-6, which tells us that W must be zero if $V_f = V_i$. Because of the path independence of work (and thus of potential energy and potential), $W = 0$ for any path connecting points i and f on a given equipotential surface regardless of whether that path lies entirely on that surface.

Figure 24-4 shows a *family* of equipotential surfaces associated with the electric field due to some distribution of charges. The work done by the electric field on a charged particle as the particle moves from one end to the other of paths

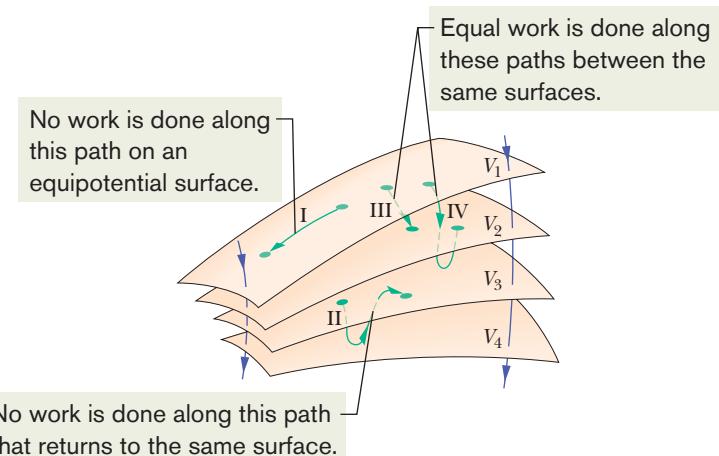


Figure 24-4 Portions of four equipotential surfaces at electric potentials $V_1 = 100$ V, $V_2 = 80$ V, $V_3 = 60$ V, and $V_4 = 40$ V. Four paths along which a test charge may move are shown. Two electric field lines are also indicated.

I and II is zero because each of these paths begins and ends on the same equipotential surface and thus there is no net change in potential. The work done as the charged particle moves from one end to the other of paths III and IV is not zero but has the same value for both these paths because the initial and final potentials are identical for the two paths; that is, paths III and IV connect the same pair of equipotential surfaces.

From symmetry, the equipotential surfaces produced by a charged particle or a spherically symmetrical charge distribution are a family of concentric spheres. For a uniform electric field, the surfaces are a family of planes perpendicular to the field lines. In fact, equipotential surfaces are always perpendicular to electric field lines and thus to \vec{E} , which is always tangent to these lines. If \vec{E} were not perpendicular to an equipotential surface, it would have a component lying along that surface. This component would then do work on a charged particle as it moved along the surface. However, by Eq. 24-6 work cannot be done if the surface is truly an equipotential surface; the only possible conclusion is that \vec{E} must be everywhere perpendicular to the surface. Figure 24-5 shows electric field lines and cross sections of the equipotential surfaces for a uniform electric field and for the field associated with a charged particle and with an electric dipole.

Calculating the Potential from the Field

We can calculate the potential difference between any two points i and f in an electric field if we know the electric field vector \vec{E} all along any path connecting those points. To make the calculation, we find the work done on a positive test charge by the field as the charge moves from i to f , and then use Eq. 24-6.

Consider an arbitrary electric field, represented by the field lines in Fig. 24-6, and a positive test charge q_0 that moves along the path shown from point i to point f . At any point on the path, an electric force $q_0\vec{E}$ acts on the charge as it moves through a differential displacement $d\vec{s}$. From Chapter 7, we know that the differential work dW done on a particle by a force \vec{F} during a displacement $d\vec{s}$ is given by the dot product of the force and the displacement:

$$dW = \vec{F} \cdot d\vec{s}. \quad (24-15)$$

For the situation of Fig. 24-6, $\vec{F} = q_0\vec{E}$ and Eq. 24-15 becomes

$$dW = q_0\vec{E} \cdot d\vec{s}. \quad (24-16)$$

To find the total work W done on the particle by the field as the particle moves from point i to point f , we sum—via integration—the differential works done on the charge as it moves through all the displacements $d\vec{s}$ along the path:

$$W = q_0 \int_i^f \vec{E} \cdot d\vec{s}. \quad (24-17)$$

If we substitute the total work W from Eq. 24-17 into Eq. 24-6, we find

$$V_f - V_i = - \int_i^f \vec{E} \cdot d\vec{s}. \quad (24-18)$$

Figure 24-6 A test charge q_0 moves from point i to point f along the path shown in a nonuniform electric field. During a displacement $d\vec{s}$, an electric force $q_0\vec{E}$ acts on the test charge. This force points in the direction of the field line at the location of the test charge.

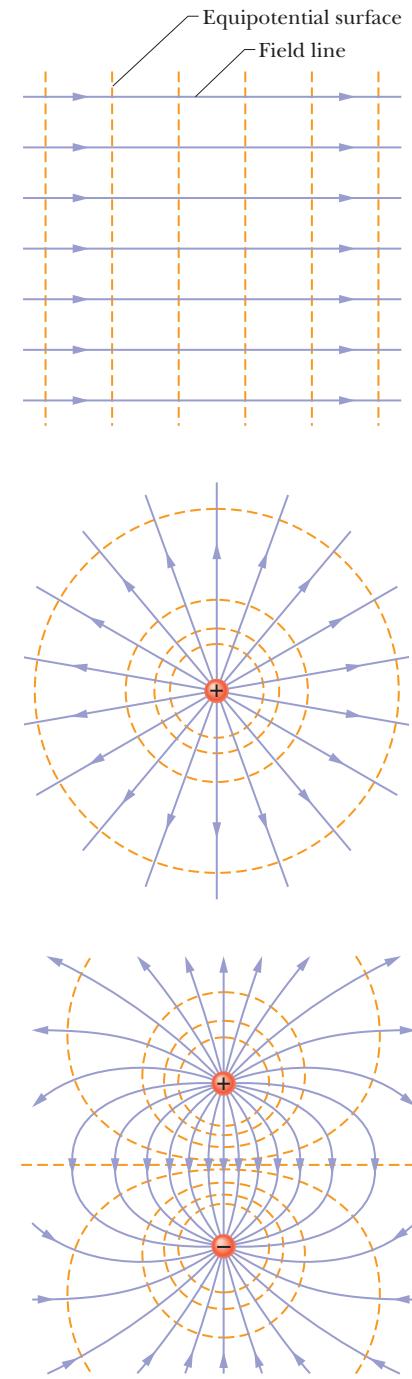
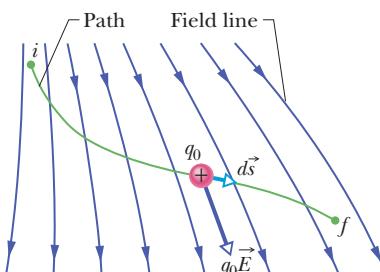


Figure 24-5 Electric field lines (purple) and cross sections of equipotential surfaces (gold) for (a) a uniform electric field, (b) the field due to a charged particle, and (c) the field due to an electric dipole.

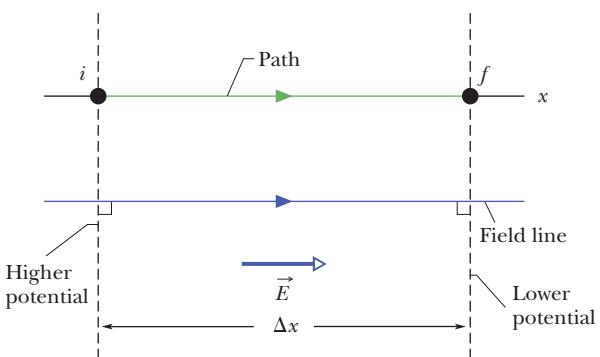


Figure 24-7 We move between points i and f , between adjacent equipotential lines in a uniform electric field \vec{E} , parallel to a field line.

Thus, the potential difference $V_f - V_i$ between any two points i and f in an electric field is equal to the negative of the *line integral* (meaning the integral along a particular path) of $\vec{E} \cdot d\vec{s}$ from i to f . However, because the electric force is conservative, all paths (whether easy or difficult to use) yield the same result.

Equation 24-18 allows us to calculate the difference in potential between any two points in the field. If we set potential $V_i = 0$, then Eq. 24-18 becomes

$$V = - \int_i^f \vec{E} \cdot d\vec{s}, \quad (24-19)$$

in which we have dropped the subscript f on V_f . Equation 24-19 gives us the potential V at any point f in the electric field *relative to the zero potential* at point i . If we let point i be at infinity, then Eq. 24-19 gives us the potential V at any point f relative to the zero potential at infinity.

Uniform Field. Let's apply Eq. 24-18 for a uniform field as shown in Fig. 24-7. We start at point i on an equipotential line with potential V_i and move to point f on an equipotential line with a lower potential V_f . The separation between the two equipotential lines is Δx . Let's also move along a path that is parallel to the electric field \vec{E} (and thus perpendicular to the equipotential lines). The angle between \vec{E} and $d\vec{s}$ in Eq. 24-18 is zero, and the dot product gives us

$$\vec{E} \cdot d\vec{s} = E ds \cos 0 = E ds.$$

Because E is constant for a uniform field, Eq. 24-18 becomes

$$V_f - V_i = -E \int_i^f ds. \quad (24-20)$$

The integral is simply an instruction for us to add all the displacement elements ds from i to f , but we already know that the sum is length Δx . Thus we can write the change in potential $V_f - V_i$ in this uniform field as

$$\Delta V = -E \Delta x \quad (\text{uniform field}). \quad (24-21)$$

This is the change in voltage ΔV between two equipotential lines in a uniform field of magnitude E , separated by distance Δx . If we move in the direction of the field by distance Δx , the potential decreases. In the opposite direction, it increases.

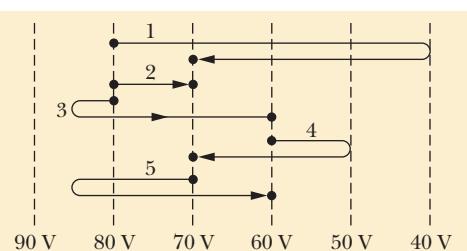


The electric field vector points from higher potential toward lower potential.



Checkpoint 2

The figure here shows a family of parallel equipotential surfaces (in cross section) and five paths along which we shall move an electron from one surface to another. (a) What is the direction of the electric field associated with the surfaces? (b) For each path, is the work we do positive, negative, or zero? (c) Rank the paths according to the work we do, greatest first.



Sample Problem 24.02 Finding the potential change from the electric field

(a) Figure 24-8a shows two points i and f in a uniform electric field \vec{E} . The points lie on the same electric field line (not shown) and are separated by a distance d . Find the potential difference $V_f - V_i$ by moving a positive test charge q_0 from i to f along the path shown, which is parallel to the field direction.

KEY IDEA

We can find the potential difference between any two points in an electric field by integrating $\vec{E} \cdot d\vec{s}$ along a path connecting those two points according to Eq. 24-18.

Calculations: We have actually already done the calculation for such a path in the direction of an electric field line in a uniform field when we derived Eq. 24-21. With slight changes in notation, Eq. 24-21 gives us

$$V_f - V_i = -Ed. \quad (\text{Answer})$$

(b) Now find the potential difference $V_f - V_i$ by moving the positive test charge q_0 from i to f along the path icf shown in Fig. 24-8b.

Calculations: The Key Idea of (a) applies here too, except now we move the test charge along a path that consists of two lines: ic and cf . At all points along line ic , the displace-

ment $d\vec{s}$ of the test charge is perpendicular to \vec{E} . Thus, the angle θ between \vec{E} and $d\vec{s}$ is 90° , and the dot product $\vec{E} \cdot d\vec{s}$ is 0. Equation 24-18 then tells us that points i and c are at the same potential: $V_c - V_i = 0$. Ah, we should have seen this coming. The points are on the same equipotential surface, which is perpendicular to the electric field lines.

For line cf we have $\theta = 45^\circ$ and, from Eq. 24-18,

$$\begin{aligned} V_f - V_i &= - \int_c^f \vec{E} \cdot d\vec{s} = - \int_c^f E(\cos 45^\circ) ds \\ &= -E(\cos 45^\circ) \int_c^f ds. \end{aligned}$$

The integral in this equation is just the length of line cf ; from Fig. 24-8b, that length is $d/\cos 45^\circ$. Thus,

$$V_f - V_i = -E(\cos 45^\circ) \frac{d}{\cos 45^\circ} = -Ed. \quad (\text{Answer})$$

This is the same result we obtained in (a), as it must be; the potential difference between two points does not depend on the path connecting them. Moral: When you want to find the potential difference between two points by moving a test charge between them, you can save time and work by choosing a path that simplifies the use of Eq. 24-18.

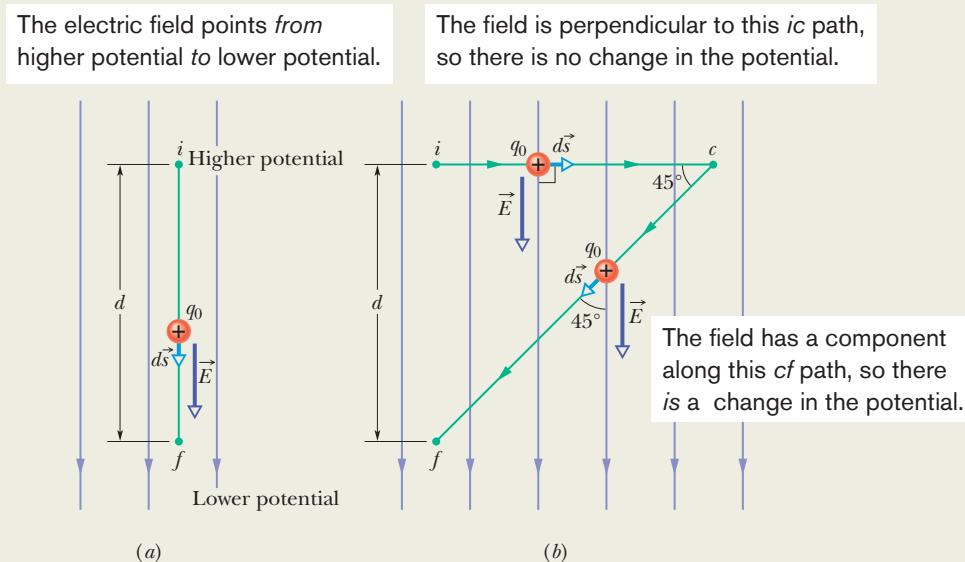


Figure 24-8 (a) A test charge q_0 moves in a straight line from point i to point f , along the direction of a uniform external electric field. (b) Charge q_0 moves along path icf in the same electric field.

24-3 POTENTIAL DUE TO A CHARGED PARTICLE

Learning Objectives

After reading this module, you should be able to . . .

24.14 For a given point in the electric field of a charged particle, apply the relationship between the electric potential V , the charge of the particle q , and the distance r from the particle.

24.15 Identify the correlation between the algebraic signs of the potential set up by a particle and the charge of the particle.

24.16 For points outside or on the surface of a spherically

symmetric charge distribution, calculate the electric potential as if all the charge is concentrated as a particle at the center of the sphere.

24.17 Calculate the net potential at any given point due to several charged particles, identifying that algebraic addition is used, not vector addition.

24.18 Draw equipotential lines for a charged particle.

Key Ideas

- The electric potential due to a single charged particle at a distance r from that charged particle is

$$V = \frac{1}{4\pi\epsilon_0} \frac{q}{r},$$

where V has the same sign as q .

- The potential due to a collection of charged particles is

$$V = \sum_{i=1}^n V_i = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_i}.$$

Thus, the potential is the algebraic sum of the individual potentials, with no consideration of directions.

Potential Due to a Charged Particle

We now use Eq. 24-18 to derive, for the space around a charged particle, an expression for the electric potential V relative to the zero potential at infinity. Consider a point P at distance R from a fixed particle of positive charge q (Fig. 24-9). To use Eq. 24-18, we imagine that we move a positive test charge q_0 from point P to infinity. Because the path we take does not matter, let us choose the simplest one—a line that extends radially from the fixed particle through P to infinity.

To use Eq. 24-18, we must evaluate the dot product

$$\vec{E} \cdot d\vec{s} = E \cos \theta ds. \quad (24-22)$$

The electric field \vec{E} in Fig. 24-9 is directed radially outward from the fixed particle. Thus, the differential displacement $d\vec{s}$ of the test particle along its path has the same direction as \vec{E} . That means that in Eq. 24-22, angle $\theta = 0$ and $\cos \theta = 1$. Because the path is radial, let us write ds as dr . Then, substituting the limits R and ∞ , we can write Eq. 24-18 as

$$V_f - V_i = - \int_R^\infty E dr. \quad (24-23)$$

Next, we set $V_f = 0$ (at ∞) and $V_i = V$ (at R). Then, for the magnitude of the electric field at the site of the test charge, we substitute from Eq. 22-3:

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2}. \quad (24-24)$$

With these changes, Eq. 24-23 then gives us

$$\begin{aligned} 0 - V &= - \frac{q}{4\pi\epsilon_0} \int_R^\infty \frac{1}{r^2} dr = \frac{q}{4\pi\epsilon_0} \left[\frac{1}{r} \right]_R^\infty \\ &= - \frac{1}{4\pi\epsilon_0} \frac{q}{R}. \end{aligned} \quad (24-25)$$

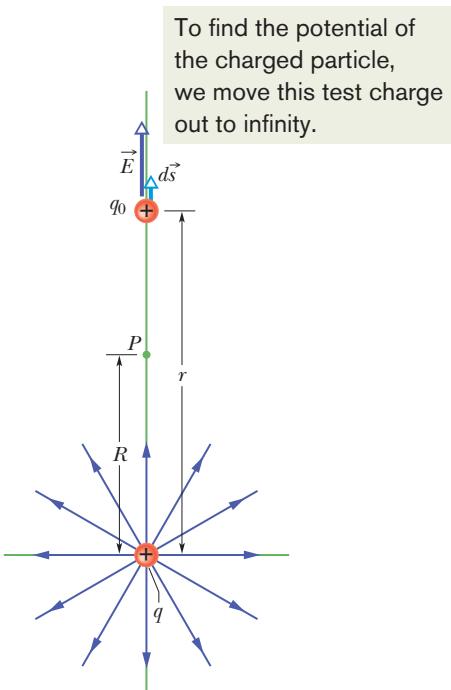


Figure 24-9 The particle with positive charge q produces an electric field \vec{E} and an electric potential V at point P . We find the potential by moving a test charge q_0 from P to infinity. The test charge is shown at distance r from the particle, during differential displacement $d\vec{s}$.

Solving for V and switching R to r , we then have

$$V = \frac{1}{4\pi\epsilon_0} \frac{q}{r} \quad (24-26)$$

as the electric potential V due to a particle of charge q at any radial distance r from the particle.

Although we have derived Eq. 24-26 for a positively charged particle, the derivation holds also for a negatively charged particle, in which case, q is a negative quantity. Note that the sign of V is the same as the sign of q :



A positively charged particle produces a positive electric potential. A negatively charged particle produces a negative electric potential.

Figure 24-10 shows a computer-generated plot of Eq. 24-26 for a positively charged particle; the magnitude of V is plotted vertically. Note that the magnitude increases as $r \rightarrow 0$. In fact, according to Eq. 24-26, V is infinite at $r = 0$, although Fig. 24-10 shows a finite, smoothed-off value there.

Equation 24-26 also gives the electric potential either *outside or on the external surface* of a spherically symmetric charge distribution. We can prove this by using one of the shell theorems of Modules 21-1 and 23-6 to replace the actual spherical charge distribution with an equal charge concentrated at its center. Then the derivation leading to Eq. 24-26 follows, provided we do not consider a point within the actual distribution.

Potential Due to a Group of Charged Particles

We can find the net electric potential at a point due to a group of charged particles with the help of the superposition principle. Using Eq. 24-26 with the plus or minus sign of the charge included, we calculate separately the potential resulting from each charge at the given point. Then we sum the potentials. Thus, for n charges, the net potential is

$$V = \sum_{i=1}^n V_i = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_i} \quad (n \text{ charged particles}). \quad (24-27)$$

Here q_i is the value of the i th charge and r_i is the radial distance of the given point from the i th charge. The sum in Eq. 24-27 is an *algebraic sum*, not a vector sum like the sum that would be used to calculate the electric field resulting from a group of charged particles. Herein lies an important computational advantage of potential over electric field: It is a lot easier to sum several scalar quantities than to sum several vector quantities whose directions and components must be considered.



Checkpoint 3

The figure here shows three arrangements of two protons. Rank the arrangements according to the net electric potential produced at point P by the protons, greatest first.

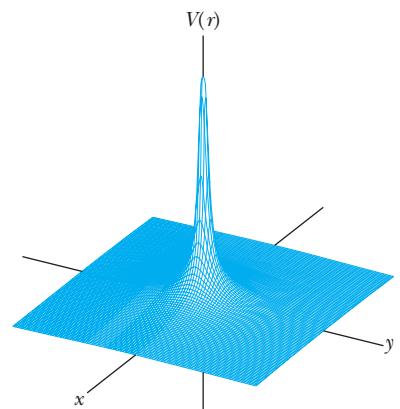
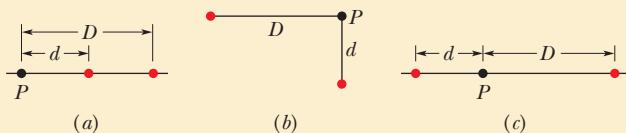


Figure 24-10 A computer-generated plot of the electric potential $V(r)$ due to a positively charged particle located at the origin of an xy plane. The potentials at points in the xy plane are plotted vertically. (Curved lines have been added to help you visualize the plot.) The infinite value of V predicted by Eq. 24-26 for $r = 0$ is not plotted.

Sample Problem 24.03 Net potential of several charged particles

What is the electric potential at point P , located at the center of the square of charged particles shown in Fig. 24-11a? The distance d is 1.3 m, and the charges are

$$\begin{aligned} q_1 &= +12 \text{ nC}, & q_3 &= +31 \text{ nC}, \\ q_2 &= -24 \text{ nC}, & q_4 &= +17 \text{ nC}. \end{aligned}$$

KEY IDEA

The electric potential V at point P is the algebraic sum of the electric potentials contributed by the four particles.

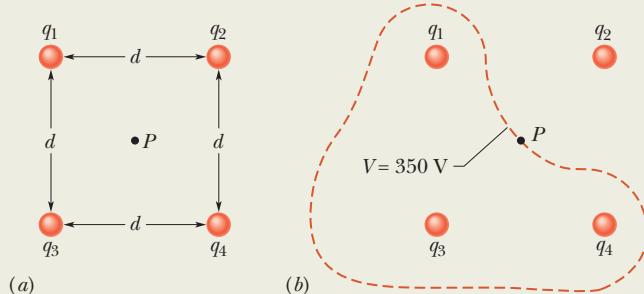


Figure 24-11 (a) Four charged particles. (b) The closed curve is a (roughly drawn) cross section of the equipotential surface that contains point P .

(Because electric potential is a scalar, the orientations of the particles do not matter.)

Calculations: From Eq. 24-27, we have

$$V = \sum_{i=1}^4 V_i = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1}{r} + \frac{q_2}{r} + \frac{q_3}{r} + \frac{q_4}{r} \right).$$

The distance r is $d/\sqrt{2}$, which is 0.919 m, and the sum of the charges is

$$\begin{aligned} q_1 + q_2 + q_3 + q_4 &= (12 - 24 + 31 + 17) \times 10^{-9} \text{ C} \\ &= 36 \times 10^{-9} \text{ C}. \end{aligned}$$

$$\text{Thus, } V = \frac{(8.99 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2)(36 \times 10^{-9} \text{ C})}{0.919 \text{ m}} \approx 350 \text{ V.} \quad (\text{Answer})$$

Close to any of the three positively charged particles in Fig. 24-11a, the potential has very large positive values. Close to the single negative charge, the potential has very large negative values. Therefore, there must be points within the square that have the same intermediate potential as that at point P . The curve in Fig. 24-11b shows the intersection of the plane of the figure with the equipotential surface that contains point P .

Sample Problem 24.04 Potential is not a vector, orientation is irrelevant

(a) In Fig. 24-12a, 12 electrons (of charge $-e$) are equally spaced and fixed around a circle of radius R . Relative to $V = 0$ at infinity, what are the electric potential and electric field at the center C of the circle due to these electrons?

KEY IDEAS

(1) The electric potential V at C is the algebraic sum of the electric potentials contributed by all the electrons. Because

electric potential is a scalar, the orientations of the electrons do not matter. (2) The electric field at C is a vector quantity and thus the orientation of the electrons is important.

Calculations: Because the electrons all have the same negative charge $-e$ and are all the same distance R from C , Eq. 24-27 gives us

$$V = -12 \frac{1}{4\pi\epsilon_0} \frac{e}{R}. \quad (\text{Answer}) \quad (24-28)$$

Because of the symmetry of the arrangement in Fig. 24-12a, the electric field vector at C due to any given electron is canceled by the field vector due to the electron that is diametrically opposite it. Thus, at C ,

$$\vec{E} = 0. \quad (\text{Answer})$$

(b) The electrons are moved along the circle until they are nonuniformly spaced over a 120° arc (Fig. 24-12b). At C , find the electric potential and describe the electric field.

Reasoning: The potential is still given by Eq. 24-28, because the distance between C and each electron is unchanged and orientation is irrelevant. The electric field is no longer zero, however, because the arrangement is no longer symmetric. A net field is now directed toward the charge distribution.

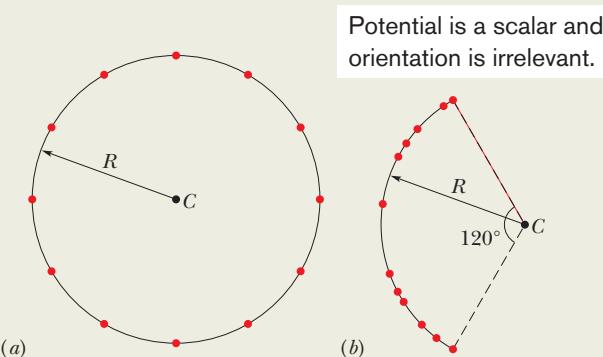


Figure 24-12 (a) Twelve electrons uniformly spaced around a circle. (b) The electrons nonuniformly spaced along an arc of the original circle.

24-4 POTENTIAL DUE TO AN ELECTRIC DIPOLE

Learning Objectives

After reading this module, you should be able to . . .

24.19 Calculate the potential V at any given point due to an electric dipole, in terms of the magnitude p of the dipole moment or the product of the charge separation d and the magnitude q of either charge.

24.20 For an electric dipole, identify the locations of positive potential, negative potential, and zero potential.

24.21 Compare the decrease in potential with increasing distance for a single charged particle and an electric dipole.

Key Idea

- At a distance r from an electric dipole with dipole moment magnitude $p = qd$, the electric potential of the dipole is

$$V = \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2}$$

for $r \gg d$; the angle θ lies between the dipole moment vector and a line extending from the dipole midpoint to the point of measurement.

Potential Due to an Electric Dipole

Now let us apply Eq. 24-27 to an electric dipole to find the potential at an arbitrary point P in Fig. 24-13a. At P , the positively charged particle (at distance $r_{(+)}$) sets up potential $V_{(+)}$ and the negatively charged particle (at distance $r_{(-)}$) sets up potential $V_{(-)}$. Then the net potential at P is given by Eq. 24-27 as

$$\begin{aligned} V &= \sum_{i=1}^2 V_i = V_{(+)} + V_{(-)} = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{r_{(+)}} + \frac{-q}{r_{(-)}} \right) \\ &= \frac{q}{4\pi\epsilon_0} \frac{r_{(-)} - r_{(+)}}{r_{(-)}r_{(+)}}. \end{aligned} \quad (24-29)$$

Naturally occurring dipoles—such as those possessed by many molecules—are quite small; so we are usually interested only in points that are relatively far from the dipole, such that $r \gg d$, where d is the distance between the charges and r is the distance from the dipole's midpoint to P . In that case, we can approximate the two lines to P as being parallel and their length difference as being the leg of a right triangle with hypotenuse d (Fig. 24-13b). Also, that difference is so small that the product of the lengths is approximately r^2 . Thus,

$$r_{(-)} - r_{(+)} \approx d \cos \theta \quad \text{and} \quad r_{(-)}r_{(+)} \approx r^2.$$

If we substitute these quantities into Eq. 24-29, we can approximate V to be

$$V = \frac{q}{4\pi\epsilon_0} \frac{d \cos \theta}{r^2},$$

where θ is measured from the dipole axis as shown in Fig. 24-13a. We can now write V as

$$V = \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2} \quad (\text{electric dipole}), \quad (24-30)$$

in which p ($= qd$) is the magnitude of the electric dipole moment \vec{p} defined in Module 22-3. The vector \vec{p} is directed along the dipole axis, from the negative to the positive charge. (Thus, θ is measured from the direction of \vec{p} .) We use this vector to report the orientation of an electric dipole.

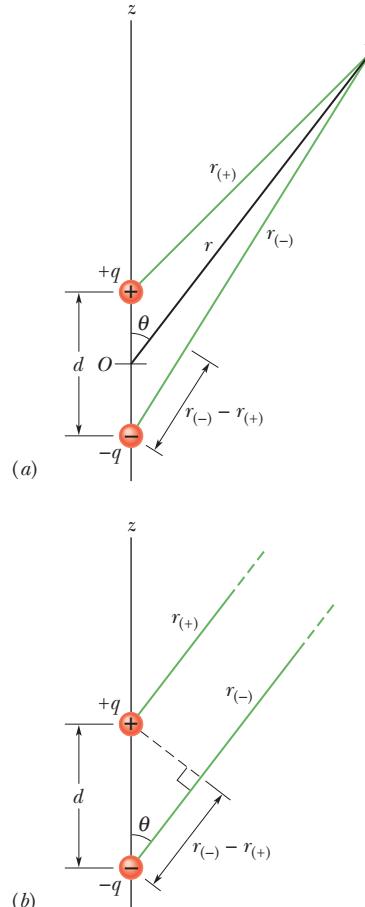


Figure 24-13 (a) Point P is a distance r from the midpoint O of a dipole. The line OP makes an angle θ with the dipole axis. (b) If P is far from the dipole, the lines of lengths $r_{(+)}$ and $r_{(-)}$ are approximately parallel to the line of length r , and the dashed black line is approximately perpendicular to the line of length $r_{(-)}$.

The electric field shifts the positive and negative charges, creating a dipole.

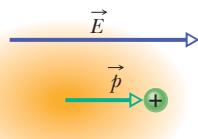


Figure 24-14 (a) An atom, showing the positively charged nucleus (green) and the negatively charged electrons (gold shading). The centers of positive and negative charge coincide. (b) If the atom is placed in an external electric field \vec{E} , the electron orbits are distorted so that the centers of positive and negative charge no longer coincide. An induced dipole moment \vec{p} appears. The distortion is greatly exaggerated here.



Checkpoint 4

Suppose that three points are set at equal (large) distances r from the center of the dipole in Fig. 24-13: Point a is on the dipole axis above the positive charge, point b is on the axis below the negative charge, and point c is on a perpendicular bisector through the line connecting the two charges. Rank the points according to the electric potential of the dipole there, greatest (most positive) first.

Induced Dipole Moment

Many molecules, such as water, have *permanent* electric dipole moments. In other molecules (called *nonpolar molecules*) and in every isolated atom, the centers of the positive and negative charges coincide (Fig. 24-14a) and thus no dipole moment is set up. However, if we place an atom or a nonpolar molecule in an external electric field, the field distorts the electron orbits and separates the centers of positive and negative charge (Fig. 24-14b). Because the electrons are negatively charged, they tend to be shifted in a direction opposite the field. This shift sets up a dipole moment \vec{p} that points in the direction of the field. This dipole moment is said to be *induced* by the field, and the atom or molecule is then said to be *polarized* by the field (that is, it has a positive side and a negative side). When the field is removed, the induced dipole moment and the polarization disappear.

24-5 POTENTIAL DUE TO A CONTINUOUS CHARGE DISTRIBUTION

Learning Objective

After reading this module, you should be able to . . .

24.22 For charge that is distributed uniformly along a line or over a surface, find the net potential at a given point by splitting the distribution up into charge elements and summing (by integration) the potential due to each one.

Key Ideas

- For a continuous distribution of charge (over an extended object), the potential is found by (1) dividing the distribution into charge elements dq that can be treated as particles and then (2) summing the potential due to each element by integrating over the full distribution:

$$V = \frac{1}{4\pi\epsilon_0} \int \frac{dq}{r}.$$

- In order to carry out the integration, dq is replaced with the product of either a linear charge density λ and a length element (such as dx), or a surface charge density σ and area element (such as $dx dy$).
- In some cases where the charge is symmetrically distributed, a two-dimensional integration can be reduced to a one-dimensional integration.

Potential Due to a Continuous Charge Distribution

When a charge distribution q is continuous (as on a uniformly charged thin rod or disk), we cannot use the summation of Eq. 24-27 to find the potential V at a point P . Instead, we must choose a differential element of charge dq , determine the potential dV at P due to dq , and then integrate over the entire charge distribution.

Let us again take the zero of potential to be at infinity. If we treat the element of charge dq as a particle, then we can use Eq. 24-26 to express the potential dV at point P due to dq :

$$dV = \frac{1}{4\pi\epsilon_0} \frac{dq}{r} \quad (\text{positive or negative } dq). \quad (24-31)$$

Here r is the distance between P and dq . To find the total potential V at P , we

integrate to sum the potentials due to all the charge elements:

$$V = \int dV = \frac{1}{4\pi\epsilon_0} \int \frac{dq}{r}. \quad (24-32)$$

The integral must be taken over the entire charge distribution. Note that because the electric potential is a scalar, there are *no vector components* to consider in Eq. 24-32.

We now examine two continuous charge distributions, a line and a disk.

Line of Charge

In Fig. 24-15a, a thin nonconducting rod of length L has a positive charge of uniform linear density λ . Let us determine the electric potential V due to the rod at point P , a perpendicular distance d from the left end of the rod.

We consider a differential element dx of the rod as shown in Fig. 24-15b. This (or any other) element of the rod has a differential charge of

$$dq = \lambda dx. \quad (24-33)$$

This element produces an electric potential dV at point P , which is a distance $r = (x^2 + d^2)^{1/2}$ from the element (Fig. 24-15c). Treating the element as a point charge, we can use Eq. 24-31 to write the potential dV as

$$dV = \frac{1}{4\pi\epsilon_0} \frac{dq}{r} = \frac{1}{4\pi\epsilon_0} \frac{\lambda dx}{(x^2 + d^2)^{1/2}}. \quad (24-34)$$

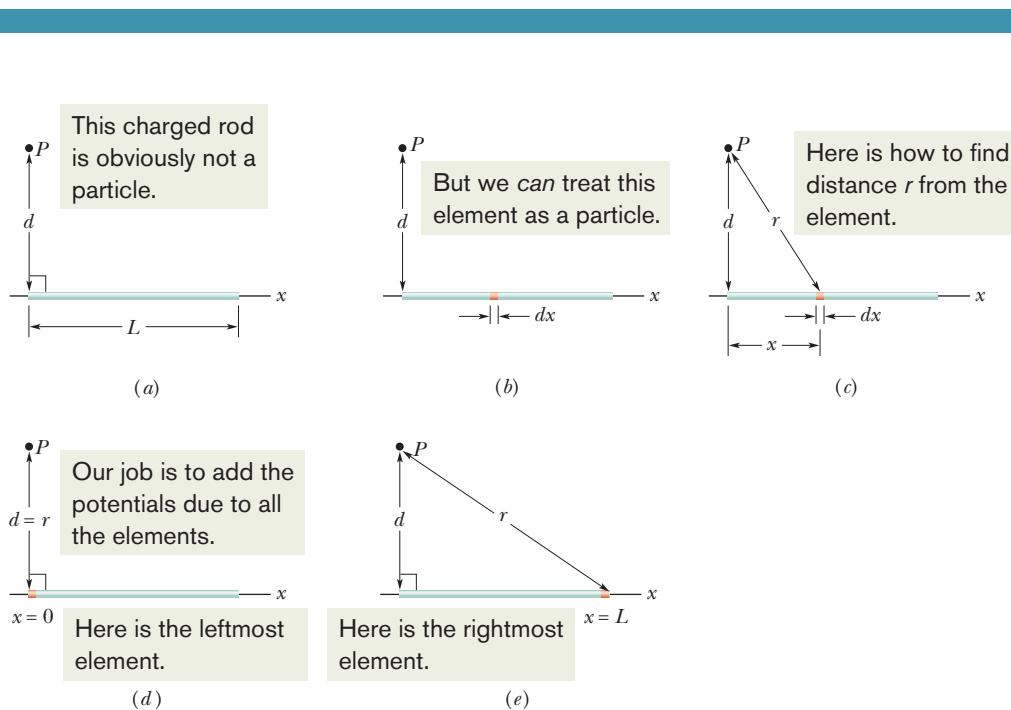


Figure 24-15 (a) A thin, uniformly charged rod produces an electric potential V at point P . (b) An element can be treated as a particle. (c) The potential at P due to the element depends on the distance r . We need to sum the potentials due to all the elements, from the left side (d) to the right side (e).

Since the charge on the rod is positive and we have taken $V = 0$ at infinity, we know from Module 24-3 that dV in Eq. 24-34 must be positive.

We now find the total potential V produced by the rod at point P by integrating Eq. 24-34 along the length of the rod, from $x = 0$ to $x = L$ (Figs. 24-15d and e), using integral 17 in Appendix E. We find

$$\begin{aligned} V &= \int dV = \int_0^L \frac{1}{4\pi\epsilon_0} \frac{\lambda}{(x^2 + d^2)^{1/2}} dx \\ &= \frac{\lambda}{4\pi\epsilon_0} \int_0^L \frac{dx}{(x^2 + d^2)^{1/2}} \\ &= \frac{\lambda}{4\pi\epsilon_0} \left[\ln(x + (x^2 + d^2)^{1/2}) \right]_0^L \\ &= \frac{\lambda}{4\pi\epsilon_0} \left[\ln(L + (L^2 + d^2)^{1/2}) - \ln d \right]. \end{aligned}$$

We can simplify this result by using the general relation $\ln A - \ln B = \ln(A/B)$. We then find

$$V = \frac{\lambda}{4\pi\epsilon_0} \ln \left[\frac{L + (L^2 + d^2)^{1/2}}{d} \right]. \quad (24-35)$$

Because V is the sum of positive values of dV , it too is positive, consistent with the logarithm being positive for an argument greater than 1.

Charged Disk

In Module 22-5, we calculated the magnitude of the electric field at points on the central axis of a plastic disk of radius R that has a uniform charge density σ on one surface. Here we derive an expression for $V(z)$, the electric potential at any point on the central axis. Because we have a circular distribution of charge on the disk, we could start with a differential element that occupies angle $d\theta$ and radial distance dr . We would then need to set up a two-dimensional integration. However, let's do something easier.

In Fig. 24-16, consider a differential element consisting of a flat ring of radius R' and radial width dR' . Its charge has magnitude

$$dq = \sigma(2\pi R')(dR'),$$

in which $(2\pi R')(dR')$ is the upper surface area of the ring. All parts of this charged element are the same distance r from point P on the disk's axis. With the aid of Fig. 24-16, we can use Eq. 24-31 to write the contribution of this ring to the electric potential at P as

$$dV = \frac{1}{4\pi\epsilon_0} \frac{dq}{r} = \frac{1}{4\pi\epsilon_0} \frac{\sigma(2\pi R')(dR')}{\sqrt{z^2 + R'^2}}. \quad (24-36)$$

We find the net potential at P by adding (via integration) the contributions of all the rings from $R' = 0$ to $R' = R$:

$$V = \int dV = \frac{\sigma}{2\epsilon_0} \int_0^R \frac{R' dR'}{\sqrt{z^2 + R'^2}} = \frac{\sigma}{2\epsilon_0} (\sqrt{z^2 + R^2} - z). \quad (24-37)$$

Note that the variable in the second integral of Eq. 24-37 is R' and not z , which remains constant while the integration over the surface of the disk is carried out. (Note also that, in evaluating the integral, we have assumed that $z \geq 0$.)

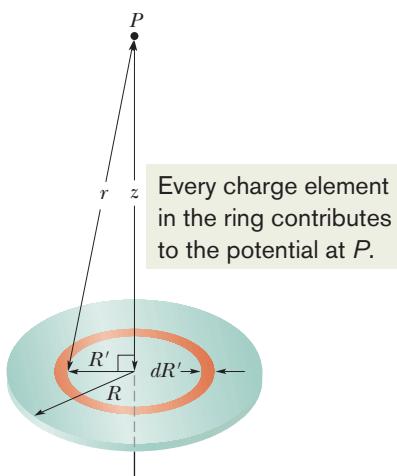


Figure 24-16 A plastic disk of radius R , charged on its top surface to a uniform surface charge density σ . We wish to find the potential V at point P on the central axis of the disk.

24-6 CALCULATING THE FIELD FROM THE POTENTIAL

Learning Objectives

After reading this module, you should be able to . . .

- 24.23** Given an electric potential as a function of position along an axis, find the electric field along that axis.
- 24.24** Given a graph of electric potential versus position along an axis, determine the electric field along the axis.
- 24.25** For a uniform electric field, relate the field magnitude E

and the separation Δx and potential difference ΔV between adjacent equipotential lines.

- 24.26** Relate the direction of the electric field and the directions in which the potential decreases and increases.

Key Ideas

- The component of \vec{E} in any direction is the negative of the rate at which the potential changes with distance in that direction:

$$E_s = -\frac{\partial V}{\partial s}.$$

- The x , y , and z components of \vec{E} may be found from

$$E_x = -\frac{\partial V}{\partial x}; \quad E_y = -\frac{\partial V}{\partial y}; \quad E_z = -\frac{\partial V}{\partial z}.$$

When \vec{E} is uniform, all this reduces to

$$E = -\frac{\Delta V}{\Delta s},$$

where s is perpendicular to the equipotential surfaces.

- The electric field is zero parallel to an equipotential surface.

Calculating the Field from the Potential

In Module 24-2, you saw how to find the potential at a point f if you know the electric field along a path from a reference point to point f . In this module, we propose to go the other way—that is, to find the electric field when we know the potential. As Fig. 24-5 shows, solving this problem graphically is easy: If we know the potential V at all points near an assembly of charges, we can draw in a family of equipotential surfaces. The electric field lines, sketched perpendicular to those surfaces, reveal the variation of \vec{E} . What we are seeking here is the mathematical equivalent of this graphical procedure.

Figure 24-17 shows cross sections of a family of closely spaced equipotential surfaces, the potential difference between each pair of adjacent surfaces being dV . As the figure suggests, the field \vec{E} at any point P is perpendicular to the equipotential surface through P .

Suppose that a positive test charge q_0 moves through a displacement $d\vec{s}$ from one equipotential surface to the adjacent surface. From Eq. 24-6, we see that the work the electric field does on the test charge during the move is $-q_0 dV$. From Eq. 24-16 and Fig. 24-17, we see that the work done by the electric field may also be written as the scalar product $(q_0 \vec{E}) \cdot d\vec{s}$, or $q_0 E(\cos \theta) ds$. Equating these two expressions for the work yields

$$-q_0 dV = q_0 E(\cos \theta) ds, \quad (24-38)$$

or

$$E \cos \theta = -\frac{dV}{ds}. \quad (24-39)$$

Since $E \cos \theta$ is the component of \vec{E} in the direction of $d\vec{s}$, Eq. 24-39 becomes

$$E_s = -\frac{\partial V}{\partial s}. \quad (24-40)$$

We have added a subscript to E and switched to the partial derivative symbols to emphasize that Eq. 24-40 involves only the variation of V along a specified axis (here called the s axis) and only the component of \vec{E} along that axis. In words, Eq. 24-40 (which is essentially the reverse operation of Eq. 24-18) states:

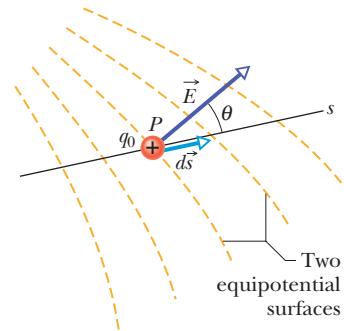


Figure 24-17 A test charge q_0 moves a distance $d\vec{s}$ from one equipotential surface to another. (The separation between the surfaces has been exaggerated for clarity.) The displacement $d\vec{s}$ makes an angle θ with the direction of the electric field \vec{E} .



The component of \vec{E} in any direction is the negative of the rate at which the electric potential changes with distance in that direction.

If we take the s axis to be, in turn, the x , y , and z axes, we find that the x , y , and z components of \vec{E} at any point are

$$E_x = -\frac{\partial V}{\partial x}; \quad E_y = -\frac{\partial V}{\partial y}; \quad E_z = -\frac{\partial V}{\partial z}. \quad (24-41)$$

Thus, if we know V for all points in the region around a charge distribution—that is, if we know the function $V(x, y, z)$ —we can find the components of \vec{E} , and thus \vec{E} itself, at any point by taking partial derivatives.

For the simple situation in which the electric field \vec{E} is uniform, Eq. 24-40 becomes

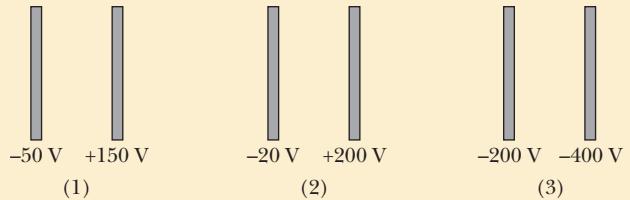
$$E = -\frac{\Delta V}{\Delta s}, \quad (24-42)$$

where s is perpendicular to the equipotential surfaces. The component of the electric field is zero in any direction parallel to the equipotential surfaces because there is no change in potential along the surfaces.



Checkpoint 5

The figure shows three pairs of parallel plates with the same separation, and the electric potential of each plate. The electric field between the plates is uniform and perpendicular to the plates. (a) Rank the pairs according to the magnitude of the electric field between the plates, greatest first. (b) For which pair is the electric field pointing rightward? (c) If an electron is released midway between the third pair of plates, does it remain there, move rightward at constant speed, move leftward at constant speed, accelerate rightward, or accelerate leftward?



Sample Problem 24.05 Finding the field from the potential

The electric potential at any point on the central axis of a uniformly charged disk is given by Eq. 24-37,

$$V = \frac{\sigma}{2\epsilon_0} (\sqrt{z^2 + R^2} - z).$$

Starting with this expression, derive an expression for the electric field at any point on the axis of the disk.

KEY IDEAS

We want the electric field \vec{E} as a function of distance z along the axis of the disk. For any value of z , the direction of \vec{E} must be along that axis because the disk has circular symmetry

about that axis. Thus, we want the component E_z of \vec{E} in the direction of z . This component is the negative of the rate at which the electric potential changes with distance z .

Calculation: Thus, from the last of Eqs. 24-41, we can write

$$\begin{aligned} E_z &= -\frac{\partial V}{\partial z} = -\frac{\sigma}{2\epsilon_0} \frac{d}{dz} (\sqrt{z^2 + R^2} - z) \\ &= \frac{\sigma}{2\epsilon_0} \left(1 - \frac{z}{\sqrt{z^2 + R^2}}\right). \end{aligned} \quad (\text{Answer})$$

This is the same expression that we derived in Module 22-5 by integration, using Coulomb's law.



Additional examples, video, and practice available at WileyPLUS

24-7 ELECTRIC POTENTIAL ENERGY OF A SYSTEM OF CHARGED PARTICLES

Learning Objectives

After reading this module, you should be able to . . .

- 24.27** Identify that the total potential energy of a system of charged particles is equal to the work an applied force must do to assemble the system, starting with the particles infinitely far apart.
- 24.28** Calculate the potential energy of a pair of charged particles.
- 24.29** Identify that if a system has more than two charged parti-

cles, then the system's total potential energy is equal to the sum of the potential energies of every pair of the particles.

- 24.30** Apply the principle of the conservation of mechanical energy to a system of charged particles.
- 24.31** Calculate the escape speed of a charged particle from a system of charged particles (the minimum initial speed required to move infinitely far from the system).

Key Idea

- The electric potential energy of a system of charged particles is equal to the work needed to assemble the system with the particles initially at rest and infinitely distant from each other. For two particles at separation r ,

$$U = W = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}.$$

Electric Potential Energy of a System of Charged Particles

In this module we are going to calculate the potential energy of a system of two charged particles and then briefly discuss how to expand the result to a system of more than two particles. Our starting point is to examine the work we must do (as an external agent) to bring together two charged particles that are initially infinitely far apart and that end up near each other and stationary. If the two particles have the same sign of charge, we must fight against their mutual repulsion. Our work is then positive and results in a positive potential energy for the final two-particle system. If, instead, the two particles have opposite signs of charge, our job is easy because of the mutual attraction of the particles. Our work is then negative and results in a negative potential energy for the system.

Let's follow this procedure to build the two-particle system in Fig. 24-18, where particle 1 (with positive charge q_1) and particle 2 (with positive charge q_2) have separation r . Although both particles are positively charged, our result will apply also to situations where they are both negatively charged or have different signs.

We start with particle 2 fixed in place and particle 1 infinitely far away, with an initial potential energy U_i for the two-particle system. Next we bring particle 1 to its final position, and then the system's potential energy is U_f . Our work changes the system's potential energy by $\Delta U = U_f - U_i$.

With Eq. 24-4 ($\Delta U = q(V_f - V_i)$), we can relate ΔU to the change in potential through which we move particle 1:

$$U_f - U_i = q_1(V_f - V_i). \quad (24-43)$$

Let's evaluate these terms. The initial potential energy is $U_i = 0$ because the particles are in the reference configuration (as discussed in Module 24-1). The two potentials in Eq. 24-43 are due to particle 2 and are given by Eq. 24-26:

$$V = \frac{1}{4\pi\epsilon_0} \frac{q_2}{r}. \quad (24-44)$$

This tells us that when particle 1 is initially at distance $r = \infty$, the potential at its location is $V_i = 0$. When we move it to the final position at distance r , the potential at its location is

$$V_f = \frac{1}{4\pi\epsilon_0} \frac{q_2}{r}. \quad (24-45)$$

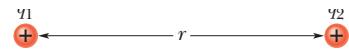


Figure 24-18 Two charges held a fixed distance r apart.

Substituting these results into Eq. 24-43 and dropping the subscript f , we find that the final configuration has a potential energy of

$$U = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r} \quad (\text{two-particle system}). \quad (24-46)$$

Equation 24-46 includes the signs of the two charges. If the two charges have the same sign, U is positive. If they have opposite signs, U is negative.

If we next bring in a third particle, with charge q_3 , we repeat our calculation, starting with particle 3 at an infinite distance and then bringing it to a final position at distance r_{31} from particle 1 and distance r_{32} from particle 2. At the final position, the potential V_f at the location of particle 3 is the algebraic sum of the potential V_1 due to particle 1 and the potential V_2 of particle 2. When we work out the algebra, we find that



The total potential energy of a system of particles is the sum of the potential energies for every pair of particles in the system.

This result applies to a system for any given number of particles.

Now that we have an expression for the potential energy of a system of particles, we can apply the principle of the conservation of energy to the system as expressed in Eq. 24-10. For example, if the system consists of many particles, we might consider the kinetic energy (and the associated *escape speed*) required of one of the particles to escape from the rest of the particles.

Sample Problem 24.06 Potential energy of a system of three charged particles

Figure 24-19 shows three charged particles held in fixed positions by forces that are not shown. What is the electric potential energy U of this system of charges? Assume that $d = 12 \text{ cm}$ and that

$$q_1 = +q, \quad q_2 = -4q, \quad \text{and} \quad q_3 = +2q,$$

in which $q = 150 \text{ nC}$.

KEY IDEA

The potential energy U of the system is equal to the work we must do to assemble the system, bringing in each charge from an infinite distance.

Calculations: Let's mentally build the system of Fig. 24-19, starting with one of the charges, say q_1 , in place and the others at infinity. Then we bring another one, say q_2 , in from infinity and put it in place. From Eq. 24-46 with d substituted for r , the potential energy U_{12} associated with the pair of charges q_1 and q_2 is

$$U_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{d}.$$

We then bring the last charge q_3 in from infinity and put it in

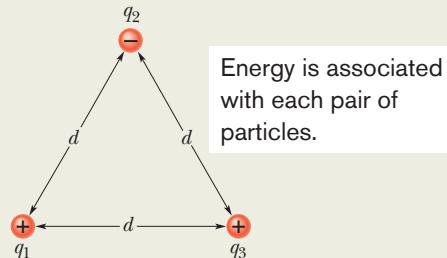


Figure 24-19 Three charges are fixed at the vertices of an equilateral triangle. What is the electric potential energy of the system?

place. The work that we must do in this last step is equal to the sum of the work we must do to bring q_3 near q_1 and the work we must do to bring it near q_2 . From Eq. 24-46, with d substituted for r , that sum is

$$W_{13} + W_{23} = U_{13} + U_{23} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_3}{d} + \frac{1}{4\pi\epsilon_0} \frac{q_2 q_3}{d}.$$

The total potential energy U of the three-charge system is the sum of the potential energies associated with the three pairs of charges. This sum (which is actually independent of the order in which the charges are brought together) is

$$\begin{aligned}
 U &= U_{12} + U_{13} + U_{23} \\
 &= \frac{1}{4\pi\epsilon_0} \left(\frac{(+q)(-4q)}{d} + \frac{(+q)(+2q)}{d} + \frac{(-4q)(+2q)}{d} \right) \\
 &= -\frac{10q^2}{4\pi\epsilon_0 d} \\
 &= -\frac{(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2)(10)(150 \times 10^{-9} \text{ C})^2}{0.12 \text{ m}} \\
 &= -1.7 \times 10^{-2} \text{ J} = -17 \text{ mJ.}
 \end{aligned}$$

(Answer)

The negative potential energy means that negative work would have to be done to assemble this structure, starting with the three charges infinitely separated and at rest. Put another way, an external agent would have to do 17 mJ of positive work to disassemble the structure completely, ending with the three charges infinitely far apart.

The lesson here is this: If you are given an assembly of charged particles, you can find the potential energy of the assembly by finding the potential of every possible pair of the particles and then summing the results.

Sample Problem 24.07 Conservation of mechanical energy with electric potential energy

An alpha particle (two protons, two neutrons) moves into a stationary gold atom (79 protons, 118 neutrons), passing through the electron region that surrounds the gold nucleus like a shell and headed directly toward the nucleus (Fig. 24-20). The alpha particle slows until it momentarily stops when its center is at radial distance $r = 9.23 \text{ fm}$ from the nuclear center. Then it moves back along its incoming path. (Because the gold nucleus is much more massive than the alpha particle, we can assume the gold nucleus does not move.) What was the kinetic energy K_i of the alpha particle when it was initially far away (hence external to the gold atom)? Assume that the only force acting between the alpha particle and the gold nucleus is the (electrostatic) Coulomb force and treat each as a single charged particle.

KEY IDEA

During the entire process, the mechanical energy of the *alpha particle + gold atom* system is conserved.

Reasoning: When the alpha particle is outside the atom, the system's initial electric potential energy U_i is zero because the atom has an equal number of electrons and protons, which produce a *net* electric field of zero. However, once the alpha particle passes through the electron region surrounding the nucleus on its way to the nucleus, the electric field due to the electrons goes to zero. The reason is that the electrons act like a closed spherical shell of uniform negative charge and, as discussed in Module 23-6, such a shell produces zero electric field in the space it encloses. The alpha particle still experiences the electric field of the protons in the nucleus, which produces a repulsive force on the protons within the alpha particle.

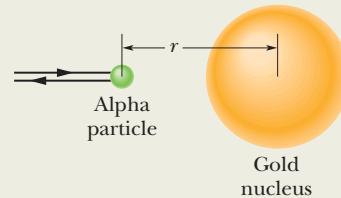


Figure 24-20 An alpha particle, traveling head-on toward the center of a gold nucleus, comes to a momentary stop (at which time all its kinetic energy has been transferred to electric potential energy) and then reverses its path.

As the incoming alpha particle is slowed by this repulsive force, its kinetic energy is transferred to electric potential energy of the system. The transfer is complete when the alpha particle momentarily stops and the kinetic energy is $K_f = 0$.

Calculations: The principle of conservation of mechanical energy tells us that

$$K_i + U_i = K_f + U_f. \quad (24-47)$$

We know two values: $U_i = 0$ and $K_f = 0$. We also know that the potential energy U_f at the stopping point is given by the right side of Eq. 24-46, with $q_1 = 2e$, $q_2 = 79e$ (in which e is the elementary charge, $1.60 \times 10^{-19} \text{ C}$), and $r = 9.23 \text{ fm}$. Thus, we can rewrite Eq. 24-47 as

$$\begin{aligned}
 K_i &= \frac{1}{4\pi\epsilon_0} \frac{(2e)(79e)}{9.23 \text{ fm}} \\
 &= \frac{(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2)(158)(1.60 \times 10^{-19} \text{ C})^2}{9.23 \times 10^{-15} \text{ m}} \\
 &= 3.94 \times 10^{-12} \text{ J} = 24.6 \text{ MeV.}
 \end{aligned}$$

(Answer)

24-8 POTENTIAL OF A CHARGED ISOLATED CONDUCTOR

Learning Objectives

After reading this module, you should be able to . . .

24.32 Identify that an excess charge placed on an isolated conductor (or connected isolated conductors) will distribute itself on the surface of the conductor so that all points of the conductor come to the same potential.

24.33 For an isolated spherical conducting shell, sketch graphs of the potential and the electric field magnitude versus distance from the center, both inside and outside the shell.

24.34 For an isolated spherical conducting shell, identify that internally the electric field is zero and the electric potential

has the same value as the surface and that externally the electric field and the electric potential have values as though all of the shell's charge is concentrated as a particle at its center.

24.35 For an isolated cylindrical conducting shell, identify that internally the electric field is zero and the electric potential has the same value as the surface and that externally the electric field and the electric potential have values as though all of the cylinder's charge is concentrated as a line of charge on the central axis.

Key Ideas

- An excess charge placed on a conductor will, in the equilibrium state, be located entirely on the outer surface of the conductor.
- The entire conductor, including interior points, is at a uniform potential.
- If an isolated charged conductor is placed in an external

electric field, then at every internal point, the electric field due to the charge cancels the external electric field that otherwise would have been there.

- Also, the net electric field at every point on the surface is perpendicular to the surface.

Potential of a Charged Isolated Conductor

In Module 23-3, we concluded that $\vec{E} = 0$ for all points inside an isolated conductor. We then used Gauss' law to prove that an excess charge placed on an isolated conductor lies entirely on its surface. (This is true even if the conductor has an empty internal cavity.) Here we use the first of these facts to prove an extension of the second:



An excess charge placed on an isolated conductor will distribute itself on the surface of that conductor so that all points of the conductor—whether on the surface or inside—come to the same potential. This is true even if the conductor has an internal cavity and even if that cavity contains a net charge.

Our proof follows directly from Eq. 24-18, which is

$$V_f - V_i = - \int_i^f \vec{E} \cdot d\vec{s}$$

Since $\vec{E} = 0$ for all points within a conductor, it follows directly that $V_f = V_i$ for all possible pairs of points i and f in the conductor.

Figure 24-21a is a plot of potential against radial distance r from the center for an isolated spherical conducting shell of 1.0 m radius, having a charge of $1.0 \mu\text{C}$. For points outside the shell, we can calculate $V(r)$ from Eq. 24-26 because the charge q behaves for such external points as if it were concentrated at the center of the shell. That equation holds right up to the surface of the shell. Now let us push a small test charge through the shell—assuming a small hole exists—to its center. No extra work is needed to do this because no net electric force acts on the test charge once it is inside the shell. Thus, the potential at all points inside the shell has the same value as that on the surface, as Fig. 24-21a shows.

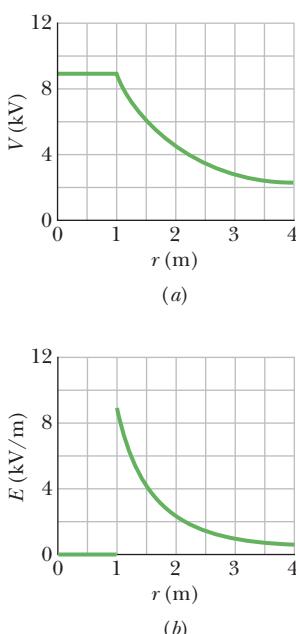


Figure 24-21 (a) A plot of $V(r)$ both inside and outside a charged spherical shell of radius 1.0 m. (b) A plot of $E(r)$ for the same shell.

Figure 24-21b shows the variation of electric field with radial distance for the same shell. Note that $E = 0$ everywhere inside the shell. The curves of Fig. 24-21b can be derived from the curve of Fig. 24-21a by differentiating with respect to r , using Eq. 24-40 (recall that the derivative of any constant is zero). The curve of Fig. 24-21a can be derived from the curves of Fig. 24-21b by integrating with respect to r , using Eq. 24-19.

Spark Discharge from a Charged Conductor

On nonspherical conductors, a surface charge does not distribute itself uniformly over the surface of the conductor. At sharp points or sharp edges, the surface charge density—and thus the external electric field, which is proportional to it—may reach very high values. The air around such sharp points or edges may become ionized, producing the corona discharge that golfers and mountaineers see on the tips of bushes, golf clubs, and rock hammers when thunderstorms threaten. Such corona discharges, like hair that stands on end, are often the precursors of lightning strikes. In such circumstances, it is wise to enclose yourself in a cavity inside a conducting shell, where the electric field is guaranteed to be zero. A car (unless it is a convertible or made with a plastic body) is almost ideal (Fig. 24-22).

Isolated Conductor in an External Electric Field

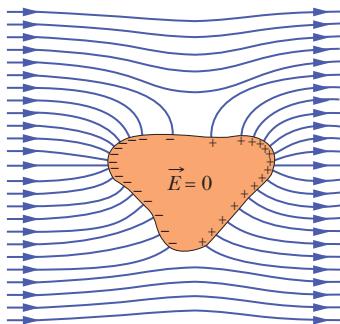
If an isolated conductor is placed in an *external electric field*, as in Fig. 24-23, all points of the conductor still come to a single potential regardless of whether the conductor has an excess charge. The free conduction electrons distribute themselves on the surface in such a way that the electric field they produce at interior points cancels the external electric field that would otherwise be there. Furthermore, the electron distribution causes the net electric field at all points on the surface to be perpendicular to the surface. If the conductor in Fig. 24-23 could be somehow removed, leaving the surface charges frozen in place, the internal and external electric field would remain absolutely unchanged.

Figure 24-23 An uncharged conductor is suspended in an external electric field. The free electrons in the conductor distribute themselves on the surface as shown, so as to reduce the net electric field inside the conductor to zero and make the net field at the surface perpendicular to the surface.



Courtesy Westinghouse Electric Corporation

Figure 24-22 A large spark jumps to a car's body and then exits by moving across the insulating left front tire (note the flash there), leaving the person inside unharmed.



Review & Summary

Electric Potential The electric potential V at a point P in the electric field of a charged object is

$$V = \frac{-W_\infty}{q_0} = \frac{U}{q_0}, \quad (24-2)$$

where W_∞ is the work that would be done by the electric force on a positive test charge were it brought from an infinite distance to P , and U is the potential energy that would then be stored in the test charge-object system.

Electric Potential Energy If a particle with charge q is placed at a point where the electric potential of a charged object is V , the electric potential energy U of the particle-object system is

$$U = qV. \quad (24-3)$$

If the particle moves through a potential difference ΔV , the change in the electric potential energy is

$$\Delta U = q \Delta V = q(V_f - V_i). \quad (24-4)$$

Mechanical Energy If a particle moves through a change ΔV in electric potential without an applied force acting on it, applying the conservation of mechanical energy gives the change in kinetic energy as

$$\Delta K = -q \Delta V. \quad (24-9)$$

If, instead, an applied force acts on the particle, doing work W_{app} , the change in kinetic energy is

$$\Delta K = -q \Delta V + W_{\text{app}}. \quad (24-11)$$

In the special case when $\Delta K = 0$, the work of an applied force

involves only the motion of the particle through a potential difference:

$$W_{\text{app}} = q \Delta V \quad (\text{for } K_i = K_f). \quad (24-12)$$

Equipotential Surfaces The points on an **equipotential surface** all have the same electric potential. The work done on a test charge in moving it from one such surface to another is independent of the locations of the initial and final points on these surfaces and of the path that joins the points. The electric field \vec{E} is always directed perpendicularly to corresponding equipotential surfaces.

Finding V from \vec{E} The electric potential difference between two points i and f is

$$V_f - V_i = - \int_i^f \vec{E} \cdot d\vec{s}, \quad (24-18)$$

where the integral is taken over any path connecting the points. If the integration is difficult along any particular path, we can choose a different path along which the integration might be easier. If we choose $V_i = 0$, we have, for the potential at a particular point,

$$V = - \int_i^f \vec{E} \cdot d\vec{s}. \quad (24-19)$$

In the special case of a uniform field of magnitude E , the potential change between two adjacent (parallel) equipotential lines separated by distance Δx is

$$\Delta V = -E \Delta x. \quad (24-21)$$

Potential Due to a Charged Particle The electric potential due to a single charged particle at a distance r from that particle is

$$V = \frac{1}{4\pi\epsilon_0} \frac{q}{r}, \quad (24-26)$$

where V has the same sign as q . The potential due to a collection of charged particles is

$$V = \sum_{i=1}^n V_i = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_i}. \quad (24-27)$$

Potential Due to an Electric Dipole At a distance r from an electric dipole with dipole moment magnitude $p = qd$, the electric potential of the dipole is

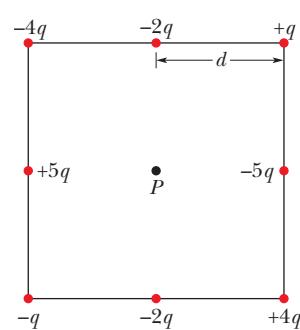


Figure 24-24 Question 1.

- 1 Figure 24-24 shows eight particles that form a square, with distance d between adjacent particles. What is the net electric potential at point P at the center of the square if we take the electric potential to be zero at infinity?

- 2 Figure 24-25 shows three sets of cross sections of equipotential surfaces in uniform electric fields; all three cover the same size region of space. The electric potential is indicated for each equipotential surface. (a) Rank the arrangements according to the magnitude of the electric field present in the region, greatest first. (b) In which is the electric field directed down the page?

$$V = \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2} \quad (24-30)$$

for $r \gg d$; the angle θ is defined in Fig. 24-13.

Potential Due to a Continuous Charge Distribution For a continuous distribution of charge, Eq. 24-27 becomes

$$V = \frac{1}{4\pi\epsilon_0} \int \frac{dq}{r}, \quad (24-32)$$

in which the integral is taken over the entire distribution.

Calculating \vec{E} from V The component of \vec{E} in any direction is the negative of the rate at which the potential changes with distance in that direction:

$$E_s = -\frac{\partial V}{\partial s}. \quad (24-40)$$

The x , y , and z components of \vec{E} may be found from

$$E_x = -\frac{\partial V}{\partial x}; \quad E_y = -\frac{\partial V}{\partial y}; \quad E_z = -\frac{\partial V}{\partial z}. \quad (24-41)$$

When \vec{E} is uniform, Eq. 24-40 reduces to

$$E = -\frac{\Delta V}{\Delta s}, \quad (24-42)$$

where s is perpendicular to the equipotential surfaces.

Electric Potential Energy of a System of Charged Particles The electric potential energy of a system of charged particles is equal to the work needed to assemble the system with the particles initially at rest and infinitely distant from each other. For two particles at separation r ,

$$U = W = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}. \quad (24-46)$$

Potential of a Charged Conductor An excess charge placed on a conductor will, in the equilibrium state, be located entirely on the outer surface of the conductor. The charge will distribute itself so that the following occur: (1) The entire conductor, including interior points, is at a uniform potential. (2) At every internal point, the electric field due to the charge cancels the external electric field that otherwise would have been there. (3) The net electric field at every point on the surface is perpendicular to the surface.

cated for each equipotential surface. (a) Rank the arrangements according to the magnitude of the electric field present in the region, greatest first. (b) In which is the electric field directed down the page?

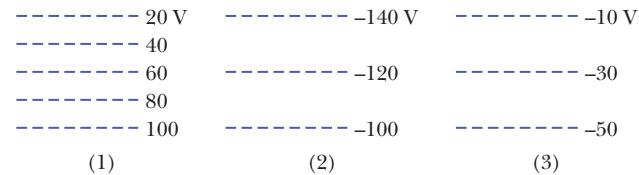


Figure 24-25 Question 2.

- 3** Figure 24-26 shows four pairs of charged particles. For each pair, let $V = 0$ at infinity and consider V_{net} at points on the x axis. For which pairs is there a point at which $V_{\text{net}} = 0$ (a) between the particles and (b) to the right of the particles? (c) At such a point is \vec{E}_{net} due to the particles equal to zero? (d) For each pair, are there off-axis points (other than at infinity) where $V_{\text{net}} = 0$?

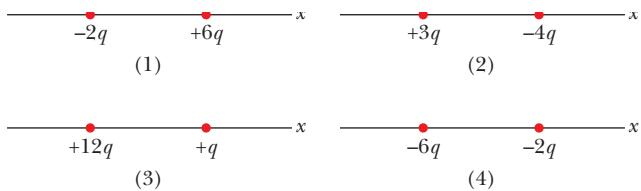


Figure 24-26 Questions 3 and 9.

- 4** Figure 24-27 gives the electric potential V as a function of x . (a) Rank the five regions according to the magnitude of the x component of the electric field within them, greatest first. What is the direction of the field along the x axis in (b) region 2 and (c) region 4?

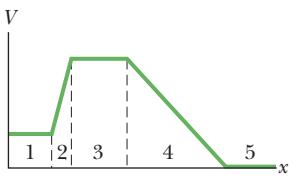


Figure 24-27 Question 4.

- 5** Figure 24-28 shows three paths along which we can move the positively charged sphere A closer to positively charged sphere B , which is held fixed in place. (a) Would sphere A be moved to a higher or lower electric potential? Is the work done (b) by our force and (c) by the electric field due to B positive, negative, or zero? (d) Rank the paths according to the work our force does, greatest first.

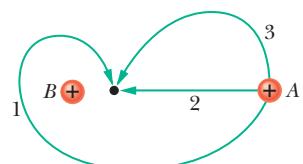


Figure 24-28 Question 5.

- 6** Figure 24-29 shows four arrangements of charged particles, all the same distance from the origin. Rank the situations according to the net electric potential at the origin, most positive first. Take the potential to be zero at infinity.

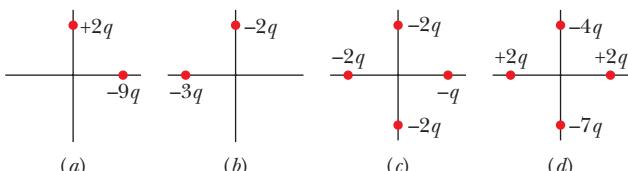


Figure 24-29 Question 6.

- 7** Figure 24-30 shows a system of three charged particles. If you move the particle of charge $+q$ from point A to point D , are the following quantities positive, negative, or zero: (a) the change in the electric potential energy of the three-particle system, (b) the work done by the net electric force on the particle you moved (that is, the net force due to the other two particles), and (c) the work done by your force? (d) What are the answers to (a) through (c) if, instead, the particle is moved from B to C ?

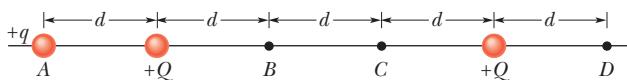
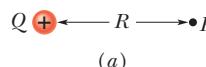


Figure 24-30 Questions 7 and 8.

- 8** In the situation of Question 7, is the work done by your force positive, negative, or zero if the particle is moved (a) from A to B , (b) from A to C , and (c) from B to D ? (d) Rank those moves according to the magnitude of the work done by your force, greatest first.

- 9** Figure 24-26 shows four pairs of charged particles with identical separations. (a) Rank the pairs according to their electric potential energy (that is, the energy of the two-particle system), greatest (most positive) first. (b) For each pair, if the separation between the particles is increased, does the potential energy of the pair increase or decrease?



- 10** (a) In Fig. 24-31a, what is the potential at point P due to charge Q at distance R from P ? Set $V = 0$ at infinity. (b) In Fig. 24-31b, the same charge Q has been spread uniformly over a circular arc of radius R and central angle 40° . What is the potential at point P , the center of curvature of the arc? (c) In Fig. 24-31c, the same charge Q has been spread uniformly over a circle of radius R . What is the potential at point P , the center of the circle? (d) Rank the three situations according to the magnitude of the electric field that is set up at P , greatest first.

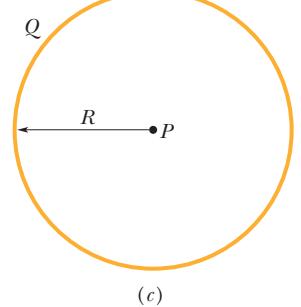
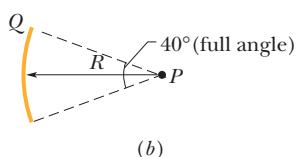


Figure 24-31 Question 10.

- 11** Figure 24-32 shows a thin, uniformly charged rod and three points at the same distance d from the rod. Rank the magnitude of the electric potential the rod produces at those three points, greatest first.

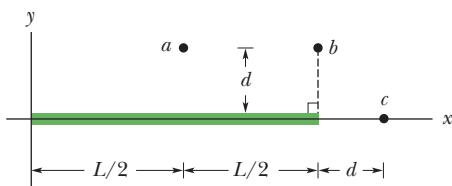


Figure 24-32 Question 11.

- 12** In Fig. 24-33, a particle is to be released at rest at point A and then is to be accelerated directly through point B by an electric field. The potential difference between points A and B is 100 V. Which point should be at higher electric potential if the particle is (a) an electron, (b) a proton, and (c) an alpha particle (a nucleus of two protons and two neutrons)? (d) Rank the kinetic energies of the particles at point B , greatest first.

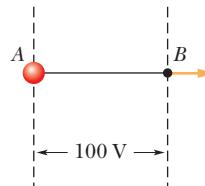


Figure 24-33
Question 12.

Problems



Tutoring problem available (at instructor's discretion) in WileyPLUS and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Worked-out solution is at <http://www.wiley.com/college/halliday>



Interactive solution is at <http://www.wiley.com/college/halliday>



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com

Module 24-1 Electric Potential

- 1 SSM** A particular 12 V car battery can send a total charge of $84 \text{ A} \cdot \text{h}$ (ampere-hours) through a circuit, from one terminal to the other. (a) How many coulombs of charge does this represent? (*Hint:* See Eq. 21-3.) (b) If this entire charge undergoes a change in electric potential of 12 V, how much energy is involved?

- 2** The electric potential difference between the ground and a cloud in a particular thunderstorm is $1.2 \times 10^9 \text{ V}$. In the unit electron-volts, what is the magnitude of the change in the electric potential energy of an electron that moves between the ground and the cloud?

- 3** Suppose that in a lightning flash the potential difference between a cloud and the ground is $1.0 \times 10^9 \text{ V}$ and the quantity of charge transferred is 30 C . (a) What is the change in energy of that transferred charge? (b) If all the energy released could be used to accelerate a 1000 kg car from rest, what would be its final speed?

Module 24-2 Equipotential Surfaces and the Electric Field

- 4** Two large, parallel, conducting plates are 12 cm apart and have charges of equal magnitude and opposite sign on their facing surfaces. An electric force of $3.9 \times 10^{-15} \text{ N}$ acts on an electron placed anywhere between the two plates. (Neglect fringing.) (a) Find the electric field at the position of the electron. (b) What is the potential difference between the plates?

- 5 SSM** An infinite nonconducting sheet has a surface charge density $\sigma = 0.10 \mu\text{C}/\text{m}^2$ on one side. How far apart are equipotential surfaces whose potentials differ by 50 V ?

- 6** When an electron moves from A to B along an electric field line in Fig. 24-34, the electric field does $3.94 \times 10^{-19} \text{ J}$ of work on it. What are the electric potential differences (a) $V_B - V_A$, (b) $V_C - V_A$, and (c) $V_C - V_B$?

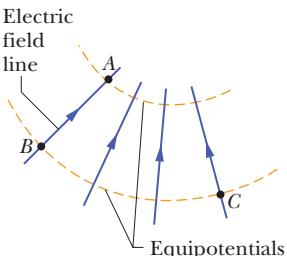


Figure 24-34 Problem 6.

- 7** The electric field in a region of space has the components $E_y = E_z = 0$ and $E_x = (4.00 \text{ N/C})x$. Point A is on the y axis at $y = 3.00 \text{ m}$, and point B is on the x axis at $x = 4.00 \text{ m}$. What is the potential difference $V_B - V_A$?

- 8** A graph of the x component of the electric field as a function of x in a region of space is shown in Fig. 24-35. The scale of the vertical axis is set by $E_{xs} = 20.0 \text{ N/C}$. The y and z components of the electric field are zero in this region. If the electric potential at the origin is 10 V , (a) what is the electric potential at $x = 2.0 \text{ m}$, (b) what is the greatest positive value of the electric potential for points on the x axis for which $0 \leq x \leq 6.0 \text{ m}$, and (c) for what value of x is the electric potential zero?

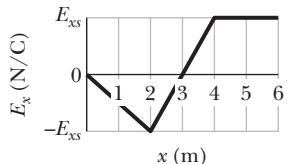


Figure 24-35 Problem 8.

- 9** An infinite nonconducting sheet has a surface charge density $\sigma = +5.80 \text{ pC/m}^2$. (a) How much work is done by the electric field due to the sheet if a particle of charge $q = +1.60 \times 10^{-19} \text{ C}$ is moved from the sheet to a point P at distance $d = 3.56 \text{ cm}$ from the sheet? (b) If the electric potential V is defined to be zero on the sheet, what is V at P ?

- 10 GO** Two uniformly charged, infinite, nonconducting planes are parallel to a yz plane and positioned at $x = -50 \text{ cm}$ and $x = +50 \text{ cm}$. The charge densities on the planes are -50 nC/m^2 and $+25 \text{ nC/m}^2$, respectively. What is the magnitude of the potential difference between the origin and the point on the x axis at $x = +80 \text{ cm}$? (*Hint:* Use Gauss' law.)

- 11** A nonconducting sphere has radius $R = 2.31 \text{ cm}$ and uniformly distributed charge $q = +3.50 \text{ fC}$. Take the electric potential at the sphere's center to be $V_0 = 0$. What is V at radial distance (a) $r = 1.45 \text{ cm}$ and (b) $r = R$. (*Hint:* See Module 23-6.)

Module 24-3 Potential Due to a Charged Particle

- 12** As a space shuttle moves through the dilute ionized gas of Earth's ionosphere, the shuttle's potential is typically changed by -1.0 V during one revolution. Assuming the shuttle is a sphere of radius 10 m , estimate the amount of charge it collects.

- 13** What are (a) the charge and (b) the charge density on the surface of a conducting sphere of radius 0.15 m whose potential is 200 V (with $V = 0$ at infinity)?

- 14** Consider a particle with charge $q = 1.0 \mu\text{C}$, point A at distance $d_1 = 2.0 \text{ m}$ from q , and point B at distance $d_2 = 1.0 \text{ m}$. (a) If A and B are diametrically opposite each other, as in Fig. 24-36a, what is the electric potential difference $V_A - V_B$? (b) What is that electric potential difference if A and B are located as in Fig. 24-36b?

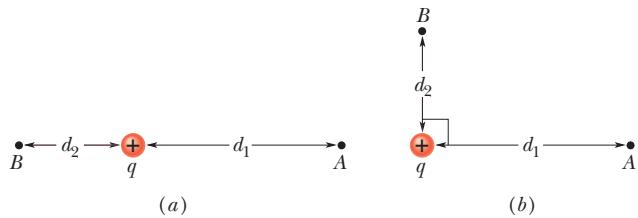


Figure 24-36 Problem 14.

- 15 SSM ILW** A spherical drop of water carrying a charge of 30 pC has a potential of 500 V at its surface (with $V = 0$ at infinity). (a) What is the radius of the drop? (b) If two such drops of the same charge and radius combine to form a single spherical drop, what is the potential at the surface of the new drop?

- 16 GO** Figure 24-37 shows a rectangular array of charged particles fixed in place, with distance $a = 39.0 \text{ cm}$ and the charges shown as integer multiples of $q_1 = 3.40 \text{ pC}$ and $q_2 = 6.00 \text{ pC}$. With $V = 0$ at infinity, what

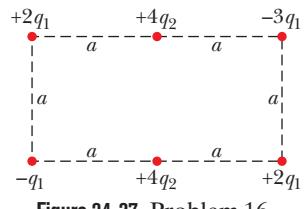


Figure 24-37 Problem 16.

is the net electric potential at the rectangle's center? (Hint: Thoughtful examination of the arrangement can reduce the calculation.)

- 17 **GO** In Fig. 24-38, what is the net electric potential at point P due to the four particles if $V = 0$ at infinity, $q = 5.00 \text{ fC}$, and $d = 4.00 \text{ cm}$?

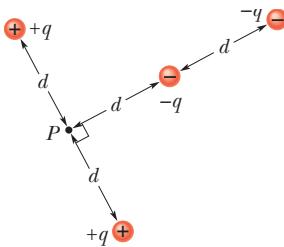


Figure 24-38 Problem 17.

- 18 **GO** Two charged particles are shown in Fig. 24-39a. Particle 1, with charge q_1 , is fixed in place at distance d . Particle 2, with charge q_2 , can be moved along the x axis. Figure 24-39b gives the net electric potential V at the origin due to the two particles as a function of the x coordinate of particle 2. The scale of the x axis is set by $x_s = 16.0 \text{ cm}$. The plot has an asymptote of $V = 5.76 \times 10^{-7} \text{ V}$ as $x \rightarrow \infty$. What is q_2 in terms of e ?

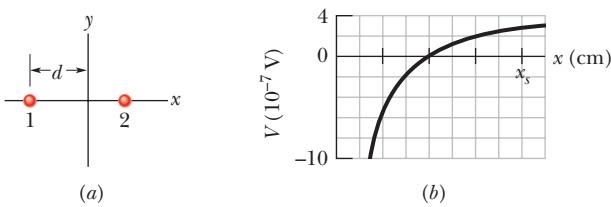


Figure 24-39 Problem 18.

- 19 In Fig. 24-40, particles with the charges $q_1 = +5e$ and $q_2 = -15e$ are fixed in place with a separation of $d = 24.0 \text{ cm}$. With electric potential defined to be $V = 0$ at infinity, what are the finite (a) positive and (b) negative values of x at which the net electric potential on the x axis is zero?

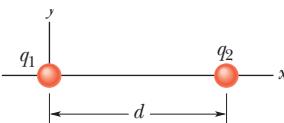


Figure 24-40 Problems 19 and 20.

- 20 Two particles, of charges q_1 and q_2 , are separated by distance d in Fig. 24-40. The net electric field due to the particles is zero at $x = d/4$. With $V = 0$ at infinity, locate (in terms of d) any point on the x axis (other than at infinity) at which the electric potential due to the two particles is zero.

Module 24-4 Potential Due to an Electric Dipole

- 21 **LW** The ammonia molecule NH_3 has a permanent electric dipole moment equal to 1.47 D , where $1 \text{ D} = 1 \text{ debye unit} = 3.34 \times 10^{-30} \text{ C} \cdot \text{m}$. Calculate the electric potential due to an ammonia molecule at a point 52.0 nm away along the axis of the dipole. (Set $V = 0$ at infinity.)

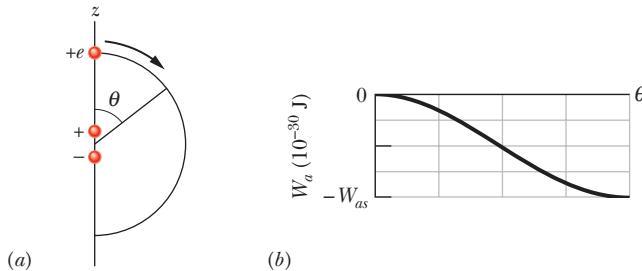


Figure 24-41 Problem 22.

- 22 In Fig. 24-41a, a particle of elementary charge $+e$ is initially at coordinate $z = 20 \text{ nm}$ on the dipole axis (here a z axis) through

an electric dipole, on the positive side of the dipole. (The origin of z is at the center of the dipole.) The particle is then moved along a circular path around the dipole center until it is at coordinate $z = -20 \text{ nm}$, on the negative side of the dipole axis. Figure 24-41b gives the work W_a done by the force moving the particle versus the angle θ that locates the particle relative to the positive direction of the z axis. The scale of the vertical axis is set by $W_{as} = 4.0 \times 10^{-30} \text{ J}$. What is the magnitude of the dipole moment?

Module 24-5 Potential Due to a Continuous Charge Distribution

- 23 (a) Figure 24-42a shows a nonconducting rod of length $L = 6.00 \text{ cm}$ and uniform linear charge density $\lambda = +3.68 \text{ pC/m}$. Assume that the electric potential is defined to be $V = 0$ at infinity. What is V at point P at distance $d = 8.00 \text{ cm}$ along the rod's perpendicular bisector? (b) Figure 24-42b shows an identical rod except that one half is now negatively charged. Both halves have a linear charge density of magnitude 3.68 pC/m . With $V = 0$ at infinity, what is V at P ?

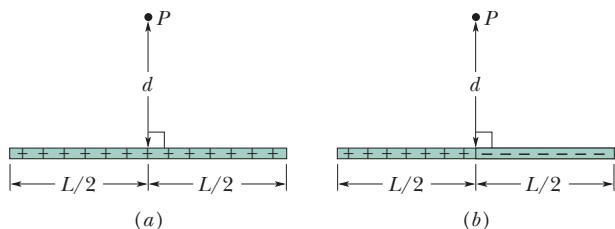


Figure 24-42 Problem 23.

- 24 In Fig. 24-43, a plastic rod having a uniformly distributed charge $Q = -25.6 \text{ pC}$ has been bent into a circular arc of radius $R = 3.71 \text{ cm}$ and central angle $\phi = 120^\circ$. With $V = 0$ at infinity, what is the electric potential at P , the center of curvature of the rod?

- 25 A plastic rod has been bent into a circle of radius $R = 8.20 \text{ cm}$. It has a charge $Q_1 = +4.20 \text{ pC}$ uniformly distributed along one-quarter of its circumference and a charge $Q_2 = -6Q_1$ uniformly distributed along the rest of the circumference (Fig. 24-44). With $V = 0$ at infinity, what is the electric potential at (a) the center C of the circle and (b) point P , on the central axis of the circle at distance $D = 6.71 \text{ cm}$ from the center?

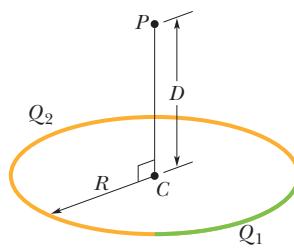


Figure 24-44 Problem 25.

- 26 **GO** Figure 24-45 shows a thin rod with a uniform charge density of $2.00 \mu\text{C/m}$. Evaluate the electric potential at point P if $d = D = L/4.00$. Assume that the potential is zero at infinity.

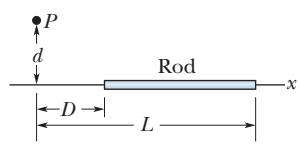


Figure 24-45 Problem 26.

••27 In Fig. 24-46, three thin plastic rods form quarter-circles with a common center of curvature at the origin. The uniform charges on the three rods are $Q_1 = +30 \text{ nC}$, $Q_2 = +3.0Q_1$, and $Q_3 = -8.0Q_1$. What is the net electric potential at the origin due to the rods?

••28 GO Figure 24-47 shows a thin plastic rod of length $L = 12.0 \text{ cm}$ and uniform positive charge $Q = 56.1 \text{ fC}$ lying on an x axis. With $V = 0$ at infinity, find the electric potential at point P_1 on the axis, at distance $d = 2.50 \text{ cm}$ from the rod.

••29 In Fig. 24-48, what is the net electric potential at the origin due to the circular arc of charge $Q_1 = +7.21 \text{ pC}$ and the two particles of charges $Q_2 = 4.00Q_1$ and $Q_3 = -2.00Q_1$? The arc's center of curvature is at the origin and its radius is $R = 2.00 \text{ m}$; the angle indicated is $\theta = 20.0^\circ$.

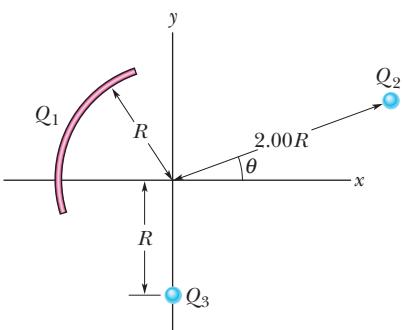


Figure 24-48 Problem 29.

••30 GO The smiling face of Fig. 24-49 consists of three items:

1. a thin rod of charge $-3.0 \mu\text{C}$ that forms a full circle of radius 6.0 cm ;
2. a second thin rod of charge $2.0 \mu\text{C}$ that forms a circular arc of radius 4.0 cm , subtending an angle of 90° about the center of the full circle;
3. an electric dipole with a dipole moment that is perpendicular to a radial line and has a magnitude of $1.28 \times 10^{-21} \text{ C} \cdot \text{m}$.

What is the net electric potential at the center?

••31 SSM WWW A plastic disk of radius $R = 64.0 \text{ cm}$ is charged on one side with a uniform surface charge density $\sigma = 7.73 \text{ fC/m}^2$, and then three quadrants of the disk are removed. The remaining quadrant is shown in Fig. 24-50. With $V = 0$ at infinity, what is the potential due to the remaining quadrant at point P , which is on the central axis of the original disk at distance $D = 25.9 \text{ cm}$ from the original center?

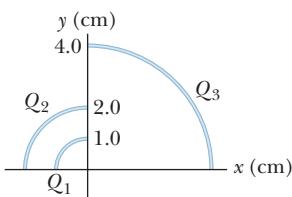


Figure 24-46 Problem 27.

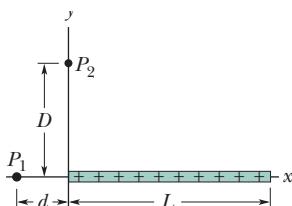


Figure 24-47 Problems 28, 33, 38, and 40.

••32 GO A nonuniform linear charge distribution given by $\lambda = bx$, where b is a constant, is located along an x axis from $x = 0$ to $x = 0.20 \text{ m}$. If $b = 20 \text{ nC/m}^2$ and $V = 0$ at infinity, what is the electric potential at (a) the origin and (b) the point $y = 0.15 \text{ m}$ on the y axis?

••33 GO The thin plastic rod shown in Fig. 24-47 has length $L = 12.0 \text{ cm}$ and a nonuniform linear charge density $\lambda = cx$, where $c = 28.9 \text{ pC/m}^2$. With $V = 0$ at infinity, find the electric potential at point P_1 on the axis, at distance $d = 3.00 \text{ cm}$ from one end.

Module 24-6 Calculating the Field from the Potential

•34 Two large parallel metal plates are 1.5 cm apart and have charges of equal magnitudes but opposite signs on their facing surfaces. Take the potential of the negative plate to be zero. If the potential halfway between the plates is then $+5.0 \text{ V}$, what is the electric field in the region between the plates?

•35 The electric potential at points in an xy plane is given by $V = (2.0 \text{ V/m}^2)x^2 - (3.0 \text{ V/m}^2)y^2$. In unit-vector notation, what is the electric field at the point $(3.0 \text{ m}, 2.0 \text{ m})$?

•36 The electric potential V in the space between two flat parallel plates 1 and 2 is given (in volts) by $V = 1500x^2$, where x (in meters) is the perpendicular distance from plate 1. At $x = 1.3 \text{ cm}$, (a) what is the magnitude of the electric field and (b) is the field directed toward or away from plate 1?

•37 SSM What is the magnitude of the electric field at the point $(3.00\hat{i} - 2.00\hat{j} + 4.00\hat{k}) \text{ m}$ if the electric potential in the region is given by $V = 2.00xyz^2$, where V is in volts and coordinates x , y , and z are in meters?

•38 Figure 24-47 shows a thin plastic rod of length $L = 13.5 \text{ cm}$ and uniform charge 43.6 fC . (a) In terms of distance d , find an expression for the electric potential at point P_1 . (b) Next, substitute variable x for d and find an expression for the magnitude of the component E_x of the electric field at P_1 . (c) What is the direction of E_x relative to the positive direction of the x axis? (d) What is the value of E_x at P_1 for $x = d = 6.20 \text{ cm}$? (e) From the symmetry in Fig. 24-47, determine E_y at P_1 .

•39 An electron is placed in an xy plane where the electric potential depends on x and y as shown, for the coordinate axes, in Fig. 24-51 (the potential does not depend on z). The scale of the vertical axis is set by $V_s = 500 \text{ V}$. In unit-vector notation, what is the electric force on the electron?

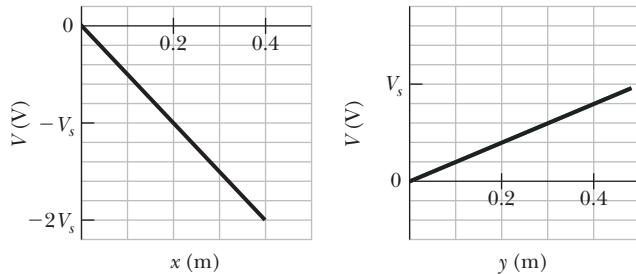


Figure 24-51 Problem 39.

••40 GO The thin plastic rod of length $L = 10.0 \text{ cm}$ in Fig. 24-47 has a nonuniform linear charge density $\lambda = cx$, where $c = 49.9 \text{ pC/m}^2$. (a) With $V = 0$ at infinity, find the electric potential at point P_2 on the y axis at $y = D = 3.56 \text{ cm}$. (b) Find the electric field component E_y at P_2 . (c) Why cannot the field component E_x at P_2 be found using the result of (a)?

Module 24-7 Electric Potential Energy of a System of Charged Particles

•41 A particle of charge $+7.5 \mu\text{C}$ is released from rest at the point $x = 60 \text{ cm}$ on an x axis. The particle begins to move due to the presence of a charge Q that remains fixed at the origin. What is the kinetic energy of the particle at the instant it has moved 40 cm if (a) $Q = +20 \mu\text{C}$ and (b) $Q = -20 \mu\text{C}$?

•42 (a) What is the electric potential energy of two electrons separated by 2.00 nm? (b) If the separation increases, does the potential energy increase or decrease?

•43 SSM ILW WWW How much work is required to set up the arrangement of Fig. 24-52 if $q = 2.30 \text{ pC}$, $a = 64.0 \text{ cm}$, and the particles are initially infinitely far apart and at rest?

•44 In Fig. 24-53, seven charged particles are fixed in place to form a square with an edge length of 4.0 cm. How much work must we do to bring a particle of charge $+6e$ initially at rest from an infinite distance to the center of the square?

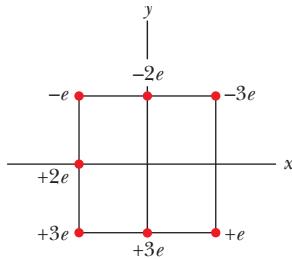


Figure 24-53 Problem 44.

•45 ILW A particle of charge q is fixed at point P , and a second particle of mass m and the same charge q is initially held a distance r_1 from P . The second particle is then released. Determine its speed when it is a distance r_2 from P . Let $q = 3.1 \mu\text{C}$, $m = 20 \text{ mg}$, $r_1 = 0.90 \text{ mm}$, and $r_2 = 2.5 \text{ mm}$.

•46 A charge of -9.0 nC is uniformly distributed around a thin plastic ring lying in a yz plane with the ring center at the origin. A -6.0 pC particle is located on the x axis at $x = 3.0 \text{ m}$. For a ring radius of 1.5 m, how much work must an external force do on the particle to move it to the origin?

•47 GO What is the *escape speed* for an electron initially at rest on the surface of a sphere with a radius of 1.0 cm and a uniformly distributed charge of $1.6 \times 10^{-15} \text{ C}$? That is, what initial speed must the electron have in order to reach an infinite distance from the sphere and have zero kinetic energy when it gets there?

•48 A thin, spherical, conducting shell of radius R is mounted on an isolating support and charged to a potential of -125 V . An electron is then fired directly toward the center of the shell, from point P at distance r from the center of the shell ($r \gg R$). What initial speed v_0 is needed for the electron to just reach the shell before reversing direction?

•49 GO Two electrons are fixed 2.0 cm apart. Another electron is shot from infinity and stops midway between the two. What is its initial speed?

•50 In Fig. 24-54, how much work must we do to bring a particle, of charge $Q = +16e$ and initially at rest, along the dashed line from

infinity to the indicated point near two fixed particles of charges $q_1 = +4e$ and $q_2 = -q_1/2$? Distance $d = 1.40 \text{ cm}$, $\theta_1 = 43^\circ$, and $\theta_2 = 60^\circ$.

•51 GO In the rectangle of Fig. 24-55, the sides have lengths 5.0 cm and 15 cm, $q_1 = -5.0 \mu\text{C}$, and $q_2 = +2.0 \mu\text{C}$. With $V = 0$ at infinity, what is the electric potential at (a) corner A and (b) corner B ? (c) How much work is required to move a charge $q_3 = +3.0 \mu\text{C}$ from B to A along a diagonal of the rectangle? (d) Does this work increase or decrease the electric potential energy of the three-charge system? Is more, less, or the same work required if q_3 is moved along a path that is (e) inside the rectangle but not on a diagonal and (f) outside the rectangle?

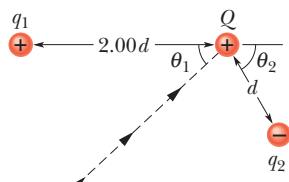


Figure 24-54 Problem 50.



Figure 24-55 Problem 51.

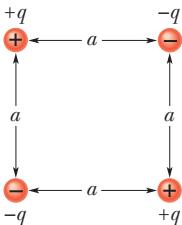


Figure 24-52
Problem 43.

•52 Figure 24-56a shows an electron moving along an electric dipole axis toward the negative side of the dipole. The dipole is fixed in place. The electron was initially very far from the dipole, with kinetic energy 100 eV. Figure 24-56b gives the kinetic energy K of the electron versus its distance r from the dipole center. The scale of the horizontal axis is set by $r_s = 0.10 \text{ m}$. What is the magnitude of the dipole moment?

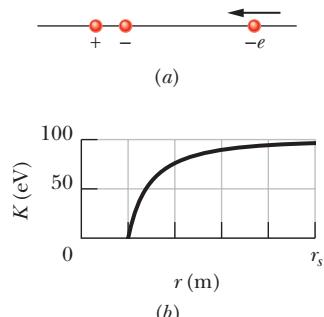


Figure 24-56 Problem 52.

•53 Two tiny metal spheres A and B , mass $m_A = 5.00 \text{ g}$ and $m_B = 10.0 \text{ g}$, have equal positive charge $q = 5.00 \mu\text{C}$. The spheres are connected by a massless nonconducting string of length $d = 1.00 \text{ m}$, which is much greater than the radii of the spheres. (a) What is the electric potential energy of the system? (b) Suppose you cut the string. At that instant, what is the acceleration of each sphere? (c) A long time after you cut the string, what is the speed of each sphere?

•54 GO A positron (charge $+e$, mass equal to the electron mass) is moving at $1.0 \times 10^7 \text{ m/s}$ in the positive direction of an x axis when, at $x = 0$, it encounters an electric field directed along the x axis. The electric potential V associated with the field is given in Fig. 24-57. The scale of the vertical axis is set by $V_s = 500.0 \text{ V}$. (a) Does the positron emerge from the field at $x = 0$ (which means its motion is reversed) or at $x = 0.50 \text{ m}$ (which means its motion is not reversed)? (b) What is its speed when it emerges?

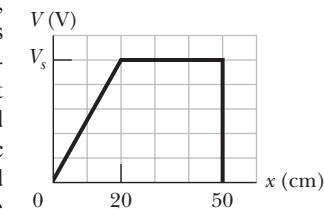


Figure 24-57 Problem 54.

•55 An electron is projected with an initial speed of $3.2 \times 10^5 \text{ m/s}$ directly toward a proton that is fixed in place. If the electron is initially a great distance from the proton, at what distance from the proton is the speed of the electron instantaneously equal to twice the initial value?

•56 Particle 1 (with a charge of $+5.0 \mu\text{C}$) and particle 2 (with a charge of $+3.0 \mu\text{C}$) are fixed in place with separation $d = 4.0 \text{ cm}$

on the x axis shown in Fig. 24-58a. Particle 3 can be moved along the x axis to the right of particle 2. Figure 24-58b gives the electric potential energy U of the three-particle system as a function of the x coordinate of particle 3. The scale of the vertical axis is set by $U_s = 5.0 \text{ J}$. What is the charge of particle 3?

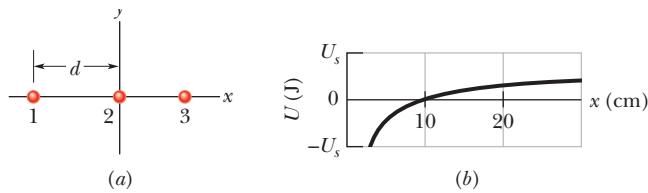


Figure 24-58 Problem 56.

- 57 SSM** Identical $50 \mu\text{C}$ charges are fixed on an x axis at $x = \pm 3.0 \text{ m}$. A particle of charge $q = -15 \mu\text{C}$ is then released from rest at a point on the positive part of the y axis. Due to the symmetry of the situation, the particle moves along the y axis and has kinetic energy 1.2 J as it passes through the point $x = 0, y = 4.0 \text{ m}$. (a) What is the kinetic energy of the particle as it passes through the origin? (b) At what negative value of y will the particle momentarily stop?

- 58 GO Proton in a well.** Figure 24-59 shows electric potential V along an x axis. The scale of the vertical axis is set by $V_s = 10.0 \text{ V}$. A proton is to be released at $x = 3.5 \text{ cm}$ with

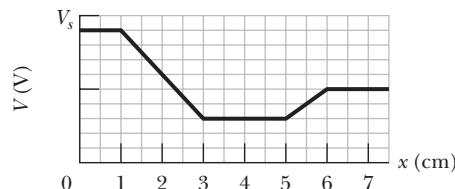


Figure 24-59 Problem 58.

- initial kinetic energy 4.00 eV . (a) If it is initially moving in the negative direction of the axis, does it reach a turning point (if so, what is the x coordinate of that point) or does it escape from the plotted region (if so, what is its speed at $x = 0$)? (b) If it is initially moving in the positive direction of the axis, does it reach a turning point (if so, what is the x coordinate of that point) or does it escape from the plotted region (if so, what is its speed at $x = 6.0 \text{ cm}$)? What are the (c) magnitude F and (d) direction (positive or negative direction of the x axis) of the electric force on the proton if the proton moves just to the left of $x = 3.0 \text{ cm}$? What are (e) F and (f) the direction if the proton moves just to the right of $x = 5.0 \text{ cm}$?

- 59** In Fig. 24-60, a charged particle (either an electron or a proton) is moving rightward between two parallel charged plates separated by distance $d = 2.00 \text{ mm}$. The plate potentials are $V_1 = -70.0 \text{ V}$ and $V_2 = -50.0 \text{ V}$. The particle is slowing from an initial speed of 90.0 km/s at the left plate. (a) Is the particle an electron or a proton? (b) What is its speed just as it reaches plate 2?

- 60** In Fig. 24-61a, we move an electron from an infinite distance to a point at distance $R = 8.00 \text{ cm}$ from a tiny charged ball. The move requires work $W = 2.16 \times 10^{-13} \text{ J}$ by us. (a) What is the charge Q on the ball? In Fig. 24-61b, the ball has been sliced up and the slices spread out so that an equal amount of charge is at the hour positions on a circular clock face of radius $R = 8.00 \text{ cm}$. Now the electron is brought from an infinite distance to the center of

the circle. (b) With that addition of the electron to the system of 12 charged particles, what is the change in the electric potential energy of the system?

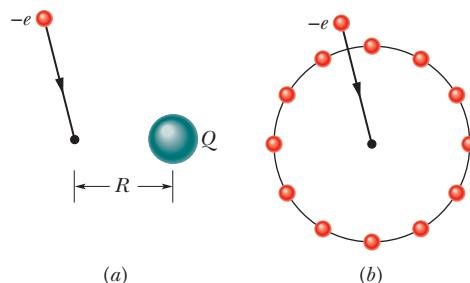


Figure 24-61 Problem 60.

- 61** Suppose N electrons can be placed in either of two configurations. In configuration 1, they are all placed on the circumference of a narrow ring of radius R and are uniformly distributed so that the distance between adjacent electrons is the same everywhere. In configuration 2, $N - 1$ electrons are uniformly distributed on the ring and one electron is placed in the center of the ring. (a) What is the smallest value of N for which the second configuration is less energetic than the first? (b) For that value of N , consider any one circumference electron—call it e_0 . How many other circumference electrons are closer to e_0 than the central electron is?

Module 24-8 Potential of a Charged Isolated Conductor

- 62** Sphere 1 with radius R_1 has positive charge q . Sphere 2 with radius $2.00R_1$ is far from sphere 1 and initially uncharged. After the separated spheres are connected with a wire thin enough to retain only negligible charge, (a) is potential V_1 of sphere 1 greater than, less than, or equal to potential V_2 of sphere 2? What fraction of q ends up on (b) sphere 1 and (c) sphere 2? (d) What is the ratio σ_1/σ_2 of the surface charge densities of the spheres?

- 63 SSM WWW** Two metal spheres, each of radius 3.0 cm , have a center-to-center separation of 2.0 m . Sphere 1 has charge $+1.0 \times 10^{-8} \text{ C}$; sphere 2 has charge $-3.0 \times 10^{-8} \text{ C}$. Assume that the separation is large enough for us to say that the charge on each sphere is uniformly distributed (the spheres do not affect each other). With $V = 0$ at infinity, calculate (a) the potential at the point halfway between the centers and the potential on the surface of (b) sphere 1 and (c) sphere 2.

- 64** A hollow metal sphere has a potential of $+400 \text{ V}$ with respect to ground (defined to be at $V = 0$) and a charge of $5.0 \times 10^{-9} \text{ C}$. Find the electric potential at the center of the sphere.

- 65 SSM** What is the excess charge on a conducting sphere of radius $r = 0.15 \text{ m}$ if the potential of the sphere is 1500 V and $V = 0$ at infinity?

- 66** Two isolated, concentric, conducting spherical shells have radii $R_1 = 0.500 \text{ m}$ and $R_2 = 1.00 \text{ m}$, uniform charges $q_1 = +2.00 \mu\text{C}$ and $q_2 = +1.00 \mu\text{C}$, and negligible thicknesses. What is the magnitude of the electric field E at radial distance (a) $r = 4.00 \text{ m}$, (b) $r = 0.700 \text{ m}$, and (c) $r = 0.200 \text{ m}$? With $V = 0$ at infinity, what is V at (d) $r = 4.00 \text{ m}$, (e) $r = 1.00 \text{ m}$, (f) $r = 0.700 \text{ m}$, (g) $r = 0.500 \text{ m}$, (h) $r = 0.200 \text{ m}$, and (i) $r = 0$? (j) Sketch $E(r)$ and $V(r)$.

- 67** A metal sphere of radius 15 cm has a net charge of $3.0 \times 10^{-8} \text{ C}$. (a) What is the electric field at the sphere's surface? (b) If $V = 0$ at infinity, what is the electric potential at the sphere's surface? (c) At what distance from the sphere's surface has the electric potential decreased by 500 V ?

Additional Problems

68 Here are the charges and coordinates of two charged particles located in an xy plane: $q_1 = +3.00 \times 10^{-6} \text{ C}$, $x = +3.50 \text{ cm}$, $y = +0.500 \text{ cm}$ and $q_2 = -4.00 \times 10^{-6} \text{ C}$, $x = -2.00 \text{ cm}$, $y = +1.50 \text{ cm}$. How much work must be done to locate these charges at their given positions, starting from infinite separation?

69 SSM A long, solid, conducting cylinder has a radius of 2.0 cm. The electric field at the surface of the cylinder is 160 N/C, directed radially outward. Let A , B , and C be points that are 1.0 cm, 2.0 cm, and 5.0 cm, respectively, from the central axis of the cylinder. What are (a) the magnitude of the electric field at C and the electric potential differences (b) $V_B - V_C$ and (c) $V_A - V_B$?

70 ~~SSM~~ *The chocolate crumb mystery.* This story begins with Problem 60 in Chapter 23. (a) From the answer to part (a) of that problem, find an expression for the electric potential as a function of the radial distance r from the center of the pipe. (The electric potential is zero on the grounded pipe wall.) (b) For the typical volume charge density $\rho = -1.1 \times 10^{-3} \text{ C/m}^3$, what is the difference in the electric potential between the pipe's center and its inside wall? (The story continues with Problem 60 in Chapter 25.)

71 SSM Starting from Eq. 24-30, derive an expression for the electric field due to a dipole at a point on the dipole axis.

72 The magnitude E of an electric field depends on the radial distance r according to $E = A/r^4$, where A is a constant with the unit volt-cubic meter. As a multiple of A , what is the magnitude of the electric potential difference between $r = 2.00 \text{ m}$ and $r = 3.00 \text{ m}$?

73 (a) If an isolated conducting sphere 10 cm in radius has a net charge of $4.0 \mu\text{C}$ and if $V = 0$ at infinity, what is the potential on the surface of the sphere? (b) Can this situation actually occur, given that the air around the sphere undergoes electrical breakdown when the field exceeds 3.0 MV/m ?

74 Three particles, charge $q_1 = +10 \mu\text{C}$, $q_2 = -20 \mu\text{C}$, and $q_3 = +30 \mu\text{C}$, are positioned at the vertices of an isosceles triangle as shown in Fig. 24-62. If $a = 10 \text{ cm}$ and $b = 6.0 \text{ cm}$, how much work must an external agent do to exchange the positions of (a) q_1 and q_3 and, instead, (b) q_1 and q_2 ?

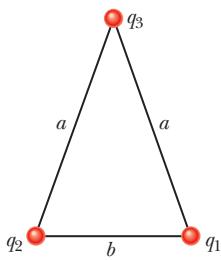


Figure 24-62
Problem 74.

75 An electric field of approximately 100 V/m is often observed near the surface of Earth. If this were the field over the entire surface, what would be the electric potential of a point on the surface? (Set $V = 0$ at infinity.)

76 A Gaussian sphere of radius 4.00 cm is centered on a ball that has a radius of 1.00 cm and a uniform charge distribution. The total (net) electric flux through the surface of the Gaussian sphere is $+5.60 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}$. What is the electric potential 12.0 cm from the center of the ball?

77 In a Millikan oil-drop experiment (Module 22-6), a uniform electric field of $1.92 \times 10^5 \text{ N/C}$ is maintained in the region between two plates separated by 1.50 cm . Find the potential difference between the plates.

78 Figure 24-63 shows three circular, nonconducting arcs of radius $R = 8.50 \text{ cm}$. The charges on the arcs are $q_1 = 4.52$

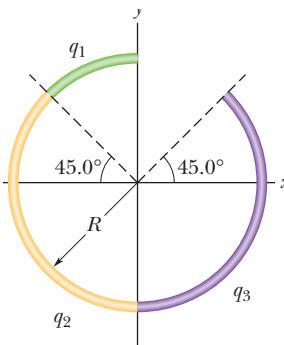


Figure 24-63
Problem 78.

pC , $q_2 = -2.00q_1$, $q_3 = +3.00q_1$. With $V = 0$ at infinity, what is the net electric potential of the arcs at the common center of curvature?

79 An electron is released from rest on the axis of an electric dipole that has charge e and charge separation $d = 20 \text{ pm}$ and that is fixed in place. The release point is on the positive side of the dipole, at distance $7.0d$ from the dipole center. What is the electron's speed when it reaches a point $5.0d$ from the dipole center?

80 Figure 24-64 shows a ring of outer radius $R = 13.0 \text{ cm}$, inner radius $r = 0.200R$, and uniform surface charge density $\sigma = 6.20 \text{ pC/m}^2$. With $V = 0$ at infinity, find the electric potential at point P on the central axis of the ring, at distance $z = 2.00R$ from the center of the ring.

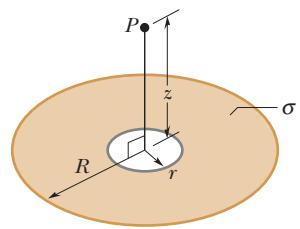


Figure 24-64 Problem 80.

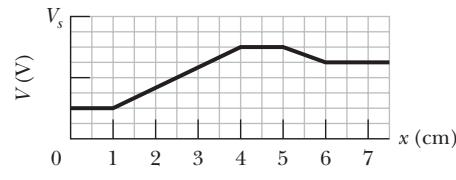


Figure 24-65 Problem 81.

81 GO *Electron in a well.* Figure 24-65 shows electric potential V along an x axis. The scale of the vertical axis is set by $V_s = 8.0 \text{ V}$. An electron is to be released at $x = 4.5 \text{ cm}$ with initial kinetic energy 3.00 eV . (a) If it is initially moving in the negative direction of the axis, does it reach a turning point (if so, what is the x coordinate of that point) or does it escape from the plotted region (if so, what is its speed at $x = 0$)? (b) If it is initially moving in the positive direction of the axis, does it reach a turning point (if so, what is the x coordinate of that point) or does it escape from the plotted region (if so, what is its speed at $x = 7.0 \text{ cm}$)? What are the (c) magnitude F and (d) direction (positive or negative direction of the x axis) of the electric force on the electron if the electron moves just to the left of $x = 4.0 \text{ cm}$? What are (e) F and (f) the direction if it moves just to the right of $x = 5.0 \text{ cm}$?

82 (a) If Earth had a uniform surface charge density of 1.0 electron/m^2 (a very artificial assumption), what would its potential be? (Set $V = 0$ at infinity.) What would be the (b) magnitude and (c) direction (radially inward or outward) of the electric field due to Earth just outside its surface?

83 In Fig. 24-66, point P is at distance $d_1 = 4.00 \text{ m}$ from particle 1 ($q_1 = -2e$) and distance $d_2 = 2.00 \text{ m}$ from particle 2 ($q_2 = +2e$), with both particles fixed in place. (a) With $V = 0$ at infinity, what is V at P ? If we bring a particle of charge $q_3 = +2e$ from infinity to P , (b) how much work do we do and (c) what is the potential energy of the three-particle system?

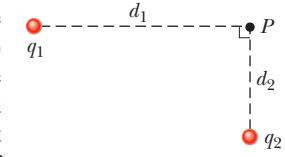


Figure 24-66 Problem 83.

84 A solid conducting sphere of radius 3.0 cm has a charge of 30 nC distributed uniformly over its surface. Let A be a point 1.0 cm from the center of the sphere, S be a point on the surface of the sphere, and B be a point 5.0 cm from the center of the sphere. What are the electric potential differences (a) $V_S - V_B$ and (b) $V_A - V_B$?

85 In Fig. 24-67, we move a particle of charge $+2e$ in from infinity to the x axis. How much work do we do? Distance D is 4.00 m .

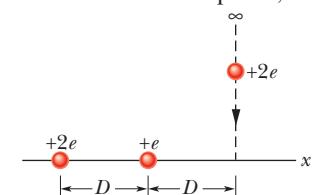


Figure 24-67 Problem 85.

86 Figure 24-68 shows a hemisphere with a charge of $4.00 \mu\text{C}$ distributed uniformly through its volume. The hemisphere lies on an xy plane the way half a grapefruit might lie face down on a kitchen table. Point P is located on the plane, along a radial line from the hemisphere's center of curvature, at radial distance 15 cm. What is the electric potential at point P due to the hemisphere?

87 SSM Three $+0.12 \text{ C}$ charges form an equilateral triangle 1.7 m on a side. Using energy supplied at the rate of 0.83 kW , how many days would be required to move one of the charges to the midpoint of the line joining the other two charges?

88 Two charges $q = +2.0 \mu\text{C}$ are fixed a distance $d = 2.0 \text{ cm}$ apart (Fig. 24-69). (a) With $V = 0$ at infinity, what is the electric potential at point C ? (b) You bring a third charge $q = +2.0 \mu\text{C}$ from infinity to C . How much work must you do? (c) What is the potential energy U of the three-charge configuration when the third charge is in place?

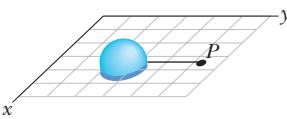


Figure 24-68 Problem 86.

89 Initially two electrons are fixed in place with a separation of $2.00 \mu\text{m}$. How much work must we do to bring a third electron in from infinity to complete an equilateral triangle?

90 A particle of positive charge Q is fixed at point P . A second particle of mass m and negative charge $-q$ moves at constant speed in a circle of radius r_1 , centered at P . Derive an expression for the work W that must be done by an external agent on the second particle to increase the radius of the circle of motion to r_2 .

91 Two charged, parallel, flat conducting surfaces are spaced $d = 1.00 \text{ cm}$ apart and produce a potential difference $\Delta V = 625 \text{ V}$ between them. An electron is projected from one surface directly toward the second. What is the initial speed of the electron if it stops just at the second surface?

92 In Fig. 24-70, point P is at the center of the rectangle. With $V = 0$ at infinity, $q_1 = 5.00 \text{ fC}$, $q_2 = 2.00 \text{ fC}$, $q_3 = 3.00 \text{ fC}$, and $d = 2.54 \text{ cm}$, what is the net electric potential at P due to the six charged particles?

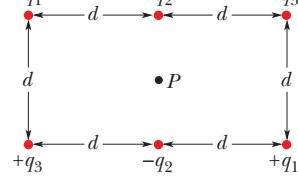


Figure 24-70 Problem 92.

93 SSM A uniform charge of $+16.0 \mu\text{C}$ is on a thin circular ring lying in an xy plane and centered on the origin. The ring's radius is 3.00 cm . If point A is at the origin and point B is on the z axis at $z = 4.00 \text{ cm}$, what is $V_B - V_A$?

94 Consider a particle with charge $q = 1.50 \times 10^{-8} \text{ C}$, and take $V = 0$ at infinity. (a) What are the shape and dimensions of an equipotential surface having a potential of 30.0 V due to q alone? (b) Are surfaces whose potentials differ by a constant amount (1.0 V , say) evenly spaced?

95 SSM A thick spherical shell of charge Q and uniform volume charge density ρ is bounded by radii r_1 and $r_2 > r_1$. With $V = 0$ at infinity, find the electric potential V as a function of distance r from the center of the distribution, considering regions (a) $r > r_2$, (b) $r_2 > r > r_1$, and (c) $r < r_1$. (d) Do these solutions agree with each other at $r = r_2$ and $r = r_1$? (Hint: See Module 23-6.)

96 A charge q is distributed uniformly throughout a spherical volume of radius R . Let $V = 0$ at infinity. What are (a) V at radial distance $r < R$ and (b) the potential difference between points at $r = R$ and the point at $r = 0$?

97 SSM A solid copper sphere whose radius is 1.0 cm has a very thin surface coating of nickel. Some of the nickel atoms are radioactive, each atom emitting an electron as it decays. Half of these electrons enter the copper sphere, each depositing 100 keV of energy there. The other half of the electrons escape, each carrying away a charge $-e$. The nickel coating has an activity of 3.70×10^8 radioactive decays per second. The sphere is hung from a long, non-conducting string and isolated from its surroundings. (a) How long will it take for the potential of the sphere to increase by 1000 V ? (b) How long will it take for the temperature of the sphere to increase by 5.0 K due to the energy deposited by the electrons? The heat capacity of the sphere is 14 J/K .

98 In Fig. 24-71, a metal sphere with charge $q = 5.00 \mu\text{C}$ and radius $r = 3.00 \text{ cm}$ is concentric with a larger metal sphere with charge $Q = 15.0 \mu\text{C}$ and radius $R = 6.00 \text{ cm}$. (a) What is the potential difference between the spheres? If we connect the spheres with a wire, what then is the charge on (b) the smaller sphere and (c) the larger sphere?

99 (a) Using Eq. 24-32, show that the electric potential at a point on the central axis of a thin ring (of charge q and radius R) and at distance z from the ring is

$$V = \frac{1}{4\pi\epsilon_0} \frac{q}{\sqrt{z^2 + R^2}}.$$

(b) From this result, derive an expression for the electric field magnitude E at points on the ring's axis; compare your result with the calculation of E in Module 22-4.

100 An alpha particle (which has two protons) is sent directly toward a target nucleus containing 92 protons. The alpha particle has an initial kinetic energy of 0.48 pJ . What is the least center-to-center distance the alpha particle will be from the target nucleus, assuming the nucleus does not move?

101 In the quark model of fundamental particles, a proton is composed of three quarks: two “up” quarks, each having charge $+2e/3$, and one “down” quark, having charge $-e/3$. Suppose that the three quarks are equidistant from one another. Take that separation distance to be $1.32 \times 10^{-15} \text{ m}$ and calculate the electric potential energy of the system of (a) only the two up quarks and (b) all three quarks.

102 A charge of $1.50 \times 10^{-8} \text{ C}$ lies on an isolated metal sphere of radius 16.0 cm . With $V = 0$ at infinity, what is the electric potential at points on the sphere's surface?

103 In Fig. 24-72, two particles of charges q_1 and q_2 are fixed to an x axis. If a third particle, of charge $+6.0 \mu\text{C}$, is brought from an infinite distance to point P , the three-particle system has the same electric potential energy as the original two-particle system. What is the charge ratio q_1/q_2 ?

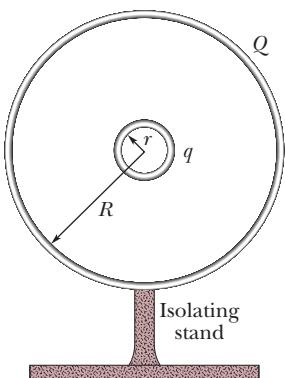


Figure 24-71 Problem 98.

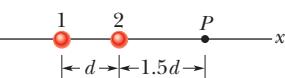


Figure 24-72 Problem 103.

Interference

35-1 LIGHT AS A WAVE

Learning Objectives

After reading this module, you should be able to...

- 35.01 Using a sketch, explain Huygens' principle.
- 35.02 With a few simple sketches, explain refraction in terms of the gradual change in the speed of a wavefront as it passes through an interface at an angle to the normal.
- 35.03 Apply the relationship between the speed of light in vacuum c , the speed of light in a material v , and the index of refraction of the material n .
- 35.04 Apply the relationship between a distance L in a material, the speed of light in that material, and the time required for a pulse of the light to travel through L .
- 35.05 Apply Snell's law of refraction.
- 35.06 When light refracts through an interface, identify that the frequency does not change but the wavelength and effective speed do.
- 35.07 Apply the relationship between the wavelength in vacuum λ , the wavelength λ_n in a material (the internal wavelength), and the index of refraction n of the material.
- 35.08 For light in a certain length of a material, calculate the number of internal wavelengths that fit into the length.
- 35.09 If two light waves travel through different materials with different indexes of refraction and then reach a common point, determine their phase difference and interpret the resulting interference in terms of maximum brightness, intermediate brightness, and darkness.
- 35.10 Apply the learning objectives of Module 17-3 (sound waves there, light waves here) to find the phase difference and interference of two waves that reach a common point after traveling paths of different lengths.
- 35.11 Given the initial phase difference between two waves with the same wavelength, determine their phase difference after they travel through different path lengths and through different indexes of refraction.
- 35.12 Identify that rainbows are examples of optical interference.

Key Ideas

- The three-dimensional transmission of waves, including light, may often be predicted by Huygens' principle, which states that all points on a wavefront serve as point sources of spherical secondary wavelets. After a time t , the new position of the wavefront will be that of a surface tangent to these secondary wavelets.
- The law of refraction can be derived from Huygens' principle by assuming that the index of refraction of any medium is $n = c/v$, in which v is the speed of light in the medium and c is the speed of light in vacuum.

- The wavelength λ_n of light in a medium depends on the index of refraction n of the medium:

$$\lambda_n = \frac{\lambda}{n},$$

in which λ is the wavelength in vacuum.

- Because of this dependency, the phase difference between two waves can change if they pass through different materials with different indexes of refraction.

What Is Physics?

One of the major goals of physics is to understand the nature of light. This goal has been difficult to achieve (and has not yet fully been achieved) because light is complicated. However, this complication means that light offers many opportunities for applications, and some of the richest opportunities involve the interference of light waves—**optical interference**.

Nature has long used optical interference for coloring. For example, the wings of a *Morpho* butterfly are a dull, uninspiring brown, as can be seen on the



Philippe Colombe/PhotoDisc/Getty Images, Inc.

Figure 35-1 The blue of the top surface of a *Morpho* butterfly wing is due to optical interference and shifts in color as your viewing perspective changes.

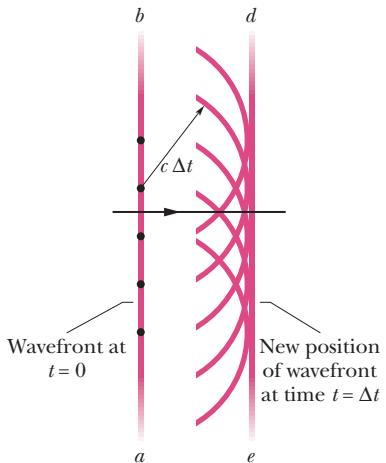


Figure 35-2 The propagation of a plane wave in vacuum, as portrayed by Huygens' principle.

bottom wing surface, but the brown is hidden on the top surface by an arresting blue due to the interference of light reflecting from that surface (Fig. 35-1). Moreover, the top surface is color-shifting; if you change your perspective or if the wing moves, the tint of the color changes. Similar color shifting is used in the inks on many currencies to thwart counterfeiters, whose copy machines can duplicate color from only one perspective and therefore cannot duplicate any shift in color caused by a change in perspective.

To understand the basic physics of optical interference, we must largely abandon the simplicity of geometrical optics (in which we describe light as rays) and return to the wave nature of light.

Light as a Wave

The first convincing wave theory for light was in 1678 by Dutch physicist Christian Huygens. Mathematically simpler than the electromagnetic theory of Maxwell, it nicely explained reflection and refraction in terms of waves and gave physical meaning to the index of refraction.

Huygens' wave theory is based on a geometrical construction that allows us to tell where a given wavefront will be at any time in the future if we know its present position. **Huygens' principle** is:



All points on a wavefront serve as point sources of spherical secondary wavelets. After a time t , the new position of the wavefront will be that of a surface tangent to these secondary wavelets.

Here is a simple example. At the left in Fig. 35-2, the present location of a wavefront of a plane wave traveling to the right in vacuum is represented by plane *ab*, perpendicular to the page. Where will the wavefront be at time Δt later? We let several points on plane *ab* (the dots) serve as sources of spherical secondary wavelets that are emitted at $t = 0$. At time Δt , the radius of all these spherical wavelets will have grown to $c \Delta t$, where c is the speed of light in vacuum. We draw plane *de* tangent to these wavelets at time Δt . This plane represents the wavefront of the plane wave at time Δt ; it is parallel to plane *ab* and a perpendicular distance $c \Delta t$ from it.

The Law of Refraction

We now use Huygens' principle to derive the law of refraction, Eq. 33-40 (Snell's law). Figure 35-3 shows three stages in the refraction of several wavefronts at a flat interface between air (medium 1) and glass (medium 2). We arbitrarily choose the wavefronts in the incident light beam to be separated by λ_1 , the wavelength in medium 1. Let the speed of light in air be v_1 and that in glass be v_2 . We assume that $v_2 < v_1$, which happens to be true.

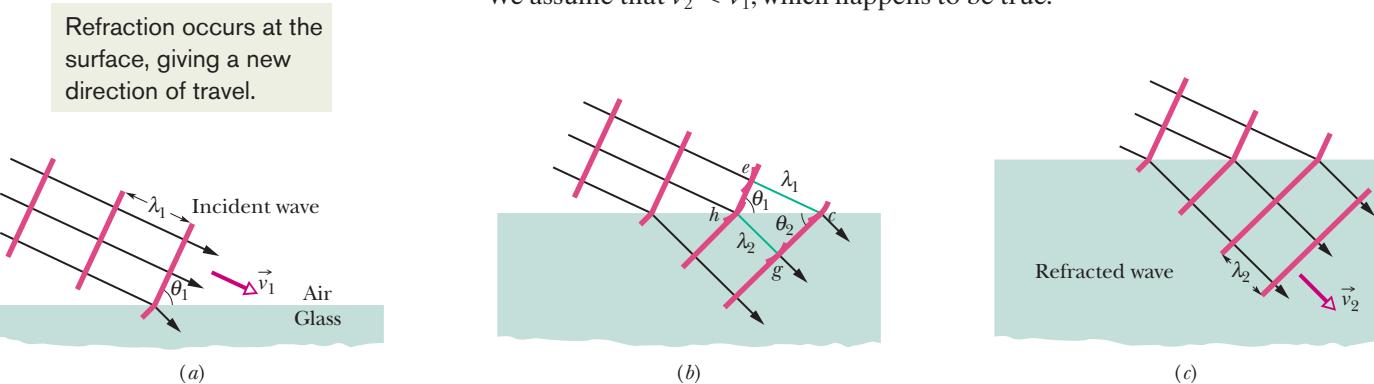


Figure 35-3 The refraction of a plane wave at an air–glass interface, as portrayed by Huygens' principle. The wavelength in glass is smaller than that in air. For simplicity, the reflected wave is not shown. Parts (a) through (c) represent three successive stages of the refraction.

Angle θ_1 in Fig. 35-3a is the angle between the wavefront and the interface; it has the same value as the angle between the *normal* to the wavefront (that is, the incident ray) and the *normal* to the interface. Thus, θ_1 is the angle of incidence.

As the wave moves into the glass, a Huygens wavelet at point *e* in Fig. 35-3b will expand to pass through point *c*, at a distance of λ_1 from point *e*. The time interval required for this expansion is that distance divided by the speed of the wavelet, or λ_1/v_1 . Now note that in this same time interval, a Huygens wavelet at point *h* will expand to pass through point *g*, at the reduced speed v_2 and with wavelength λ_2 . Thus, this time interval must also be equal to λ_2/v_2 . By equating these times of travel, we obtain the relation

$$\frac{\lambda_1}{\lambda_2} = \frac{v_1}{v_2}, \quad (35-1)$$

which shows that the wavelengths of light in two media are proportional to the speeds of light in those media.

By Huygens' principle, the refracted wavefront must be tangent to an arc of radius λ_2 centered on *h*, say at point *g*. The refracted wavefront must also be tangent to an arc of radius λ_1 centered on *e*, say at *c*. Then the refracted wavefront must be oriented as shown. Note that θ_2 , the angle between the refracted wavefront and the interface, is actually the angle of refraction.

For the right triangles *hce* and *hcg* in Fig. 35-3b we may write

$$\sin \theta_1 = \frac{\lambda_1}{hc} \quad (\text{for triangle } hce)$$

and

$$\sin \theta_2 = \frac{\lambda_2}{hc} \quad (\text{for triangle } hcg).$$

Dividing the first of these two equations by the second and using Eq. 35-1, we find

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{\lambda_1}{\lambda_2} = \frac{v_1}{v_2}. \quad (35-2)$$

We can define the **index of refraction** *n* for each medium as the ratio of the speed of light in vacuum to the speed of light *v* in the medium. Thus,

$$n = \frac{c}{v} \quad (\text{index of refraction}). \quad (35-3)$$

In particular, for our two media, we have

$$n_1 = \frac{c}{v_1} \quad \text{and} \quad n_2 = \frac{c}{v_2}.$$

We can now rewrite Eq. 35-2 as

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{c/n_1}{c/n_2} = \frac{n_2}{n_1}$$

or

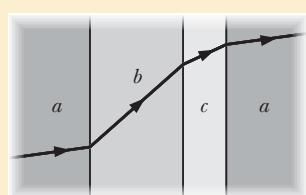
$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (\text{law of refraction}), \quad (35-4)$$

as introduced in Chapter 33.



Checkpoint 1

The figure shows a monochromatic ray of light traveling across parallel interfaces, from an original material *a*, through layers of materials *b* and *c*, and then back into material *a*. Rank the materials according to the speed of light in them, greatest first.



Wavelength and Index of Refraction

We have now seen that the wavelength of light changes when the speed of the light changes, as happens when light crosses an interface from one medium into another. Further, the speed of light in any medium depends on the index of refraction of the medium, according to Eq. 35-3. Thus, the wavelength of light in any medium depends on the index of refraction of the medium. Let a certain monochromatic light have wavelength λ and speed c in vacuum and wavelength λ_n and speed v in a medium with an index of refraction n . Now we can rewrite Eq. 35-1 as

$$\lambda_n = \lambda \frac{v}{c}. \quad (35-5)$$

Using Eq. 35-3 to substitute $1/n$ for v/c then yields

$$\lambda_n = \frac{\lambda}{n}. \quad (35-6)$$

This equation relates the wavelength of light in any medium to its wavelength in vacuum: A greater index of refraction means a smaller wavelength.

Next, let f_n represent the frequency of the light in a medium with index of refraction n . Then from the general relation of Eq. 16-13 ($v = \lambda f$), we can write

$$f_n = \frac{v}{\lambda_n}.$$

Substituting Eqs. 35-3 and 35-6 then gives us

$$f_n = \frac{c/n}{\lambda/n} = \frac{c}{\lambda} = f,$$

where f is the frequency of the light in vacuum. Thus, although the speed and wavelength of light in the medium are different from what they are in vacuum, *the frequency of the light in the medium is the same as it is in vacuum*.

Phase Difference. The fact that the wavelength of light depends on the index of refraction via Eq. 35-6 is important in certain situations involving the interference of light waves. For example, in Fig. 35-4, the *waves of the rays* (that is, the waves represented by the rays) have identical wavelengths λ and are initially in phase in air ($n \approx 1$). One of the waves travels through medium 1 of index of refraction n_1 and length L . The other travels through medium 2 of index of refraction n_2 and the same length L . When the waves leave the two media, they will have the same wavelength—their wavelength λ in air. However, because their wavelengths differed in the two media, the two waves may no longer be in phase.

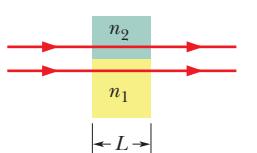


Figure 35-4 Two light rays travel through two media having different indexes of refraction.



The phase difference between two light waves can change if the waves travel through different materials having different indexes of refraction.

As we shall discuss soon, this change in the phase difference can determine how the light waves will interfere if they reach some common point.

To find their new phase difference in terms of wavelengths, we first count the number N_1 of wavelengths there are in the length L of medium 1. From Eq. 35-6, the wavelength in medium 1 is $\lambda_{n1} = \lambda/n_1$; so

$$N_1 = \frac{L}{\lambda_{n1}} = \frac{Ln_1}{\lambda}. \quad (35-7)$$

Similarly, we count the number N_2 of wavelengths there are in the length L of medium 2, where the wavelength is $\lambda_{n2} = \lambda/n_2$:

$$N_2 = \frac{L}{\lambda_{n2}} = \frac{Ln_2}{\lambda}. \quad (35-8)$$

To find the new phase difference between the waves, we subtract the smaller of N_1 and N_2 from the larger. Assuming $n_2 > n_1$, we obtain

$$N_2 - N_1 = \frac{Ln_2}{\lambda} - \frac{Ln_1}{\lambda} = \frac{L}{\lambda} (n_2 - n_1). \quad (35-9)$$

Suppose Eq. 35-9 tells us that the waves now have a phase difference of 45.6 wavelengths. That is equivalent to taking the initially in-phase waves and shifting one of them by 45.6 wavelengths. However, a shift of an integer number of wavelengths (such as 45) would put the waves back in phase; so it is only the decimal fraction (here, 0.6) that is important. A phase difference of 45.6 wavelengths is equivalent to an *effective phase difference* of 0.6 wavelength.

A phase difference of 0.5 wavelength puts two waves exactly out of phase. If the waves had equal amplitudes and were to reach some common point, they would then undergo fully destructive interference, producing darkness at that point. With a phase difference of 0.0 or 1.0 wavelength, they would, instead, undergo fully constructive interference, resulting in brightness at the common point. Our phase difference of 0.6 wavelength is an intermediate situation but closer to fully destructive interference, and the waves would produce a dimly illuminated common point.

We can also express phase difference in terms of radians and degrees, as we have done already. A phase difference of one wavelength is equivalent to phase differences of 2π rad and 360° .

Path Length Difference. As we discussed with sound waves in Module 17-3, two waves that begin with some initial phase difference can end up with a different phase difference if they travel through paths with different lengths before coming back together. The key for the waves (whatever their type might be) is the path length difference ΔL , or more to the point, how ΔL compares to the wavelength λ of the waves. From Eqs. 17-23 and 17-24, we know that, for light waves, fully constructive interference (maximum brightness) occurs when

$$\frac{\Delta L}{\lambda} = 0, 1, 2, \dots \quad (\text{fully constructive interference}), \quad (35-10)$$

and that fully destructive interference (darkness) occurs when

$$\frac{\Delta L}{\lambda} = 0.5, 1.5, 2.5, \dots \quad (\text{fully destructive interference}). \quad (35-11)$$

Intermediate values correspond to intermediate interference and thus also illumination.

Rainbows and Optical Interference

In Module 33-5, we discussed how the colors of sunlight are separated into a rainbow when sunlight travels through falling raindrops. We dealt with a simplified situation in which a single ray of white light entered a drop. Actually, light waves pass into a drop along the entire side that faces the Sun. Here we cannot discuss the details of how these waves travel through the drop and then emerge, but we can see that different parts of an incoming wave will travel different paths within the drop. That means waves will emerge from the drop with different phases. Thus, we can see that at some angles the emerging light will be in phase and give constructive interference. The rainbow is the result of such constructive interference. For example, the red of the rainbow appears because waves of red light emerge in phase from each raindrop in the direction in which you see that part of the rainbow. The light waves that emerge in other directions from each raindrop have a range of different phases because they take a

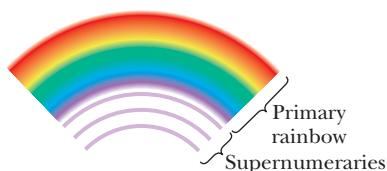


Figure 35-5 A primary rainbow and the faint supernumeraries below it are due to optical interference.

range of different paths through each drop. This light is neither bright nor colorful, and so you do not notice it.

If you are lucky and look carefully below a primary rainbow, you can see dimmer colored arcs called *supernumeraries* (Fig. 35-5). Like the main arcs of the rainbow, the supernumeraries are due to waves that emerge from each drop approximately in phase with one another to give constructive interference. If you are very lucky and look very carefully above a secondary rainbow, you might see even more (but even dimmer) supernumeraries. Keep in mind that both types of rainbows and both sets of supernumeraries are naturally occurring examples of optical interference and naturally occurring evidence that light consists of waves. 



Checkpoint 2

The light waves of the rays in Fig. 35-4 have the same wavelength and amplitude and are initially in phase. (a) If 7.60 wavelengths fit within the length of the top material and 5.50 wavelengths fit within that of the bottom material, which material has the greater index of refraction? (b) If the rays are angled slightly so that they meet at the same point on a distant screen, will the interference there result in the brightest possible illumination, bright intermediate illumination, dark intermediate illumination, or darkness?

Sample Problem 35.01 Phase difference of two waves due to difference in refractive indexes

In Fig. 35-4, the two light waves that are represented by the rays have wavelength 550.0 nm before entering media 1 and 2. They also have equal amplitudes and are in phase. Medium 1 is now just air, and medium 2 is a transparent plastic layer of index of refraction 1.600 and thickness 2.600 μm .

(a) What is the phase difference of the emerging waves in wavelengths, radians, and degrees? What is their effective phase difference (in wavelengths)?

KEY IDEA

The phase difference of two light waves can change if they travel through different media, with different indexes of refraction. The reason is that their wavelengths are different in the different media. We can calculate the change in phase difference by counting the number of wavelengths that fits into each medium and then subtracting those numbers.

Calculations: When the path lengths of the waves in the two media are identical, Eq. 35-9 gives the result of the subtraction. Here we have $n_1 = 1.000$ (for the air), $n_2 = 1.600$, $L = 2.600 \mu\text{m}$, and $\lambda = 550.0 \text{ nm}$. Thus, Eq. 35-9 yields

$$\begin{aligned} N_2 - N_1 &= \frac{L}{\lambda} (n_2 - n_1) \\ &= \frac{2.600 \times 10^{-6} \text{ m}}{5.500 \times 10^{-7} \text{ m}} (1.600 - 1.000) \\ &= 2.84. \end{aligned} \quad (\text{Answer})$$

Thus, the phase difference of the emerging waves is 2.84 wavelengths. Because 1.0 wavelength is equivalent to 2π rad and 360° , you can show that this phase difference is equivalent to

$$\text{phase difference} = 17.8 \text{ rad} \approx 1020^\circ. \quad (\text{Answer})$$

The effective phase difference is the decimal part of the actual phase difference *expressed in wavelengths*. Thus, we have

$$\text{effective phase difference} = 0.84 \text{ wavelength}. \quad (\text{Answer})$$

You can show that this is equivalent to 5.3 rad and about 300° . **Caution:** We do *not* find the effective phase difference by taking the decimal part of the actual phase difference as expressed in radians or degrees. For example, we do *not* take 0.8 rad from the actual phase difference of 17.8 rad.

(b) If the waves reached the same point on a distant screen, what type of interference would they produce?

Reasoning: We need to compare the effective phase difference of the waves with the phase differences that give the extreme types of interference. Here the effective phase difference of 0.84 wavelength is between 0.5 wavelength (for fully destructive interference, or the darkest possible result) and 1.0 wavelength (for fully constructive interference, or the brightest possible result), but closer to 1.0 wavelength. Thus, the waves would produce intermediate interference that is closer to fully constructive interference—they would produce a relatively bright spot.



Additional examples, video, and practice available at WileyPLUS

35-2 YOUNG'S INTERFERENCE EXPERIMENT

Learning Objectives

After reading this module, you should be able to . . .

35.13 Describe the diffraction of light by a narrow slit and the effect of narrowing the slit.

35.14 With sketches, describe the production of the interference pattern in a double-slit interference experiment using monochromatic light.

35.15 Identify that the phase difference between two waves can change if the waves travel along paths of different lengths, as in the case of Young's experiment.

35.16 In a double-slit experiment, apply the relationship between the path length difference ΔL and the wavelength λ , and then interpret the result in terms of interference (maximum brightness, intermediate brightness, and darkness).

35.17 For a given point in a double-slit interference pattern, express the path length difference ΔL of the rays reaching that point in terms of the slit separation d and the angle θ to that point.

35.18 In a Young's experiment, apply the relationships between the slit separation d , the light wavelength λ , and the

angles θ to the minima (dark fringes) and to the maxima (bright fringes) in the interference pattern.

35.19 Sketch the double-slit interference pattern, identifying what lies at the center and what the various bright and dark fringes are called (such as "first side maximum" and "third order").

35.20 Apply the relationship between the distance D between a double-slit screen and a viewing screen, the angle θ to a point in the interference pattern, and the distance y to that point from the pattern's center.

35.21 For a double-slit interference pattern, identify the effects of changing d or λ and also identify what determines the angular limit to the pattern.

35.22 For a transparent material placed over one slit in a Young's experiment, determine the thickness or index of refraction required to shift a given fringe to the center of the interference pattern.

Key Ideas

- In Young's interference experiment, light passing through a single slit falls on two slits in a screen. The light leaving these slits flares out (by diffraction), and interference occurs in the region beyond the screen. A fringe pattern, due to the interference, forms on a viewing screen.

- The conditions for maximum and minimum intensity are

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima—bright fringes}),$$

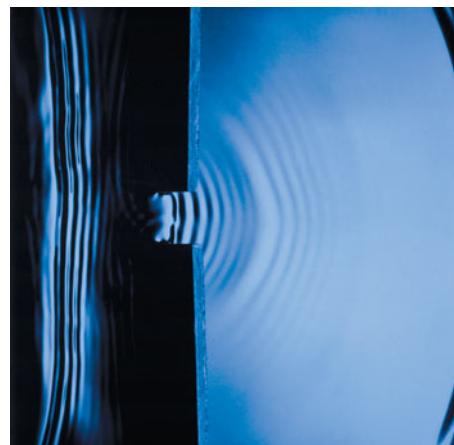
$$d \sin \theta = (m + \frac{1}{2})\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{minima—dark fringes}),$$

where θ is the angle the light path makes with a central axis and d is the slit separation.

Diffraction

In this module we shall discuss the experiment that first proved that light is a wave. To prepare for that discussion, we must introduce the idea of **diffraction** of waves, a phenomenon that we explore much more fully in Chapter 36. Its essence is this: If a wave encounters a barrier that has an opening of dimensions similar to the wavelength, the part of the wave that passes through the opening will flare (spread) out—will *diffract*—into the region beyond the barrier. The flaring is consistent with the spreading of wavelets in the Huygens construction of Fig. 35-2. Diffraction occurs for waves of all types, not just light waves; Fig. 35-6 shows the diffraction of water waves traveling across the surface of water in a shallow tank. Similar diffraction of ocean waves through openings in a barrier can actually increase the erosion of a beach the barrier is intended to protect.

Figure 35-6 Waves produced by an oscillating paddle at the left flare out through an opening in a barrier along the water surface.



George Resch/Fundamental Photographs

A wave passing through a slit flares (diffracts).

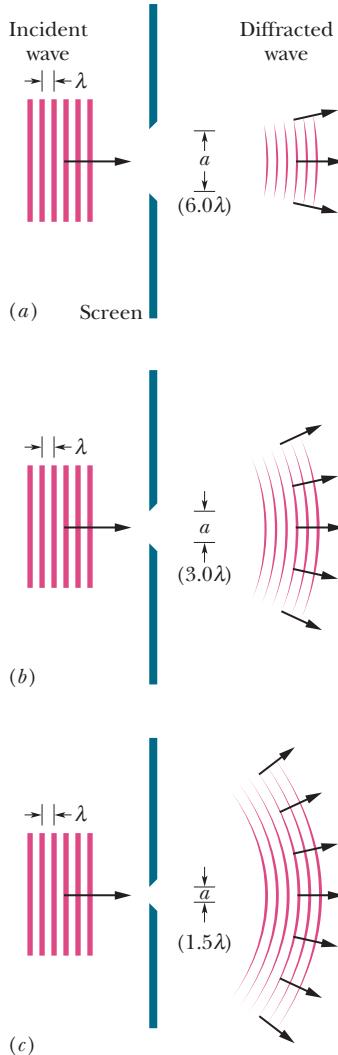


Figure 35-7 Diffraction represented schematically. For a given wavelength λ , the diffraction is more pronounced the smaller the slit width a . The figures show the cases for (a) slit width $a = 6.0\lambda$, (b) slit width $a = 3.0\lambda$, and (c) slit width $a = 1.5\lambda$. In all three cases, the screen and the length of the slit extend well into and out of the page, perpendicular to it.

Figure 35-7a shows the situation schematically for an incident plane wave of wavelength λ encountering a slit that has width $a = 6.0\lambda$ and extends into and out of the page. The part of the wave that passes through the slit flares out on the far side. Figures 35-7b (with $a = 3.0\lambda$) and 35-7c ($a = 1.5\lambda$) illustrate the main feature of diffraction: the narrower the slit, the greater the diffraction.

Diffraction limits geometrical optics, in which we represent an electromagnetic wave with a ray. If we actually try to form a ray by sending light through a narrow slit, or through a series of narrow slits, diffraction will always defeat our effort because it always causes the light to spread. Indeed, the narrower we make the slits (in the hope of producing a narrower beam), the greater the spreading is. Thus, geometrical optics holds only when slits or other apertures that might be located in the path of light do not have dimensions comparable to or smaller than the wavelength of the light.

Young's Interference Experiment

In 1801, Thomas Young experimentally proved that light is a wave, contrary to what most other scientists then thought. He did so by demonstrating that light undergoes interference, as do water waves, sound waves, and waves of all other types. In addition, he was able to measure the average wavelength of sunlight; his value, 570 nm, is impressively close to the modern accepted value of 555 nm. We shall here examine Young's experiment as an example of the interference of light waves.

Figure 35-8 gives the basic arrangement of Young's experiment. Light from a distant monochromatic source illuminates slit S_0 in screen A. The emerging light then spreads via diffraction to illuminate two slits S_1 and S_2 in screen B. Diffraction of the light by these two slits sends overlapping circular waves into

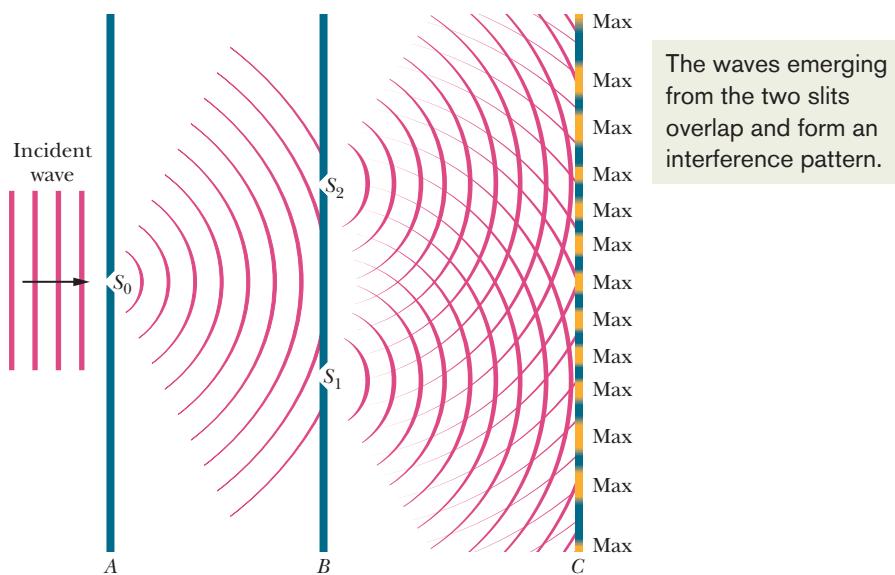


Figure 35-8 In Young's interference experiment, incident monochromatic light is diffracted by slit S_0 , which then acts as a point source of light that emits semicircular wavefronts. As that light reaches screen B, it is diffracted by slits S_1 and S_2 , which then act as two point sources of light. The light waves traveling from slits S_1 and S_2 overlap and undergo interference, forming an interference pattern of maxima and minima on viewing screen C. This figure is a cross section; the screens, slits, and interference pattern extend into and out of the page. Between screens B and C, the semicircular wavefronts centered on S_2 depict the waves that would be there if only S_2 were open. Similarly, those centered on S_1 depict waves that would be there if only S_1 were open.

the region beyond screen B , where the waves from one slit interfere with the waves from the other slit.

The “snapshot” of Fig. 35-8 depicts the interference of the overlapping waves. However, we cannot see evidence for the interference except where a viewing screen C intercepts the light. Where it does so, points of interference maxima form visible bright rows—called *bright bands*, *bright fringes*, or (loosely speaking) *maxima*—that extend across the screen (into and out of the page in Fig. 35-8). Dark regions—called *dark bands*, *dark fringes*, or (loosely speaking) *minima*—result from fully destructive interference and are visible between adjacent pairs of bright fringes. (*Maxima* and *minima* more properly refer to the center of a band.) The pattern of bright and dark fringes on the screen is called an **interference pattern**. Figure 35-9 is a photograph of part of the interference pattern that would be seen by an observer standing to the left of screen C in the arrangement of Fig. 35-8.

Locating the Fringes

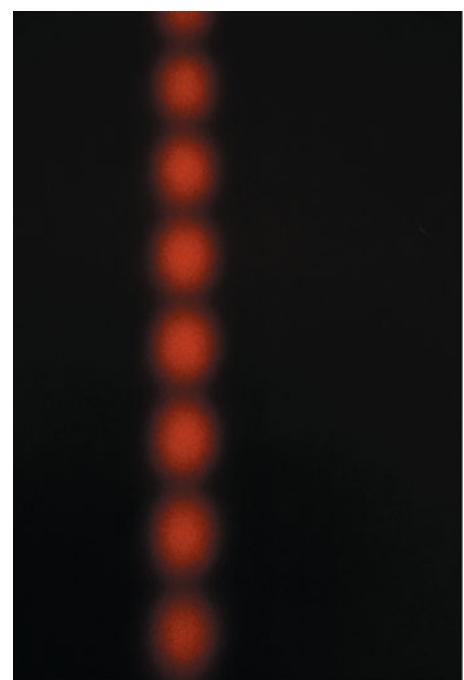
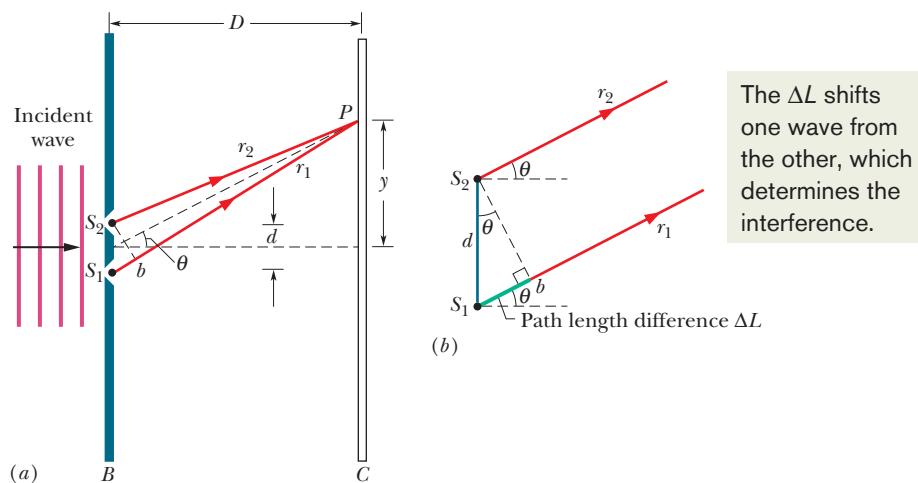
Light waves produce fringes in a *Young's double-slit interference experiment*, as it is called, but what exactly determines the locations of the fringes? To answer, we shall use the arrangement in Fig. 35-10a. There, a plane wave of monochromatic light is incident on two slits S_1 and S_2 in screen B ; the light diffracts through the slits and produces an interference pattern on screen C . We draw a central axis from the point halfway between the slits to screen C as a reference. We then pick, for discussion, an arbitrary point P on the screen, at angle θ to the central axis. This point intercepts the wave of ray r_1 from the bottom slit and the wave of ray r_2 from the top slit.

Path Length Difference. These waves are in phase when they pass through the two slits because there they are just portions of the same incident wave. However, once they have passed the slits, the two waves must travel different distances to reach P . We saw a similar situation in Module 17-3 with sound waves and concluded that



The phase difference between two waves can change if the waves travel paths of different lengths.

The change in phase difference is due to the *path length difference* ΔL in the paths taken by the waves. Consider two waves initially exactly in phase, traveling along paths with a path length difference ΔL , and then passing through some common point. When ΔL is zero or an integer number of wavelengths, the waves arrive at the common point exactly in phase and they interfere fully constructively there. If that is true for the waves of rays r_1 and r_2 in Fig. 35-10, then



Courtesy Jearl Walker

Figure 35-9 A photograph of the interference pattern produced by the arrangement shown in Fig. 35-8, but with short slits. (The photograph is a front view of part of screen C .) The alternating maxima and minima are called *interference fringes* (because they resemble the decorative fringe sometimes used on clothing and rugs).

Figure 35-10 (a) Waves from slits S_1 and S_2 (which extend into and out of the page) combine at P , an arbitrary point on screen C at distance y from the central axis. The angle θ serves as a convenient locator for P . (b) For $D \gg d$, we can approximate rays r_1 and r_2 as being parallel, at angle θ to the central axis.

point P is part of a bright fringe. When, instead, ΔL is an odd multiple of half a wavelength, the waves arrive at the common point exactly out of phase and they interfere fully destructively there. If that is true for the waves of rays r_1 and r_2 , then point P is part of a dark fringe. (And, of course, we can have intermediate situations of interference and thus intermediate illumination at P .) Thus,



What appears at each point on the viewing screen in a Young's double-slit interference experiment is determined by the path length difference ΔL of the rays reaching that point.

Angle. We can specify where each bright fringe and each dark fringe is located on the screen by giving the angle θ from the central axis to that fringe. To find θ , we must relate it to ΔL . We start with Fig. 35-10a by finding a point b along ray r_1 such that the path length from b to P equals the path length from S_2 to P . Then the path length difference ΔL between the two rays is the distance from S_1 to b .

The relation between this S_1 -to- b distance and θ is complicated, but we can simplify it considerably if we arrange for the distance D from the slits to the screen to be much greater than the slit separation d . Then we can approximate rays r_1 and r_2 as being parallel to each other and at angle θ to the central axis (Fig. 35-10b). We can also approximate the triangle formed by S_1 , S_2 , and b as being a right triangle, and approximate the angle inside that triangle at S_2 as being θ . Then, for that triangle, $\sin \theta = \Delta L/d$ and thus

$$\Delta L = d \sin \theta \quad (\text{path length difference}). \quad (35-12)$$

For a bright fringe, we saw that ΔL must be either zero or an integer number of wavelengths. Using Eq. 35-12, we can write this requirement as

$$\Delta L = d \sin \theta = (\text{integer})(\lambda), \quad (35-13)$$

or as

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima—bright fringes}). \quad (35-14)$$

For a dark fringe, ΔL must be an odd multiple of half a wavelength. Again using Eq. 35-12, we can write this requirement as

$$\Delta L = d \sin \theta = (\text{odd number})\left(\frac{1}{2}\lambda\right), \quad (35-15)$$

or as

$$d \sin \theta = (m + \frac{1}{2})\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{minima—dark fringes}). \quad (35-16)$$

With Eqs. 35-14 and 35-16, we can find the angle θ to any fringe and thus locate that fringe; further, we can use the values of m to label the fringes. For the value and label $m = 0$, Eq. 35-14 tells us that a bright fringe is at $\theta = 0$ and thus on the central axis. This *central maximum* is the point at which waves arriving from the two slits have a path length difference $\Delta L = 0$, hence zero phase difference.

For, say, $m = 2$, Eq. 35-14 tells us that *bright* fringes are at the angle

$$\theta = \sin^{-1}\left(\frac{2\lambda}{d}\right)$$

above and below the central axis. Waves from the two slits arrive at these two fringes with $\Delta L = 2\lambda$ and with a phase difference of two wavelengths. These fringes are said to be the *second-order bright fringes* (meaning $m = 2$) or the *second side maxima* (the second maxima to the side of the central maximum), or

they are described as being the second bright fringes from the central maximum.

For $m = 1$, Eq. 35-16 tells us that *dark* fringes are at the angle

$$\theta = \sin^{-1}\left(\frac{1.5\lambda}{d}\right)$$

above and below the central axis. Waves from the two slits arrive at these two fringes with $\Delta L = 1.5\lambda$ and with a phase difference, in wavelengths, of 1.5. These fringes are called the *second-order dark fringes* or *second minima* because they are the second dark fringes to the side of the central axis. (The first dark fringes, or first minima, are at locations for which $m = 0$ in Eq. 35-16.)

Nearby Screen. We derived Eqs. 35-14 and 35-16 for the situation $D \gg d$. However, they also apply if we place a converging lens between the slits and the viewing screen and then move the viewing screen closer to the slits, to the focal point of the lens. (The screen is then said to be in the *focal plane* of the lens; that is, it is in the plane perpendicular to the central axis at the focal point.) One property of a converging lens is that it focuses all rays that are parallel to one another to the same point on its focal plane. Thus, the rays that now arrive at any point on the screen (in the focal plane) were exactly parallel (rather than approximately) when they left the slits. They are like the initially parallel rays in Fig. 34-14a that are directed to a point (the focal point) by a lens.



Checkpoint 3

In Fig. 35-10a, what are ΔL (as a multiple of the wavelength) and the phase difference (in wavelengths) for the two rays if point P is (a) a third side maximum and (b) a third minimum?

Sample Problem 35.02 Double-slit interference pattern

What is the distance on screen C in Fig. 35-10a between adjacent maxima near the center of the interference pattern? The wavelength λ of the light is 546 nm, the slit separation d is 0.12 mm, and the slit–screen separation D is 55 cm. Assume that θ in Fig. 35-10 is small enough to permit use of the approximations $\sin \theta \approx \tan \theta \approx \theta$, in which θ is expressed in radian measure.

KEY IDEAS

(1) First, let us pick a maximum with a low value of m to ensure that it is near the center of the pattern. Then, from the geometry of Fig. 35-10a, the maximum's vertical distance y_m from the center of the pattern is related to its angle θ from the central axis by

$$\tan \theta \approx \theta = \frac{y_m}{D}.$$

(2) From Eq. 35-14, this angle θ for the m th maximum is given by

$$\sin \theta \approx \theta = \frac{m\lambda}{d}.$$

Calculations: If we equate our two expressions for angle θ and then solve for y_m , we find

$$y_m = \frac{m\lambda D}{d}. \quad (35-17)$$

For the next maximum as we move away from the pattern's center, we have

$$y_{m+1} = \frac{(m+1)\lambda D}{d}. \quad (35-18)$$

We find the distance between these adjacent maxima by subtracting Eq. 35-17 from Eq. 35-18:

$$\begin{aligned} \Delta y &= y_{m+1} - y_m = \frac{\lambda D}{d} \\ &= \frac{(546 \times 10^{-9} \text{ m})(55 \times 10^{-2} \text{ m})}{0.12 \times 10^{-3} \text{ m}} \\ &= 2.50 \times 10^{-3} \text{ m} \approx 2.5 \text{ mm.} \end{aligned} \quad (\text{Answer})$$

As long as d and θ in Fig. 35-10a are small, the separation of the interference fringes is independent of m ; that is, the fringes are evenly spaced.



Additional examples, video, and practice available at WileyPLUS

Sample Problem 35.03 Double-slit interference pattern with plastic over one slit

A double-slit interference pattern is produced on a screen, as in Fig. 35-10; the light is monochromatic at a wavelength of 600 nm. A strip of transparent plastic with index of refraction $n = 1.50$ is to be placed over one of the slits. Its presence changes the interference between light waves from the two slits, causing the interference pattern to be shifted across the screen from the original pattern. Figure 35-11a shows the original locations of the central bright fringe ($m = 0$) and the first bright fringes ($m = 1$) above and below the central fringe. The purpose of the plastic is to shift the pattern upward so that the lower $m = 1$ bright fringe is shifted to the center of the pattern. Should the plastic be placed over the top slit (as arbitrarily drawn in Fig. 35-11b) or the bottom slit, and what thickness L should it have?

KEY IDEA

The interference at a point on the screen depends on the phase difference of the light rays arriving from the two slits. The light rays are in phase at the slits because they derive from the same wave, but their relative phase can shift on the way to the screen due to (1) a difference in the length of the paths they follow and (2) a difference in the number of their internal wavelengths λ_n in the materials through which they pass. The first condition applies to any off-center point, and the second condition applies when the plastic covers one of the slits.

Path length difference: Figure 35-11a shows rays r_1 and r_2 along which waves from the two slits travel to reach the lower $m = 1$ bright fringe. Those waves start in phase at the slits but arrive at the fringe with a phase difference of exactly 1 wavelength. To remind ourselves of this main characteristic of the fringe, let us call it the 1λ fringe. The one-wavelength phase difference is due to the one-wavelength path length difference between the rays reaching the fringe; that is, there is exactly one more wavelength along ray r_2 than along r_1 .

Figure 35-11b shows the 1λ fringe shifted up to the center of the pattern with the plastic strip over the top slit (we still do not know whether the plastic should be there or over the bottom slit). The figure also shows the new orientations of rays r_1 and r_2 to reach that fringe. There still must be one more wavelength along r_2 than along r_1 (because they still produce the 1λ fringe), but now the path length difference between those rays is zero, as we can tell from the geometry of Fig. 35-11b. However, r_2 now passes through the plastic.

Internal wavelength: The wavelength λ_n of light in a material with index of refraction n is smaller than the wavelength in vacuum, as given by Eq. 35-6 ($\lambda_n = \lambda/n$). Here, this means that the wavelength of the light is smaller in the plastic than in the air. Thus, the ray that passes through the plastic will have more wavelengths along it than the ray that passes through only air—so we do get the one extra wavelength we need along ray r_2 by placing the plastic over the top slit, as drawn in Fig. 35-11b.

Thickness: To determine the required thickness L of the plastic, we first note that the waves are initially in phase and travel equal distances L through different materials (plastic and air). Because we know the phase difference and require L , we use Eq. 35-9,

$$N_2 - N_1 = \frac{L}{\lambda} (n_2 - n_1). \quad (35-19)$$

We know that $N_2 - N_1$ is 1 for a phase difference of one wavelength, n_2 is 1.50 for the plastic in front of the top slit, n_1 is 1.00 for the air in front of the bottom slit, and λ is 600×10^{-9} m. Then Eq. 35-19 tells us that, to shift the lower $m = 1$ bright fringe up to the center of the interference pattern, the plastic must have the thickness

$$L = \frac{\lambda(N_2 - N_1)}{n_2 - n_1} = \frac{(600 \times 10^{-9} \text{ m})(1)}{1.50 - 1.00} = 1.2 \times 10^{-6} \text{ m}. \quad (\text{Answer})$$

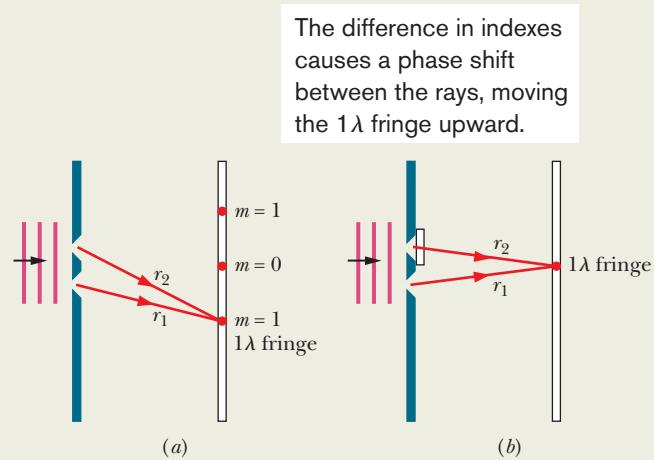


Figure 35-11 (a) Arrangement for two-slit interference (not to scale). The locations of three bright fringes (or maxima) are indicated. (b) A strip of plastic covers the top slit. We want the 1λ fringe to be at the center of the pattern.

35-3 INTERFERENCE AND DOUBLE-SLIT INTENSITY

Learning Objectives

After reading this module, you should be able to . . .

35.23 Distinguish between coherent and incoherent light.

35.24 For two light waves arriving at a common point, write expressions for their electric field components as functions of time and a phase constant.

35.25 Identify that the phase difference between two waves determines their interference.

35.26 For a point in a double-slit interference pattern, calculate the intensity in terms of the phase difference of

the arriving waves and relate that phase difference to the angle θ locating that point in the pattern.

35.27 Use a phasor diagram to find the resultant wave (amplitude and phase constant) of two or more light waves arriving at a common point and use that result to determine the intensity.

35.28 Apply the relationship between a light wave's angular frequency ω and the angular speed ω of the phasor representing the wave.

Key Ideas

- If two light waves that meet at a point are to interfere perceptibly, the phase difference between them must remain constant with time; that is, the waves must be coherent.

When two coherent waves meet, the resulting intensity may be found by using phasors.

- In Young's interference experiment, two waves, each with intensity I_0 , yield a resultant wave of intensity I at the viewing screen, with

$$I = 4I_0 \cos^2 \frac{1}{2}\phi, \quad \text{where } \phi = \frac{2\pi d}{\lambda} \sin \theta.$$

Coherence

For the interference pattern to appear on viewing screen C in Fig. 35-8, the light waves reaching any point P on the screen must have a phase difference that does not vary in time. That is the case in Fig. 35-8 because the waves passing through slits S_1 and S_2 are portions of the single light wave that illuminates the slits. Because the phase difference remains constant, the light from slits S_1 and S_2 is said to be completely **coherent**.

Sunlight and Fingernails. Direct sunlight is partially coherent; that is, sunlight waves intercepted at two points have a constant phase difference only if the points are very close. If you look closely at your fingernail in bright sunlight, you can see a faint interference pattern called *speckle* that causes the nail to appear to be covered with specks. You see this effect because light waves scattering from very close points on the nail are sufficiently coherent to interfere with one another at your eye. The slits in a double-slit experiment, however, are not close enough, and in direct sunlight, the light at the slits would be **incoherent**. To get coherent light, we would have to send the sunlight through a single slit as in Fig. 35-8; because that single slit is small, light that passes through it is coherent. In addition, the smallness of the slit causes the coherent light to spread via diffraction to illuminate both slits in the double-slit experiment.

Incoherent Sources. If we replace the double slits with two similar but independent monochromatic light sources, such as two fine incandescent wires, the phase difference between the waves emitted by the sources varies rapidly and randomly. (This occurs because the light is emitted by vast numbers of atoms in the wires, acting randomly and independently for extremely short times—of the order of nanoseconds.) As a result, at any given point on the viewing screen, the interference between the waves from the two sources varies rapidly and randomly between fully constructive and fully destructive. The eye (and most common optical detectors) cannot follow such changes, and no interference pattern can be seen. The fringes disappear, and the screen is seen as being uniformly illuminated.

Coherent Source. A *laser* differs from common light sources in that its atoms emit light in a cooperative manner, thereby making the light coherent. Moreover, the light is almost monochromatic, is emitted in a thin beam with little spreading, and can be focused to a width that almost matches the wavelength of the light.

Intensity in Double-Slit Interference

Equations 35-14 and 35-16 tell us how to locate the maxima and minima of the double-slit interference pattern on screen *C* of Fig. 35-10 as a function of the angle θ in that figure. Here we wish to derive an expression for the intensity I of the fringes as a function of θ .

The light leaving the slits is in phase. However, let us assume that the light waves from the two slits are not in phase when they arrive at point *P*. Instead, the electric field components of those waves at point *P* are not in phase and vary with time as

$$E_1 = E_0 \sin \omega t \quad (35-20)$$

and

$$E_2 = E_0 \sin(\omega t + \phi), \quad (35-21)$$

where ω is the angular frequency of the waves and ϕ is the phase constant of wave E_2 . Note that the two waves have the same amplitude E_0 and a phase difference of ϕ . Because that phase difference does not vary, the waves are coherent. We shall show that these two waves will combine at *P* to produce an intensity I given by

$$I = 4I_0 \cos^2 \frac{1}{2}\phi, \quad (35-22)$$

and that

$$\phi = \frac{2\pi d}{\lambda} \sin \theta. \quad (35-23)$$

In Eq. 35-22, I_0 is the intensity of the light that arrives on the screen from one slit when the other slit is temporarily covered. We assume that the slits are so narrow in comparison to the wavelength that this single-slit intensity is essentially uniform over the region of the screen in which we wish to examine the fringes.

Equations 35-22 and 35-23, which together tell us how the intensity I of the fringe pattern varies with the angle θ in Fig. 35-10, necessarily contain information about the location of the maxima and minima. Let us see if we can extract that information to find equations about those locations.

Maxima. Study of Eq. 35-22 shows that intensity maxima will occur when

$$\frac{1}{2}\phi = m\pi, \quad \text{for } m = 0, 1, 2, \dots \quad (35-24)$$

If we put this result into Eq. 35-23, we find

$$2m\pi = \frac{2\pi d}{\lambda} \sin \theta, \quad \text{for } m = 0, 1, 2, \dots$$

or

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima}), \quad (35-25)$$

which is exactly Eq. 35-14, the expression that we derived earlier for the locations of the maxima.

Minima. The minima in the fringe pattern occur when

$$\frac{1}{2}\phi = (m + \frac{1}{2})\pi, \quad \text{for } m = 0, 1, 2, \dots \quad (35-26)$$

If we combine this relation with Eq. 35-23, we are led at once to

$$d \sin \theta = (m + \frac{1}{2})\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{minima}), \quad (35-27)$$

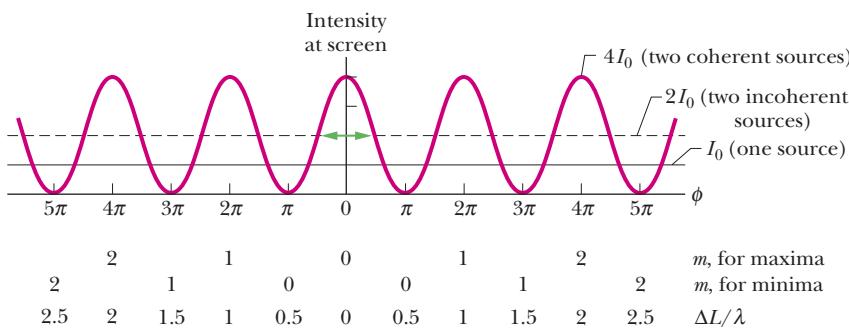


Figure 35-12 A plot of Eq. 35-22, showing the intensity of a double-slit interference pattern as a function of the phase difference between the waves when they arrive from the two slits. I_0 is the (uniform) intensity that would appear on the screen if one slit were covered. The average intensity of the fringe pattern is $2I_0$, and the *maximum* intensity (for coherent light) is $4I_0$.

which is just Eq. 35-16, the expression we derived earlier for the locations of the fringe minima.

Figure 35-12, which is a plot of Eq. 35-22, shows the intensity of double-slit interference patterns as a function of the phase difference ϕ between the waves at the screen. The horizontal solid line is I_0 , the (uniform) intensity on the screen when one of the slits is covered up. Note in Eq. 35-22 and the graph that the intensity I varies from zero at the fringe minima to $4I_0$ at the fringe maxima.

If the waves from the two sources (slits) were *incoherent*, so that no enduring phase relation existed between them, there would be no fringe pattern and the intensity would have the uniform value $2I_0$ for all points on the screen; the horizontal dashed line in Fig. 35-12 shows this uniform value.

Interference cannot create or destroy energy but merely redistributes it over the screen. Thus, the *average* intensity on the screen must be the same $2I_0$ regardless of whether the sources are coherent. This follows at once from Eq. 35-22; if we substitute $\frac{1}{2}$, the average value of the cosine-squared function, this equation reduces to $I_{\text{avg}} = 2I_0$.

Proof of Eqs. 35-22 and 35-23

We shall combine the electric field components E_1 and E_2 , given by Eqs. 35-20 and 35-21, respectively, by the method of phasors as is discussed in Module 16-6. In Fig. 35-13a, the waves with components E_1 and E_2 are represented by phasors of magnitude E_0 that rotate around the origin at angular speed ω . The values of E_1 and E_2 at any time are the projections of the corresponding phasors on the vertical axis. Figure 35-13a shows the phasors and their projections at an arbitrary time t . Consistent with Eqs. 35-20 and 35-21, the phasor for E_1 has a rotation angle ωt and the phasor for E_2 has a rotation angle $\omega t + \phi$ (it is phase-shifted ahead of E_1). As each phasor rotates, its projection on the vertical axis varies with time in the same way that the sinusoidal functions of Eqs. 35-20 and 35-21 vary with time.

To combine the field components E_1 and E_2 at any point P in Fig. 35-10, we add their phasors vectorially, as shown in Fig. 35-13b. The magnitude of the vector sum is the amplitude E of the resultant wave at point P , and that wave has a certain phase constant β . To find the amplitude E in Fig. 35-13b, we first note that the two angles marked β are equal because they are opposite equal-length sides of a triangle. From the theorem (for triangles) that an exterior angle (here ϕ , as shown in Fig. 35-13b) is equal to the sum of the two opposite interior angles (here that sum is $\beta + \beta$), we see that $\beta = \frac{1}{2}\phi$. Thus, we have

$$\begin{aligned} E &= 2(E_0 \cos \beta) \\ &= 2E_0 \cos \frac{1}{2}\phi. \end{aligned} \quad (35-28)$$

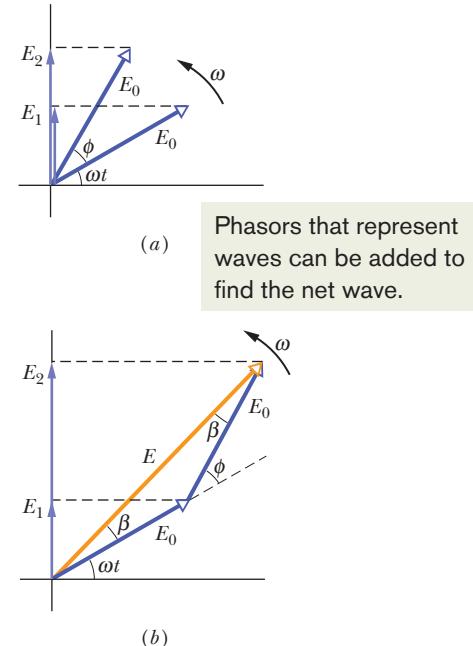


Figure 35-13 (a) Phasors representing, at time t , the electric field components given by Eqs. 35-20 and 35-21. Both phasors have magnitude E_0 and rotate with angular speed ω . Their phase difference is ϕ . (b) Vector addition of the two phasors gives the phasor representing the resultant wave, with amplitude E and phase constant β .

If we square each side of this relation, we obtain

$$E^2 = 4E_0^2 \cos^2 \frac{1}{2}\phi. \quad (35-29)$$

Intensity. Now, from Eq. 33-24, we know that the intensity of an electromagnetic wave is proportional to the square of its amplitude. Therefore, the waves we are combining in Fig. 35-13b, whose amplitudes are E_0 , each has an intensity I_0 that is proportional to E_0^2 , and the resultant wave, with amplitude E , has an intensity I that is proportional to E^2 . Thus,

$$\frac{I}{I_0} = \frac{E^2}{E_0^2}.$$

Substituting Eq. 35-29 into this equation and rearranging then yield

$$I = 4I_0 \cos^2 \frac{1}{2}\phi,$$

which is Eq. 35-22, which we set out to prove.

We still must prove Eq. 35-23, which relates the phase difference ϕ between the waves arriving at any point P on the screen of Fig. 35-10 to the angle θ that serves as a locator of that point.

The phase difference ϕ in Eq. 35-21 is associated with the path length difference S_1b in Fig. 35-10b. If S_1b is $\frac{1}{2}\lambda$, then ϕ is π ; if S_1b is λ , then ϕ is 2π , and so on. This suggests

$$\left(\begin{array}{c} \text{phase} \\ \text{difference} \end{array} \right) = \frac{2\pi}{\lambda} \left(\begin{array}{c} \text{path length} \\ \text{difference} \end{array} \right). \quad (35-30)$$

The path length difference S_1b in Fig. 35-10b is $d \sin \theta$ (a leg of the right triangle); so Eq. 35-30 for the phase difference between the two waves arriving at point P on the screen becomes

$$\phi = \frac{2\pi d}{\lambda} \sin \theta,$$

which is Eq. 35-23, the other equation that we set out to prove to relate ϕ to the angle θ that locates P .

Combining More Than Two Waves

In a more general case, we might want to find the resultant of more than two sinusoidally varying waves at a point. Whatever the number of waves is, our general procedure is this:

1. Construct a series of phasors representing the waves to be combined. Draw them end to end, maintaining the proper phase relations between adjacent phasors.
2. Construct the vector sum of this array. The length of this vector sum gives the amplitude of the resultant phasor. The angle between the vector sum and the first phasor is the phase of the resultant with respect to this first phasor. The projection of this vector-sum phasor on the vertical axis gives the time variation of the resultant wave.



Checkpoint 4

Each of four pairs of light waves arrives at a certain point on a screen. The waves have the same wavelength. At the arrival point, their amplitudes and phase differences are (a) $2E_0$, $6E_0$, and π rad; (b) $3E_0$, $5E_0$, and π rad; (c) $9E_0$, $7E_0$, and 3π rad; (d) $2E_0$, $2E_0$, and 0 rad. Rank the four pairs according to the intensity of the light at the arrival point, greatest first. (Hint: Draw phasors.)



Sample Problem 35.04 Combining three light waves by using phasors

Three light waves combine at a certain point where their electric field components are

$$\begin{aligned}E_1 &= E_0 \sin \omega t, \\E_2 &= E_0 \sin(\omega t + 60^\circ), \\E_3 &= E_0 \sin(\omega t - 30^\circ).\end{aligned}$$

Find their resultant component $E(t)$ at that point.

KEY IDEA

The resultant wave is

$$E(t) = E_1(t) + E_2(t) + E_3(t).$$

We can use the method of phasors to find this sum, and we are free to evaluate the phasors at any time t .

Calculations: To simplify the solution, we choose $t = 0$, for which the phasors representing the three waves are shown in Fig. 35-14. We can add these three phasors either directly on a vector-capable calculator or by components. For the component approach, we first write the sum of their horizontal components as

$$\sum E_h = E_0 \cos 0 + E_0 \cos 60^\circ + E_0 \cos(-30^\circ) = 2.37E_0.$$

The sum of their vertical components, which is the value of E at $t = 0$, is

$$\sum E_v = E_0 \sin 0 + E_0 \sin 60^\circ + E_0 \sin(-30^\circ) = 0.366E_0.$$

The resultant wave $E(t)$ thus has an amplitude E_R of

$$E_R = \sqrt{(2.37E_0)^2 + (0.366E_0)^2} = 2.4E_0,$$

and a phase angle β relative to the phasor representing E_1 of

$$\beta = \tan^{-1}\left(\frac{0.366E_0}{2.37E_0}\right) = 8.8^\circ.$$

We can now write, for the resultant wave $E(t)$,

$$\begin{aligned}E &= E_R \sin(\omega t + \beta) \\&= 2.4E_0 \sin(\omega t + 8.8^\circ).\end{aligned} \quad (\text{Answer})$$

Be careful to interpret the angle β correctly in Fig. 35-14: It is the constant angle between E_R and the phasor representing E_1 as the four phasors rotate as a single unit around the origin. The angle between E_R and the horizontal axis in Fig. 35-14 does not remain equal to β .

Phasors that represent waves can be added to find the net wave.

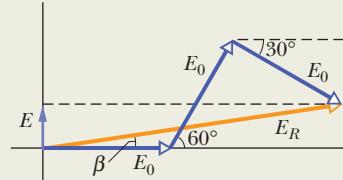


Figure 35-14 Three phasors, representing waves with equal amplitudes E_0 and with phase constants 0° , 60° , and -30° , shown at time $t = 0$. The phasors combine to give a resultant phasor with magnitude E_R , at angle β .



Additional examples, video, and practice available at WileyPLUS

35-4 INTERFERENCE FROM THIN FILMS

Learning Objectives

After reading this module, you should be able to . . .

35.29 Sketch the setup for thin-film interference, showing the incident ray and reflected rays (perpendicular to the film but drawn slightly slanted for clarity) and identifying the thickness and the three indexes of refraction.

35.30 Identify the condition in which a reflection can result in a phase shift, and give the value of that phase shift.

35.31 Identify the three factors that determine the interference of the reflected waves: reflection shifts, path length difference, and internal wavelength (set by the film's index of refraction).

35.32 For a thin film, use the reflection shifts and the desired result (the reflected waves are in phase or out of phase, or

the transmitted waves are in phase or out of phase) to determine and then apply the necessary equation relating the thickness L , the wavelength λ (measured in air), and the index of refraction n of the film.

35.33 For a very thin film in air (with thickness much less than the wavelength of visible light), explain why the film is always dark.

35.34 At each end of a thin film in the form of a wedge, determine and then apply the necessary equation relating the thickness L , the wavelength λ (measured in air), and the index of refraction n of the film, and then count the number of bright bands and dark bands across the film.

Key Ideas

- When light is incident on a thin transparent film, the light waves reflected from the front and back surfaces interfere. For near-normal incidence, the wavelength conditions for maximum and minimum intensity of the light reflected from a *film in air* are

$$2L = (m + \frac{1}{2}) \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots$$

(maxima—bright film in air),

$$2L = m \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots$$

(minima—dark film in air),

where n_2 is the index of refraction of the film, L is its thickness, and λ is the wavelength of the light in air.

- If a film is sandwiched between media other than air, these equations for bright and dark films may be interchanged, depending on the relative indexes of refraction.
- If the light incident at an interface between media with different indexes of refraction is in the medium with the smaller index of refraction, the reflection causes a phase change of π rad, or half a wavelength, in the reflected wave. Otherwise, there is no phase change due to the reflection. Refraction causes no phase shift.

The interference depends on the reflections and the path lengths.

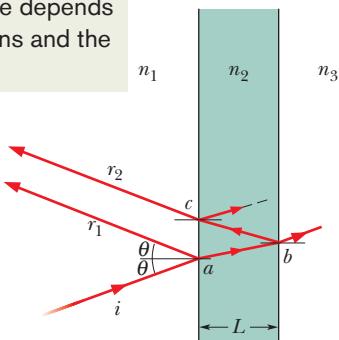


Figure 35-15 Light waves, represented with ray i , are incident on a thin film of thickness L and index of refraction n_2 . Rays r_1 and r_2 represent light waves that have been reflected by the front and back surfaces of the film, respectively. (All three rays are actually nearly perpendicular to the film.) The interference of the waves of r_1 and r_2 with each other depends on their phase difference. The index of refraction n_1 of the medium at the left can differ from the index of refraction n_3 of the medium at the right, but for now we assume that both media are air, with $n_1 = n_3 = 1.0$, which is less than n_2 .

Interference from Thin Films

The colors on a sunlit soap bubble or an oil slick are caused by the interference of light waves reflected from the front and back surfaces of a thin transparent film. The thickness of the soap or oil film is typically of the order of magnitude of the wavelength of the (visible) light involved. (Greater thicknesses spoil the coherence of the light needed to produce the colors due to interference.)

Figure 35-15 shows a thin transparent film of uniform thickness L and index of refraction n_2 , illuminated by bright light of wavelength λ from a distant point source. For now, we assume that air lies on both sides of the film and thus that $n_1 = n_3$ in Fig. 35-15. For simplicity, we also assume that the light rays are almost perpendicular to the film ($\theta \approx 0$). We are interested in whether the film is bright or dark to an observer viewing it almost perpendicularly. (Since the film is brightly illuminated, how could it possibly be dark? You will see.)

The incident light, represented by ray i , intercepts the front (left) surface of the film at point a and undergoes both reflection and refraction there. The reflected ray r_1 is intercepted by the observer's eye. The refracted light crosses the film to point b on the back surface, where it undergoes both reflection and refraction. The light reflected at b crosses back through the film to point c , where it undergoes both reflection and refraction. The light refracted at c , represented by ray r_2 , is intercepted by the observer's eye.

If the light waves of rays r_1 and r_2 are exactly in phase at the eye, they produce an interference maximum and region ac on the film is bright to the observer. If they are exactly out of phase, they produce an interference minimum and region ac is dark to the observer, *even though it is illuminated*. If there is some intermediate phase difference, there are intermediate interference and brightness.

The Key. Thus, the key to what the observer sees is the phase difference between the waves of rays r_1 and r_2 . Both rays are derived from the same ray i , but the path involved in producing r_2 involves light traveling twice across the film (a to b , and then b to c), whereas the path involved in producing r_1 involves no travel through the film. Because θ is about zero, we approximate the path length difference between the waves of r_1 and r_2 as $2L$. However, to find the phase difference between the waves, we cannot just find the number of wavelengths λ that is equivalent to a path length difference of $2L$. This simple approach is impossible for two reasons: (1) the path length difference occurs in a medium other than air, and (2) reflections are involved, which can change the phase.



The phase difference between two waves can change if one or both are reflected.

Let's next discuss changes in phase that are caused by reflections.

Reflection Phase Shifts

Refraction at an interface never causes a phase change—but reflection can, depending on the indexes of refraction on the two sides of the interface. Figure 35-16 shows what happens when reflection causes a phase change, using as an example pulses on a denser string (along which pulse travel is relatively slow) and a lighter string (along which pulse travel is relatively fast).

When a pulse traveling relatively slowly along the denser string in Fig. 35-16a reaches the interface with the lighter string, the pulse is partially transmitted and partially reflected, with no change in orientation. For light, this situation corresponds to the incident wave traveling in the medium of greater index of refraction n (recall that greater n means slower speed). In that case, the wave that is reflected at the interface does not undergo a change in phase; that is, its *reflection phase shift* is zero.

When a pulse traveling more quickly along the lighter string in Fig. 35-16b reaches the interface with the denser string, the pulse is again partially transmitted and partially reflected. The transmitted pulse again has the same orientation as the incident pulse, but now the reflected pulse is inverted. For a sinusoidal wave, such an inversion involves a phase change of π rad, or half a wavelength. For light, this situation corresponds to the incident wave traveling in the medium of lesser index of refraction (with greater speed). In that case, the wave that is reflected at the interface undergoes a phase shift of π rad, or half a wavelength.

We can summarize these results for light in terms of the index of refraction of the medium off which (or from which) the light reflects:



Reflection	Reflection phase shift
Off lower index	0
Off higher index	0.5 wavelength

This might be remembered as “higher means half.”

Equations for Thin-Film Interference

In this chapter we have now seen three ways in which the phase difference between two waves can change:

1. by reflection
2. by the waves traveling along paths of different lengths
3. by the waves traveling through media of different indexes of refraction

When light reflects from a thin film, producing the waves of rays r_1 and r_2 shown in Fig. 35-15, all three ways are involved. Let us consider them one by one.

Reflection Shift. We first reexamine the two reflections in Fig. 35-15. At point a on the front interface, the incident wave (in air) reflects from the medium having the higher of the two indexes of refraction; so the wave of reflected ray r_1 has its phase shifted by 0.5 wavelength. At point b on the back interface, the incident wave reflects from the medium (air) having the lower of the two indexes of refraction; so the wave reflected there is not shifted in phase by the reflection, and thus neither is the portion of it that exits the film as ray r_2 . We can organize this information with the first line in Table 35-1, which refers to the simplified drawing in Fig. 35-17 for a thin film in air. So far, as a result of the reflection phase shifts, the waves of r_1 and r_2 have a phase difference of 0.5 wavelength and thus are exactly out of phase.

Path Length Difference. Now we must consider the path length difference $2L$ that occurs because the wave of ray r_2 crosses the film twice. (This difference

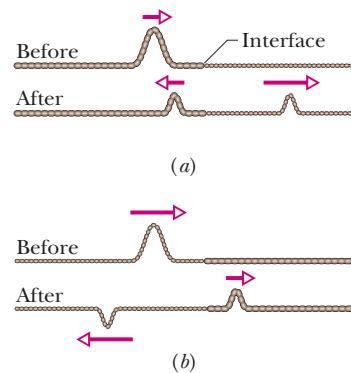


Figure 35-16 Phase changes when a pulse is reflected at the interface between two stretched strings of different linear densities. The wave speed is greater in the lighter string. (a) The incident pulse is in the denser string. (b) The incident pulse is in the lighter string. Only here is there a phase change, and only in the reflected wave.

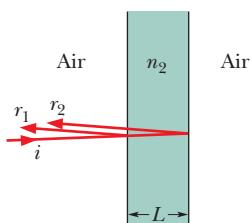


Figure 35-17 Reflections from a thin film in air.

Table 35-1 An Organizing Table for Thin-Film Interference in Air (Fig. 35-17)^a

Reflection phase shifts	r_1	r_2
0.5 wavelength	0.5	0
Path length difference		$2L$
Index in which path length difference occurs		n_2
In phase ^a :	$2L = \frac{\text{odd number}}{2} \times \frac{\lambda}{n_2}$	
Out of phase ^a :	$2L = \text{integer} \times \frac{\lambda}{n_2}$	

^aValid for $n_2 > n_1$ and $n_2 > n_3$.

$2L$ is shown on the second line in Table 35-1.) If the waves of r_1 and r_2 are to be exactly in phase so that they produce fully constructive interference, the path length $2L$ must cause an additional phase difference of 0.5, 1.5, 2.5, . . . wavelengths. Only then will the net phase difference be an integer number of wavelengths. Thus, for a bright film, we must have

$$2L = \frac{\text{odd number}}{2} \times \text{wavelength} \quad (\text{in-phase waves}). \quad (35-31)$$

The wavelength we need here is the wavelength λ_{n2} of the light in the medium containing path length $2L$ —that is, in the medium with index of refraction n_2 . Thus, we can rewrite Eq. 35-31 as

$$2L = \frac{\text{odd number}}{2} \times \lambda_{n2} \quad (\text{in-phase waves}). \quad (35-32)$$

If, instead, the waves are to be exactly out of phase so that there is fully destructive interference, the path length $2L$ must cause either no additional phase difference or a phase difference of 1, 2, 3, . . . wavelengths. Only then will the net phase difference be an odd number of half-wavelengths. For a dark film, we must have

$$2L = \text{integer} \times \text{wavelength} \quad (\text{out-of-phase waves}). \quad (35-33)$$

where, again, the wavelength is the wavelength λ_{n2} in the medium containing $2L$. Thus, this time we have

$$2L = \text{integer} \times \lambda_{n2} \quad (\text{out-of-phase waves}). \quad (35-34)$$

Now we can use Eq. 35-6 ($\lambda_n = \lambda/n$) to write the wavelength of the wave of ray r_2 inside the film as

$$\lambda_{n2} = \frac{\lambda}{n_2}, \quad (35-35)$$

where λ is the wavelength of the incident light in vacuum (and approximately also in air). Substituting Eq. 35-35 into Eq. 35-32 and replacing “odd number/2” with $(m + \frac{1}{2})$ give us

$$2L = (m + \frac{1}{2}) \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima—bright film in air}). \quad (35-36)$$

Similarly, with m replacing “integer,” Eq. 35-34 yields

$$2L = m \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{minima—dark film in air}). \quad (35-37)$$

For a given film thickness L , Eqs. 35-36 and 35-37 tell us the wavelengths of light for which the film appears bright and dark, respectively, one wavelength for each value of m . Intermediate wavelengths give intermediate brightnesses. For a given wavelength λ , Eqs. 35-36 and 35-37 tell us the thicknesses of the films that appear bright and dark in that light, respectively, one thickness for each value of m . Intermediate thicknesses give intermediate brightnesses.

Heads Up. (1) For a thin film surrounded by air, Eq. 35-36 corresponds to bright reflections and Eq. 35-37 corresponds to no reflections. For transmissions, the roles of the equations are reversed (after all, if the light is brightly reflected, then it is not transmitted, and vice versa). (2) If we have a different set of values of the indexes of refraction, the roles of the equations may be reversed. For any given set of indexes, you must go through the thought process behind Table 35-1 and, in particular, determine the reflection shifts to see which equation applies to bright reflections and which applies to no reflections. (3) The index of refraction in the equations is that of the thin film, where the path length difference occurs.

Film Thickness Much Less Than λ

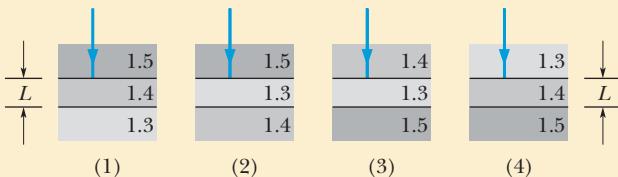
A special situation arises when a film is so thin that L is much less than λ , say, $L < 0.1\lambda$. Then the path length difference $2L$ can be neglected, and the phase difference between r_1 and r_2 is due *only* to reflection phase shifts. If the film of Fig. 35-17, where the reflections cause a phase difference of 0.5 wavelength, has thickness $L < 0.1\lambda$, then r_1 and r_2 are exactly out of phase, and thus the film is dark, regardless of the wavelength and intensity of the light. This special situation corresponds to $m = 0$ in Eq. 35-37. We shall count *any* thickness $L < 0.1\lambda$ as being the least thickness specified by Eq. 35-37 to make the film of Fig. 35-17 dark. (Every such thickness will correspond to $m = 0$.) The next greater thickness that will make the film dark is that corresponding to $m = 1$.

In Fig. 35-18, bright white light illuminates a vertical soap film whose thickness increases from top to bottom. However, the top portion is so thin that it is dark. In the (somewhat thicker) middle we see fringes, or bands, whose color depends primarily on the wavelength at which reflected light undergoes fully constructive interference for a particular thickness. Toward the (thickest) bottom the fringes become progressively narrower and the colors begin to overlap and fade.

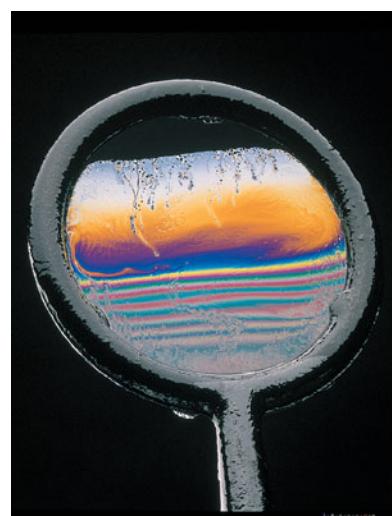


Checkpoint 5

The figure shows four situations in which light reflects perpendicularly from a thin film of thickness L ,



with indexes of refraction as given. (a) For which situations does reflection at the film interfaces cause a zero phase difference for the two reflected rays? (b) For which situations will the film be dark if the path length difference $2L$ causes a phase difference of 0.5 wavelength?



Richard Megna/Fundamental Photographs

Figure 35-18 The reflection of light from a soapy water film spanning a vertical loop. The top portion is so thin (due to gravitational slumping) that the light reflected there undergoes destructive interference, making that portion dark. Colored interference fringes, or bands, decorate the rest of the film but are marred by circulation of liquid within the film as the liquid is gradually pulled downward by gravitation.

Sample Problem 35.05 Thin-film interference of a water film in air

White light, with a uniform intensity across the visible wavelength range of 400 to 690 nm, is perpendicularly incident on a water film, of index of refraction $n_2 = 1.33$ and thickness $L = 320$ nm, that is suspended in air. At what wavelength λ is the light reflected by the film brightest to an observer?

KEY IDEA

The reflected light from the film is brightest at the wavelengths λ for which the reflected rays are in phase with one another. The equation relating these wavelengths λ to the given film thickness L and film index of refraction n_2 is either Eq. 35-36 or Eq. 35-37, depending on the reflection phase shifts for this particular film.

Calculations: To determine which equation is needed, we should fill out an organizing table like Table 35-1. However, because there is air on both sides of the water film, the situation here is exactly like that in Fig. 35-17, and thus the table would be exactly like Table 35-1. Then from Table 35-1, we

see that the reflected rays are in phase (and thus the film is brightest) when

$$2L = \frac{\text{odd number}}{2} \times \frac{\lambda}{n_2},$$

which leads to Eq. 35-36:

$$2L = (m + \frac{1}{2}) \frac{\lambda}{n_2}.$$

Solving for λ and substituting for L and n_2 , we find

$$\lambda = \frac{2n_2 L}{m + \frac{1}{2}} = \frac{(2)(1.33)(320 \text{ nm})}{m + \frac{1}{2}} = \frac{851 \text{ nm}}{m + \frac{1}{2}}.$$

For $m = 0$, this gives us $\lambda = 1700$ nm, which is in the infrared region. For $m = 1$, we find $\lambda = 567$ nm, which is yellow-green light, near the middle of the visible spectrum. For $m = 2$, $\lambda = 340$ nm, which is in the ultraviolet region. Thus, the wavelength at which the light seen by the observer is brightest is

$$\lambda = 567 \text{ nm.}$$

(Answer)



Additional examples, video, and practice available at WileyPLUS

Sample Problem 35.06 Thin-film interference of a coating on a glass lens

In Fig. 35-19, a glass lens is coated on one side with a thin film of magnesium fluoride (MgF_2) to reduce reflection from the lens surface. The index of refraction of MgF_2 is 1.38; that of the glass is 1.50. What is the least coating thickness that eliminates (via interference) the reflections at the middle of the visible spectrum ($\lambda = 550 \text{ nm}$)? Assume that the light is approximately perpendicular to the lens surface.

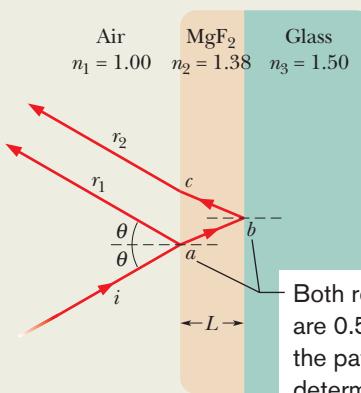
KEY IDEA

Reflection is eliminated if the film thickness L is such that light waves reflected from the two film interfaces are exactly out of phase. The equation relating L to the given wavelength λ and the index of refraction n_2 of the thin film is either Eq. 35-36 or Eq. 35-37, depending on the reflection phase shifts at the interfaces.

Calculations: To determine which equation is needed, we fill out an organizing table like Table 35-1. At the first interface, the incident light is in air, which has a lesser index of refraction than the MgF_2 (the thin film). Thus, we fill in 0.5 wavelength under r_1 in our organizing table (meaning that the waves of ray r_1 are shifted by 0.5λ at the first interface). At the second interface, the incident light is in the MgF_2 , which has a lesser index of refraction than the glass on the other side of the interface. Thus, we fill in 0.5 wavelength under r_2 in our table.

Because both reflections cause the same phase shift, they tend to put the waves of r_1 and r_2 in phase. Since we want those waves to be *out of phase*, their path length difference $2L$ must be an odd number of half-wavelengths:

$$2L = \frac{\text{odd number}}{2} \times \frac{\lambda}{n_2}.$$



Both reflection phase shifts are 0.5 wavelength. So, only the path length difference determines the interference.

Figure 35-19 Unwanted reflections from glass can be suppressed (at a chosen wavelength) by coating the glass with a thin transparent film of magnesium fluoride of the properly chosen thickness.

This leads to Eq. 35-36 (for a bright film sandwiched in air but for a dark film in the arrangement here). Solving that equation for L then gives us the film thicknesses that will eliminate reflection from the lens and coating:

$$L = (m + \frac{1}{2}) \frac{\lambda}{2n_2}, \quad \text{for } m = 0, 1, 2, \dots \quad (35-38)$$

We want the least thickness for the coating—that is, the smallest value of L . Thus, we choose $m = 0$, the smallest possible value of m . Substituting it and the given data in Eq. 35-38, we obtain

$$L = \frac{\lambda}{4n_2} = \frac{550 \text{ nm}}{(4)(1.38)} = 99.6 \text{ nm}. \quad (\text{Answer})$$

Sample Problem 35.07 Thin-film interference of a transparent wedge

Figure 35-20a shows a transparent plastic block with a thin wedge of air at the right. (The wedge thickness is exaggerated in the figure.) A broad beam of red light, with wavelength $\lambda = 632.8 \text{ nm}$, is directed downward through the top of the block (at an incidence angle of 0°). Some of the light that passes into the plastic is reflected back up from the top and bottom surfaces of the wedge, which acts as a thin film (of air) with a thickness that varies uniformly and gradually from L_L at the left-hand end to L_R at the right-hand end. (The plastic layers above and below the wedge of air are too thick to act as thin films.) An observer looking down on the block sees an interference pattern consisting of six dark fringes and five bright red fringes along the wedge. What is the change in thickness $\Delta L (= L_R - L_L)$ along the wedge?

KEY IDEAS

- (1) The brightness at any point along the left-right length of the air wedge is due to the interference of the waves reflected at the top and bottom interfaces of the wedge.
- (2) The variation of brightness in the pattern of bright and dark fringes is due to the variation in the thickness of the wedge. In some regions, the thickness puts the reflected waves in phase and thus produces a bright reflection (a bright red fringe). In other regions, the thickness puts the reflected waves out of phase and thus produces no reflection (a dark fringe).

Organizing the reflections: Because the observer sees more dark fringes than bright fringes, we can assume that a dark fringe is produced at both the left and right ends of

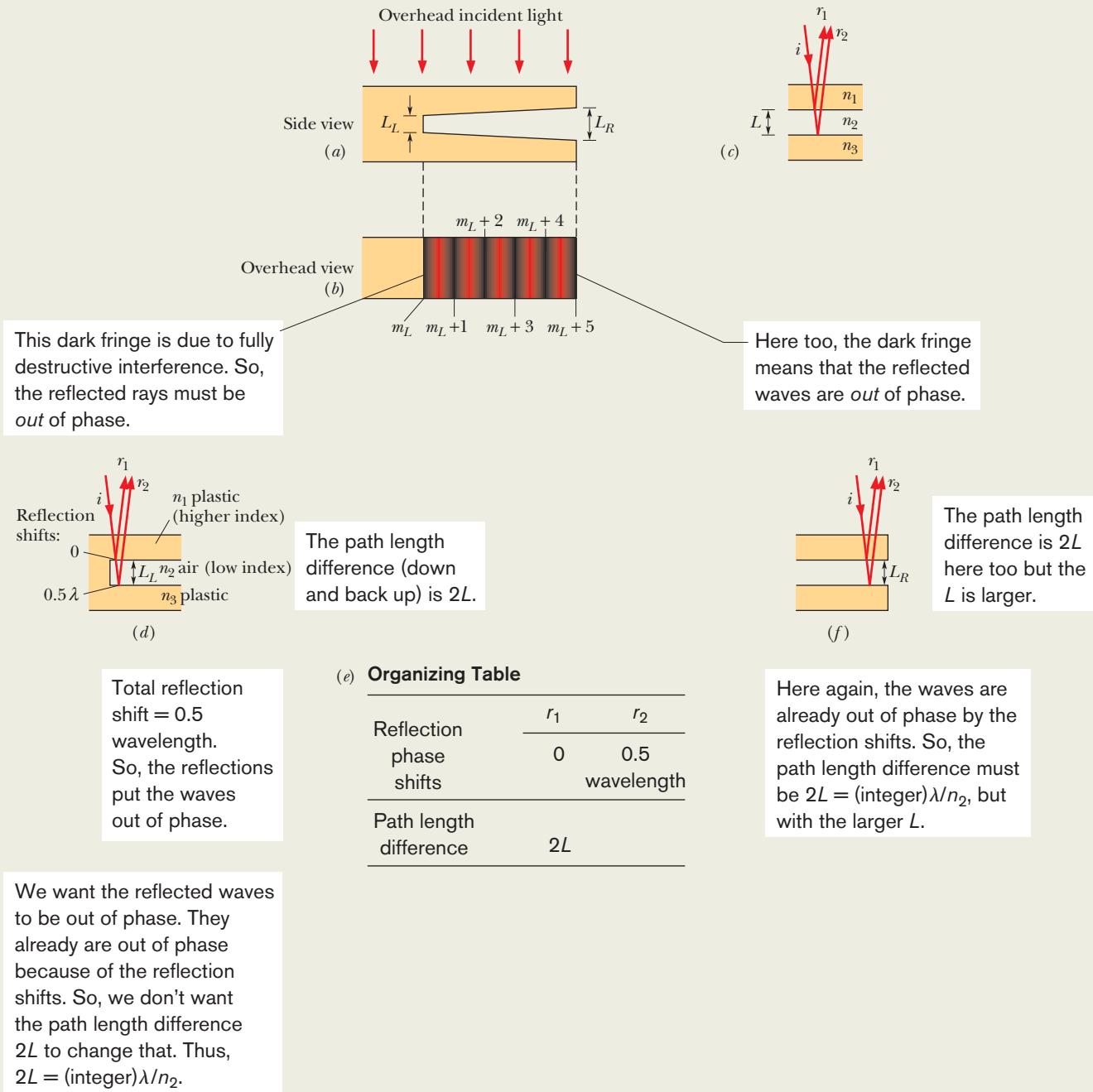


Figure 35-20 (a) Red light is incident on a thin, air-filled wedge in the side of a transparent plastic block. The thickness of the wedge is L_L at the left end and L_R at the right end. (b) The view from above the block: an interference pattern of six dark fringes and five bright red fringes lies over the region of the wedge. (c) A representation of the incident ray i , reflected rays r_1 and r_2 , and thickness L of the wedge anywhere along the length of the wedge. The reflection rays at the (d) left and (f) right ends of the wedge and (e) their organizing table.

the wedge. Thus, the interference pattern is that shown in Fig. 35-20b.

We can represent the reflection of light at the top and bottom interfaces of the wedge, at any point along its length, with Fig. 35-20c, in which L is the wedge thickness at that point. Let us apply this figure to the left end of the wedge, where the reflections give a dark fringe.

We know that, for a dark fringe, the waves of rays r_1 and r_2 in Fig. 35-20d must be out of phase. We also know that the equation relating the film thickness L to the light's wavelength λ and the film's index of refraction n_2 is either Eq. 35-36 or Eq. 35-37, depending on the reflection phase shifts. To determine which equation gives a dark fringe at the left end of the wedge, we should fill out an organizing table like Table 35-1, as shown in Fig. 35-20e.

At the top interface of the wedge, the incident light is in the plastic, which has a greater n than the air beneath that interface. So, we fill in 0 under r_1 in our organizing table. At the bottom interface of the wedge, the incident light is in air, which has a lesser n than the plastic beneath that interface. So we fill in 0.5 wavelength under r_2 . So, the phase difference due to the reflection shifts is 0.5 wavelength. Thus the reflections alone tend to put the waves of r_1 and r_2 out of phase.

Reflections at left end (Fig. 35-20d): Because we see a dark fringe at the left end of the wedge, which the reflection phase shifts alone would produce, we don't want the path length difference to alter that condition. So, the path length difference $2L$ at the left end must be given by

$$2L = \text{integer} \times \frac{\lambda}{n_2},$$



Additional examples, video, and practice available at WileyPLUS

which leads to Eq. 35-37:

$$2L = m \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots \quad (35-39)$$

Reflections at right end (Fig. 35-20f): Equation 35-39 holds not only for the left end of the wedge but also for any point along the wedge where a dark fringe is observed, including the right end, with a different integer value of m for each fringe. The least value of m is associated with the least thickness of the wedge where a dark fringe is observed. Progressively greater values of m are associated with progressively greater thicknesses of the wedge where a dark fringe is observed. Let m_L be the value at the left end. Then the value at the right end must be $m_L + 5$ because, from Fig. 35-20b, the right end is located at the fifth dark fringe from the left end.

Thickness difference: To find ΔL , we first solve Eq. 35-39 twice—once for the thickness L_L at the left end and once for the thickness L_R at the right end:

$$L_L = (m_L) \frac{\lambda}{2n_2}, \quad L_R = (m_L + 5) \frac{\lambda}{2n_2}. \quad (35-40)$$

We can now subtract L_L from L_R and substitute $n_2 = 1.00$ for the air within the wedge and $\lambda = 632.8 \times 10^{-9}$ m:

$$\begin{aligned} \Delta L &= L_R - L_L = \frac{(m_L + 5)\lambda}{2n_2} - \frac{m_L\lambda}{2n_2} = \frac{5}{2} \frac{\lambda}{n_2} \\ &= 1.58 \times 10^{-6} \text{ m.} \end{aligned} \quad (\text{Answer})$$

35-5 MICHELSON'S INTERFEROMETER

Learning Objectives

After reading this module, you should be able to . . .

35.35 With a sketch, explain how an interferometer works.

35.36 When a transparent material is inserted into one of the beams in an interferometer, apply the relationship between the phase change of the light (in terms of

Key Ideas

- In Michelson's interferometer, a light wave is split into two beams that then recombine after traveling along different paths.

- The interference pattern they produce depends on the difference in the lengths of those paths and the indexes of refraction along the paths.

wavelength) and the material's thickness and index of refraction.

35.37 For an interferometer, apply the relationship between the distance a mirror is moved and the resulting fringe shift in the interference pattern.

- If a transparent material of index n and thickness L is in one path, the phase difference (in terms of wavelength) in the recombining beams is equal to

$$\text{phase difference} = \frac{2L}{\lambda} (n - 1),$$

where λ is the wavelength of the light.

Michelson's Interferometer

An **interferometer** is a device that can be used to measure lengths or changes in length with great accuracy by means of interference fringes. We describe the form originally devised and built by A. A. Michelson in 1881.

Consider light that leaves point P on extended source S in Fig. 35-21 and encounters *beam splitter* M . A beam splitter is a mirror that transmits half the incident light and reflects the other half. In the figure we have assumed, for convenience, that this mirror possesses negligible thickness. At M the light thus divides into two waves. One proceeds by transmission toward mirror M_1 at the end of one arm of the instrument; the other proceeds by reflection toward mirror M_2 at the end of the other arm. The waves are entirely reflected at these mirrors and are sent back along their directions of incidence, each wave eventually entering telescope T . What the observer sees is a pattern of curved or approximately straight interference fringes; in the latter case the fringes resemble the stripes on a zebra.

Mirror Shift. The path length difference for the two waves when they recombine at the telescope is $2d_2 - 2d_1$, and anything that changes this path length difference will cause a change in the phase difference between these two waves at the eye. As an example, if mirror M_2 is moved by a distance $\frac{1}{2}\lambda$, the path length difference is changed by λ and the fringe pattern is shifted by one fringe (as if each dark stripe on a zebra had moved to where the adjacent dark stripe had been). Similarly, moving mirror M_2 by $\frac{1}{4}\lambda$ causes a shift by half a fringe (each dark zebra stripe shifts to where the adjacent white stripe had been).

Insertion. A shift in the fringe pattern can also be caused by the insertion of a thin transparent material into the optical path of one of the mirrors—say, M_1 . If the material has thickness L and index of refraction n , then the number of wavelengths along the light's to-and-fro path through the material is, from Eq. 35-7,

$$N_m = \frac{2L}{\lambda_n} = \frac{2Ln}{\lambda}. \quad (35-41)$$

The number of wavelengths in the same thickness $2L$ of air before the insertion of the material is

$$N_a = \frac{2L}{\lambda}. \quad (35-42)$$

When the material is inserted, the light returned from mirror M_1 undergoes a phase change (in terms of wavelengths) of

$$N_m - N_a = \frac{2Ln}{\lambda} - \frac{2L}{\lambda} = \frac{2L}{\lambda}(n - 1). \quad (35-43)$$

For each phase change of one wavelength, the fringe pattern is shifted by one fringe. Thus, by counting the number of fringes through which the material causes the pattern to shift, and substituting that number for $N_m - N_a$ in Eq. 35-43, you can determine the thickness L of the material in terms of λ .

Standard of Length. By such techniques the lengths of objects can be expressed in terms of the wavelengths of light. In Michelson's day, the standard of length—the meter—was the distance between two fine scratches on a certain metal bar preserved at Sèvres, near Paris. Michelson showed, using his interferometer, that the standard meter was equivalent to 1 553 163.5 wavelengths of a certain monochromatic red light emitted from a light source containing cadmium. For this careful measurement, Michelson received the 1907 Nobel Prize in physics. His work laid the foundation for the eventual abandonment (in 1961) of the meter bar as a standard of length and for the redefinition of the meter in terms of the wavelength of light. By 1983, even this wavelength standard was not precise enough to meet the growing technical needs, and it was replaced with a new standard based on a defined value for the speed of light.

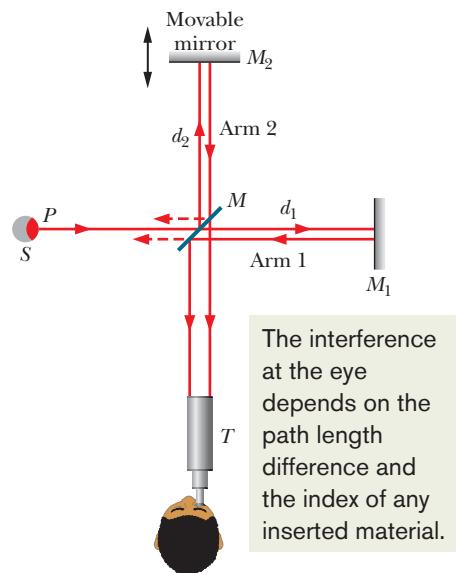


Figure 35-21 Michelson's interferometer, showing the path of light originating at point P of an extended source S . Mirror M splits the light into two beams, which reflect from mirrors M_1 and M_2 back to M and then to telescope T . In the telescope an observer sees a pattern of interference fringes.

Review & Summary

Huygens' Principle The three-dimensional transmission of waves, including light, may often be predicted by *Huygens' principle*, which states that all points on a wavefront serve as point sources of spherical secondary wavelets. After a time t , the new position of the wavefront will be that of a surface tangent to these secondary wavelets.

The law of refraction can be derived from Huygens' principle by assuming that the index of refraction of any medium is $n = c/v$, in which v is the speed of light in the medium and c is the speed of light in vacuum.

Wavelength and Index of Refraction The wavelength λ_n of light in a medium depends on the index of refraction n of the medium:

$$\lambda_n = \frac{\lambda}{n}, \quad (35-6)$$

in which λ is the wavelength in vacuum. Because of this dependency, the phase difference between two waves can change if they pass through different materials with different indexes of refraction.

Young's Experiment In **Young's interference experiment**, light passing through a single slit falls on two slits in a screen. The light leaving these slits flares out (by diffraction), and interference occurs in the region beyond the screen. A fringe pattern, due to the interference, forms on a viewing screen.

The light intensity at any point on the viewing screen depends in part on the difference in the path lengths from the slits to that point. If this difference is an integer number of wavelengths, the waves interfere constructively and an intensity maximum results. If it is an odd number of half-wavelengths, there is destructive interference and an intensity minimum occurs. The conditions for maximum and minimum intensity are

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima—bright fringes}), \quad (35-14)$$

$$d \sin \theta = (m + \frac{1}{2})\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{minima—dark fringes}), \quad (35-16)$$

where θ is the angle the light path makes with a central axis and d is the slit separation.

Coherence If two light waves that meet at a point are to interfere perceptibly, the phase difference between them must remain constant with time; that is, the waves must be **coherent**. When two coherent waves meet, the resulting intensity may be found by using phasors.

Intensity in Two-Slit Interference In Young's interference experiment, two waves, each with intensity I_0 , yield a resultant wave of intensity I at the viewing screen, with

$$I = 4I_0 \cos^2 \frac{1}{2}\phi, \quad \text{where } \phi = \frac{2\pi d}{\lambda} \sin \theta. \quad (35-22, 35-23)$$

Equations 35-14 and 35-16, which identify the positions of the fringe maxima and minima, are contained within this relation.

Thin-Film Interference When light is incident on a thin transparent film, the light waves reflected from the front and back surfaces interfere. For near-normal incidence, the wavelength conditions for maximum and minimum intensity of the light reflected from a *film in air* are

$$2L = (m + \frac{1}{2}) \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima—bright film in air}), \quad (35-36)$$

$$2L = m \frac{\lambda}{n_2}, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{minima—dark film in air}), \quad (35-37)$$

where n_2 is the index of refraction of the film, L is its thickness, and λ is the wavelength of the light in air.

If the light incident at an interface between media with different indexes of refraction is in the medium with the smaller index of refraction, the reflection causes a phase change of π rad, or half a wavelength, in the reflected wave. Otherwise, there is no phase change due to the reflection. Refraction causes no phase shift.

The Michelson Interferometer In *Michelson's interferometer* a light wave is split into two beams that, after traversing paths of different lengths, are recombined so they interfere and form a fringe pattern. Varying the path length of one of the beams allows distances to be accurately expressed in terms of wavelengths of light, by counting the number of fringes through which the pattern shifts because of the change.

Questions

1 Does the spacing between fringes in a two-slit interference pattern increase, decrease, or stay the same if (a) the slit separation is increased, (b) the color of the light is switched from red to blue, and (c) the whole apparatus is submerged in cooking sherry? (d) If the slits are illuminated with white light, then at any side maximum, does the blue component or the red component peak closer to the central maximum?

2 (a) If you move from one bright fringe in a two-slit interference pattern to the next one farther out, (b) does the path length difference ΔL increase or decrease and (c) by how much does it change, in wavelengths λ ?

3 Figure 35-22 shows two light rays that are initially exactly in phase and that reflect from several glass surfaces. Neglect the

slight slant in the path of the light in the second arrangement. (a) What is the path length difference of the rays? In wavelengths λ , (b) what should that path length difference equal if the rays are to be exactly out of phase when they emerge, and (c) what is the smallest value of d that will allow that final phase difference?

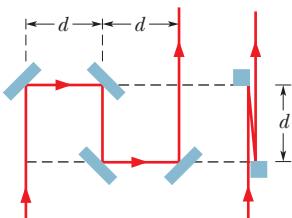


Figure 35-22 Question 3.

4 In Fig. 35-23, three pulses of light—*a*, *b*, and *c*—of the same wavelength are sent through layers

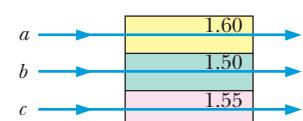


Figure 35-23 Question 4.

of plastic having the given indexes of refraction and along the paths indicated. Rank the pulses according to their travel time through the plastic layers, greatest first.

5 Is there an interference maximum, a minimum, an intermediate state closer to a maximum, or an intermediate state closer to a minimum at point *P* in Fig. 35-10 if the path length difference of the two rays is (a) 2.2λ , (b) 3.5λ , (c) 1.8λ , and (d) 1.0λ ? For each situation, give the value of *m* associated with the maximum or minimum involved.

6 Figure 35-24a gives intensity *I* versus position *x* on the viewing screen for the central portion of a two-slit interference pattern. The other parts of the figure give phasor diagrams for the electric field components of the waves arriving at the screen from the two slits (as in Fig. 35-13a). Which numbered points on the screen best correspond to which phasor diagram?

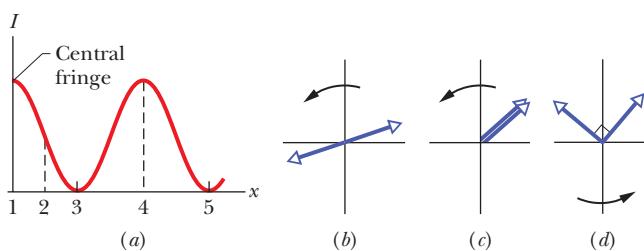


Figure 35-24 Question 6.

7 Figure 35-25 shows two sources S_1 and S_2 that emit radio waves of wavelength λ in all directions. The sources are exactly in phase and are separated by a distance equal to 1.5λ . The vertical broken line is the perpendicular bisector of the distance between the sources. (a) If we start at the indicated start point and travel along path 1, does the interference produce a maximum all along the path, a minimum all along the path, or alternating maxima and minima? Repeat for (b) path 2 (along an axis through the sources) and (c) path 3 (along a perpendicular to that axis).

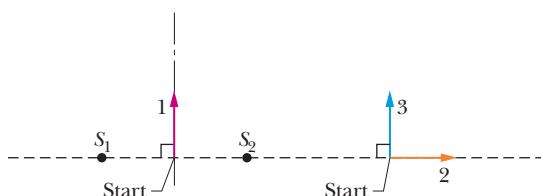


Figure 35-25 Question 7.

8 Figure 35-26 shows two rays of light, of wavelength 600 nm, that reflect from glass surfaces separated by 150 nm. The rays are initially in phase. (a) What is the path length difference of the rays? (b) When they have cleared the reflection region, are the rays exactly in phase, exactly out of phase, or in some intermediate state?

9 Light travels along the length of a 1500-nm-long nanostructure. When a peak of the wave is at one end of the nanostructure, is there a peak or a valley at the other end if the wavelength is (a) 500 nm and (b) 1000 nm?

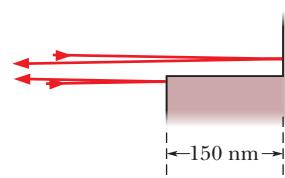


Figure 35-26 Question 8.

10 Figure 35-27a shows the cross section of a vertical thin film whose width increases downward because gravitation causes slumping. Figure 35-27b is a face-on view of the film, showing four bright (red) interference fringes that result when the film is illuminated with a perpendicular beam of red light. Points in the cross section corresponding to the bright fringes are labeled. In terms of the wavelength of the light inside the film, what is the difference in film thickness between (a) points *a* and *b* and (b) points *b* and *d*?

11 Figure 35-28 shows four situations in which light reflects perpendicularly from a thin film of thickness *L* sandwiched between much thicker materials. The indexes of refraction are given. In which situations does Eq. 35-36 correspond to the reflections yielding maxima (that is, a bright film)?

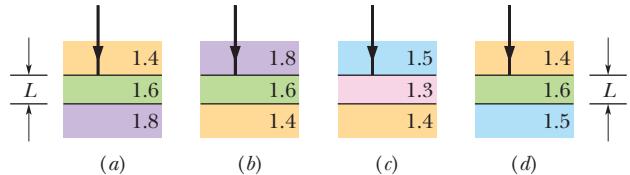


Figure 35-28 Question 11.

12 Figure 35-29 shows the transmission of light through a thin film in air by a perpendicular beam (tilted in the figure for clarity). (a) Did ray r_3 undergo a phase shift due to reflection? (b) In wavelengths, what is the reflection phase shift for ray r_4 ? (c) If the film thickness is *L*, what is the path length difference between rays r_3 and r_4 ?

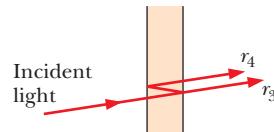


Figure 35-29 Question 12.

13 Figure 35-30 shows three situations in which two rays of sunlight penetrate slightly into and then scatter out of lunar soil. Assume that the rays are initially in phase. In which situation are the associated waves most likely to end up in phase? (Just as the Moon becomes full, its brightness suddenly peaks, becoming 25% greater than its brightness on the nights before and after, because at full Moon we intercept light waves that are scattered by lunar soil back toward the Sun and undergo constructive interference at our eyes. Before astronauts first landed on the Moon, NASA was concerned that backscatter of sunlight from the soil might blind the lunar astronauts if they did not have proper viewing shields on their helmets.)

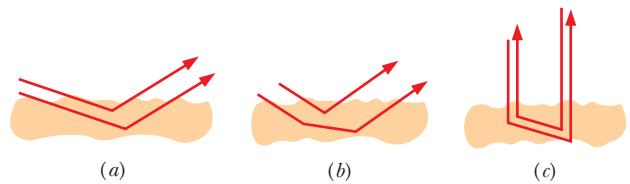


Figure 35-30 Question 13.

Problems



Tutoring problem available (at instructor's discretion) in WileyPLUS and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com

WWW Worked-out solution is at

ILW Interactive solution is at

<http://www.wiley.com/college/halliday>

Module 35-1 Light as a Wave

- 1 In Fig. 35-31, a light wave along ray r_1 reflects once from a mirror and a light wave along ray r_2 reflects twice from that same mirror and once from a tiny mirror at distance L from the bigger mirror. (Neglect the slight tilt of the rays.) The waves have wavelength 620 nm and are initially in phase. (a) What is the smallest value of L that puts the final light waves exactly out of phase? (b) With the tiny mirror initially at that value of L , how far must it be moved away from the bigger mirror to again put the final waves out of phase?

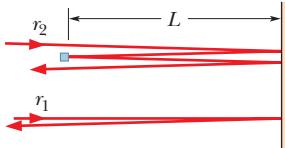


Figure 35-31 Problems 1 and 2.

- 2 In Fig. 35-31, a light wave along ray r_1 reflects once from a mirror and a light wave along ray r_2 reflects twice from that same mirror and once from a tiny mirror at distance L from the bigger mirror. (Neglect the slight tilt of the rays.) The waves have wavelength λ and are initially exactly out of phase. What are the (a) smallest, (b) second smallest, and (c) third smallest values of L/λ that result in the final waves being exactly in phase?

- 3 **SSM** In Fig. 35-4, assume that two waves of light in air, of wavelength 400 nm, are initially in phase. One travels through a glass layer of index of refraction $n_1 = 1.60$ and thickness L . The other travels through an equally thick plastic layer of index of refraction $n_2 = 1.50$. (a) What is the smallest value L should have if the waves are to end up with a phase difference of 5.65 rad? (b) If the waves arrive at some common point with the same amplitude, is their interference fully constructive, fully destructive, intermediate but closer to fully constructive, or intermediate but closer to fully destructive?

- 4 In Fig. 35-32a, a beam of light in material 1 is incident on a boundary at an angle of 30° . The extent to which the light is bent due to refraction depends, in part, on the index of refraction n_2 of material 2. Figure 35-32b gives the angle of refraction θ_2 versus n_2 for a range of possible n_2 values, from $n_a = 1.30$ to $n_b = 1.90$. What is the speed of light in material 1?

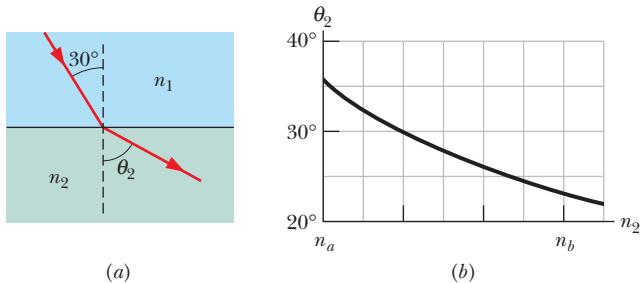


Figure 35-32 Problem 4.

- 5 How much faster, in meters per second, does light travel in sapphire than in diamond? See Table 33-1.

- 6 The wavelength of yellow sodium light in air is 589 nm. (a) What is its frequency? (b) What is its wavelength in glass whose index of refraction is 1.52? (c) From the results of (a) and (b), find its speed in this glass.

- 7 The speed of yellow light (from a sodium lamp) in a certain liquid is measured to be 1.92×10^8 m/s. What is the index of refraction of this liquid for the light?

- 8 In Fig. 35-33, two light pulses are sent through layers of plastic with thicknesses of either L or $2L$ as shown and indexes of refraction $n_1 = 1.55$, $n_2 = 1.70$, $n_3 = 1.60$, $n_4 = 1.45$, $n_5 = 1.59$, $n_6 = 1.65$, and $n_7 = 1.50$. (a) Which pulse travels through the plastic in less time? (b) What multiple of L/c gives the difference in the traversal times of the pulses?

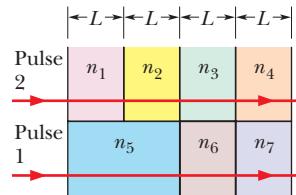


Figure 35-33 Problem 8.

- 9 In Fig. 35-4, assume that the two light waves, of wavelength 620 nm in air, are initially out of phase by π rad. The indexes of refraction of the media are $n_1 = 1.45$ and $n_2 = 1.65$. What are the (a) smallest and (b) second smallest value of L that will put the waves exactly in phase once they pass through the two media?

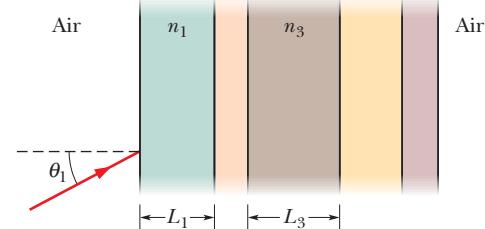


Figure 35-34 Problem 10.

- 10 In Fig. 35-34, a light ray is incident at angle $\theta_1 = 50^\circ$ on a series of five transparent layers with parallel boundaries. For layers 1 and 3, $L_1 = 20 \mu\text{m}$, $L_3 = 25 \mu\text{m}$, $n_1 = 1.6$, and $n_3 = 1.45$. (a) At

what angle does the light emerge back into air at the right? (b) How much time does the light take to travel through layer 3?

- 11 Suppose that the two waves in Fig. 35-4 have wavelength $\lambda = 500$ nm in air. What multiple of λ gives their phase difference when they emerge if (a) $n_1 = 1.50$, $n_2 = 1.60$, and $L = 8.50 \mu\text{m}$; (b) $n_1 = 1.62$, $n_2 = 1.72$, and $L = 8.50 \mu\text{m}$; and (c) $n_1 = 1.59$, $n_2 = 1.79$, and $L = 3.25 \mu\text{m}$? (d) Suppose that in each of these three situations the waves arrive at a common point (with the same amplitude) after emerging. Rank the situations according to the brightness the waves produce at the common point.

- 12 In Fig. 35-35, two light rays go through different paths by reflecting from the various flat surfaces shown. The light waves have a wavelength of 420.0 nm and are initially in phase. What are the (a) smallest and (b) second smallest value of distance L that will put the waves exactly out of phase as they emerge from the region?

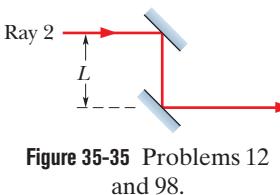
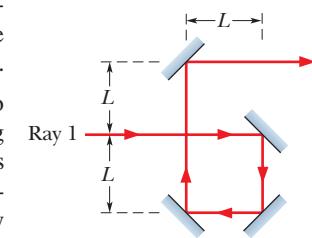


Figure 35-35 Problems 12 and 98.

both travel through a layer of plastic as shown in Fig. 35-36, with $L_1 = 4.00 \mu\text{m}$, $L_2 = 3.50 \mu\text{m}$, $n_1 = 1.40$, and $n_2 = 1.60$. (a) What multiple of λ gives their phase difference after they both have emerged from the layers? (b) If the waves later arrive at some common point with the same amplitude, is their interference fully constructive, fully destructive, intermediate but closer to fully constructive, or intermediate but closer to fully destructive?

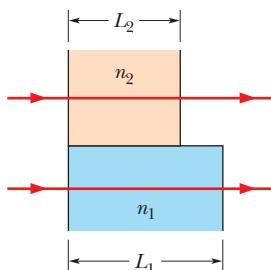


Figure 35-36 Problem 13.

Module 35-2 Young's Interference Experiment

•14 In a double-slit arrangement the slits are separated by a distance equal to 100 times the wavelength of the light passing through the slits. (a) What is the angular separation in radians between the central maximum and an adjacent maximum? (b) What is the distance between these maxima on a screen 50.0 cm from the slits?

•15 SSM A double-slit arrangement produces interference fringes for sodium light ($\lambda = 589 \text{ nm}$) that have an angular separation of $3.50 \times 10^{-3} \text{ rad}$. For what wavelength would the angular separation be 10.0% greater?

•16 A double-slit arrangement produces interference fringes for sodium light ($\lambda = 589 \text{ nm}$) that are 0.20° apart. What is the angular separation if the arrangement is immersed in water ($n = 1.33$)?

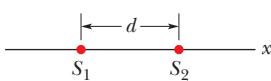


Figure 35-37 Problems 17 and 22.

•17 GO SSM In Fig. 35-37, two radio-frequency point sources S_1 and S_2 , separated by distance $d = 2.0 \text{ m}$, are radiating in phase with $\lambda = 0.50 \text{ m}$. A detector moves in a large circular path around the two sources in a plane containing them. How many maxima does it detect?

•18 In the two-slit experiment of Fig. 35-10, let angle θ be 20.0° , the slit separation be $4.24 \mu\text{m}$, and the wavelength be $\lambda = 500 \text{ nm}$. (a) What multiple of λ gives the phase difference between the waves of rays r_1 and r_2 when they arrive at point P on the distant screen? (b) What is the phase difference in radians? (c) Determine where in the interference pattern point P lies by giving the maximum or minimum on which it lies, or the maximum and minimum between which it lies.

•19 SSM ILW Suppose that Young's experiment is performed with blue-green light of wavelength 500 nm. The slits are 1.20 mm apart, and the viewing screen is 5.40 m from the slits. How far apart are the bright fringes near the center of the interference pattern?

•20 Monochromatic green light, of wavelength 550 nm, illuminates two parallel narrow slits $7.70 \mu\text{m}$ apart. Calculate the angular deviation (θ in Fig. 35-10) of the third-order ($m = 3$) bright fringe (a) in radians and (b) in degrees.

•21 In a double-slit experiment, the distance between slits is 5.0 mm and the slits are 1.0 m from the screen. Two interference patterns can be seen on the screen: one due to light of wavelength 480 nm, and the other due to light of wavelength 600 nm. What is the separation on the screen between the third-order ($m = 3$) bright fringes of the two interference patterns?

•22 In Fig. 35-37, two isotropic point sources S_1 and S_2 emit identical light waves in phase at wavelength λ . The sources lie at separation d on an x axis, and a light detector is moved in a circle of large radius around the midpoint between them. It detects 30 points of zero intensity, including two on the x axis, one of them to the left of the sources and the other to the right of the sources. What is the value of d/λ ?

•23 GO In Fig. 35-38, sources A and B emit long-range radio waves of wavelength 400 m, with the phase of the emission from A ahead of that from source B by 90° . The distance r_A from A to detector D is greater than the corresponding distance r_B by 100 m. What is the phase difference of the waves at D ?

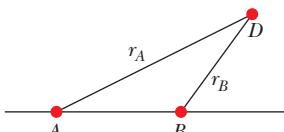


Figure 35-38 Problem 23.

•24 In Fig. 35-39, two isotropic point sources S_1 and S_2 emit light in phase at wavelength λ and at the same amplitude. The sources are separated by distance $2d = 6.00\lambda$. They lie on an axis that is parallel to an x axis, which runs along a viewing screen at distance $D = 20.0\lambda$. The origin lies on the perpendicular bisector between the sources. The figure shows two rays reaching point P on the screen, at position x_P . (a) At what value of x_P do the rays have the minimum possible phase difference? (b) What multiple of λ gives that minimum phase difference? (c) At what value of x_P do the rays have the maximum possible phase difference? What multiple of λ gives (d) that maximum phase difference and (e) the phase difference when $x_P = 6.00\lambda$? (f) When $x_P = 6.00\lambda$, is the resulting intensity at point P maximum, minimum, intermediate but closer to maximum, or intermediate but closer to minimum?

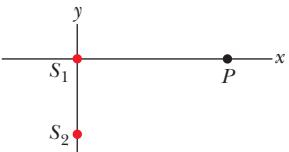


Figure 35-39 Problem 24.

•25 GO In Fig. 35-40, two isotropic point sources of light (S_1 and S_2) are separated by distance $2.70 \mu\text{m}$ along a y axis and emit in phase at wavelength 900 nm and at the same amplitude. A light detector is located at point P at coordinate x_P on the x axis. What is the greatest value of x_P at which the detected light is minimum due to destructive interference?

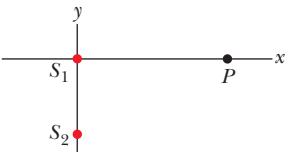


Figure 35-40 Problems 25 and 28.

•26 In a double-slit experiment, the fourth-order maximum for a wavelength of 450 nm occurs at an angle of $\theta = 90^\circ$. (a) What range of wavelengths in the visible range (400 nm to 700 nm) are not present in the third-order maxima? To eliminate all visible light in the fourth-order maximum, (b) should the slit separation be increased or decreased and (c) what least change is needed?

•27 A thin flake of mica ($n = 1.58$) is used to cover one slit of a double-slit interference arrangement. The central point on the viewing screen is now occupied by what had been the seventh bright side fringe ($m = 7$). If $\lambda = 550 \text{ nm}$, what is the thickness of the mica?

•28 GO Figure 35-40 shows two isotropic point sources of light (S_1 and S_2) that emit in phase at wavelength 400 nm and at the same amplitude. A detection point P is shown on an x axis that extends through source S_1 . The phase difference ϕ be-

tween the light arriving at point P from the two sources is to be measured as P is moved along the x axis from $x = 0$ out to $x = +\infty$. The results out to $x_s = 10 \times 10^{-7} \text{ m}$ are given in Fig. 35-41. On the way out to

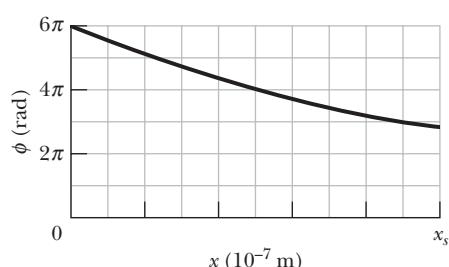


Figure 35-41 Problem 28.

$+\infty$, what is the greatest value of x at which the light arriving at P from S_1 is exactly out of phase with the light arriving at P from S_2 ?

Module 35-3 Interference and Double-Slit Intensity

•30 **SSM** Two waves of the same frequency have amplitudes 1.00 and 2.00. They interfere at a point where their phase difference is 60.0° . What is the resultant amplitude?

•30 Find the sum y of the following quantities:

$$y_1 = 10 \sin \omega t \quad \text{and} \quad y_2 = 8.0 \sin(\omega t + 30^\circ).$$

•31 **ILW** Add the quantities $y_1 = 10 \sin \omega t$, $y_2 = 15 \sin(\omega t + 30^\circ)$, and $y_3 = 5.0 \sin(\omega t - 45^\circ)$ using the phasor method.

•32 **GO** In the double-slit experiment of Fig. 35-10, the electric fields of the waves arriving at point P are given by

$$E_1 = (2.00 \mu\text{V/m}) \sin[(1.26 \times 10^{15})t]$$

$$E_2 = (2.00 \mu\text{V/m}) \sin[(1.26 \times 10^{15})t + 39.6 \text{ rad}],$$

where time t is in seconds. (a) What is the amplitude of the resultant electric field at point P ? (b) What is the ratio of the intensity I_P at point P to the intensity I_{cen} at the center of the interference pattern? (c) Describe where point P is in the interference pattern by giving the maximum or minimum on which it lies, or the maximum and minimum between which it lies. In a phasor diagram of the electric fields, (d) at what rate would the phasors rotate around the origin and (e) what is the angle between the phasors?

•33 **GO** Three electromagnetic waves travel through a certain point P along an x axis. They are polarized parallel to a y axis, with the following variations in their amplitudes. Find their resultant at P .

$$E_1 = (10.0 \mu\text{V/m}) \sin[(2.0 \times 10^{14} \text{ rad/s})t]$$

$$E_2 = (5.00 \mu\text{V/m}) \sin[(2.0 \times 10^{14} \text{ rad/s})t + 45.0^\circ]$$

$$E_3 = (5.00 \mu\text{V/m}) \sin[(2.0 \times 10^{14} \text{ rad/s})t - 45.0^\circ]$$

•34 In the double-slit experiment of Fig. 35-10, the viewing screen is at distance $D = 4.00 \text{ m}$, point P lies at distance $y = 20.5 \text{ cm}$ from the center of the pattern, the slit separation d is $4.50 \mu\text{m}$, and the wavelength λ is 580 nm . (a) Determine where point P is in the interference pattern by giving the maximum or minimum on which it lies, or the maximum and minimum between which it lies. (b) What is the ratio of the intensity I_P at point P to the intensity I_{cen} at the center of the pattern?

Module 35-4 Interference from Thin Films

•35 **SSM** We wish to coat flat glass ($n = 1.50$) with a transparent material ($n = 1.25$) so that reflection of light at wavelength 600 nm is eliminated by interference. What minimum thickness can the coating have to do this?

•36 A 600-nm -thick soap film ($n = 1.40$) in air is illuminated with white light in a direction perpendicular to the film. For how many different wavelengths in the 300 to 700 nm range is there (a) fully constructive interference and (b) fully destructive interference in the reflected light?

•37 The rhinestones in costume jewelry are glass with index of refraction 1.50. To make them more reflective, they are often coated

with a layer of silicon monoxide of index of refraction 2.00. What is the minimum coating thickness needed to ensure that light of wavelength 560 nm and of perpendicular incidence will be reflected from the two surfaces of the coating with fully constructive interference?

•38 White light is sent downward onto a horizontal thin film that is sandwiched between two materials. The indexes of refraction are 1.80 for the top material, 1.70 for the thin film, and 1.50 for the bottom material. The film thickness is $5.00 \times 10^{-7} \text{ m}$. Of the visible wavelengths (400 to 700 nm) that result in fully constructive interference at an observer above the film, which is the (a) longer and (b) shorter wavelength? The materials and film are then heated so that the film thickness increases. (c) Does the light resulting in fully constructive interference shift toward longer or shorter wavelengths?

•39 **ILW** Light of wavelength 624 nm is incident perpendicularly on a soap film ($n = 1.33$) suspended in air. What are the (a) least and (b) second least thicknesses of the film for which the reflections from the film undergo fully constructive interference?

•40 A thin film of acetone ($n = 1.25$) coats a thick glass plate ($n = 1.50$). White light is incident normal to the film. In the reflections, fully destructive interference occurs at 600 nm and fully constructive interference at 700 nm . Calculate the thickness of the acetone film.

•41 through 52 **GO** 43, 51 **SSM** 47, 51

Reflection by thin layers. In Fig. 35-42, light is incident perpendicularly on a thin layer of material 2 that lies between (thicker) materials 1 and 3. (The rays are tilted only for clarity.) The waves of rays r_1 and r_2 interfere, and here we consider the type of interference to be either maximum (max) or minimum (min). For this situation, each problem in Table 35-

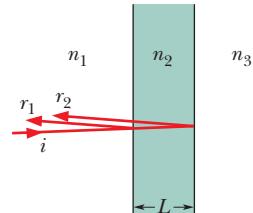


Figure 35-42 Problems 41 through 52.

2 refers to the indexes of refraction n_1 , n_2 , and n_3 , the type of interference, the thin-layer thickness L in nanometers, and the wavelength λ in nanometers of the light as measured in air. Where λ is missing, give the wavelength that is in the visible range. Where L is missing, give the second least thickness or the third least thickness as indicated.

Table 35-2 Problems 41 through 52: Reflection by Thin Layers. See the setup for these problems.

	n_1	n_2	n_3	Type	L	λ
41	1.68	1.59	1.50	min	2nd	342
42	1.55	1.60	1.33	max	285	
43	1.60	1.40	1.80	min	200	
44	1.50	1.34	1.42	max	2nd	587
45	1.55	1.60	1.33	max	3rd	612
46	1.68	1.59	1.50	min	415	
47	1.50	1.34	1.42	min	380	
48	1.60	1.40	1.80	max	2nd	632
49	1.32	1.75	1.39	max	3rd	382
50	1.40	1.46	1.75	min	2nd	482
51	1.40	1.46	1.75	min	210	
52	1.32	1.75	1.39	max	325	

••53 The reflection of perpendicularly incident white light by a soap film in air has an interference maximum at 600 nm and a minimum at 450 nm, with no minimum in between. If $n = 1.33$ for the film, what is the film thickness, assumed uniform?

••54 A plane wave of monochromatic light is incident normally on a uniform thin film of oil that covers a glass plate. The wavelength of the source can be varied continuously. Fully destructive interference of the reflected light is observed for wavelengths of 500 and 700 nm and for no wavelengths in between. If the index of refraction of the oil is 1.30 and that of the glass is 1.50, find the thickness of the oil film.

••55 SSM WWW A disabled tanker leaks kerosene ($n = 1.20$) into the Persian Gulf, creating a large slick on top of the water ($n = 1.30$). (a) If you are looking straight down from an airplane, while the Sun is overhead, at a region of the slick where its thickness is 460 nm, for which wavelength(s) of visible light is the reflection brightest because of constructive interference? (b) If you are scuba diving directly under this same region of the slick, for which wavelength(s) of visible light is the transmitted intensity strongest?

••56 A thin film, with a thickness of 272.7 nm and with air on both sides, is illuminated with a beam of white light. The beam is perpendicular to the film and consists of the full range of wavelengths for the visible spectrum. In the light reflected by the film, light with a wavelength of 600.0 nm undergoes fully constructive interference. At what wavelength does the reflected light undergo fully destructive interference? (Hint: You must make a reasonable assumption about the index of refraction.)

••57 through 68 GO 64, 65 **SSM** 59
Transmission through thin layers. In Fig. 35-43, light is incident perpendicularly on a thin layer of material 2 that lies between (thicker) materials 1 and 3. (The rays are tilted only for clarity.) Part of the light ends up in material 3 as ray r_3 (the light does not reflect inside material 2) and r_4 (the light reflects twice inside material 2). The waves of r_3 and r_4 interfere, and here we consider the type of interference to be either maximum (max) or minimum (min). For this situation, each problem in Table 35-3 refers to the indexes of refraction n_1 , n_2 , and n_3 , the type

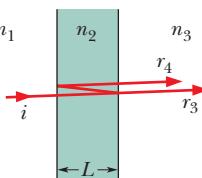


Figure 35-43
Problems 57 through 68.

of interference, the thin-layer thickness L in nanometers, and the wavelength λ in nanometers of the light as measured in air. Where λ is missing, give the wavelength that is in the visible range. Where L is missing, give the second least thickness or the third least thickness as indicated.

••69 GO In Fig. 35-44, a broad beam of light of wavelength 630 nm is incident at 90° on a thin, wedge-shaped film with index of refraction 1.50. Transmission gives 10 bright and 9 dark fringes along the film's length. What is the left-to-right change in film thickness?

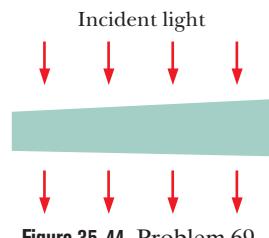


Figure 35-44 Problem 69.

••70 GO In Fig. 35-45, a broad beam of light of wavelength 620 nm is sent directly downward through the top plate of a pair of glass plates touching at the left end. The air between the plates acts as a thin film, and an interference pattern can be seen from above the plates. Initially, a dark fringe lies at the left end, a bright fringe lies at the right end, and nine dark fringes lie between those two end fringes. The plates are then very gradually squeezed together at a constant rate to decrease the angle between them. As a result, the fringe at the right side changes between being bright to being dark every 15.0 s. (a) At what rate is the spacing between the plates at the right end being changed? (b) By how much has the spacing there changed when both left and right ends have a dark fringe and there are five dark fringes between them?

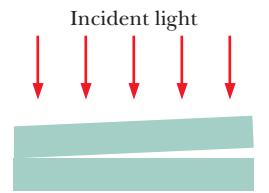


Figure 35-45 Problems 70–74.

••71 In Fig. 35-45, two microscope slides touch at one end and are separated at the other end. When light of wavelength 500 nm shines vertically down on the slides, an overhead observer sees an interference pattern on the slides with the dark fringes separated by 1.2 mm. What is the angle between the slides?

••72 In Fig. 35-45, a broad beam of monochromatic light is directed perpendicularly through two glass plates that are held together at one end to create a wedge of air between them. An observer intercepting light reflected from the wedge of air, which acts as a thin film, sees 4001 dark fringes along the length of the wedge. When the air between the plates is evacuated, only 4000 dark fringes are seen. Calculate to six significant figures the index of refraction of air from these data.

••73 SSM In Fig. 35-45, a broad beam of light of wavelength 683 nm is sent directly downward through the top plate of a pair of glass plates. The plates are 120 mm long, touch at the left end, and are separated by 48.0 μm at the right end. The air between the plates acts as a thin film. How many bright fringes will be seen by an observer looking down through the top plate?

••74 GO Two rectangular glass plates ($n = 1.60$) are in contact along one edge and are separated along the opposite edge (Fig. 35-45). Light with a wavelength of 600

Table 35-3 Problems 57 through 68: Transmission Through Thin Layers.
See the setup for these problems.

	n_1	n_2	n_3	Type	L	λ
57	1.55	1.60	1.33	min	285	
58	1.32	1.75	1.39	min	3rd	382
59	1.68	1.59	1.50	max	415	
60	1.50	1.34	1.42	max	380	
61	1.32	1.75	1.39	min	325	
62	1.68	1.59	1.50	max	2nd	342
63	1.40	1.46	1.75	max	2nd	482
64	1.40	1.46	1.75	max	210	
65	1.60	1.40	1.80	min	2nd	632
66	1.60	1.40	1.80	max	200	
67	1.50	1.34	1.42	min	2nd	587
68	1.55	1.60	1.33	min	3rd	612

nm is incident perpendicularly onto the top plate. The air between the plates acts as a thin film. Nine dark fringes and eight bright fringes are observed from above the top plate. If the distance between the two plates along the separated edges is increased by 600 nm, how many dark fringes will there then be across the top plate?

••75 SSM ILW Figure 35-46a shows a lens with radius of curvature R lying on a flat glass plate and illuminated from above by light with wavelength λ . Figure 35-46b (a photograph taken from above the lens) shows that circular interference fringes (known as *Newton's rings*) appear, associated with the variable thickness d of the air film between the lens and the plate. Find the radii r of the interference maxima assuming $r/R \ll 1$.

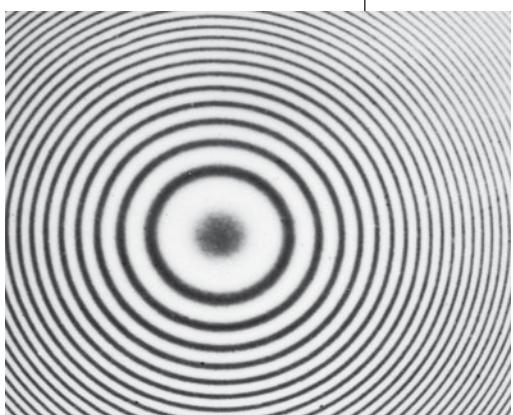


Figure 35-46
Problems
75–77.

(b) Courtesy Bausch & Lomb

••76 The lens in a Newton's rings experiment (see Problem 75) has diameter 20 mm and radius of curvature $R = 5.0$ m. For $\lambda = 589$ nm in air, how many bright rings are produced with the setup (a) in air and (b) immersed in water ($n = 1.33$)?

••77 A Newton's rings apparatus is to be used to determine the radius of curvature of a lens (see Fig. 35-46 and Problem 75). The radii of the n th and $(n + 20)$ th bright rings are found to be 0.162 and 0.368 cm, respectively, in light of wavelength 546 nm. Calculate the radius of curvature of the lower surface of the lens.

••78 A thin film of liquid is held in a horizontal circular ring, with air on both sides of the film. A beam of light at wavelength 550 nm is directed perpendicularly onto the film, and the intensity I of its reflection is monitored. Figure 35-47 gives intensity I as a function of time t ; the horizontal scale is set by $t_s = 20.0$ s. The intensity changes because of evaporation from the two sides of the film. Assume that the film is flat and has parallel sides, a radius of 1.80 cm, and an index of refraction of 1.40. Also assume that the film's volume decreases at a constant rate. Find that rate.

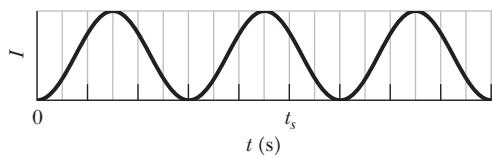


Figure 35-47 Problem 78.

Module 35-5 Michelson's Interferometer

•79 If mirror M_2 in a Michelson interferometer (Fig. 35-21) is moved through 0.233 mm, a shift of 792 bright fringes occurs. What is the wavelength of the light producing the fringe pattern?

•80 A thin film with index of refraction $n = 1.40$ is placed in one arm of a Michelson interferometer, perpendicular to the optical path. If this causes a shift of 7.0 bright fringes of the pattern produced by light of wavelength 589 nm, what is the film thickness?

••81 SSM WWW In Fig. 35-48, an airtight chamber of length $d = 5.0$ cm is placed in one of the arms of a Michelson interferometer. (The glass window on each end of the chamber has negligible thickness.) Light of wavelength $\lambda = 500$ nm is used. Evacuating the air from the chamber causes a shift of 60 bright fringes. From these data and to six significant figures, find the index of refraction of air at atmospheric pressure.

••82 The element sodium can emit light at two wavelengths, $\lambda_1 = 588.9950$ nm and $\lambda_2 = 589.5924$ nm. Light from sodium is being used in a Michelson interferometer (Fig. 35-21). Through what distance must mirror M_2 be moved if the shift in the fringe pattern for one wavelength is to be 1.00 fringe more than the shift in the fringe pattern for the other wavelength?

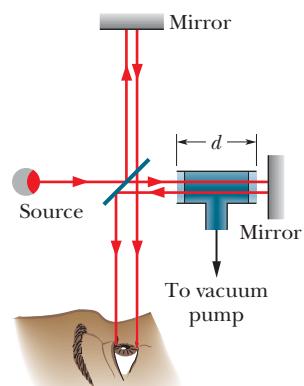


Figure 35-48 Problem 81.

Additional Problems

83 GO Two light rays, initially in phase and with a wavelength of 500 nm, go through different paths by reflecting from the various mirrors shown in Fig. 35-49. (Such a reflection does not itself produce a phase shift.) (a) What least value of distance d will put the rays exactly out of phase when they emerge from the region? (Ignore the slight tilt of the path for ray 2.) (b) Repeat the question assuming that the entire apparatus is immersed in a protein solution with an index of refraction of 1.38.

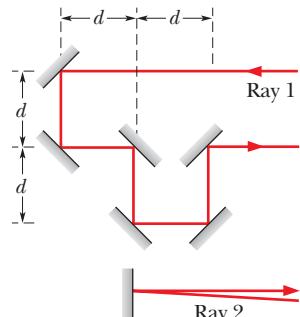


Figure 35-49 Problem 83.

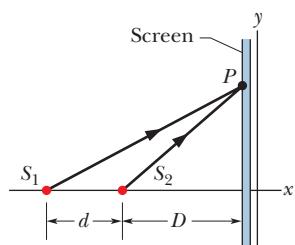


Figure 35-50 Problem 84.

84 GO In Figure 35-50, two isotropic point sources S_1 and S_2 emit light in phase at wavelength λ and at the same amplitude. The sources are separated by distance $d = 6.00\lambda$ on an x axis. A viewing screen is at distance $D = 20.0\lambda$ from S_2 and parallel to the y axis. The figure shows two rays reaching point P on the screen, at height y_P . (a) At what value of y_P do the rays have the minimum possible phase difference? (b) What multiple of λ gives that minimum phase difference? (c) At what value of y_P do the rays have the maximum possible phase difference? What multiple of λ gives (d) that maximum phase difference and (e) the phase difference when $y_P = d$? (f) When $y_P = d$, is the resulting intensity at point P maximum, mini-

mum, intermediate but closer to maximum, or intermediate but closer to minimum?

85 SSM A double-slit arrangement produces bright interference fringes for sodium light (a distinct yellow light at a wavelength of $\lambda = 589 \text{ nm}$). The fringes are angularly separated by 0.30° near the center of the pattern. What is the angular fringe separation if the entire arrangement is immersed in water, which has an index of refraction of 1.33?

86 GO In Fig. 35-51a, the waves along rays 1 and 2 are initially in phase, with the same wavelength λ in air. Ray 2 goes through a material with length L and index of refraction n . The rays are then reflected by mirrors to a common point P on a screen. Suppose that we can vary n from $n = 1.0$ to $n = 2.5$. Suppose also that, from $n = 1.0$ to $n = n_s = 1.5$, the intensity I of the light at point P varies with n as given in Fig. 35-51b. At what values of n greater than 1.4 is intensity I (a) maximum and (b) zero? (c) What multiple of λ gives the phase difference between the rays at point P when $n = 2.0$?

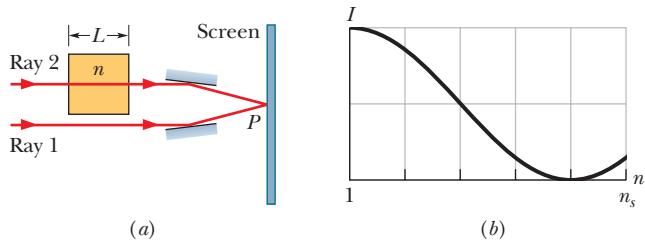


Figure 35-51 Problems 86 and 87.

87 SSM In Fig. 35-51a, the waves along rays 1 and 2 are initially in phase, with the same wavelength λ in air. Ray 2 goes through a material with length L and index of refraction n . The rays are then reflected by mirrors to a common point P on a screen. Suppose that we can vary L from 0 to 2400 nm. Suppose also that, from $L = 0$ to $L_s = 900 \text{ nm}$, the intensity I of the light at point P varies with L as given in Fig. 35-52. At what values of L greater than L_s is intensity I (a) maximum and (b) zero? (c) What multiple of λ gives the phase difference between ray 1 and ray 2 at common point P when $L = 1200 \text{ nm}$?

88 Light of wavelength 700.0 nm is sent along a route of length 2000 nm. The route is then filled with a medium having an index of refraction of 1.400. In degrees, by how much does the medium phase-shift the light? Give (a) the full shift and (b) the equivalent shift that has a value less than 360° .

89 SSM In Fig. 35-53, a microwave transmitter at height a above the water level of a wide lake transmits microwaves of wavelength λ toward a receiver on the opposite shore, a distance x above the water level. The microwaves reflecting from the water interfere with the microwaves arriving directly from the transmitter.

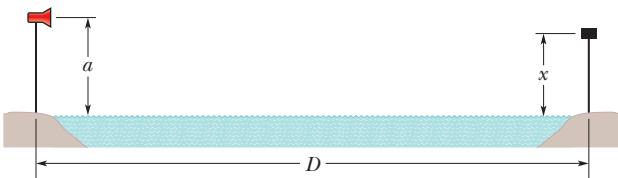


Figure 35-53 Problem 89.

Assuming that the lake width D is much greater than a and x , and that $\lambda \geq a$, find an expression that gives the values of x for which the signal at the receiver is maximum. (Hint: Does the reflection cause a phase change?)

90 In Fig. 35-54, two isotropic point sources S_1 and S_2 emit light at wavelength $\lambda = 400 \text{ nm}$. Source S_1 is located at $y = 640 \text{ nm}$; source S_2 is located at $y = -640 \text{ nm}$. At point P_1 (at $x = 720 \text{ nm}$), the wave from S_2 arrives ahead of the wave from S_1 by a phase difference of $0.600\pi \text{ rad}$. (a) What multiple of λ gives the phase difference between the waves from the two sources as the waves arrive at point P_2 , which is located at $y = 720 \text{ nm}$? (The figure is not drawn to scale.) (b) If the waves arrive at P_2 with equal amplitudes, is the interference there fully constructive, fully destructive, intermediate but closer to fully constructive, or intermediate but closer to fully destructive?

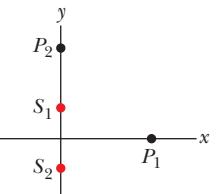


Figure 35-54
Problem 90.

91 Ocean waves moving at a speed of 4.0 m/s are approaching a beach at angle $\theta_1 = 30^\circ$ to the normal, as shown from above in Fig. 35-55. Suppose the water depth changes abruptly at a certain distance from the beach and the wave speed there drops to 3.0 m/s. (a) Close to the beach, what is the angle θ_2 between the direction of wave motion and the normal? (Assume the same law of refraction as for light.) (b) Explain why most waves come in normal to a shore even though at large distances they approach at a variety of angles.

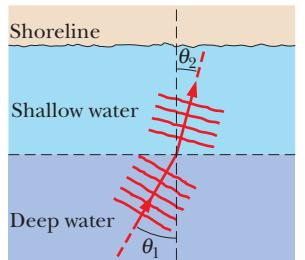


Figure 35-55 Problem 91.

92 Figure 35-56a shows two light rays that are initially in phase as they travel upward through a block of plastic, with wavelength 400 nm as measured in air. Light ray r_1 exits directly into air. However, before light ray r_2 exits into air, it travels through a liquid in a hollow cylinder within the plastic. Initially the height L_{liq} of the liquid is 40.0 μm , but then the liquid begins to evaporate. Let ϕ be the phase difference between rays r_1 and r_2 once they both exit into the air. Figure 35-56b shows ϕ versus the liquid's height L_{liq} until the liquid disappears, with ϕ given in terms of wavelength and the horizontal scale set by $L_s = 40.00 \mu\text{m}$. What are (a) the index of refraction of the plastic and (b) the index of refraction of the liquid?

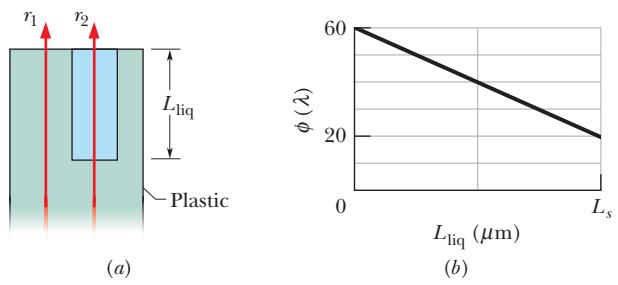


Figure 35-56 Problem 92.

93 SSM If the distance between the first and tenth minima of a double-slit pattern is 18.0 mm and the slits are separated by 0.150 mm with the screen 50.0 cm from the slits, what is the wavelength of the light used?

94 Figure 35-57 shows an optical fiber in which a central plastic core of index of refraction $n_1 = 1.58$ is surrounded by a plastic sheath of index of refraction $n_2 = 1.53$. Light can travel along different paths within the central core, leading to different travel times through the fiber. This causes an initially short pulse of light to spread as it travels along the fiber, resulting in information loss. Consider light that travels directly along the central axis of the fiber and light that is repeatedly reflected at the critical angle along the core–sheath interface, reflecting from side to side as it travels down the central core. If the fiber length is 300 m, what is the difference in the travel times along these two routes?

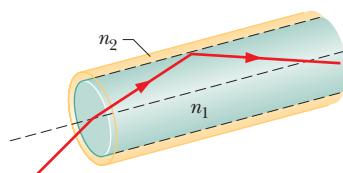


Figure 35-57 Problem 94.

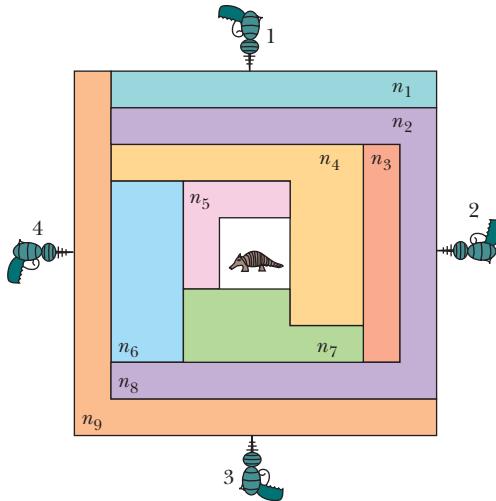
95 SSM Two parallel slits are illuminated with monochromatic light of wavelength 500 nm. An interference pattern is formed on a screen some distance from the slits, and the fourth dark band is located 1.68 cm from the central bright band on the screen. (a) What is the path length difference corresponding to the fourth dark band? (b) What is the distance on the screen between the central bright band and the first bright band on either side of the central band? (*Hint:* The angle to the fourth dark band and the angle to the first bright band are small enough that $\tan \theta \approx \sin \theta$.)

96 A camera lens with index of refraction greater than 1.30 is coated with a thin transparent film of index of refraction 1.25 to eliminate by interference the reflection of light at wavelength λ that is incident perpendicularly on the lens. What multiple of λ gives the minimum film thickness needed?

97 SSM Light of wavelength λ is used in a Michelson interferometer. Let x be the position of the movable mirror, with $x = 0$ when the arms have equal lengths $d_2 = d_1$. Write an expression for the intensity of the observed light as a function of x , letting I_m be the maximum intensity.

98 In two experiments, light is to be sent along the two paths shown in Fig. 35-35 by reflecting it from the various flat surfaces shown. In the first experiment, rays 1 and 2 are initially in phase and have a wavelength of 620.0 nm. In the second experiment, rays 1 and 2 are initially in phase and have a wavelength of 496.0 nm. What least value of distance L is required such that the 620.0 nm waves emerge from the region exactly in phase but the 496.0 nm waves emerge exactly out of phase?

99 Figure 35-58 shows the design of a Texas arcade game. Four laser pistols are pointed toward the center of an array of plastic

Figure 35-58
Problem 99.

layers where a clay armadillo is the target. The indexes of refraction of the layers are $n_1 = 1.55$, $n_2 = 1.70$, $n_3 = 1.45$, $n_4 = 1.60$, $n_5 = 1.45$, $n_6 = 1.61$, $n_7 = 1.59$, $n_8 = 1.70$, and $n_9 = 1.60$. The layer thicknesses are either 2.00 mm or 4.00 mm, as drawn. What is the travel time through the layers for the laser burst from (a) pistol 1, (b) pistol 2, (c) pistol 3, and (d) pistol 4? (e) If the pistols are fired simultaneously, which laser burst hits the target first?

100 A thin film suspended in air is $0.410 \mu\text{m}$ thick and is illuminated with white light incident perpendicularly on its surface. The index of refraction of the film is 1.50. At what wavelength will visible light that is reflected from the two surfaces of the film undergo fully constructive interference?

101 Find the slit separation of a double-slit arrangement that will produce interference fringes 0.018 rad apart on a distant screen when the light has wavelength $\lambda = 589 \text{ nm}$.

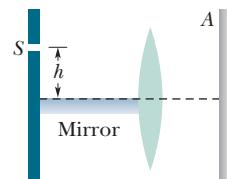
102 In a phasor diagram for any point on the viewing screen for the two-slit experiment in Fig. 35-10, the resultant-wave phasor rotates 60.0° in $2.50 \times 10^{-16} \text{ s}$. What is the wavelength?

103 In Fig. 35-59, an oil drop ($n = 1.20$) floats on the surface of water ($n = 1.33$) and is viewed from overhead when illuminated by sunlight shining vertically downward and reflected vertically upward. (a) Are the outer (thinnest) regions of the drop bright or dark? The oil film displays several spectra of colors. (b) Move from the rim inward to the third blue band and, using a wavelength of 475 nm for blue light, determine the film thickness there. (c) If the oil thickness increases, why do the colors gradually fade and then disappear?

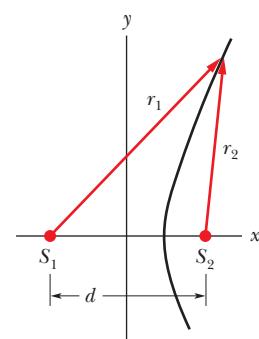


Figure 35-59 Problem 103.

104 Lloyd's Mirror. In Fig. 35-60, monochromatic light of wavelength λ diffracts through a narrow slit S in an otherwise opaque screen. On the other side, a plane mirror is perpendicular to the screen and a distance h from the slit. A viewing screen A is a distance much greater than h . (Because it sits in a plane through the focal point of the lens, screen A is effectively very distant. The lens plays no other role in the experiment and can otherwise be neglected.) Light that travels from the slit directly to A interferes with light from the slit that reflects from the mirror to A . The reflection causes a half-wavelength phase shift. (a) Is the fringe that corresponds to a zero path length difference bright or dark? Find expressions (like Eqs. 35-14 and 35-16) that locate (b) the bright fringes and (c) the dark fringes in the interference pattern. (*Hint:* Consider the image of S produced by the mirror as seen from a point on the viewing screen, and then consider Young's two-slit interference.)

Figure 35-60
Problem 104.

105 The two point sources in Fig. 35-61 emit coherent waves. Show that all curves (such as the one shown), over which the phase difference for rays r_1 and r_2 is a constant, are hyperbolas. (*Hint:* A constant phase difference implies a constant difference in length between r_1 and r_2 .)

Figure 35-61
Problem 105.

Diffraction

36-1 SINGLE-SLIT DIFFRACTION

Learning Objectives

After reading this module, you should be able to . . .

- 36.01 Describe the diffraction of light waves by a narrow opening and an edge, and also describe the resulting interference pattern.
- 36.02 Describe an experiment that demonstrates the Fresnel bright spot.
- 36.03 With a sketch, describe the arrangement for a single-slit diffraction experiment.
- 36.04 With a sketch, explain how splitting a slit width into equal zones leads to the equations giving the angles to the minima in the diffraction pattern.
- 36.05 Apply the relationships between width a of a thin,

Key Ideas

- When waves encounter an edge, an obstacle, or an aperture the size of which is comparable to the wavelength of the waves, those waves spread out as they travel and, as a result, undergo interference. This type of interference is called diffraction.
- Waves passing through a long narrow slit of width a produce, on a viewing screen, a single-slit diffraction

rectangular slit or object, the wavelength λ , the angle θ to any of the minima in the diffraction pattern, the distance to a viewing screen, and the distance between a minimum and the center of the pattern.

- 36.06 Sketch the diffraction pattern for monochromatic light, identifying what lies at the center and what the various bright and dark fringes are called (such as “first minimum”).
- 36.07 Identify what happens to a diffraction pattern when the wavelength of the light or the width of the diffracting aperture or object is varied.

pattern that includes a central maximum (bright fringe) and other maxima. They are separated by minima that are located relative to the central axis by angles θ :

$$a \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{minima}).$$

- The maxima are located approximately halfway between minima.

What Is Physics?

One focus of physics in the study of light is to understand and put to use the diffraction of light as it passes through a narrow slit or (as we shall discuss) past either a narrow obstacle or an edge. We touched on this phenomenon in Chapter 35 when we looked at how light flared—diffracted—through the slits in Young’s experiment. Diffraction through a given slit is more complicated than simple flaring, however, because the light also interferes with itself and produces an interference pattern. It is because of such complications that light is rich with application opportunities. Even though the diffraction of light as it passes through a slit or past an obstacle seems awfully academic, countless engineers and scientists make their living using this physics, and the total worth of diffraction applications worldwide is probably incalculable.

Before we can discuss some of these applications, we first must discuss why diffraction is due to the wave nature of light.

Diffraction and the Wave Theory of Light

In Chapter 35 we defined diffraction rather loosely as the flaring of light as it emerges from a narrow slit. More than just flaring occurs, however, because the



Ken Kay/Fundamental Photographs

Figure 36-1 This diffraction pattern appeared on a viewing screen when light that had passed through a narrow vertical slit reached the screen. Diffraction caused the light to flare out perpendicular to the long sides of the slit. That flaring produced an interference pattern consisting of a broad central maximum plus less intense and narrower secondary (or side) maxima, with minima between them.

light produces an interference pattern called a **diffraction pattern**. For example, when monochromatic light from a distant source (or a laser) passes through a narrow slit and is then intercepted by a viewing screen, the light produces on the screen a diffraction pattern like that in Fig. 36-1. This pattern consists of a broad and intense (very bright) central maximum plus a number of narrower and less intense maxima (called **secondary** or **side** maxima) to both sides. In between the maxima are minima. Light flares into those dark regions, but the light waves cancel out one another.

Such a pattern would be totally unexpected in geometrical optics: If light traveled in straight lines as rays, then the slit would allow some of those rays through to form a sharp rendition of the slit on the viewing screen instead of a pattern of bright and dark bands as we see in Fig. 36-1. As in Chapter 35, we must conclude that geometrical optics is only an approximation.

Edges. Diffraction is not limited to situations in which light passes through a narrow opening (such as a slit or pinhole). It also occurs when light passes an edge, such as the edges of the razor blade whose diffraction pattern is shown in Fig. 36-2. Note the lines of maxima and minima that run approximately parallel to the edges, at both the inside edges of the blade and the outside edges. As the light passes, say, the vertical edge at the left, it flares left and right and undergoes interference, producing the pattern along the left edge. The rightmost portion of that pattern actually lies behind the blade, within what would be the blade's shadow if geometrical optics prevailed.

Floaters. You encounter a common example of diffraction when you look at a clear blue sky and see tiny specks and hairlike structures floating in your view. These *floaters*, as they are called, are produced when light passes the edges of tiny deposits in the vitreous humor, the transparent material filling most of the eyeball. What you are seeing when a floater is in your field of vision is the diffraction pattern produced on the retina by one of these deposits. If you sight through a pinhole in a piece of cardboard so as to make the light entering your eye approximately a plane wave, you can distinguish individual maxima and minima in the patterns.

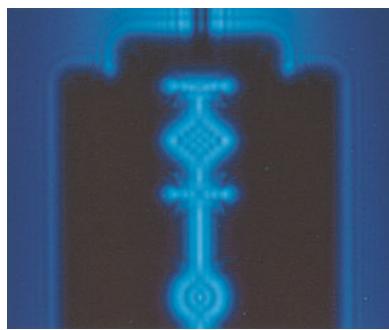
Cheerleaders. Diffraction is a wave effect. That is, it occurs because light is a wave and it occurs with other types of waves as well. For example, you have probably seen diffraction in action at football games. When a cheerleader near the playing field yells up at several thousand noisy fans, the yell can hardly be heard because the sound waves diffract when they pass through the narrow opening of the cheerleader's mouth. This flaring leaves little of the waves traveling toward the fans in front of the cheerleader. To offset the diffraction, the cheerleader can yell through a megaphone. The sound waves then emerge from the much wider opening at the end of the megaphone. The flaring is thus reduced, and much more of the sound reaches the fans in front of the cheerleader.

The Fresnel Bright Spot

Diffraction finds a ready explanation in the wave theory of light. However, this theory, originally advanced in the late 1600s by Huygens and used 123 years later by Young to explain double-slit interference, was very slow in being adopted, largely because it ran counter to Newton's theory that light was a stream of particles.

Newton's view was the prevailing view in French scientific circles of the early 19th century, when Augustin Fresnel was a young military engineer. Fresnel, who believed in the wave theory of light, submitted a paper to the French Academy of Sciences describing his experiments with light and his wave-theory explanations of them.

In 1819, the Academy, dominated by supporters of Newton and thinking to challenge the wave point of view, organized a prize competition for an essay on the subject of diffraction. Fresnel won. The Newtonians, however, were not swayed. One of them, S. D. Poisson, pointed out the "strange result" that if Fresnel's theories were correct, then light waves should flare into the shadow region of a sphere as they pass the edge of the sphere, producing a bright spot at the center of the shadow. The prize committee arranged a test of Poisson's prediction and dis-



Ken Kay/Fundamental Photographs

Figure 36-2 The diffraction pattern produced by a razor blade in monochromatic light. Note the lines of alternating maximum and minimum intensities.

covered that the predicted *Fresnel bright spot*, as we call it today, was indeed there (Fig. 36-3). Nothing builds confidence in a theory so much as having one of its unexpected and counterintuitive predictions verified by experiment.

Diffraction by a Single Slit: Locating the Minima

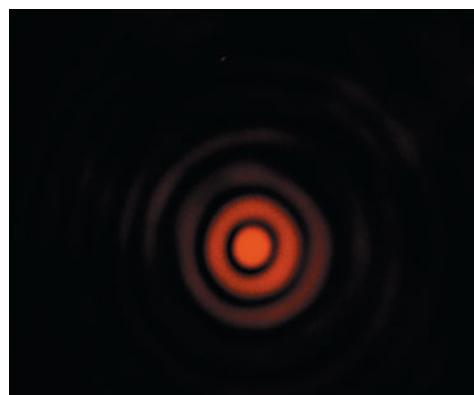
Let us now examine the diffraction pattern of plane waves of light of wavelength λ that are diffracted by a single long, narrow slit of width a in an otherwise opaque screen B , as shown in cross section in Fig. 36-4. (In that figure, the slit's length extends into and out of the page, and the incoming wavefronts are parallel to screen B .) When the diffracted light reaches viewing screen C , waves from different points within the slit undergo interference and produce a diffraction pattern of bright and dark fringes (interference maxima and minima) on the screen. To locate the fringes, we shall use a procedure somewhat similar to the one we used to locate the fringes in a two-slit interference pattern. However, diffraction is more mathematically challenging, and here we shall be able to find equations for only the dark fringes.

Before we do that, however, we can justify the central bright fringe seen in Fig. 36-1 by noting that the Huygens wavelets from all points in the slit travel about the same distance to reach the center of the pattern and thus are in phase there. As for the other bright fringes, we can say only that they are approximately halfway between adjacent dark fringes.

Pairings. To find the dark fringes, we shall use a clever (and simplifying) strategy that involves pairing up all the rays coming through the slit and then finding what conditions cause the wavelets of the rays in each pair to cancel each other. We apply this strategy in Fig. 36-4 to locate the first dark fringe, at point P_1 . First, we mentally divide the slit into two *zones* of equal widths $a/2$. Then we extend to P_1 a light ray r_1 from the top point of the top zone and a light ray r_2 from the top point of the bottom zone. We want the wavelets along these two rays to cancel each other when they arrive at P_1 . Then any similar pairing of rays from the two zones will give cancellation. A central axis is drawn from the center of the slit to screen C , and P_1 is located at an angle θ to that axis.

Path Length Difference. The wavelets of the pair of rays r_1 and r_2 are in phase within the slit because they originate from the same wavefront passing through the slit, along the width of the slit. However, to produce the first dark fringe they must be out of phase by $\lambda/2$ when they reach P_1 ; this phase difference is due to their path length difference, with the path traveled by the wavelet of r_2 to reach P_1 being longer than the path traveled by the wavelet of r_1 . To display this path length difference, we find a point b on ray r_2 such that the path length from b to P_1 matches the path length of ray r_1 . Then the path length difference between the two rays is the distance from the center of the slit to b .

When viewing screen C is near screen B , as in Fig. 36-4, the diffraction pattern on C is difficult to describe mathematically. However, we can simplify the mathematics considerably if we arrange for the screen separation D to be much larger than the slit width a . Then, as in Fig. 36-5, we can approximate rays r_1 and r_2



Courtesy Jearl Walker

Figure 36-3 A photograph of the diffraction pattern of a disk. Note the concentric diffraction rings and the Fresnel bright spot at the center of the pattern. This experiment is essentially identical to that arranged by the committee testing Fresnel's theories, because both the sphere they used and the disk used here have a cross section with a circular edge.

This pair of rays cancel each other at P_1 . So do all such pairings.

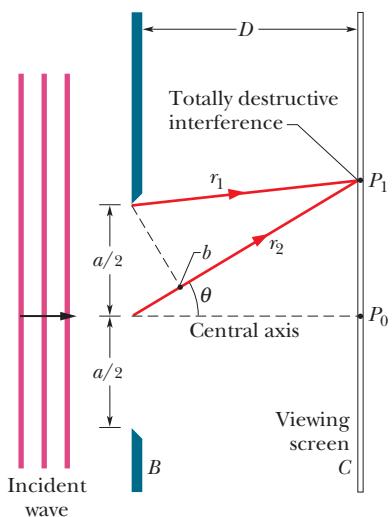


Figure 36-4 Waves from the top points of two zones of width $a/2$ undergo fully destructive interference at point P_1 on viewing screen C .

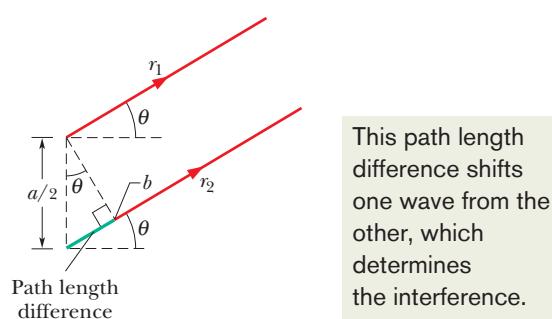
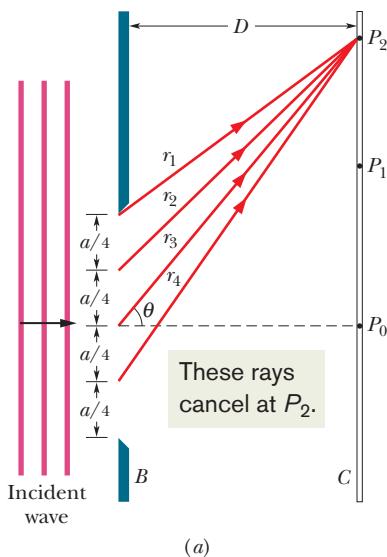
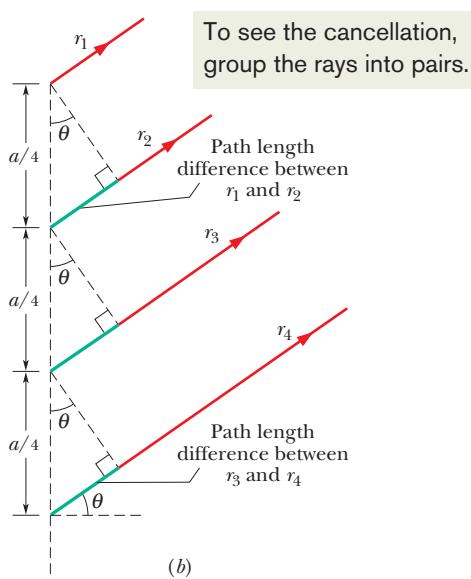


Figure 36-5 For $D \gg a$, we can approximate rays r_1 and r_2 as being parallel, at angle θ to the central axis.



(a)



(b)

Figure 36-6 (a) Waves from the top points of four zones of width $a/4$ undergo fully destructive interference at point P_2 . (b) For $D \gg a$, we can approximate rays r_1, r_2, r_3 , and r_4 as being parallel, at angle θ to the central axis.

as being parallel, at angle θ to the central axis. We can also approximate the triangle formed by point b , the top point of the slit, and the center point of the slit as being a right triangle, and one of the angles inside that triangle as being θ . The path length difference between rays r_1 and r_2 (which is still the distance from the center of the slit to point b) is then equal to $(a/2) \sin \theta$.

First Minimum. We can repeat this analysis for any other pair of rays originating at corresponding points in the two zones (say, at the midpoints of the zones) and extending to point P_1 . Each such pair of rays has the same path length difference $(a/2) \sin \theta$. Setting this common path length difference equal to $\lambda/2$ (our condition for the first dark fringe), we have

$$\frac{a}{2} \sin \theta = \frac{\lambda}{2},$$

which gives us

$$a \sin \theta = \lambda \quad (\text{first minimum}). \quad (36-1)$$

Given slit width a and wavelength λ , Eq. 36-1 tells us the angle θ of the first dark fringe above and (by symmetry) below the central axis.

Narrowing the Slit. Note that if we begin with $a > \lambda$ and then narrow the slit while holding the wavelength constant, we increase the angle at which the first dark fringes appear; that is, the extent of the diffraction (the extent of the flaring and the width of the pattern) is *greater* for a *narrower slit*. When we have reduced the slit width to the wavelength (that is, $a = \lambda$), the angle of the first dark fringes is 90° . Since the first dark fringes mark the two edges of the central bright fringe, that bright fringe must then cover the entire viewing screen.

Second Minimum. We find the second dark fringes above and below the central axis as we found the first dark fringes, except that we now divide the slit into *four* zones of equal widths $a/4$, as shown in Fig. 36-6a. We then extend rays r_1, r_2, r_3 , and r_4 from the top points of the zones to point P_2 , the location of the second dark fringe above the central axis. To produce that fringe, the path length difference between r_1 and r_2 , that between r_2 and r_3 , and that between r_3 and r_4 must all be equal to $\lambda/2$.

For $D \gg a$, we can approximate these four rays as being parallel, at angle θ to the central axis. To display their path length differences, we extend a perpendicular line through each adjacent pair of rays, as shown in Fig. 36-6b, to form a series of right triangles, each of which has a path length difference as one side. We see from the top triangle that the path length difference between r_1 and r_2 is $(a/4) \sin \theta$. Similarly, from the bottom triangle, the path length difference between r_3 and r_4 is also $(a/4) \sin \theta$. In fact, the path length difference for any two rays that originate at corresponding points in two adjacent zones is $(a/4) \sin \theta$. Since in each such case the path length difference is equal to $\lambda/2$, we have

$$\frac{a}{4} \sin \theta = \frac{\lambda}{2},$$

which gives us

$$a \sin \theta = 2\lambda \quad (\text{second minimum}). \quad (36-2)$$

All Minima. We could now continue to locate dark fringes in the diffraction pattern by splitting up the slit into more zones of equal width. We would always choose an even number of zones so that the zones (and their waves) could be paired as we have been doing. We would find that the dark fringes above and below the central axis can be located with the general equation

$$a \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{minima—dark fringes}). \quad (36-3)$$

You can remember this result in the following way. Draw a triangle like the one in Fig. 36-5, but for the full slit width a , and note that the path length difference between the top and bottom rays equals $a \sin \theta$. Thus, Eq. 36-3 says:



In a single-slit diffraction experiment, dark fringes are produced where the path length differences ($a \sin \theta$) between the top and bottom rays are equal to $\lambda, 2\lambda, 3\lambda, \dots$.

This may seem to be wrong because the waves of those two particular rays will be exactly in phase with each other when their path length difference is an integer number of wavelengths. However, they each will still be part of a pair of waves that are exactly out of phase with each other; thus, *each* wave will be canceled by some other wave, resulting in darkness. (Two light waves that are exactly out of phase will always cancel each other, giving a net wave of zero, even if they happen to be exactly in phase with other light waves.)

Using a Lens. Equations 36-1, 36-2, and 36-3 are derived for the case of $D \gg a$. However, they also apply if we place a converging lens between the slit and the viewing screen and then move the screen in so that it coincides with the focal plane of the lens. The lens ensures that rays which now reach any point on the screen are *exactly* parallel (rather than approximately) back at the slit. They are like the initially parallel rays of Fig. 34-14a that are directed to the focal point by a converging lens.



Checkpoint 1

We produce a diffraction pattern on a viewing screen by means of a long narrow slit illuminated by blue light. Does the pattern expand away from the bright center (the maxima and minima shift away from the center) or contract toward it if we
 (a) switch to yellow light or (b) decrease the slit width?

Sample Problem 36.01 Single-slit diffraction pattern with white light

A slit of width a is illuminated by white light.

(a) For what value of a will the first minimum for red light of wavelength $\lambda = 650 \text{ nm}$ appear at $\theta = 15^\circ$?

KEY IDEA

Diffraction occurs separately for each wavelength in the range of wavelengths passing through the slit, with the locations of the minima for each wavelength given by Eq. 36-3 ($a \sin \theta = m\lambda$).

Calculation: When we set $m = 1$ (for the first minimum) and substitute the given values of θ and λ , Eq. 36-3 yields

$$a = \frac{m\lambda}{\sin \theta} = \frac{(1)(650 \text{ nm})}{\sin 15^\circ} = 2511 \text{ nm} \approx 2.5 \mu\text{m} \quad (\text{Answer})$$

For the incident light to flare out that much ($\pm 15^\circ$ to the first minima) the slit has to be very fine indeed—in this case, a mere four times the wavelength. For comparison, note that a fine human hair may be about $100 \mu\text{m}$ in diameter.

(b) What is the wavelength λ' of the light whose first side diffraction maximum is at 15° , thus coinciding with the first minimum for the red light?

KEY IDEA

The first side maximum for any wavelength is about halfway between the first and second minima for that wavelength.

Calculations: Those first and second minima can be located with Eq. 36-3 by setting $m = 1$ and $m = 2$, respectively. Thus, the first side maximum can be located *approximately* by setting $m = 1.5$. Then Eq. 36-3 becomes

$$a \sin \theta = 1.5\lambda'.$$

Solving for λ' and substituting known data yield

$$\lambda' = \frac{a \sin \theta}{1.5} = \frac{(2511 \text{ nm})(\sin 15^\circ)}{1.5} = 430 \text{ nm.} \quad (\text{Answer})$$

Light of this wavelength is violet (far blue, near the short-wavelength limit of the human range of visible light). From the two equations we used, can you see that the first side maximum for light of wavelength 430 nm will always coincide with the first minimum for light of wavelength 650 nm , no matter what the slit width is? However, the angle θ at which this overlap occurs does depend on slit width. If the slit is relatively narrow, the angle will be relatively large, and conversely.



Additional examples, video, and practice available at WileyPLUS

36-2 INTENSITY IN SINGLE-SLIT DIFFRACTION

Learning Objectives

After reading this module, you should be able to . . .

- 36.08** Divide a thin slit into multiple zones of equal width and write an expression for the phase difference of the wavelets from adjacent zones in terms of the angle θ to a point on the viewing screen.
- 36.09** For single-slit diffraction, draw phasor diagrams for the central maximum and several of the minima and maxima off to one side, indicating the phase difference between adjacent phasors, explaining how the net electric field is calculated, and

identifying the corresponding part of the diffraction pattern.

- 36.10** Describe a diffraction pattern in terms of the net electric field at points in the pattern.
- 36.11** Evaluate α , the convenient connection between angle θ to a point in a diffraction pattern and the intensity I at that point.
- 36.12** For a given point in a diffraction pattern, at a given angle, calculate the intensity I in terms of the intensity I_m at the center of the pattern.

Key Idea

- The intensity of the diffraction pattern at any given angle θ is

$$I(\theta) = I_m \left(\frac{\sin \alpha}{\alpha} \right)^2,$$

where I_m is the intensity at the center of the pattern and

$$\alpha = \frac{\pi a}{\lambda} \sin \theta.$$

Intensity in Single-Slit Diffraction, Qualitatively

In Module 36-1 we saw how to find the positions of the minima and the maxima in a single-slit diffraction pattern. Now we turn to a more general problem: find an expression for the intensity I of the pattern as a function of θ , the angular position of a point on a viewing screen.

To do this, we divide the slit of Fig. 36-4 into N zones of equal widths Δx small enough that we can assume each zone acts as a source of Huygens wavelets. We wish to superimpose the wavelets arriving at an arbitrary point P on the viewing screen, at angle θ to the central axis, so that we can determine the amplitude E_θ of the electric component of the resultant wave at P . The intensity of the light at P is then proportional to the square of that amplitude.

To find E_θ , we need the phase relationships among the arriving wavelets. The point here is that in general they have different phases because they travel different distances to reach P . The phase difference between wavelets from adjacent zones is given by

$$\left(\begin{array}{c} \text{phase} \\ \text{difference} \end{array} \right) = \left(\frac{2\pi}{\lambda} \right) \left(\begin{array}{c} \text{path length} \\ \text{difference} \end{array} \right).$$

For point P at angle θ , the path length difference between wavelets from adjacent zones is $\Delta x \sin \theta$. Thus, we can write the phase difference $\Delta\phi$ between wavelets from adjacent zones as

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \right) (\Delta x \sin \theta). \quad (36-4)$$

We assume that the wavelets arriving at P all have the same amplitude ΔE . To find the amplitude E_θ of the resultant wave at P , we add the amplitudes ΔE via phasors. To do this, we construct a diagram of N phasors, one corresponding to the wavelet from each zone in the slit.

Central Maximum. For point P_0 at $\theta = 0$ on the central axis of Fig. 36-4, Eq. 36-4 tells us that the phase difference $\Delta\phi$ between the wavelets is zero; that is, the wavelets all arrive in phase. Figure 36-7a is the corresponding phasor diagram; adjacent phasors represent wavelets from adjacent zones and are arranged head to tail. Because there is zero phase difference between the wavelets, there is zero angle between each pair of adjacent phasors. The amplitude E_θ of the net

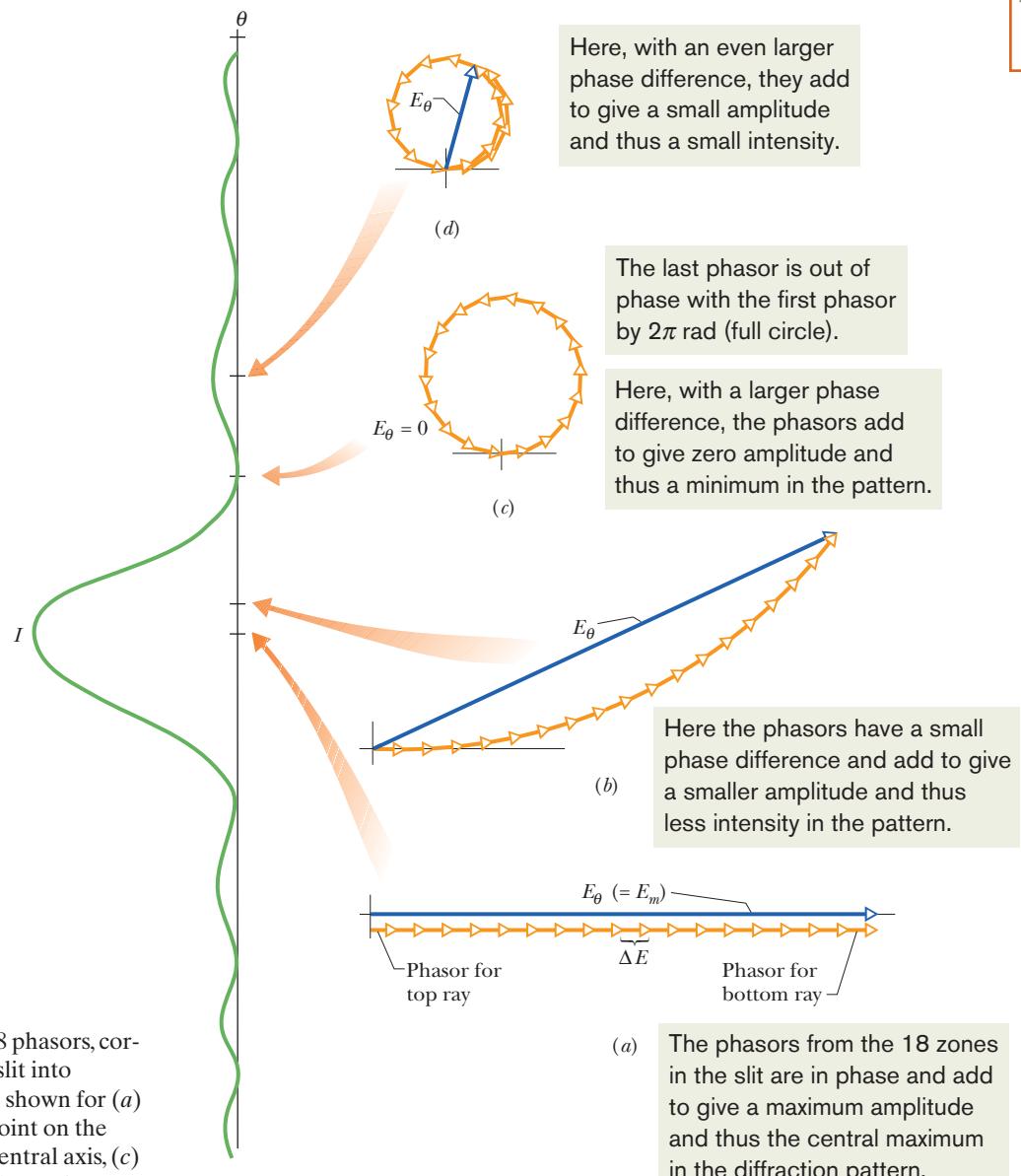


Figure 36-7 Phasor diagrams for $N = 18$ phasors, corresponding to the division of a single slit into 18 zones. Resultant amplitudes E_θ are shown for (a) the central maximum at $\theta = 0$, (b) a point on the screen lying at a small angle θ to the central axis, (c) the first minimum, and (d) the first side maximum.

wave at P_0 is the vector sum of these phasors. This arrangement of the phasors turns out to be the one that gives the greatest value for the amplitude E_θ . We call this value E_m ; that is, E_m is the value of E_θ for $\theta = 0$.

We next consider a point P that is at a small angle θ to the central axis. Equation 36-4 now tells us that the phase difference $\Delta\phi$ between wavelets from adjacent zones is no longer zero. Figure 36-7b shows the corresponding phasor diagram; as before, the phasors are arranged head to tail, but now there is an angle $\Delta\phi$ between adjacent phasors. The amplitude E_θ at this new point is still the vector sum of the phasors, but it is smaller than that in Fig. 36-7a, which means that the intensity of the light is less at this new point P than at P_0 .

First Minimum. If we continue to increase θ , the angle $\Delta\phi$ between adjacent phasors increases, and eventually the chain of phasors curls completely around so that the head of the last phasor just reaches the tail of the first phasor (Fig. 36-7c). The ampli-

tude E_θ is now zero, which means that the intensity of the light is also zero. We have reached the first minimum, or dark fringe, in the diffraction pattern. The first and last phasors now have a phase difference of 2π rad, which means that the path length difference between the top and bottom rays through the slit equals one wavelength. Recall that this is the condition we determined for the first diffraction minimum.

First Side Maximum. As we continue to increase θ , the angle $\Delta\phi$ between adjacent phasors continues to increase, the chain of phasors begins to wrap back on itself, and the resulting coil begins to shrink. Amplitude E_θ now increases until it reaches a maximum value in the arrangement shown in Fig. 36-7d. This arrangement corresponds to the first side maximum in the diffraction pattern.

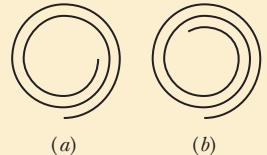
Second Minimum. If we increase θ a bit more, the resulting shrinkage of the coil decreases E_θ , which means that the intensity also decreases. When θ is increased enough, the head of the last phasor again meets the tail of the first phasor. We have then reached the second minimum.

We could continue this qualitative method of determining the maxima and minima of the diffraction pattern but, instead, we shall now turn to a quantitative method.



Checkpoint 2

The figures represent, in smoother form (with more phasors) than Fig. 36-7, the phasor diagrams for two points of a diffraction pattern that are on opposite sides of a certain diffraction maximum. (a) Which maximum is it? (b) What is the approximate value of m (in Eq. 36-3) that corresponds to this maximum?



(a)

(b)

Intensity in Single-Slit Diffraction, Quantitatively

Equation 36-3 tells us how to locate the minima of the single-slit diffraction pattern on screen C of Fig. 36-4 as a function of the angle θ in that figure. Here we wish to derive an expression for the intensity $I(\theta)$ of the pattern as a function of θ . We state, and shall prove below, that the intensity is given by

$$I(\theta) = I_m \left(\frac{\sin \alpha}{\alpha} \right)^2, \quad (36-5)$$

$$\text{where } \alpha = \frac{1}{2}\phi = \frac{\pi a}{\lambda} \sin \theta. \quad (36-6)$$

The symbol α is just a convenient connection between the angle θ that locates a point on the viewing screen and the light intensity $I(\theta)$ at that point. The intensity I_m is the greatest value of the intensities $I(\theta)$ in the pattern and occurs at the central maximum (where $\theta = 0$), and ϕ is the phase difference (in radians) between the top and bottom rays from the slit of width a .

Study of Eq. 36-5 shows that intensity minima will occur where

$$\alpha = m\pi, \quad \text{for } m = 1, 2, 3, \dots \quad (36-7)$$

If we put this result into Eq. 36-6, we find

$$m\pi = \frac{\pi a}{\lambda} \sin \theta, \quad \text{for } m = 1, 2, 3, \dots,$$

$$\text{or } a \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{minima—dark fringes}), \quad (36-8)$$

which is exactly Eq. 36-3, the expression that we derived earlier for the location of the minima.

Plots. Figure 36-8 shows plots of the intensity of a single-slit diffraction pattern, calculated with Eqs. 36-5 and 36-6 for three slit widths: $a = \lambda$, $a = 5\lambda$, and $a = 10\lambda$. Note that as the slit width increases (relative to the wavelength), the width of the *central diffraction maximum* (the central hill-like region of the graphs) decreases; that is, the light undergoes less flaring by the slit. The secondary maxima also decrease in width (and become weaker). In the limit of slit width a being much greater than wavelength λ , the secondary maxima due to the slit disappear; we then no longer have single-slit diffraction (but we still have diffraction due to the edges of the wide slit, like that produced by the edges of the razor blade in Fig. 36-2).

Proof of Eqs. 36-5 and 36-6

To find an expression for the intensity at a point in the diffraction pattern, we need to divide the slit into many zones and then add the phasors corresponding to those zones, as we did in Fig. 36-7. The arc of phasors in Fig. 36-9 represents the wavelets that reach an arbitrary point P on the viewing screen of Fig. 36-4, corresponding to a particular small angle θ . The amplitude E_θ of the resultant wave at P is the vector sum of these phasors. If we divide the slit of Fig. 36-4 into infinitesimal zones of width Δx , the arc of phasors in Fig. 36-9 approaches the arc of a circle; we call its radius R as indicated in that figure. The length of the arc must be E_m , the amplitude at the center of the diffraction pattern, because if we straightened out the arc we would have the phasor arrangement of Fig. 36-7a (shown lightly in Fig. 36-9).

The angle ϕ in the lower part of Fig. 36-9 is the difference in phase between the infinitesimal vectors at the left and right ends of arc E_m . From the geometry, ϕ is also the angle between the two radii marked R in Fig. 36-9. The dashed line in that figure, which bisects ϕ , then forms two congruent right triangles. From either triangle we can write

$$\sin \frac{1}{2}\phi = \frac{E_\theta}{2R}. \quad (36-9)$$

In radian measure, ϕ is (with E_m considered to be a circular arc)

$$\phi = \frac{E_m}{R}.$$

Solving this equation for R and substituting in Eq. 36-9 lead to

$$E_\theta = \frac{E_m}{\frac{1}{2}\phi} \sin \frac{1}{2}\phi. \quad (36-10)$$

Intensity. In Module 33-2 we saw that the intensity of an electromagnetic wave is proportional to the square of the amplitude of its electric field. Here, this means that the maximum intensity I_m (at the center of the pattern) is proportional to E_m^2 and the intensity $I(\theta)$ at angle θ is proportional to E_θ^2 . Thus,

$$\frac{I(\theta)}{I_m} = \frac{E_\theta^2}{E_m^2}. \quad (36-11)$$

Substituting for E_θ with Eq. 36-10 and then substituting $\alpha = \frac{1}{2}\phi$, we are led to Eq. 36-5 for the intensity as a function of θ :

$$I(\theta) = I_m \left(\frac{\sin \alpha}{\alpha} \right)^2.$$

The second equation we wish to prove relates α to θ . The phase difference ϕ between the rays from the top and bottom of the entire slit may be related to a path length difference with Eq. 36-4; it tells us that

$$\phi = \left(\frac{2\pi}{\lambda} \right) (a \sin \theta),$$

where a is the sum of the widths Δx of the infinitesimal zones. However, $\phi = 2\alpha$, so this equation reduces to Eq. 36-6.

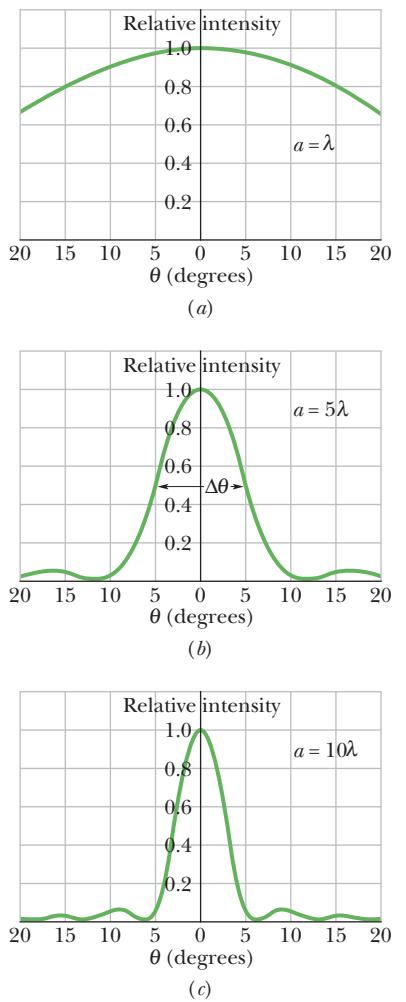


Figure 36-8 The relative intensity in single-slit diffraction for three values of the ratio a/λ . The wider the slit is, the narrower is the central diffraction maximum.

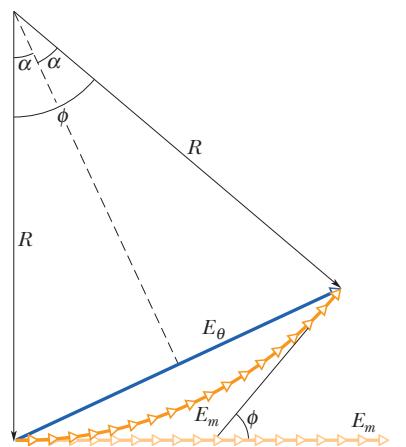
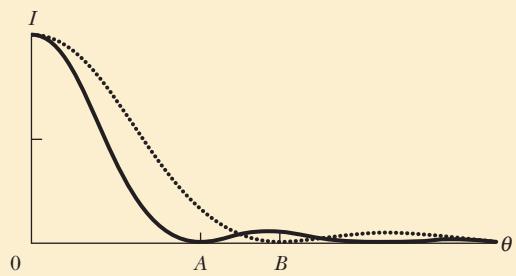


Figure 36-9 A construction used to calculate the intensity in single-slit diffraction. The situation shown corresponds to that of Fig. 36-7b.

**Checkpoint 3**

Two wavelengths, 650 and 430 nm, are used separately in a single-slit diffraction experiment. The figure shows the results as graphs of intensity I versus angle θ for the two diffraction patterns. If both wavelengths are then used simultaneously, what color will be seen in the combined diffraction pattern at (a) angle A and (b) angle B ?

**Sample Problem 36.02 Intensities of the maxima in a single-slit interference pattern**

Find the intensities of the first three secondary maxima (side maxima) in the single-slit diffraction pattern of Fig. 36-1, measured as a percentage of the intensity of the central maximum.

KEY IDEAS

The secondary maxima lie approximately halfway between the minima, whose angular locations are given by Eq. 36-7 ($\alpha = m\pi$). The locations of the secondary maxima are then given (approximately) by

$$\alpha = (m + \frac{1}{2})\pi, \quad \text{for } m = 1, 2, 3, \dots,$$

with α in radian measure. We can relate the intensity I at any point in the diffraction pattern to the intensity I_m of the central maximum via Eq. 36-5.

Calculations: Substituting the approximate values of α for the secondary maxima into Eq. 36-5 to obtain the relative

intensities at those maxima, we get

$$\frac{I}{I_m} = \left(\frac{\sin \alpha}{\alpha} \right)^2 = \left(\frac{\sin(m + \frac{1}{2})\pi}{(m + \frac{1}{2})\pi} \right)^2, \quad \text{for } m = 1, 2, 3, \dots$$

The first of the secondary maxima occurs for $m = 1$, and its relative intensity is

$$\begin{aligned} \frac{I_1}{I_m} &= \left(\frac{\sin(1 + \frac{1}{2})\pi}{(1 + \frac{1}{2})\pi} \right)^2 = \left(\frac{\sin 1.5\pi}{1.5\pi} \right)^2 \\ &= 4.50 \times 10^{-2} \approx 4.5\%. \end{aligned} \quad (\text{Answer})$$

For $m = 2$ and $m = 3$ we find that

$$\frac{I_2}{I_m} = 1.6\% \quad \text{and} \quad \frac{I_3}{I_m} = 0.83\%. \quad (\text{Answer})$$

As you can see from these results, successive secondary maxima decrease rapidly in intensity. Figure 36-1 was deliberately overexposed to reveal them.



Additional examples, video, and practice available at WileyPLUS

36-3 DIFFRACTION BY A CIRCULAR APERTURE

Learning Objectives

After reading this module, you should be able to ...

36.13 Describe and sketch the diffraction pattern from a small circular aperture or obstacle.

36.14 For diffraction by a small circular aperture or obstacle, apply the relationships between the angle θ to the first minimum, the wavelength λ of the light, the diameter d of the aperture, the distance D to a viewing screen, and the distance y between the minimum and the center of the diffraction pattern.

36.15 By discussing the diffraction patterns of point objects,

explain how diffraction limits visual resolution of objects.

36.16 Identify that Rayleigh's criterion for resolvability gives the (approximate) angle at which two point objects are just barely resolvable.

36.17 Apply the relationships between the angle θ_R in Rayleigh's criterion, the wavelength λ of the light, the diameter d of the aperture (for example, the diameter of the pupil of an eye), the angle θ subtended by two distant point objects, and the distance L to those objects.

Key Ideas

- Diffraction by a circular aperture or a lens with diameter d produces a central maximum and concentric maxima and minima, with the first minimum at an angle θ given by

$$\sin \theta = 1.22 \frac{\lambda}{d} \quad (\text{first minimum—circular aperture}).$$

- Rayleigh's criterion suggests that two objects are on the

verge of resolvability if the central diffraction maximum of one is at the first minimum of the other. Their angular separation can then be no less than

$$\theta_R = 1.22 \frac{\lambda}{d} \quad (\text{Rayleigh's criterion}),$$

in which d is the diameter of the aperture through which the light passes.

Diffraction by a Circular Aperture

Here we consider diffraction by a circular aperture—that is, a circular opening, such as a circular lens, through which light can pass. Figure 36-10 shows the image formed by light from a laser that was directed onto a circular aperture with a very small diameter. This image is not a point, as geometrical optics would suggest, but a circular disk surrounded by several progressively fainter secondary rings. Comparison with Fig. 36-1 leaves little doubt that we are dealing with a diffraction phenomenon. Here, however, the aperture is a circle of diameter d rather than a rectangular slit.

The (complex) analysis of such patterns shows that the first minimum for the diffraction pattern of a circular aperture of diameter d is located by

$$\sin \theta = 1.22 \frac{\lambda}{d} \quad (\text{first minimum—circular aperture}). \quad (36-12)$$

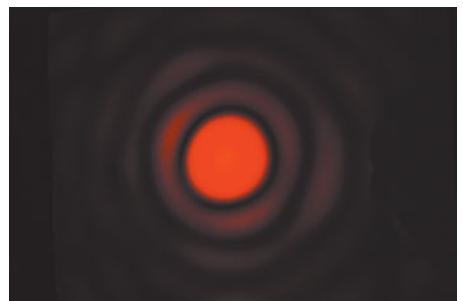
The angle θ here is the angle from the central axis to any point on that (circular) minimum. Compare this with Eq. 36-1,

$$\sin \theta = \frac{\lambda}{a} \quad (\text{first minimum—single slit}), \quad (36-13)$$

which locates the first minimum for a long narrow slit of width a . The main difference is the factor 1.22, which enters because of the circular shape of the aperture.

Resolvability

The fact that lens images are diffraction patterns is important when we wish to *resolve* (distinguish) two distant point objects whose angular separation is small. Figure 36-11 shows, in three different cases, the visual appearance and corresponding intensity pattern for two distant point objects (stars, say) with small



Courtesy Jearl Walker

Figure 36-10 The diffraction pattern of a circular aperture. Note the central maximum and the circular secondary maxima. The figure has been overexposed to bring out these secondary maxima, which are much less intense than the central maximum.

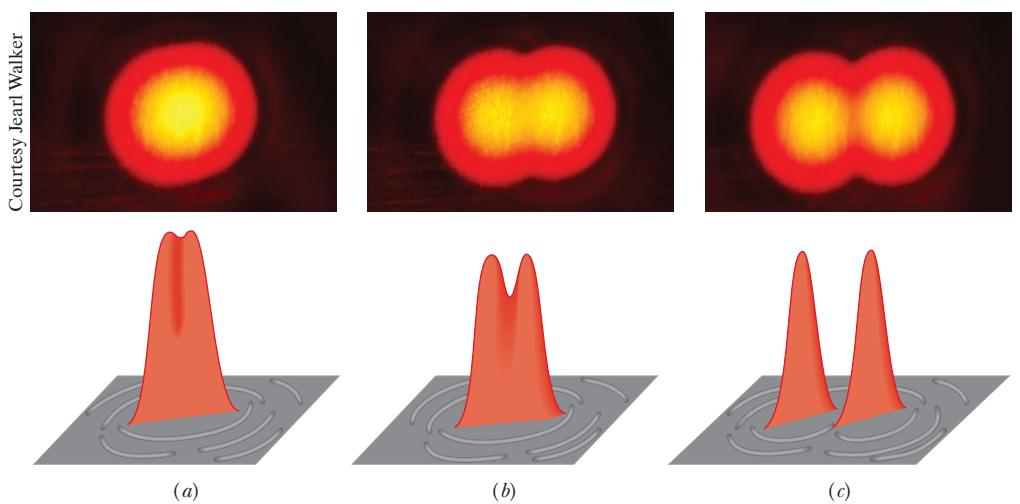


Figure 36-11 At the top, the images of two point sources (stars) formed by a converging lens. At the bottom, representations of the image intensities. In (a) the angular separation of the sources is too small for them to be distinguished, in (b) they can be marginally distinguished, and in (c) they are clearly distinguished. Rayleigh's criterion is satisfied in (b), with the central maximum of one diffraction pattern coinciding with the first minimum of the other.

angular separation. In Figure 36-11a, the objects are not resolved because of diffraction; that is, their diffraction patterns (mainly their central maxima) overlap so much that the two objects cannot be distinguished from a single point object. In Fig. 36-11b the objects are barely resolved, and in Fig. 36-11c they are fully resolved.

In Fig. 36-11b the angular separation of the two point sources is such that the central maximum of the diffraction pattern of one source is centered on the first minimum of the diffraction pattern of the other, a condition called **Rayleigh's criterion** for resolvability. From Eq. 36-12, two objects that are barely resolvable by this criterion must have an angular separation θ_R of

$$\theta_R = \sin^{-1} \frac{1.22\lambda}{d}.$$

Since the angles are small, we can replace $\sin \theta_R$ with θ_R expressed in radians:

$$\theta_R = 1.22 \frac{\lambda}{d} \quad (\text{Rayleigh's criterion}). \quad (36-14)$$

Human Vision. Applying Rayleigh's criterion for resolvability to human vision is only an approximation because visual resolvability depends on many factors, such as the relative brightness of the sources and their surroundings, turbulence in the air between the sources and the observer, and the functioning of the observer's visual system. Experimental results show that the least angular separation that can actually be resolved by a person is generally somewhat greater than the value given by Eq. 36-14. However, for calculations here, we shall take Eq. 36-14 as being a precise criterion: If the angular separation θ between the sources is greater than θ_R , we can visually resolve the sources; if it is less, we cannot.

Pointillism. Rayleigh's criterion can explain the arresting illusions of color in the style of painting known as pointillism (Fig. 36-12). In this style, a painting is made not with brush strokes in the usual sense but rather with a myriad of small colored dots. One fascinating aspect of a pointillistic painting is that when you change your distance from it, the colors shift in subtle, almost subconscious ways. This color shifting has to do with whether you can resolve the colored dots. When you stand close enough to the painting, the angular separations θ of adjacent dots are greater than θ_R and thus the dots can be seen individually. Their colors are the true colors of the paints used. However, when

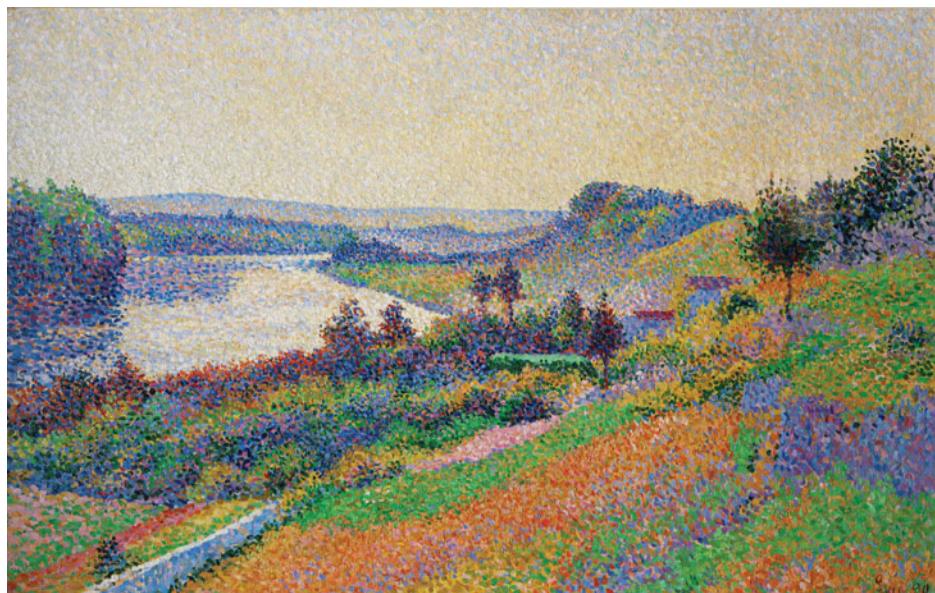


Figure 36-12 The pointillistic painting *The Seine at Herblay* by Maximilien Luce consists of thousands of colored dots. With the viewer very close to the canvas, the dots and their true colors are visible. At normal viewing distances, the dots are irresolvable and thus blend.

Maximilien Luce, *The Seine at Herblay*, 1890. Musée d'Orsay, Paris, France. Photo by Erich Lessing/Art Resource

you stand far enough from the painting, the angular separations θ are less than θ_R and the dots cannot be seen individually. The resulting blend of colors coming into your eye from any group of dots can then cause your brain to “make up” a color for that group—a color that may not actually exist in the group. In this way, a pointillistic painter uses your visual system to create the colors of the art.

When we wish to use a lens instead of our visual system to resolve objects of small angular separation, it is desirable to make the diffraction pattern as small as possible. According to Eq. 36-14, this can be done either by increasing the lens diameter or by using light of a shorter wavelength. For this reason ultraviolet light is often used with microscopes because its wavelength is shorter than a visible light wavelength.



Checkpoint 4

Suppose that you can barely resolve two red dots because of diffraction by the pupil of your eye. If we increase the general illumination around you so that the pupil decreases in diameter, does the resolvability of the dots improve or diminish? Consider only diffraction. (You might experiment to check your answer.)

Sample Problem 36.03 Pointillistic paintings use the diffraction of your eye

Figure 36-13a is a representation of the colored dots on a pointillistic painting. Assume that the average center-to-center separation of the dots is $D = 2.0 \text{ mm}$. Also assume that the diameter of the pupil of your eye is $d = 1.5 \text{ mm}$ and that the least angular separation between dots you can resolve is set only by Rayleigh’s criterion. What is the least viewing distance from which you cannot distinguish any dots on the painting?

KEY IDEA

Consider any two adjacent dots that you can distinguish when you are close to the painting. As you move away, you continue to distinguish the dots until their angular separation θ (in your view) has decreased to the angle given by

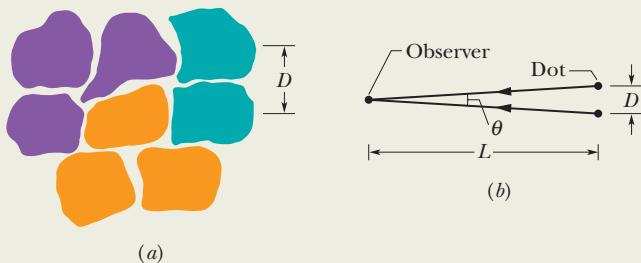


Figure 36-13 (a) Representation of some dots on a pointillistic painting, showing an average center-to-center separation D . (b) The arrangement of separation D between two dots, their angular separation θ , and the viewing distance L .

Rayleigh’s criterion:

$$\theta_R = 1.22 \frac{\lambda}{d}. \quad (36-15)$$

Calculations: Figure 36-13b shows, from the side, the angular separation θ of the dots, their center-to-center separation D , and your distance L from them. Because D/L is small, angle θ is also small and we can make the approximation

$$\theta = \frac{D}{L}. \quad (36-16)$$

Setting θ of Eq. 36-16 equal to θ_R of Eq. 36-15 and solving for L , we then have

$$L = \frac{Dd}{1.22\lambda}. \quad (36-17)$$

Equation 36-17 tells us that L is larger for smaller λ . Thus, as you move away from the painting, adjacent red dots (long wavelengths) become indistinguishable before adjacent blue dots do. To find the least distance L at which no colored dots are distinguishable, we substitute $\lambda = 400 \text{ nm}$ (blue or violet light) into Eq. 36-17:

$$L = \frac{(2.0 \times 10^{-3} \text{ m})(1.5 \times 10^{-3} \text{ m})}{(1.22)(400 \times 10^{-9} \text{ m})} = 6.1 \text{ m. (Answer)}$$

At this or a greater distance, the color you perceive at any given spot on the painting is a blended color that may not actually exist there.



Additional examples, video, and practice available at WileyPLUS

Sample Problem 36.04 Rayleigh's criterion for resolving two distant objects

A circular converging lens, with diameter $d = 32$ mm and focal length $f = 24$ cm, forms images of distant point objects in the focal plane of the lens. The wavelength is $\lambda = 550$ nm.

(a) Considering diffraction by the lens, what angular separation must two distant point objects have to satisfy Rayleigh's criterion?

KEY IDEA

Figure 36-14 shows two distant point objects P_1 and P_2 , the lens, and a viewing screen in the focal plane of the lens. It also shows, on the right, plots of light intensity I versus position on the screen for the central maxima of the images formed by the lens. Note that the angular separation θ_o of the objects equals the angular separation θ_i of the images. Thus, if the images are to satisfy Rayleigh's criterion, these separations must be given by Eq. 36-14 (for small angles).

Calculations: From Eq. 36-14, we obtain

$$\begin{aligned}\theta_o &= \theta_i = \theta_R = 1.22 \frac{\lambda}{d} \\ &= \frac{(1.22)(550 \times 10^{-9} \text{ m})}{32 \times 10^{-3} \text{ m}} = 2.1 \times 10^{-5} \text{ rad. (Answer)}\end{aligned}$$

Each central maximum in the two intensity curves of Fig. 36-14 is centered on the first minimum of the other curve.

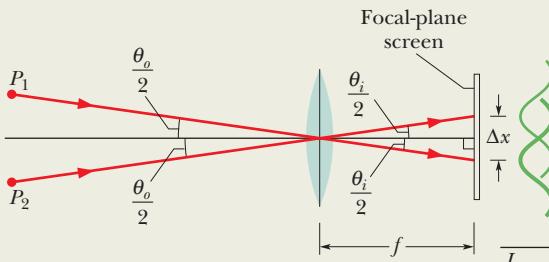


Figure 36-14 Light from two distant point objects P_1 and P_2 passes through a converging lens and forms images on a viewing screen in the focal plane of the lens. Only one representative ray from each object is shown. The images are not points but diffraction patterns, with intensities approximately as plotted at the right.

(b) What is the separation Δx of the centers of the images in the focal plane? (That is, what is the separation of the *central* peaks in the two intensity-versus-position curves?)

Calculations: From either triangle between the lens and the screen in Fig. 36-14, we see that $\tan \theta_i/2 = \Delta x/2f$. Rearranging this equation and making the approximation $\tan \theta \approx \theta$, we find

$$\Delta x = f\theta_i, \quad (36-18)$$

where θ_i is in radian measure. We then find

$$\Delta x = (0.24 \text{ m})(2.1 \times 10^{-5} \text{ rad}) = 5.0 \mu\text{m. (Answer)}$$



Additional examples, video, and practice available at WileyPLUS

36-4 DIFFRACTION BY A DOUBLE SLIT

Learning Objectives

After reading this module, you should be able to . . .

36.18 In a sketch of a double-slit experiment, explain how the diffraction through each slit modifies the two-slit interference pattern, and identify the diffraction envelope, the central peak, and the side peaks of that envelope.

36.19 For a given point in a double-slit diffraction pattern, calculate the intensity I in terms of the intensity I_m at the center of the pattern.

36.20 In the intensity equation for a double-slit diffraction

pattern, identify what part corresponds to the interference between the two slits and what part corresponds to the diffraction by each slit.

36.21 For double-slit diffraction, apply the relationship between the ratio d/a and the locations of the diffraction minima in the single-slit diffraction pattern, and then count the number of two-slit maxima that are contained in the central peak and in the side peaks of the diffraction envelope.

where I_m is the intensity at the center of the pattern,

$$\beta = \left(\frac{\pi d}{\lambda} \right) \sin \theta,$$

and

$$\alpha = \left(\frac{\pi a}{\lambda} \right) \sin \theta.$$

$$I(\theta) = I_m (\cos^2 \beta) \left(\frac{\sin \alpha}{\alpha} \right)^2 \quad (\text{double slit}),$$

Diffraction by a Double Slit

In the double-slit experiments of Chapter 35, we implicitly assumed that the slits were much narrower than the wavelength of the light illuminating them; that is, $a \ll \lambda$. For such narrow slits, the central maximum of the diffraction pattern of either slit covers the entire viewing screen. Moreover, the interference of light from the two slits produces bright fringes with approximately the same intensity (Fig. 35-12).

In practice with visible light, however, the condition $a \ll \lambda$ is often not met. For relatively wide slits, the interference of light from two slits produces bright fringes that do not all have the same intensity. That is, the intensities of the fringes produced by double-slit interference (as discussed in Chapter 35) are modified by diffraction of the light passing through each slit (as discussed in this chapter).

Plots. As an example, the intensity plot of Fig. 36-15a suggests the double-slit interference pattern that would occur if the slits were infinitely narrow (and thus $a \ll \lambda$); all the bright interference fringes would have the same intensity. The intensity plot of Fig. 36-15b is that for diffraction by a single actual slit; the diffraction pattern has a broad central maximum and weaker secondary maxima at $\pm 17^\circ$. The plot of Fig. 36-15c suggests the interference pattern for two actual slits. That plot was constructed by using the curve of Fig. 36-15b as an *envelope* on the intensity plot in Fig. 36-15a. The positions of the fringes are not changed; only the intensities are affected.

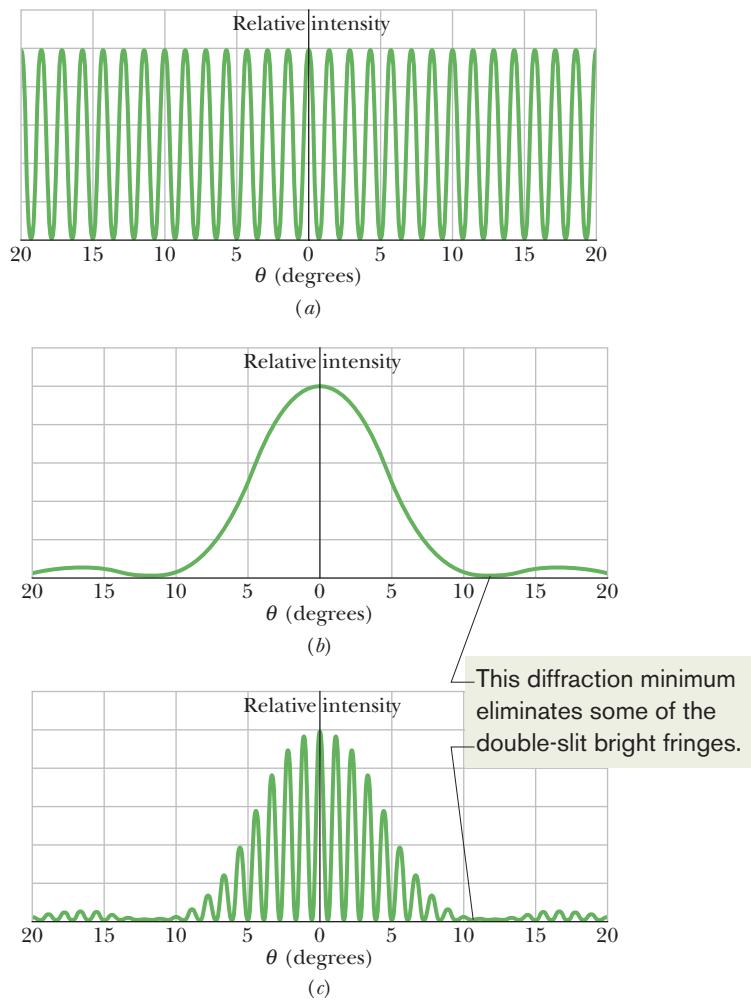


Figure 36-15 (a) The intensity plot to be expected in a double-slit interference experiment with vanishingly narrow slits. (b) The intensity plot for diffraction by a typical slit of width a (not vanishingly narrow). (c) The intensity plot to be expected for two slits of width a . The curve of (b) acts as an envelope, limiting the intensity of the double-slit fringes in (a). Note that the first minima of the diffraction pattern of (b) eliminate the double-slit fringes that would occur near 12° in (c).

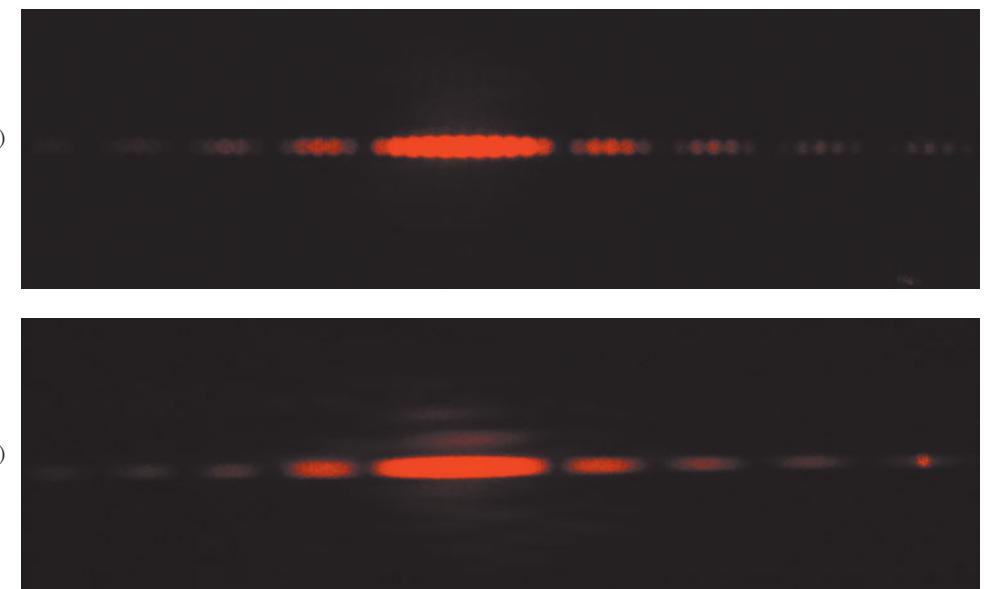


Figure 36-16 (a) Interference fringes for an actual double-slit system; compare with Fig. 36-15c. (b) The diffraction pattern of a single slit; compare with Fig. 36-15b.

Courtesy Jearl Walker

Photos. Figure 36-16a shows an actual pattern in which both double-slit interference and diffraction are evident. If one slit is covered, the single-slit diffraction pattern of Fig. 36-16b results. Note the correspondence between Figs. 36-16a and 36-15c, and between Figs. 36-16b and 36-15b. In comparing these figures, bear in mind that Fig. 36-16 has been deliberately overexposed to bring out the faint secondary maxima and that several secondary maxima (rather than one) are shown.

Intensity. With diffraction effects taken into account, the intensity of a double-slit interference pattern is given by

$$I(\theta) = I_m \left(\cos^2 \beta \right) \left(\frac{\sin \alpha}{\alpha} \right)^2 \quad (\text{double slit}), \quad (36-19)$$

in which

$$\beta = \frac{\pi d}{\lambda} \sin \theta \quad (36-20)$$

and

$$\alpha = \frac{\pi a}{\lambda} \sin \theta. \quad (36-21)$$

Here d is the distance between the centers of the slits and a is the slit width. Note carefully that the right side of Eq. 36-19 is the product of I_m and two factors. (1) The *interference factor* $\cos^2 \beta$ is due to the interference between two slits with slit separation d (as given by Eqs. 35-22 and 35-23). (2) The *diffraction factor* $[(\sin \alpha)/\alpha]^2$ is due to diffraction by a single slit of width a (as given by Eqs. 36-5 and 36-6).

Let us check these factors. If we let $a \rightarrow 0$ in Eq. 36-21, for example, then $\alpha \rightarrow 0$ and $(\sin \alpha)/\alpha \rightarrow 1$. Equation 36-19 then reduces, as it must, to an equation describing the interference pattern for a pair of vanishingly narrow slits with slit separation d . Similarly, putting $d = 0$ in Eq. 36-20 is equivalent physically to causing the two slits to merge into a single slit of width a . Then Eq. 36-20 yields $\beta = 0$ and $\cos^2 \beta = 1$. In this case Eq. 36-19 reduces, as it must, to an equation describing the diffraction pattern for a single slit of width a .

Language. The double-slit pattern described by Eq. 36-19 and displayed in Fig. 36-16a combines interference and diffraction in an intimate way. Both are superposition effects, in that they result from the combining of waves with different phases at a given point. If the combining waves originate from a small number of elementary coherent sources—as in a double-slit experiment with $a \ll \lambda$ —we call the

process *interference*. If the combining waves originate in a single wavefront—as in a single-slit experiment—we call the process *diffraction*. This distinction between interference and diffraction (which is somewhat arbitrary and not always adhered to) is a convenient one, but we should not forget that both are superposition effects and usually both are present simultaneously (as in Fig. 36-16a).

Sample Problem 36.05 Double-slit experiment with diffraction of each slit included

In a double-slit experiment, the wavelength λ of the light source is 405 nm, the slit separation d is 19.44 μm , and the slit width a is 4.050 μm . Consider the interference of the light from the two slits and also the diffraction of the light through each slit.

- (a) How many bright interference fringes are within the central peak of the diffraction envelope?

KEY IDEAS

We first analyze the two basic mechanisms responsible for the optical pattern produced in the experiment:

- Single-slit diffraction:** The limits of the central peak are the first minima in the diffraction pattern due to either slit individually. (See Fig. 36-15.) The angular locations of those minima are given by Eq. 36-3 ($a \sin \theta = m_1 \lambda$). Here let us rewrite this equation as $a \sin \theta = m_1 \lambda$, with the subscript 1 referring to the one-slit diffraction. For the first minima in the diffraction pattern, we substitute $m_1 = 1$, obtaining

$$a \sin \theta = \lambda. \quad (36-22)$$

- Double-slit interference:** The angular locations of the bright fringes of the double-slit interference pattern are given by Eq. 35-14, which we can write as

$$d \sin \theta = m_2 \lambda, \quad \text{for } m_2 = 0, 1, 2, \dots \quad (36-23)$$

Here the subscript 2 refers to the double-slit interference.

Calculations: We can locate the first diffraction minimum within the double-slit fringe pattern by dividing Eq. 36-23 by Eq. 36-22 and solving for m_2 . By doing so and then substituting the given data, we obtain

$$m_2 = \frac{d}{a} = \frac{19.44 \mu\text{m}}{4.050 \mu\text{m}} = 4.8.$$

This tells us that the bright interference fringe for $m_2 = 4$ fits into the central peak of the one-slit diffraction pattern, but the fringe for $m_2 = 5$ does not fit. Within the central diffraction peak we have the central bright fringe ($m_2 = 0$), and four bright fringes (up to $m_2 = 4$) on each side of it. Thus, a total of nine bright fringes of the double-slit interference pattern are within the central peak of the diffraction

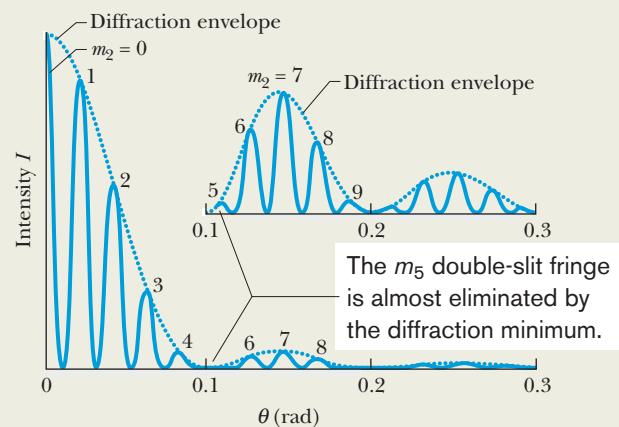


Figure 36-17 One side of the intensity plot for a two-slit interference experiment. The inset shows (vertically expanded) the plot within the first and second side peaks of the diffraction envelope.

envelope. The bright fringes to one side of the central bright fringe are shown in Fig. 36-17.

- (b) How many bright fringes are within either of the first side peaks of the diffraction envelope?

KEY IDEA

The outer limits of the first side diffraction peaks are the second diffraction minima, each of which is at the angle θ given by $a \sin \theta = m_1 \lambda$ with $m_1 = 2$:

$$a \sin \theta = 2\lambda. \quad (36-24)$$

Calculation: Dividing Eq. 36-23 by Eq. 36-24, we find

$$m_2 = \frac{2d}{a} = \frac{(2)(19.44 \mu\text{m})}{4.050 \mu\text{m}} = 9.6.$$

This tells us that the second diffraction minimum occurs just before the bright interference fringe for $m_2 = 10$ in Eq. 36-23. Within either first side diffraction peak we have the fringes from $m_2 = 5$ to $m_2 = 9$, for a total of five bright fringes of the double-slit interference pattern (shown in the inset of Fig. 36-17). However, if the $m_2 = 5$ bright fringe, which is almost eliminated by the first diffraction minimum, is considered too dim to count, then only four bright fringes are in the first side diffraction peak.



Additional examples, video, and practice available at WileyPLUS

36-5 DIFFRACTION GRATINGS

Learning Objectives

After reading this module, you should be able to ...

36.22 Describe a diffraction grating and sketch the interference pattern it produces in monochromatic light.

36.23 Distinguish the interference patterns of a diffraction grating and a double-slit arrangement.

36.24 Identify the terms line and order number.

36.25 For a diffraction grating, relate order number m to the path length difference of rays that give a bright fringe.

36.26 For a diffraction grating, relate the slit separation d , the angle θ to a bright fringe in the pattern, the order number

m of that fringe, and the wavelength λ of the light.

36.27 Identify the reason why there is a maximum order number for a given diffraction grating.

36.28 Explain the derivation of the equation for a line's half-width in a diffraction-grating pattern.

36.29 Calculate the half-width of a line at a given angle in a diffraction-grating pattern.

36.30 Explain the advantage of increasing the number of slits in a diffraction grating.

36.31 Explain how a grating spectroscope works.

Key Idea

- A diffraction grating is a series of "slits" used to separate an incident wave into its component wavelengths by separating and displaying their diffraction maxima. Diffraction by N (multiple) slits results in maxima (lines) at angles θ such that

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima}).$$

- A line's half-width is the angle from its center to the point where it disappears into the darkness and is given by

$$\Delta\theta_{hw} = \frac{\lambda}{Nd \cos \theta} \quad (\text{half-width}).$$

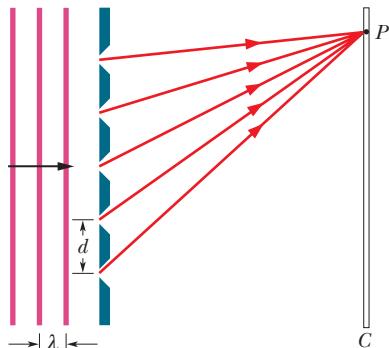


Figure 36-18 An idealized diffraction grating, consisting of only five rulings, that produces an interference pattern on a distant viewing screen C .

Diffraction Gratings

One of the most useful tools in the study of light and of objects that emit and absorb light is the **diffraction grating**. This device is somewhat like the double-slit arrangement of Fig. 35-10 but has a much greater number N of slits, often called *rulings*, perhaps as many as several thousand per millimeter. An idealized grating consisting of only five slits is represented in Fig. 36-18. When monochromatic light is sent through the slits, it forms narrow interference fringes that can be analyzed to determine the wavelength of the light. (Diffraction gratings can also be opaque surfaces with narrow parallel grooves arranged like the slits in Fig. 36-18. Light then scatters back from the grooves to form interference fringes rather than being transmitted through open slits.)

Pattern. With monochromatic light incident on a diffraction grating, if we gradually increase the number of slits from two to a large number N , the intensity plot changes from the typical double-slit plot of Fig. 36-15c to a much more complicated one and then eventually to a simple graph like that shown in Fig. 36-19a. The pattern you would see on a viewing screen using monochromatic red light from,

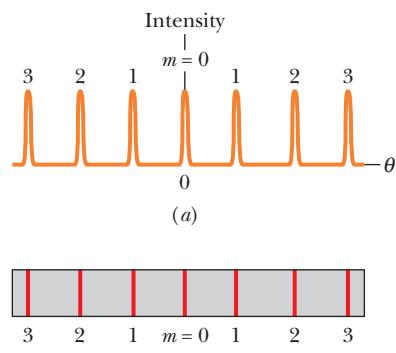


Figure 36-19 (a) The intensity plot produced by a diffraction grating with a great many rulings consists of narrow peaks, here labeled with their order numbers m . (b) The corresponding bright fringes seen on the screen are called lines and are here also labeled with order numbers m .

say, a helium-neon laser is shown in Fig. 36-19b. The maxima are now very narrow (and so are called *lines*); they are separated by relatively wide dark regions.

Equation. We use a familiar procedure to find the locations of the bright lines on the viewing screen. We first assume that the screen is far enough from the grating so that the rays reaching a particular point P on the screen are approximately parallel when they leave the grating (Fig. 36-20). Then we apply to each pair of adjacent rulings the same reasoning we used for double-slit interference. The separation d between rulings is called the *grating spacing*. (If N rulings occupy a total width w , then $d = w/N$.) The path length difference between adjacent rays is again $d \sin \theta$ (Fig. 36-20), where θ is the angle from the central axis of the grating (and of the diffraction pattern) to point P . A line will be located at P if the path length difference between adjacent rays is an integer number of wavelengths:

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima—lines}), \quad (36-25)$$

where λ is the wavelength of the light. Each integer m represents a different line; hence these integers can be used to label the lines, as in Fig. 36-19. The integers are then called the *order numbers*, and the lines are called the zeroth-order line (the central line, with $m = 0$), the first-order line ($m = 1$), the second-order line ($m = 2$), and so on.

Determining Wavelength. If we rewrite Eq. 36-25 as $\theta = \sin^{-1}(m\lambda/d)$, we see that, for a given diffraction grating, the angle from the central axis to any line (say, the third-order line) depends on the wavelength of the light being used. Thus, when light of an unknown wavelength is sent through a diffraction grating, measurements of the angles to the higher-order lines can be used in Eq. 36-25 to determine the wavelength. Even light of several unknown wavelengths can be distinguished and identified in this way. We cannot do that with the double-slit arrangement of Module 35-2, even though the same equation and wavelength dependence apply there. In double-slit interference, the bright fringes due to different wavelengths overlap too much to be distinguished.

Width of the Lines

A grating's ability to resolve (separate) lines of different wavelengths depends on the width of the lines. We shall here derive an expression for the *half-width* of the central line (the line for which $m = 0$) and then state an expression for the half-widths of the higher-order lines. We define the **half-width** of the central line as being the angle $\Delta\theta_{hw}$ from the center of the line at $\theta = 0$ outward to where the line effectively ends and darkness effectively begins with the first minimum (Fig. 36-21). At such a minimum, the N rays from the N slits of the grating cancel one another. (The actual width of the central line is, of course, $2(\Delta\theta_{hw})$, but line widths are usually compared via half-widths.)

In Module 36-1 we were also concerned with the cancellation of a great many rays, there due to diffraction through a single slit. We obtained Eq. 36-3, which, because of the similarity of the two situations, we can use to find the first minimum here. It tells us that the first minimum occurs where the path length difference between the top and bottom rays equals λ . For single-slit diffraction, this difference is $a \sin \theta$. For a grating of N rulings, each separated from the next by distance d , the distance between the top and bottom rulings is Nd (Fig. 36-22), and so the path length difference between the top and bottom rays here is $Nd \sin \Delta\theta_{hw}$. Thus, the first minimum occurs where

$$Nd \sin \Delta\theta_{hw} = \lambda. \quad (36-26)$$

Because $\Delta\theta_{hw}$ is small, $\sin \Delta\theta_{hw} = \Delta\theta_{hw}$ (in radian measure). Substituting this in Eq. 36-26 gives the half-width of the central line as

$$\Delta\theta_{hw} = \frac{\lambda}{Nd} \quad (\text{half-width of central line}). \quad (36-27)$$

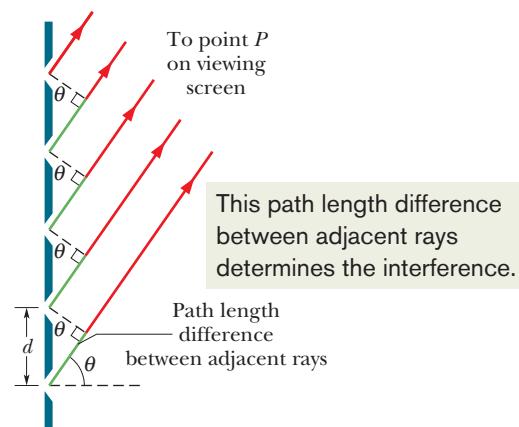


Figure 36-20 The rays from the rulings in a diffraction grating to a distant point P are approximately parallel. The path length difference between each two adjacent rays is $d \sin \theta$, where θ is measured as shown. (The rulings extend into and out of the page.)

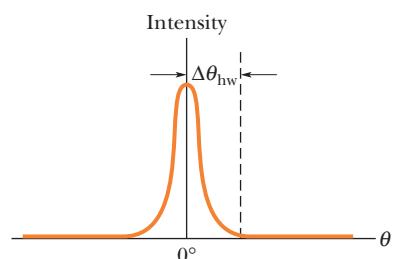


Figure 36-21 The half-width $\Delta\theta_{hw}$ of the central line is measured from the center of that line to the adjacent minimum on a plot of I versus θ like Fig. 36-19a.

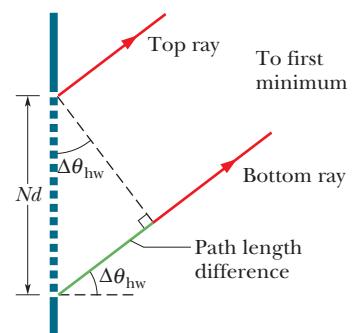


Figure 36-22 The top and bottom rulings of a diffraction grating of N rulings are separated by Nd . The top and bottom rays passing through these rulings have a path length difference of $Nd \sin \Delta\theta_{hw}$, where $\Delta\theta_{hw}$ is the angle to the first minimum. (The angle is here greatly exaggerated for clarity.)

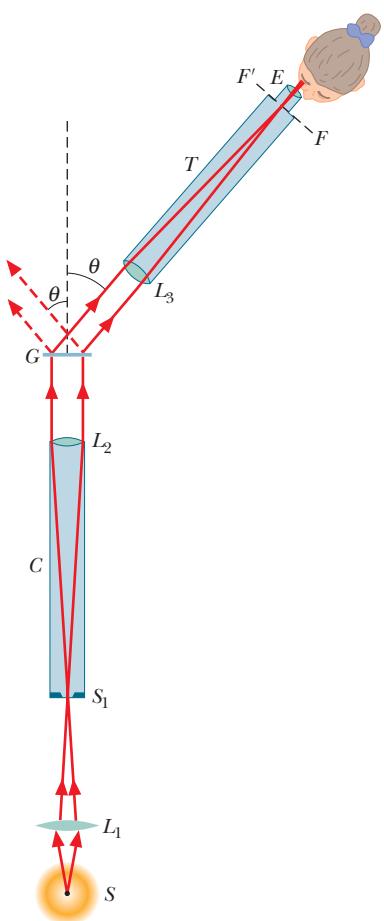


Figure 36-23 A simple type of grating spectroscope used to analyze the wavelengths of light emitted by source S .

We state without proof that the half-width of any other line depends on its location relative to the central axis and is

$$\Delta\theta_{hw} = \frac{\lambda}{Nd \cos \theta} \quad (\text{half-width of line at } \theta). \quad (36-28)$$

Note that for light of a given wavelength λ and a given ruling separation d , the widths of the lines decrease with an increase in the number N of rulings. Thus, of two diffraction gratings, the grating with the larger value of N is better able to distinguish between wavelengths because its diffraction lines are narrower and so produce less overlap.

Grating Spectroscope

Diffraction gratings are widely used to determine the wavelengths that are emitted by sources of light ranging from lamps to stars. Figure 36-23 shows a simple *grating spectroscope* in which a grating is used for this purpose. Light from source S is focused by lens L_1 on a vertical slit S_1 placed in the focal plane of lens L_2 . The light emerging from tube C (called a *collimator*) is a plane wave and is incident perpendicularly on grating G , where it is diffracted into a diffraction pattern, with the $m = 0$ order diffracted at angle $\theta = 0$ along the central axis of the grating.

We can view the diffraction pattern that would appear on a viewing screen at any angle θ simply by orienting telescope T in Fig. 36-23 to that angle. Lens L_3 of the telescope then focuses the light diffracted at angle θ (and at slightly smaller and larger angles) onto a focal plane FF' within the telescope. When we look through eyepiece E , we see a magnified view of this focused image.

By changing the angle θ of the telescope, we can examine the entire diffraction pattern. For any order number other than $m = 0$, the original light is spread out according to wavelength (or color) so that we can determine, with Eq. 36-25, just what wavelengths are being emitted by the source. If the source emits discrete wavelengths, what we see as we rotate the telescope horizontally through the angles corresponding to an order m is a vertical line of color for each wavelength, with the shorter-wavelength line at a smaller angle θ than the longer-wavelength line.

Hydrogen. For example, the light emitted by a hydrogen lamp, which contains hydrogen gas, has four discrete wavelengths in the visible range. If our eyes intercept this light directly, it appears to be white. If, instead, we view it through a grating spectroscope, we can distinguish, in several orders, the lines of the four colors corresponding to these visible wavelengths. (Such lines are called *emission lines*.) Four orders are represented in Fig. 36-24. In the central order ($m = 0$), the lines corresponding to all four wavelengths are superimposed, giving a single white line at $\theta = 0$. The colors are separated in the higher orders.

The third order is not shown in Fig. 36-24 for the sake of clarity; it actually overlaps the second and fourth orders. The fourth-order red line is missing because it is not formed by the grating used here. That is, when we attempt to

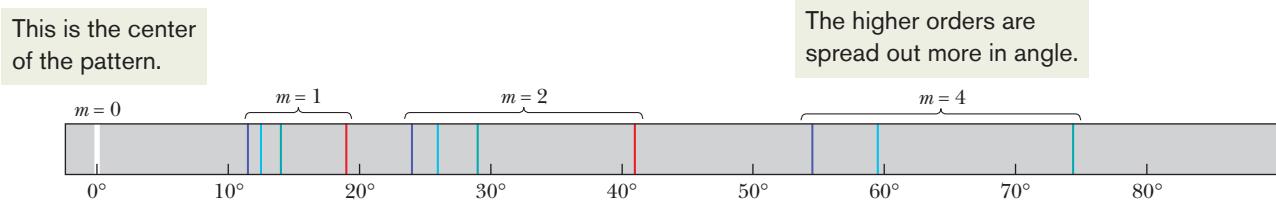


Figure 36-24 The zeroth, first, second, and fourth orders of the visible emission lines from hydrogen. Note that the lines are farther apart at greater angles. (They are also dimmer and wider, although that is not shown here.)

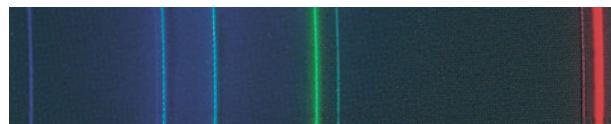


Figure 36-25 The visible emission lines of cadmium, as seen through a grating spectroscope.

Department of Physics, Imperial College/Science Photo Library/
Photo Researchers, Inc.

solve Eq. 36-25 for the angle θ for the red wavelength when $m = 4$, we find that $\sin \theta$ is greater than unity, which is not possible. The fourth order is then said to be *incomplete* for this grating; it might not be incomplete for a grating with greater spacing d , which will spread the lines less than in Fig. 36-24. Figure 36-25 is a photograph of the visible emission lines produced by cadmium.



Checkpoint 5

The figure shows lines of different orders produced by a diffraction grating in monochromatic red light. (a) Is the center of the pattern to the left or right? (b) In monochromatic green light, are the half-widths of the lines produced in the same orders greater than, less than, or the same as the half-widths of the lines shown?



36-6 GRATINGS: DISPERSION AND RESOLVING POWER

Learning Objectives

After reading this module, you should be able to . . .

- 36.32 Identify dispersion as the spreading apart of the diffraction lines associated with different wavelengths.
- 36.33 Apply the relationships between dispersion D , wavelength difference $\Delta\lambda$, angular separation $\Delta\theta$, slit separation d , order number m , and the angle θ corresponding to the order number.
- 36.34 Identify the effect on the dispersion of a diffraction

grating if the slit separation is varied.

- 36.35 Identify that for us to resolve lines, a diffraction grating must make them distinguishable.

36.36 Apply the relationship between resolving power R , wavelength difference $\Delta\lambda$, average wavelength λ_{avg} , number of rulings N , and order number m .

- 36.37 Identify the effect on the resolving power R if the number of slits N is increased.

Key Ideas

- The dispersion D of a diffraction grating is a measure of the angular separation $\Delta\theta$ of the lines it produces for two wavelengths differing by $\Delta\lambda$. For order number m , at angle θ , the dispersion is given by

$$D = \frac{\Delta\theta}{\Delta\lambda} = \frac{m}{d \cos \theta} \quad (\text{dispersion}).$$

- The resolving power R of a diffraction grating is a measure of its ability to make the emission lines of two close wavelengths distinguishable. For two wavelengths differing by $\Delta\lambda$ and with an average value of λ_{avg} , the resolving power is given by

$$R = \frac{\lambda_{\text{avg}}}{\Delta\lambda} = Nm \quad (\text{resolving power}).$$

Gratings: Dispersion and Resolving Power

Dispersion

To be useful in distinguishing wavelengths that are close to each other (as in a grating spectroscope), a grating must spread apart the diffraction lines associated with the various wavelengths. This spreading, called **dispersion**, is defined as

$$D = \frac{\Delta\theta}{\Delta\lambda} \quad (\text{dispersion defined}). \quad (36-29)$$



Kristen Brochmann/Fundamental Photographs

The fine rulings, each $0.5 \mu\text{m}$ wide, on a compact disc function as a diffraction grating. When a small source of white light illuminates a disc, the diffracted light forms colored “lanes” that are the composite of the diffraction patterns from the rulings. 

Here $\Delta\theta$ is the angular separation of two lines whose wavelengths differ by $\Delta\lambda$. The greater D is, the greater is the distance between two emission lines whose wavelengths differ by $\Delta\lambda$. We show below that the dispersion of a grating at angle θ is given by

$$D = \frac{m}{d \cos \theta} \quad (\text{dispersion of a grating}). \quad (36-30)$$

Thus, to achieve higher dispersion we must use a grating of smaller grating spacing d and work in a higher-order m . Note that the dispersion does not depend on the number of rulings N in the grating. The SI unit for D is the degree per meter or the radian per meter.

Resolving Power

To *resolve* lines whose wavelengths are close together (that is, to make the lines distinguishable), the line should also be as narrow as possible. Expressed otherwise, the grating should have a high **resolving power** R , defined as

$$R = \frac{\lambda_{\text{avg}}}{\Delta\lambda} \quad (\text{resolving power defined}). \quad (36-31)$$

Here λ_{avg} is the mean wavelength of two emission lines that can barely be recognized as separate, and $\Delta\lambda$ is the wavelength difference between them. The greater R is, the closer two emission lines can be and still be resolved. We shall show below that the resolving power of a grating is given by the simple expression

$$R = Nm \quad (\text{resolving power of a grating}). \quad (36-32)$$

To achieve high resolving power, we must use many rulings (large N).

Proof of Eq. 36-30

Let us start with Eq. 36-25, the expression for the locations of the lines in the diffraction pattern of a grating:

$$d \sin \theta = m\lambda.$$

Let us regard θ and λ as variables and take differentials of this equation. We find

$$d(\cos \theta) d\theta = m d\lambda.$$

For small enough angles, we can write these differentials as small differences, obtaining

$$d(\cos \theta) \Delta\theta = m \Delta\lambda \quad (36-33)$$

or
$$\frac{\Delta\theta}{\Delta\lambda} = \frac{m}{d \cos \theta}.$$

The ratio on the left is simply D (see Eq. 36-29), and so we have indeed derived Eq. 36-30.

Proof of Eq. 36-33

We start with Eq. 36-33, which was derived from Eq. 36-25, the expression for the locations of the lines in the diffraction pattern formed by a grating. Here $\Delta\lambda$ is the small wavelength difference between two waves that are diffracted by the grating, and $\Delta\theta$ is the angular separation between them in the diffraction pattern. If $\Delta\theta$ is to be the smallest angle that will permit the two lines to be resolved, it must (by Rayleigh's criterion) be equal to the half-width of each line, which is given by Eq. 36-28:

$$\Delta\theta_{\text{hw}} = \frac{\lambda}{Nd \cos \theta}.$$

Table 36-1 Three Gratings^a

Grating	N	d (nm)	θ	D ($^{\circ}/\mu\text{m}$)	R
<i>A</i>	10 000	2540	13.4°	23.2	10 000
<i>B</i>	20 000	2540	13.4°	23.2	20 000
<i>C</i>	10 000	1360	25.5°	46.3	10 000

^aData are for $\lambda = 589$ nm and $m = 1$.

If we substitute $\Delta\theta_{\text{hw}}$ as given here for $\Delta\theta$ in Eq. 36-33, we find that

$$\frac{\lambda}{N} = m \Delta\lambda,$$

from which it readily follows that

$$R = \frac{\lambda}{\Delta\lambda} = Nm.$$

This is Eq. 36-32, which we set out to derive.

Dispersion and Resolving Power Compared

The resolving power of a grating must not be confused with its dispersion. Table 36-1 shows the characteristics of three gratings, all illuminated with light of wavelength $\lambda = 589$ nm, whose diffracted light is viewed in the first order ($m = 1$ in Eq. 36-25). You should verify that the values of D and R as given in the table can be calculated with Eqs. 36-30 and 36-32, respectively. (In the calculations for D , you will need to convert radians per meter to degrees per micrometer.)

For the conditions noted in Table 36-1, gratings *A* and *B* have the same dispersion D and *A* and *C* have the same resolving power R .

Figure 36-26 shows the intensity patterns (also called *line shapes*) that would be produced by these gratings for two lines of wavelengths λ_1 and λ_2 , in the vicinity of $\lambda = 589$ nm. Grating *B*, with the higher resolving power, produces narrower lines and thus is capable of distinguishing lines that are much closer together in wavelength than those in the figure. Grating *C*, with the higher dispersion, produces the greater angular separation between the lines.

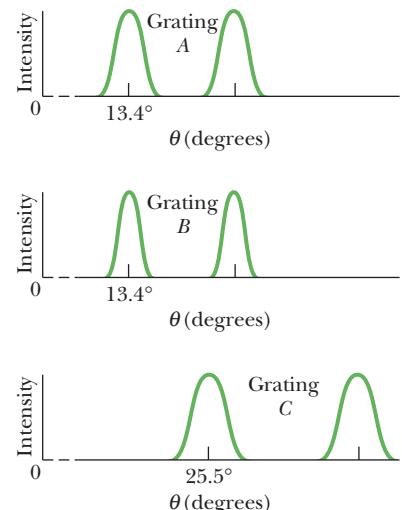


Figure 36-26 The intensity patterns for light of two wavelengths sent through the gratings of Table 36-1. Grating *B* has the highest resolving power, and grating *C* the highest dispersion.

Sample Problem 36.06 Dispersion and resolving power of a diffraction grating

A diffraction grating has 1.26×10^4 rulings uniformly spaced over width $w = 25.4$ mm. It is illuminated at normal incidence by yellow light from a sodium vapor lamp. This light contains two closely spaced emission lines (known as the sodium doublet) of wavelengths 589.00 nm and 589.59 nm.

(a) At what angle does the first-order maximum occur (on either side of the center of the diffraction pattern) for the wavelength of 589.00 nm?

KEY IDEA

The maxima produced by the diffraction grating can be determined with Eq. 36-25 ($d \sin \theta = m\lambda$).

Calculations: The grating spacing d is

$$d = \frac{w}{N} = \frac{25.4 \times 10^{-3} \text{ m}}{1.26 \times 10^4} = 2.016 \times 10^{-6} \text{ m} = 2016 \text{ nm}.$$

The first-order maximum corresponds to $m = 1$. Substituting these values for d and m into Eq. 36-25 leads to

$$\theta = \sin^{-1} \frac{m\lambda}{d} = \sin^{-1} \frac{(1)(589.00 \text{ nm})}{2016 \text{ nm}} = 16.99^\circ \approx 17.0^\circ. \quad (\text{Answer})$$

(b) Using the dispersion of the grating, calculate the angular separation between the two lines in the first order.

KEY IDEAS

(1) The angular separation $\Delta\theta$ between the two lines in the first order depends on their wavelength difference $\Delta\lambda$ and the dispersion D of the grating, according to Eq. 36-29 ($D = \Delta\theta/\Delta\lambda$). (2) The dispersion D depends on the angle θ at which it is to be evaluated.

Calculations: We can assume that, in the first order, the two sodium lines occur close enough to each other for us to evaluate D at the angle $\theta = 16.99^\circ$ we found in part (a) for one of those lines. Then Eq. 36-30 gives the dispersion as

$$D = \frac{m}{d \cos \theta} = \frac{1}{(2016 \text{ nm})(\cos 16.99^\circ)} \\ = 5.187 \times 10^{-4} \text{ rad/nm.}$$

From Eq. 36-29 and with $\Delta\lambda$ in nanometers, we then have

$$\Delta\theta = D \Delta\lambda = (5.187 \times 10^{-4} \text{ rad/nm})(589.59 - 589.00) \\ = 3.06 \times 10^{-4} \text{ rad} = 0.0175^\circ. \quad (\text{Answer})$$

You can show that this result depends on the grating spacing d but not on the number of rulings there are in the grating.

(c) What is the least number of rulings a grating can have and still be able to resolve the sodium doublet in the first order?

KEY IDEAS

(1) The resolving power of a grating in any order m is physically set by the number of rulings N in the grating according to Eq. 36-32 ($R = Nm$). (2) The smallest wavelength difference $\Delta\lambda$ that can be resolved depends on the average wavelength involved and on the resolving power R of the grating, according to Eq. 36-31 ($R = \lambda_{\text{avg}}/\Delta\lambda$).

Calculation: For the sodium doublet to be barely resolved, $\Delta\lambda$ must be their wavelength separation of 0.59 nm, and λ_{avg} must be their average wavelength of 589.30 nm. Thus, we find that the smallest number of rulings for a grating to resolve the sodium doublet is

$$N = \frac{R}{m} = \frac{\lambda_{\text{avg}}}{m \Delta\lambda} \\ = \frac{589.30 \text{ nm}}{(1)(0.59 \text{ nm})} = 999 \text{ rulings.} \quad (\text{Answer})$$



Additional examples, video, and practice available at WileyPLUS

36-7 X-RAY DIFFRACTION

Learning Objectives

After reading this module, you should be able to ...

36.38 Identify approximately where x rays are located in the electromagnetic spectrum.

36.39 Define a unit cell.

36.40 Define reflecting planes (or crystal planes) and interplanar spacing.

36.41 Sketch two rays that scatter from adjacent planes, showing the angle that is used in calculations.

36.42 For the intensity maxima in x-ray scattering by a crystal, apply the relationship between the interplanar spacing d , the angle θ of scattering, the order number m , and the wavelength λ of the x rays.

36.43 Given a drawing of a unit cell, demonstrate how an interplanar spacing can be determined.

Key Ideas

- If x rays are directed toward a crystal structure, they undergo Bragg scattering, which is easiest to visualize if the crystal atoms are considered to be in parallel planes.

- For x rays of wavelength λ scattering from crystal planes

with separation d , the angles θ at which the scattered intensity is maximum are given by

$$2d \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{Bragg's law}).$$

X-Ray Diffraction

X rays are electromagnetic radiation whose wavelengths are of the order of 1 Å ($= 10^{-10} \text{ m}$). Compare this with a wavelength of 550 nm ($= 5.5 \times 10^{-7} \text{ m}$) at the

center of the visible spectrum. Figure 36-27 shows that x rays are produced when electrons escaping from a heated filament F are accelerated by a potential difference V and strike a metal target T .

A standard optical diffraction grating cannot be used to discriminate between different wavelengths in the x-ray wavelength range. For $\lambda = 1 \text{ \AA}$ ($= 0.1 \text{ nm}$) and $d = 3000 \text{ nm}$, for example, Eq. 36-25 shows that the first-order maximum occurs at

$$\theta = \sin^{-1} \frac{m\lambda}{d} = \sin^{-1} \frac{(1)(0.1 \text{ nm})}{3000 \text{ nm}} = 0.0019^\circ.$$

This is too close to the central maximum to be practical. A grating with $d \approx \lambda$ is desirable, but, because x-ray wavelengths are about equal to atomic diameters, such gratings cannot be constructed mechanically.

In 1912, it occurred to German physicist Max von Laue that a crystalline solid, which consists of a regular array of atoms, might form a natural three-dimensional “diffraction grating” for x rays. The idea is that, in a crystal such as sodium chloride (NaCl), a basic unit of atoms (called the *unit cell*) repeats itself throughout the array. Figure 36-28a represents a section through a crystal of NaCl and identifies this basic unit. The unit cell is a cube measuring a_0 on each side.

When an x-ray beam enters a crystal such as NaCl, x rays are *scattered*—that is, redirected—in all directions by the crystal structure. In some directions the scattered waves undergo destructive interference, resulting in intensity minima; in other directions the interference is constructive, resulting in intensity maxima. This process of scattering and interference is a form of diffraction.

Fictional Planes. Although the process of diffraction of x rays by a crystal is complicated, the maxima turn out to be in directions *as if* the x rays were

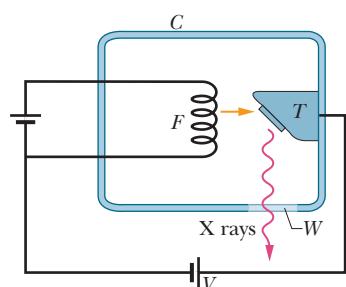


Figure 36-27 X rays are generated when electrons leaving heated filament F are accelerated through a potential difference V and strike a metal target T . The “window” W in the evacuated chamber C is transparent to x rays.

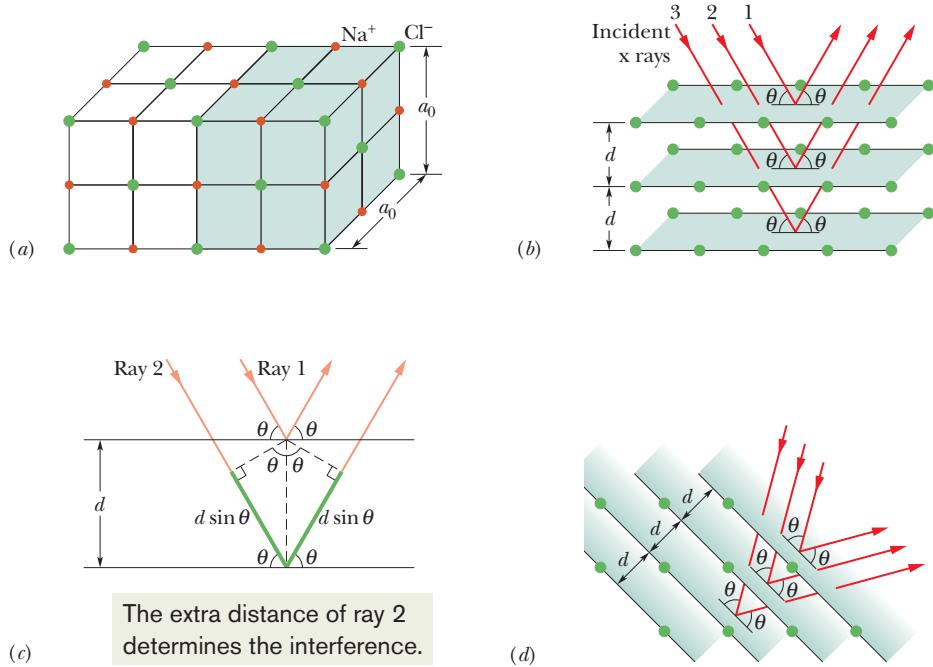


Figure 36-28 (a) The cubic structure of NaCl, showing the sodium and chlorine ions and a unit cell (shaded). (b) Incident x rays undergo diffraction by the structure of (a). The x rays are diffracted as if they were reflected by a family of parallel planes, with angles measured relative to the planes (not relative to a normal as in optics). (c) The path length difference between waves effectively reflected by two adjacent planes is $2d \sin \theta$. (d) A different orientation of the incident x rays relative to the structure. A different family of parallel planes now effectively reflects the x rays.

reflected by a family of parallel *reflecting planes* (or *crystal planes*) that extend through the atoms within the crystal and that contain regular arrays of the atoms. (The x rays are not actually reflected; we use these fictional planes only to simplify the analysis of the actual diffraction process.)

Figure 36-28b shows three reflecting planes (part of a family containing many parallel planes) with *interplanar spacing* d , from which the incident rays shown are said to reflect. Rays 1, 2, and 3 reflect from the first, second, and third planes, respectively. At each reflection the angle of incidence and the angle of reflection are represented with θ . Contrary to the custom in optics, these angles are defined relative to the *surface* of the reflecting plane rather than a normal to that surface. For the situation of Fig. 36-28b, the interplanar spacing happens to be equal to the unit cell dimension a_0 .

Figure 36-28c shows an edge-on view of reflection from an adjacent pair of planes. The waves of rays 1 and 2 arrive at the crystal in phase. After they are reflected, they must again be in phase because the reflections and the reflecting planes have been defined solely to explain the intensity maxima in the diffraction of x rays by a crystal. Unlike light rays, the x rays do not refract upon entering the crystal; moreover, we do not define an index of refraction for this situation. Thus, the relative phase between the waves of rays 1 and 2 as they leave the crystal is set solely by their path length difference. For these rays to be in phase, the path length difference must be equal to an integer multiple of the wavelength λ of the x rays.

Diffraction Equation. By drawing the dashed perpendiculars in Fig. 36-28c, we find that the path length difference is $2d \sin \theta$. In fact, this is true for any pair of adjacent planes in the family of planes represented in Fig. 36-28b. Thus, we have, as the criterion for intensity maxima for x-ray diffraction,

$$2d \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{Bragg's law}), \quad (36-34)$$

where m is the order number of an intensity maximum. Equation 36-34 is called **Bragg's law** after British physicist W. L. Bragg, who first derived it. (He and his father shared the 1915 Nobel Prize in physics for their use of x rays to study the structures of crystals.) The angle of incidence and reflection in Eq. 36-34 is called a *Bragg angle*.

Regardless of the angle at which x rays enter a crystal, there is always a family of planes from which they can be said to reflect so that we can apply Bragg's law. In Fig. 36-28d, notice that the crystal structure has the same orientation as it does in Fig. 36-28a, but the angle at which the beam enters the structure differs from that shown in Fig. 36-28b. This new angle requires a new family of reflecting planes, with a different interplanar spacing d and different Bragg angle θ , in order to explain the x-ray diffraction via Bragg's law.

Determining a Unit Cell. Figure 36-29 shows how the interplanar spacing d can be related to the unit cell dimension a_0 . For the particular family of planes shown there, the Pythagorean theorem gives

$$5d = \sqrt{\frac{5}{4}a_0^2},$$

or

$$d = \frac{a_0}{\sqrt{20}} = 0.2236a_0. \quad (36-35)$$

Figure 36-29 suggests how the dimensions of the unit cell can be found once the interplanar spacing has been measured by means of x-ray diffraction.

X-ray diffraction is a powerful tool for studying both x-ray spectra and the arrangement of atoms in crystals. To study spectra, a particular set of crystal planes, having a known spacing d , is chosen. These planes effectively reflect different wavelengths at different angles. A detector that can discriminate one angle from another can then be used to determine the wavelength of radiation reaching it. The crystal itself can be studied with a monochromatic x-ray beam, to determine not only the spacing of various crystal planes but also the structure of the unit cell.

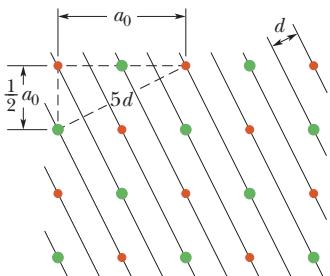


Figure 36-29 A family of planes through the structure of Fig. 36-28a, and a way to relate the edge length a_0 of a unit cell to the interplanar spacing d .

Review & Summary

Diffraction When waves encounter an edge, an obstacle, or an aperture the size of which is comparable to the wavelength of the waves, those waves spread out as they travel and, as a result, undergo interference. This is called **diffraction**.

Single-Slit Diffraction Waves passing through a long narrow slit of width a produce, on a viewing screen, a **single-slit diffraction pattern** that includes a central maximum and other maxima, separated by minima located at angles θ to the central axis that satisfy

$$a \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{minima}). \quad (36-3)$$

The intensity of the diffraction pattern at any given angle θ is

$$I(\theta) = I_m \left(\frac{\sin \alpha}{\alpha} \right)^2, \quad \text{where } \alpha = \frac{\pi a}{\lambda} \sin \theta \quad (36-5, 36-6)$$

and I_m is the intensity at the center of the pattern.

Circular-Aperture Diffraction Diffraction by a circular aperture or a lens with diameter d produces a central maximum and concentric maxima and minima, with the first minimum at an angle θ given by

$$\sin \theta = 1.22 \frac{\lambda}{d} \quad (\text{first minimum—circular aperture}). \quad (36-12)$$

Rayleigh's Criterion Rayleigh's criterion suggests that two objects are on the verge of resolvability if the central diffraction maximum of one is at the first minimum of the other. Their angular separation can then be no less than

$$\theta_R = 1.22 \frac{\lambda}{d} \quad (\text{Rayleigh's criterion}), \quad (36-14)$$

in which d is the diameter of the aperture through which the light passes.

Questions

1 You are conducting a single-slit diffraction experiment with light of wavelength λ . What appears, on a distant viewing screen, at a point at which the top and bottom rays through the slit have a path length difference equal to (a) 5λ and (b) 4.5λ ?

2 In a single-slit diffraction experiment, the top and bottom rays through the slit arrive at a certain point on the viewing screen with a path length difference of 4.0 wavelengths. In a phasor representation like those in Fig. 36-7, how many overlapping circles does the chain of phasors make?

3 For three experiments, Fig. 36-30 gives the parameter β of Eq. 36-20 versus angle θ for two-slit interference using light of wavelength 500 nm. The slit separations in the three experiments differ. Rank the experiments according to (a) the slit separations and (b) the total number of two-slit interference maxima in the pattern, greatest first.

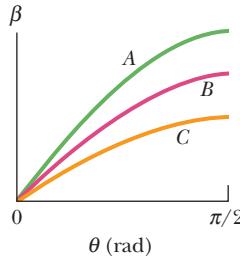


Figure 36-30 Question 3.

Double-Slit Diffraction Waves passing through two slits, each of width a , whose centers are a distance d apart, display diffraction patterns whose intensity I at angle θ is

$$I(\theta) = I_m (\cos^2 \beta) \left(\frac{\sin \alpha}{\alpha} \right)^2 \quad (\text{double slit}), \quad (36-19)$$

with $\beta = (\pi d / \lambda) \sin \theta$ and α as for single-slit diffraction.

Diffraction Gratings A **diffraction grating** is a series of “slits” used to separate an incident wave into its component wavelengths by separating and displaying their diffraction maxima. Diffraction by N (multiple) slits results in maxima (lines) at angles θ such that

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \dots \quad (\text{maxima}), \quad (36-25)$$

with the **half-widths** of the lines given by

$$\Delta\theta_{hw} = \frac{\lambda}{Nd \cos \theta} \quad (\text{half-widths}). \quad (36-28)$$

The dispersion D and resolving power R are given by

$$D = \frac{\Delta\theta}{\Delta\lambda} = \frac{m}{d \cos \theta} \quad (36-29, 36-30)$$

and

$$R = \frac{\lambda_{avg}}{\Delta\lambda} = Nm. \quad (36-31, 36-32)$$

X-Ray Diffraction The regular array of atoms in a crystal is a three-dimensional diffraction grating for short-wavelength waves such as x rays. For analysis purposes, the atoms can be visualized as being arranged in planes with characteristic interplanar spacing d . Diffraction maxima (due to constructive interference) occur if the incident direction of the wave, measured from the surfaces of these planes, and the wavelength λ of the radiation satisfy **Bragg's law**:

$$2d \sin \theta = m\lambda, \quad \text{for } m = 1, 2, 3, \dots \quad (\text{Bragg's law}). \quad (36-34)$$

4 For three experiments, Fig. 36-31 gives α versus angle θ in one-slit diffraction using light of wavelength 500 nm. Rank the experiments according to (a) the slit widths and (b) the total number of diffraction minima in the pattern, greatest first.

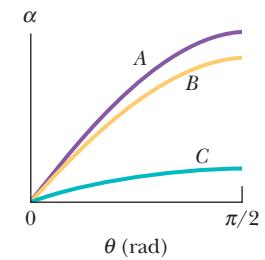


Figure 36-31 Question 4.

5 Figure 36-32 shows four choices for the rectangular opening of a source of either sound waves or light waves. The sides have lengths of either L or $2L$, with L being 3.0 times the wavelength of the waves. Rank the openings according to the extent of (a) left-right spreading and (b) up-down spreading of the waves due to diffraction, greatest first.

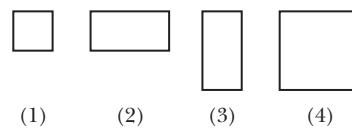


Figure 36-32 Question 5.

6 Light of frequency f illuminating a long narrow slit produces a diffraction pattern. (a) If we switch to light of frequency $1.3f$, does the pattern expand away from the center or contract toward the center? (b) Does the pattern expand or contract if, instead, we submerge the equipment in clear corn syrup?

7 At night many people see rings (called *entoptic halos*) surrounding bright outdoor lamps in otherwise dark surroundings. The rings are the first of the side maxima in diffraction patterns produced by structures that are thought to be within the cornea (or possibly the lens) of the observer's eye. (The central maxima of such patterns overlap the lamp.) (a) Would a particular ring become smaller or larger if the lamp were switched from blue to red light? (b) If a lamp emits white light, is blue or red on the outside edge of the ring?

8 (a) For a given diffraction grating, does the smallest difference $\Delta\lambda$ in two wavelengths that can be resolved increase, decrease, or remain the same as the wavelength increases? (b) For a given wavelength region (say, around 500 nm), is $\Delta\lambda$ greater in the first order or in the third order?

9 Figure 36-33 shows a red line and a green line of the same order in the pattern produced by a diffraction grating. If we increased the number of rulings in the grating—say, by removing tape that had covered the outer half of the rulings—would (a) the half-widths of the lines and (b) the separation of the lines increase, decrease, or remain the same? (c) Would the lines shift to the right, shift to the left, or remain in place?

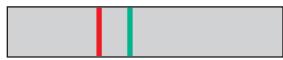


Figure 36-33 Questions 9 and 10.

10 For the situation of Question 9 and Fig. 36-33, if instead we increased the grating spacing, would (a) the half-widths of the lines and (b) the separation of the lines increase, decrease, or remain the same? (c) Would the lines shift to the right, shift to the left, or remain in place?

11 (a) Figure 36-34a shows the lines produced by diffraction gratings *A* and *B* using light of the same wavelength; the lines are of the same order and appear at the same angles θ . Which grating

has the greater number of rulings? (b) Figure 36-34b shows lines of two orders produced by a single diffraction grating using light of two wavelengths, both in the red region of the spectrum. Which lines, the left pair or right pair, are in the order with greater m ? Is the center of the diffraction pattern located to the left or to the right in (c) Fig. 36-34a and (d) Fig. 36-34b?

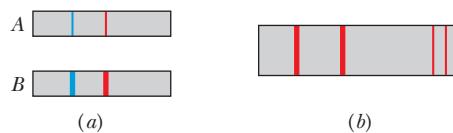


Figure 36-34 Question 11.

12 Figure 36-35 shows the bright fringes that lie within the central diffraction envelope in two double-slit diffraction experiments using the same wavelength of light. Are (a) the slit width a , (b) the slit separation d , and (c) the ratio d/a in experiment *B* greater than, less than, or the same as those quantities in experiment *A*?

13 In three arrangements you view two closely spaced small objects that are the same large distance from you. The angles that the objects occupy in your field of view and their distances from you are the following: (1) 2ϕ and R ; (2) 2ϕ and $2R$; (3) $\phi/2$ and $R/2$. (a) Rank the arrangements according to the separation between the objects, greatest first. If you can just barely resolve the two objects in arrangement 2, can you resolve them in (b) arrangement 1 and (c) arrangement 3?

14 For a certain diffraction grating, the ratio λ/a of wavelength to ruling spacing is 1/3.5. Without written calculation or use of a calculator, determine which of the orders beyond the zeroth order appear in the diffraction pattern.

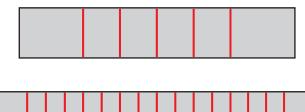


Figure 36-35 Question 12.

Problems



Tutoring problem available (at instructor's discretion) in WileyPLUS and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com

WWW Worked-out solution is at

ILW Interactive solution is at

<http://www.wiley.com/college/halliday>

Module 36-1 Single-Slit Diffraction

•1 The distance between the first and fifth minima of a single-slit diffraction pattern is 0.35 mm with the screen 40 cm away from the slit, when light of wavelength 550 nm is used. (a) Find the slit width. (b) Calculate the angle θ of the first diffraction minimum.

•2 What must be the ratio of the slit width to the wavelength for a single slit to have the first diffraction minimum at $\theta = 45.0^\circ$?

•3 A plane wave of wavelength 590 nm is incident on a slit with a width of $a = 0.40$ mm. A thin converging lens of focal length +70 cm is placed between the slit and a viewing screen and focuses the light on the screen. (a) How far is the screen from the lens? (b) What is the distance on the screen from the center of the diffraction pattern to the first minimum?

•4 In conventional television, signals are broadcast from towers to home receivers. Even when a receiver is not in direct view of a

tower because of a hill or building, it can still intercept a signal if the signal diffracts enough around the obstacle, into the obstacle's "shadow region." Previously, television signals had a wavelength of about 50 cm, but digital television signals that are transmitted from towers have a wavelength of about 10 mm. (a) Did this change in wavelength increase or decrease the diffraction of the signals into the shadow regions of obstacles? Assume that a signal passes through an opening of 5.0 m width between two adjacent buildings. What is the angular spread of the central diffraction maximum (out to the first minima) for wavelengths of (b) 50 cm and (c) 10 mm?

•5 A single slit is illuminated by light of wavelengths λ_a and λ_b , chosen so that the first diffraction minimum of the λ_a component coincides with the second minimum of the λ_b component. (a) If $\lambda_b = 350$ nm, what is λ_a ? For what order number m_b (if any) does a

minimum of the λ_b component coincide with the minimum of the λ_a component in the order number (b) $m_a = 2$ and (c) $m_a = 3$?

- 6 Monochromatic light of wavelength 441 nm is incident on a narrow slit. On a screen 2.00 m away, the distance between the second diffraction minimum and the central maximum is 1.50 cm. (a) Calculate the angle of diffraction θ of the second minimum. (b) Find the width of the slit.

•7 Light of wavelength 633 nm is incident on a narrow slit. The angle between the first diffraction minimum on one side of the central maximum and the first minimum on the other side is 1.20° . What is the width of the slit?

- 8 Sound waves with frequency 3000 Hz and speed 343 m/s diffract through the rectangular opening of a speaker cabinet and into a large auditorium of length $d = 100$ m. The opening, which has a horizontal width of 30.0 cm, faces a wall 100 m away (Fig. 36-36). Along that wall, how far from the central axis will a listener be at the first diffraction minimum and thus have difficulty hearing the sound? (Neglect reflections.)

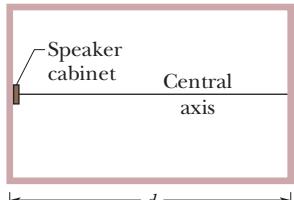


Figure 36-36 Problem 8.

•9 **SSM ILW** A slit 1.00 mm wide is illuminated by light of wavelength 589 nm. We see a diffraction pattern on a screen 3.00 m away. What is the distance between the first two diffraction minima on the same side of the central diffraction maximum?

- 10 **GO** Manufacturers of wire (and other objects of small dimension) sometimes use a laser to continually monitor the thickness of the product. The wire intercepts the laser beam, producing a diffraction pattern like that of a single slit of the same width as the wire diameter (Fig. 36-37). Suppose a helium-neon laser, of wavelength 632.8 nm, illuminates a wire, and the diffraction pattern appears on a screen at distance $L = 2.60$ m. If the desired wire diameter is 1.37 mm, what is the observed distance between the two tenth-order minima (one on each side of the central maximum)?

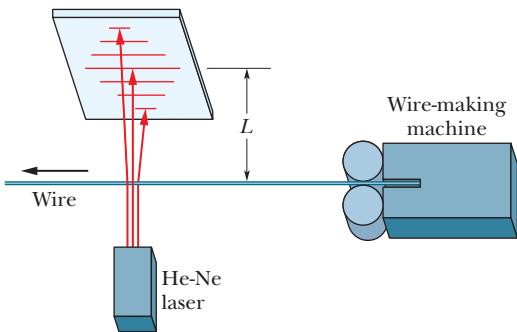


Figure 36-37 Problem 10.

Module 36-2 Intensity in Single-Slit Diffraction

- 11 A 0.10-mm-wide slit is illuminated by light of wavelength 589 nm. Consider a point P on a viewing screen on which the diffraction pattern of the slit is viewed; the point is at 30° from the central axis of the slit. What is the phase difference between the Huygens wavelets arriving at point P from the top and midpoint of the slit? (Hint: See Eq. 36-4.)

- 12 Figure 36-38 gives α versus the sine of the angle θ in a single-slit diffraction experiment using light of wavelength 610 nm. The vertical axis

scale is set by $\alpha_s = 12$ rad. What are (a) the slit width, (b) the total number of diffraction minima in the pattern (count them on both sides of the center of the diffraction pattern), (c) the least angle for a minimum, and (d) the greatest angle for a minimum?

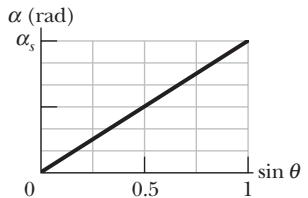


Figure 36-38 Problem 12.

- 13 Monochromatic light with wavelength 538 nm is incident on a slit with width 0.025 mm. The distance from the slit to a screen is 3.5 m. Consider a point on the screen 1.1 cm from the central maximum. Calculate (a) θ for that point, (b) α , and (c) the ratio of the intensity at that point to the intensity at the central maximum.

- 14 In the single-slit diffraction experiment of Fig. 36-4, let the wavelength of the light be 500 nm, the slit width be $6.00 \mu\text{m}$, and the viewing screen be at distance $D = 3.00$ m. Let a y axis extend upward along the viewing screen, with its origin at the center of the diffraction pattern. Also let I_P represent the intensity of the diffracted light at point P at $y = 15.0$ cm. (a) What is the ratio of I_P to the intensity I_m at the center of the pattern? (b) Determine where point P is in the diffraction pattern by giving the maximum and minimum between which it lies, or the two minima between which it lies.

- 15 **SSM WWW** The full width at half-maximum (FWHM) of a central diffraction maximum is defined as the angle between the two points in the pattern where the intensity is one-half that at the center of the pattern. (See Fig. 36-8b.) (a) Show that the intensity drops to one-half the maximum value when $\sin^2 \alpha = \alpha^2/2$. (b) Verify that $\alpha = 1.39$ rad (about 80°) is a solution to the transcendental equation of (a). (c) Show that the FWHM is $\Delta\theta = 2 \sin^{-1}(0.443\lambda/a)$, where a is the slit width. Calculate the FWHM of the central maximum for slit width (d) 1.00λ , (e) 5.00λ , and (f) 10.0λ .

- 16 **Babinet's principle.** A monochromatic beam of parallel light is incident on a “collimating” hole of diameter $x \gg \lambda$. Point P lies in the geometrical shadow region on a distant screen (Fig. 36-39a). Two diffracting objects, shown in Fig. 36-39b, are placed in turn over the collimating hole. Object A is an opaque circle with a hole in it, and B is the “photographic negative” of A . Using superposition concepts, show that the intensity at P is identical for the two diffracting objects A and B .

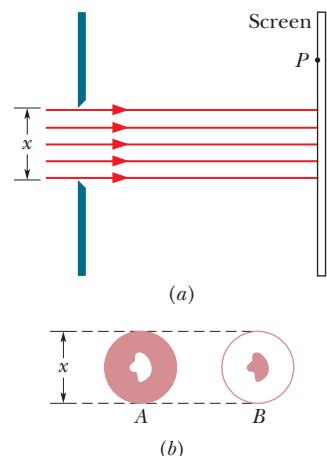


Figure 36-39 Problem 16.

- 17 (a) Show that the values of α at which intensity maxima for single-slit diffraction occur can be found exactly by differentiating Eq. 36-5 with respect to α and equating the result to zero, obtaining the condition $\tan \alpha = \alpha$. To find values of α satisfying this relation, plot the curve $y = \tan \alpha$ and the straight line $y = \alpha$ and then find their intersections, or use a calculator to find an appropriate value of α by trial and error. Next, from $\alpha = (m + \frac{1}{2})\pi$, determine the values of m associated with the maxima in the single-slit pattern. (These m values are *not* integers because secondary maxima do not lie exactly halfway between minima.) What are the (b) smallest α and (c) associated m , the (d) second smallest α and (e) associated m , and the (f) third smallest α and (g) associated m ?

Module 36-3 Diffraction by a Circular Aperture

•18 The wall of a large room is covered with acoustic tile in which small holes are drilled 5.0 mm from center to center. How far can a person be from such a tile and still distinguish the individual holes, assuming ideal conditions, the pupil diameter of the observer's eye to be 4.0 mm, and the wavelength of the room light to be 550 nm?

•19 (a) How far from grains of red sand must you be to position yourself just at the limit of resolving the grains if your pupil diameter is 1.5 mm, the grains are spherical with radius 50 μm , and the light from the grains has wavelength 650 nm? (b) If the grains were blue and the light from them had wavelength 400 nm, would the answer to (a) be larger or smaller?

•20 The radar system of a navy cruiser transmits at a wavelength of 1.6 cm, from a circular antenna with a diameter of 2.3 m. At a range of 6.2 km, what is the smallest distance that two speedboats can be from each other and still be resolved as two separate objects by the radar system?

•21 SSM WWW Estimate the linear separation of two objects on Mars that can just be resolved under ideal conditions by an observer on Earth (a) using the naked eye and (b) using the 200 in. (= 5.1 m) Mount Palomar telescope. Use the following data: distance to Mars = 8.0×10^7 km, diameter of pupil = 5.0 mm, wavelength of light = 550 nm.

•22 Assume that Rayleigh's criterion gives the limit of resolution of an astronaut's eye looking down on Earth's surface from a typical space shuttle altitude of 400 km. (a) Under that idealized assumption, estimate the smallest linear width on Earth's surface that the astronaut can resolve. Take the astronaut's pupil diameter to be 5 mm and the wavelength of visible light to be 550 nm. (b) Can the astronaut resolve the Great Wall of China (Fig. 36-40), which is more than 3000 km long, 5 to 10 m thick at its base, 4 m thick at its top, and 8 m in height? (c) Would the astronaut be able to resolve any unmistakable sign of intelligent life on Earth's surface?



©AP/Wide World Photos

Figure 36-40 Problem 22. The Great Wall of China.

•23 SSM The two headlights of an approaching automobile are 1.4 m apart. At what (a) angular separation and (b) maximum distance will the eye resolve them? Assume that the pupil diameter is 5.0 mm, and use a wavelength of 550 nm for the light. Also assume that diffraction effects alone limit the resolution so that Rayleigh's criterion can be applied.

•24 *Entoptic halos.* If someone looks at a bright outdoor lamp in otherwise dark surroundings, the lamp appears to be surrounded by bright and dark rings (hence *halos*) that are actually a circular diffraction pattern as in Fig. 36-10, with the central maximum overlapping the direct light from the lamp. The diffraction is produced by structures within the cornea or lens of the eye (hence *entoptic*). If the lamp is monochromatic at wavelength 550 nm and the first dark ring subtends angular diameter 2.5° in the observer's view, what is the (linear) diameter of the structure producing the diffraction?

•25 Find the separation of two points on the Moon's surface that can just be resolved by the 200 in. (= 5.1 m) telescope at Mount Palomar, assuming that this separation is determined by diffraction effects. The distance from Earth to the Moon is 3.8×10^5 km. Assume a wavelength of 550 nm for the light.

•26 The telescopes on some commercial surveillance satellites can resolve objects on the ground as small as 85 cm across (see Google Earth), and the telescopes on military surveillance satellites reportedly can resolve objects as small as 10 cm across. Assume first that object resolution is determined entirely by Rayleigh's criterion and is not degraded by turbulence in the atmosphere. Also assume that the satellites are at a typical altitude of 400 km and that the wavelength of visible light is 550 nm. What would be the required diameter of the telescope aperture for (a) 85 cm resolution and (b) 10 cm resolution? (c) Now, considering that turbulence is certain to degrade resolution and that the aperture diameter of the Hubble Space Telescope is 2.4 m, what can you say about the answer to (b) and about how the military surveillance resolutions are accomplished?

•27 If Superman really had x-ray vision at 0.10 nm wavelength and a 4.0 mm pupil diameter, at what maximum altitude could he distinguish villains from heroes, assuming that he needs to resolve points separated by 5.0 cm to do this?

•28 The wings of tiger beetles (Fig. 36-41) are colored by interference due to thin cuticle-like layers. In addition, these layers are arranged in patches that are 60 μm across and produce different colors. The color you see is a pointillistic mixture of thin-film interference colors that varies with perspective. Approximately



Kjell B. Sandved/Bruce Coleman, Inc./Photoshot Holdings Ltd.

Figure 36-41 Problem 28. Tiger beetles are colored by pointillistic mixtures of thin-film interference colors.

what viewing distance from a wing puts you at the limit of resolving the different colored patches according to Rayleigh's criterion? Use 550 nm as the wavelength of light and 3.00 mm as the diameter of your pupil.

••29 (a) What is the angular separation of two stars if their images are barely resolved by the Thaw refracting telescope at the Allegheny Observatory in Pittsburgh? The lens diameter is 76 cm and its focal length is 14 m. Assume $\lambda = 550$ nm. (b) Find the distance between these barely resolved stars if each of them is 10 light-years distant from Earth. (c) For the image of a single star in this telescope, find the diameter of the first dark ring in the diffraction pattern, as measured on a photographic plate placed at the focal plane of the telescope lens. Assume that the structure of the image is associated entirely with diffraction at the lens aperture and not with lens "errors."

••30 GO *Floating*. The floaters you see when viewing a bright, featureless background are diffraction patterns of defects in the vitreous humor that fills most of your eye. Sighting through a pinhole sharpens the diffraction pattern. If you also view a small circular dot, you can approximate the defect's size. Assume that the defect diffracts light as a circular aperture does. Adjust the dot's distance L from your eye (or eye lens) until the dot and the circle of the first minimum in the diffraction pattern appear to have the same size in your view. That is, until they have the same diameter D' on the retina at distance $L' = 2.0$ cm from the front of the eye, as suggested in Fig. 36-42a, where the angles on the two sides of the eye lens are equal. Assume that the wavelength of visible light is $\lambda = 550$ nm. If the dot has diameter $D = 2.0$ mm and is distance $L = 45.0$ cm from the eye and the defect is $x = 6.0$ mm in front of the retina (Fig. 36-42b), what is the diameter of the defect?

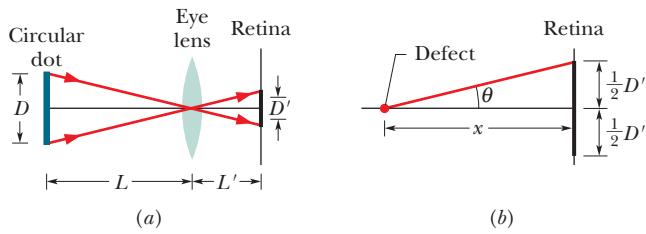


Figure 36-42 Problem 30.

••31 SSM Millimeter-wave radar generates a narrower beam than conventional microwave radar, making it less vulnerable to anti-radar missiles than conventional radar. (a) Calculate the angular width 2θ of the central maximum, from first minimum to first minimum, produced by a 220 GHz radar beam emitted by a 55.0-cm-diameter circular antenna. (The frequency is chosen to coincide with a low-absorption atmospheric "window.") (b) What is 2θ for a more conventional circular antenna that has a diameter of 2.3 m and emits at wavelength 1.6 cm?

••32 (a) A circular diaphragm 60 cm in diameter oscillates at a frequency of 25 kHz as an underwater source of sound used for submarine detection. Far from the source, the sound intensity is distributed as the diffraction pattern of a circular hole whose diameter equals that of the diaphragm. Take the speed of sound in water to be 1450 m/s and find the angle between the normal to the diaphragm and a line from the diaphragm to the first minimum. (b) Is there such a minimum for a source having an (audible) frequency of 1.0 kHz?

••33 GO Nuclear-pumped x-ray lasers are seen as a possible weapon to destroy ICBM booster rockets at ranges up to 2000 km.

One limitation on such a device is the spreading of the beam due to diffraction, with resulting dilution of beam intensity. Consider such a laser operating at a wavelength of 1.40 nm. The element that emits light is the end of a wire with diameter 0.200 mm. (a) Calculate the diameter of the central beam at a target 2000 km away from the beam source. (b) What is the ratio of the beam intensity at the target to that at the end of the wire? (The laser is fired from space, so neglect any atmospheric absorption.)

••34 GO A circular obstacle produces the same diffraction pattern as a circular hole of the same diameter (except very near $\theta = 0$). Airborne water drops are examples of such obstacles. When you see the Moon through suspended water drops, such as in a fog, you intercept the diffraction pattern from many drops. The composite of the central diffraction maxima of those drops forms a white region that surrounds the Moon and may obscure it. Figure 36-43 is a photograph in which the Moon is obscured. There are two faint, colored rings around the Moon (the larger one may be too faint to be seen in your copy of the photograph). The smaller ring is on the outer edge of the central maxima from the drops; the somewhat larger ring is on the outer edge of the smallest of the secondary maxima from the drops (see Fig. 36-10). The color is visible because the rings are adjacent to the diffraction minima (dark rings) in the patterns. (Colors in other parts of the pattern overlap too much to be visible.)

(a) What is the color of these rings on the outer edges of the diffraction maxima? (b) The colored ring around the central maxima in Fig. 36-43 has an angular diameter that is 1.35 times the angular diameter of the Moon, which is 0.50° . Assume that the drops all have about the same diameter. Approximately what is that diameter?



Pekka Parvianen/Photo Researchers, Inc.

Figure 36-43 Problem 34. The corona around the Moon is a composite of the diffraction patterns of airborne water drops.

Module 36-4 Diffraction by a Double Slit

•35 Suppose that the central diffraction envelope of a double-slit diffraction pattern contains 11 bright fringes and the first diffraction minima eliminate (are coincident with) bright fringes. How many bright fringes lie between the first and second minima of the diffraction envelope?

•36 A beam of light of a single wavelength is incident perpendicularly on a double-slit arrangement, as in Fig. 35-10. The slit widths

are each $46 \mu\text{m}$ and the slit separation is 0.30 mm . How many complete bright fringes appear between the two first-order minima of the diffraction pattern?

•37 In a double-slit experiment, the slit separation d is 2.00 times the slit width w . How many bright interference fringes are in the central diffraction envelope?

•38 In a certain two-slit interference pattern, 10 bright fringes lie within the second side peak of the diffraction envelope and diffraction minima coincide with two-slit interference maxima. What is the ratio of the slit separation to the slit width?

•39 Light of wavelength 440 nm passes through a double slit, yielding a diffraction pattern whose graph of intensity I versus angular position θ is shown in Fig. 36-44. Calculate (a) the slit width and (b) the slit separation. (c) Verify the displayed intensities of the $m = 1$ and $m = 2$ interference fringes.

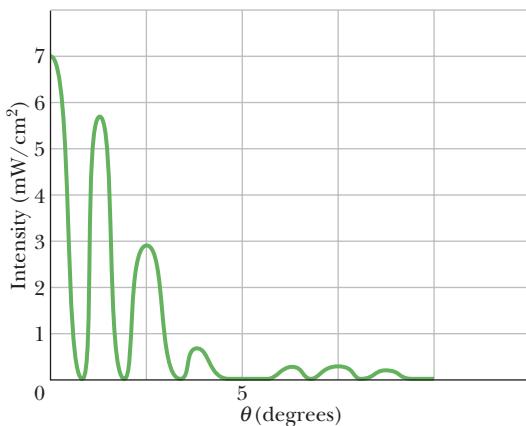


Figure 36-44 Problem 39.

•40 Figure 36-45 gives the parameter β of Eq. 36-20 versus the sine of the angle θ in a two-slit interference experiment using light of wavelength 435 nm . The vertical axis scale is set by $\beta_s = 80.0 \text{ rad}$. What are (a) the slit separation, (b) the total number of interference maxima (count them on both sides of the pattern's center), (c) the smallest angle for a maxima, and (d) the greatest angle for a minimum? Assume that none of the interference maxima are completely eliminated by a diffraction minimum.

•41 In the two-slit interference experiment of Fig. 35-10, the slit widths are each $12.0 \mu\text{m}$, their separation is $24.0 \mu\text{m}$, the wavelength is 600 nm , and the viewing screen is at a distance of 4.00 m . Let I_P represent the intensity at point P on the screen, at height $y = 70.0 \text{ cm}$. (a) What is the ratio of I_P to the intensity I_m at the center of the pattern? (b) Determine where P is in the two-slit interference pattern by giving the maximum or minimum on which it lies or the maximum and minimum between which it lies. (c) In the same way, for the diffraction that occurs, determine where point P is in the diffraction pattern.

•42 (a) In a double-slit experiment, what largest ratio of d to a causes diffraction to eliminate the fourth bright side fringe? (b) What other bright fringes are also eliminated? (c) How many other ratios of d to a cause the diffraction to (exactly) eliminate that bright fringe?

•43 (a) How many bright fringes appear between

the first diffraction-envelope minima to either side of the central maximum in a double-slit pattern if $\lambda = 550 \text{ nm}$, $d = 0.150 \text{ mm}$, and $a = 30.0 \mu\text{m}$? (b) What is the ratio of the intensity of the third bright fringe to the intensity of the central fringe?

Module 36-5 Diffraction Gratings

•44 Perhaps to confuse a predator, some tropical gyrid beetles (whirligig beetles) are colored by optical interference that is due to scales whose alignment forms a diffraction grating (which scatters light instead of transmitting it). When the incident light rays are perpendicular to the grating, the angle between the first-order maxima (on opposite sides of the zeroth-order maximum) is about 26° in light with a wavelength of 550 nm . What is the grating spacing of the beetle?

•45 A diffraction grating 20.0 mm wide has 6000 rulings. Light of wavelength 589 nm is incident perpendicularly on the grating. What are the (a) largest, (b) second largest, and (c) third largest values of θ at which maxima appear on a distant viewing screen?

•46 Visible light is incident perpendicularly on a grating with 315 rulings/mm. What is the longest wavelength that can be seen in the fifth-order diffraction?

•47 A grating has 400 lines/mm. How many orders of the entire visible spectrum (400 – 700 nm) can it produce in a diffraction experiment, in addition to the $m = 0$ order?

•48 A diffraction grating is made up of slits of width 300 nm with separation 900 nm . The grating is illuminated by monochromatic plane waves of wavelength $\lambda = 600 \text{ nm}$ at normal incidence. (a) How many maxima are there in the full diffraction pattern? (b) What is the angular width of a spectral line observed in the first order if the grating has 1000 slits?

•49 Light of wavelength 600 nm is incident normally on a diffraction grating. Two adjacent maxima occur at angles given by $\sin \theta = 0.2$ and $\sin \theta = 0.3$. The fourth-order maxima are missing. (a) What is the separation between adjacent slits? (b) What is the smallest slit width this grating can have? For that slit width, what are the (c) largest, (d) second largest, and (e) third largest values of the order number m of the maxima produced by the grating?

•50 With light from a gaseous discharge tube incident normally on a grating with slit separation $1.73 \mu\text{m}$, sharp maxima of green light are experimentally found at angles $\theta = \pm 17.6^\circ, 37.3^\circ, -37.1^\circ, 65.2^\circ$, and -65.0° . Compute the wavelength of the green light that best fits these data.

•51 A diffraction grating having 180 lines/mm is illuminated with a light signal containing only two wavelengths, $\lambda_1 = 400 \text{ nm}$ and $\lambda_2 = 500 \text{ nm}$. The signal is incident perpendicularly on the grating. (a) What is the angular separation between the second-order maxima of these two wavelengths? (b) What is the smallest angle at which two of the resulting maxima are superimposed? (c) What is the highest order for which maxima for both wavelengths are present in the diffraction pattern?

•52 A beam of light consisting of wavelengths from 460.0 nm to 640.0 nm is directed perpendicularly onto a diffraction grating with 160 lines/mm. (a) What is the lowest order that is overlapped by another order? (b) What is the highest order for which the complete wavelength range of the beam is present? In that highest order, at what angle does the light at wavelength (c) 460.0 nm and (d) 640.0 nm appear? (e) What is the greatest angle at which the light at wavelength 460.0 nm appears?

•53 A grating has 350 rulings/mm and is illuminated at normal

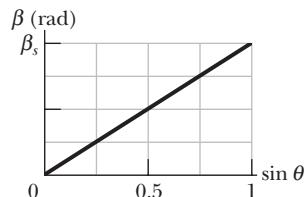


Figure 36-45 Problem 40.

incidence by white light. A spectrum is formed on a screen 30.0 cm from the grating. If a hole 10.0 mm square is cut in the screen, its inner edge being 50.0 mm from the central maximum and parallel to it, what are the (a) shortest and (b) longest wavelengths of the light that passes through the hole?

••54 Derive this expression for the intensity pattern for a three-slit “grating”:

$$I = \frac{1}{9} I_m (1 + 4 \cos \phi + 4 \cos^2 \phi),$$

where $\phi = (2\pi d \sin \theta)/\lambda$ and $a \ll \lambda$.

Module 36-6 Gratings: Dispersion and Resolving Power

•55 SSM ILW A source containing a mixture of hydrogen and deuterium atoms emits red light at two wavelengths whose mean is 656.3 nm and whose separation is 0.180 nm. Find the minimum number of lines needed in a diffraction grating that can resolve these lines in the first order.

•56 (a) How many rulings must a 4.00-cm-wide diffraction grating have to resolve the wavelengths 415.496 and 415.487 nm in the second order? (b) At what angle are the second-order maxima found?

•57 Light at wavelength 589 nm from a sodium lamp is incident perpendicularly on a grating with 40 000 rulings over width 76 nm. What are the first-order (a) dispersion D and (b) resolving power R , the second-order (c) D and (d) R , and the third-order (e) D and (f) R ?

•58 A grating has 600 rulings/mm and is 5.0 mm wide. (a) What is the smallest wavelength interval it can resolve in the third order at $\lambda = 500$ nm? (b) How many higher orders of maxima can be seen?

•59 A diffraction grating with a width of 2.0 cm contains 1000 lines/cm across that width. For an incident wavelength of 600 nm, what is the smallest wavelength difference this grating can resolve in the second order?

•60 The D line in the spectrum of sodium is a doublet with wavelengths 589.0 and 589.6 nm. Calculate the minimum number of lines needed in a grating that will resolve this doublet in the second-order spectrum.

•61 With a particular grating the sodium doublet (589.00 nm and 589.59 nm) is viewed in the third order at 10° to the normal and is barely resolved. Find (a) the grating spacing and (b) the total width of the rulings.

•62 A diffraction grating illuminated by monochromatic light normal to the grating produces a certain line at angle θ . (a) What is the product of that line's half-width and the grating's resolving power? (b) Evaluate that product for the first order of a grating of slit separation 900 nm in light of wavelength 600 nm.

•63 Assume that the limits of the visible spectrum are arbitrarily chosen as 430 and 680 nm. Calculate the number of rulings per millimeter of a grating that will spread the first-order spectrum through an angle of 20.0° .

Module 36-7 X-Ray Diffraction

•64 What is the smallest Bragg angle for x rays of wavelength 30 pm to reflect from reflecting planes spaced 0.30 nm apart in a calcite crystal?

•65 An x-ray beam of wavelength A undergoes first-order reflection (Bragg law diffraction) from a crystal when its angle of incidence to a crystal face is 23° , and an x-ray beam of wavelength 97 pm undergoes third-order reflection when its angle of incidence to that face is 60° . Assuming that the two beams reflect from the same family of reflecting planes, find (a) the interplanar spacing and (b) the wavelength A .

•66 An x-ray beam of a certain wavelength is incident on an NaCl crystal, at 30.0° to a certain family of reflecting planes of spacing 39.8 pm. If the reflection from those planes is of the first order, what is the wavelength of the x rays?

•67 Figure 36-46 is a graph of intensity versus angular position θ for the diffraction of an x-ray beam by a crystal. The horizontal scale is set by $\theta_s = 2.00^\circ$. The beam consists of two wavelengths, and the spacing between the reflecting planes is 0.94 nm. What are the (a) shorter and (b) longer wavelengths in the beam?

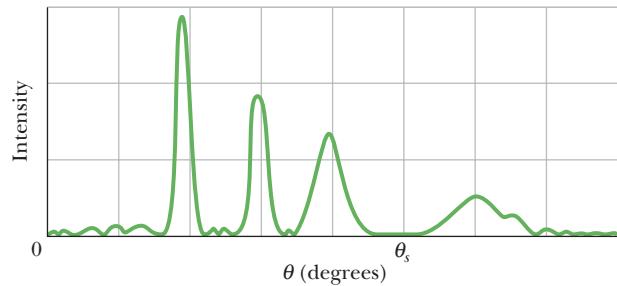


Figure 36-46 Problem 67.

•68 If first-order reflection occurs in a crystal at Bragg angle 3.4° , at what Bragg angle does second-order reflection occur from the same family of reflecting planes?

•69 X rays of wavelength 0.12 nm are found to undergo second-order reflection at a Bragg angle of 28° from a lithium fluoride crystal. What is the interplanar spacing of the reflecting planes in the crystal?

•70 GO In Fig. 36-47, first-order reflection from the reflection planes shown occurs when an x-ray beam of wavelength 0.260 nm makes an angle $\theta = 63.8^\circ$ with the top face of the crystal. What is the unit cell size a_0 ?

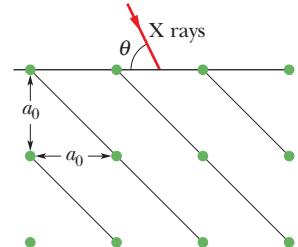


Figure 36-47 Problem 70.

•71 WWW In Fig. 36-48, let a beam of x rays of wavelength 0.125 nm be incident on an NaCl crystal at angle $\theta = 45.0^\circ$ to the top face of the crystal and a family of reflecting planes. Let the reflecting planes have separation $d = 0.252$ nm. The crystal is turned through angle ϕ around an axis perpendicular to the plane of the page until these reflecting planes give diffraction maxima. What are the (a) smaller and (b) larger value of ϕ if the crystal is turned clockwise and the (c) smaller and (d) larger value of ϕ if it is turned counter-clockwise?

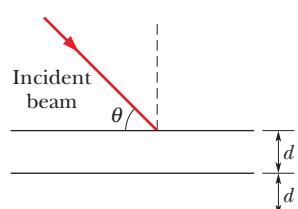


Figure 36-48 Problems 71 and 72.

•72 In Fig. 36-48, an x-ray beam of wavelengths from 95.0 to 140 pm is incident at $\theta = 45.0^\circ$ to a family of reflecting planes with spacing $d = 275$ pm. What are the (a) longest wavelength λ and (b) associated order number m and the (c) shortest λ and (d) associated m of the intensities maxima in the diffraction of the beam?

•73 Consider a two-dimensional square crystal structure, such as one side of the structure shown in Fig. 36-28a. The largest interplanar spacing of reflecting planes is the unit cell size a_0 . Calculate and sketch the (a) second largest, (b) third largest, (c) fourth largest, (d)

fifth largest, and (e) sixth largest interplanar spacing. (f) Show that your results in (a) through (e) are consistent with the general formula

$$d = \frac{a_0}{\sqrt{h^2 + k^2}},$$

where h and k are relatively prime integers (they have no common factor other than unity).

Additional Problems

74 An astronaut in a space shuttle claims she can just barely resolve two point sources on Earth's surface, 160 km below. Calculate their (a) angular and (b) linear separation, assuming ideal conditions. Take $\lambda = 540$ nm and the pupil diameter of the astronaut's eye to be 5.0 mm.

75 SSM Visible light is incident perpendicularly on a diffraction grating of 200 rulings/mm. What are the (a) longest, (b) second longest, and (c) third longest wavelengths that can be associated with an intensity maximum at $\theta = 30.0^\circ$?

76 A beam of light consists of two wavelengths, 590.159 nm and 590.220 nm, that are to be resolved with a diffraction grating. If the grating has lines across a width of 3.80 cm, what is the minimum number of lines required for the two wavelengths to be resolved in the second order?

77 SSM In a single-slit diffraction experiment, there is a minimum of intensity for orange light ($\lambda = 600$ nm) and a minimum of intensity for blue-green light ($\lambda = 500$ nm) at the same angle of 1.00 mrad. For what minimum slit width is this possible?

78 GO A double-slit system with individual slit widths of 0.030 mm and a slit separation of 0.18 mm is illuminated with 500 nm light directed perpendicular to the plane of the slits. What is the total number of complete bright fringes appearing between the two first-order minima of the diffraction pattern? (Do not count the fringes that coincide with the minima of the diffraction pattern.)

79 SSM A diffraction grating has resolving power $R = \lambda_{\text{avg}}/\Delta\lambda = Nm$. (a) Show that the corresponding frequency range Δf that can just be resolved is given by $\Delta f = c/Nm\lambda$. (b) From Fig. 36-22, show that the times required for light to travel along the ray at the bottom of the figure and the ray at the top differ by $\Delta t = (Nd/c) \sin \theta$. (c) Show that $(\Delta f)(\Delta t) = 1$, this relation being independent of the various grating parameters. Assume $N \gg 1$.

80 The pupil of a person's eye has a diameter of 5.00 mm. According to Rayleigh's criterion, what distance apart must two small objects be if their images are just barely resolved when they are 250 mm from the eye? Assume they are illuminated with light of wavelength 500 nm.

81 Light is incident on a grating at an angle ψ as shown in Fig. 36-49.

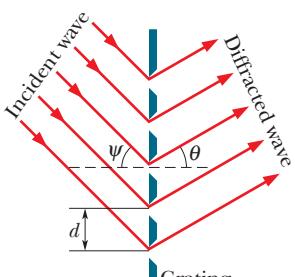


Figure 36-49 Problem 81.

Show that bright fringes occur at angles θ that satisfy the equation

$$d(\sin \psi + \sin \theta) = m\lambda, \quad \text{for } m = 0, 1, 2, \dots$$

(Compare this equation with Eq. 36-25.) Only the special case $\psi = 0$ has been treated in this chapter.

82 A grating with $d = 1.50 \mu\text{m}$ is illuminated at various angles of incidence by light of wavelength 600 nm. Plot, as a function of the angle of incidence (0 to 90°), the angular deviation of the first-order maximum from the incident direction. (See Problem 81.)

83 SSM In two-slit interference, if the slit separation is $14 \mu\text{m}$ and the slit widths are each $2.0 \mu\text{m}$, (a) how many two-slit maxima are in the central peak of the diffraction envelope and (b) how many are in either of the first side peak of the diffraction envelope?

84 GO In a two-slit interference pattern, what is the ratio of slit separation to slit width if there are 17 bright fringes within the central diffraction envelope and the diffraction minima coincide with two-slit interference maxima?

85 A beam of light with a narrow wavelength range centered on 450 nm is incident perpendicularly on a diffraction grating with a width of 1.80 cm and a line density of 1400 lines/cm across that width. For this light, what is the smallest wavelength difference this grating can resolve in the third order?

86 If you look at something 40 m from you, what is the smallest length (perpendicular to your line of sight) that you can resolve, according to Rayleigh's criterion? Assume the pupil of your eye has a diameter of 4.00 mm, and use 500 nm as the wavelength of the light reaching you.

87 Two yellow flowers are separated by 60 cm along a line perpendicular to your line of sight to the flowers. How far are you from the flowers when they are at the limit of resolution according to the Rayleigh criterion? Assume the light from the flowers has a single wavelength of 550 nm and that your pupil has a diameter of 5.5 mm.

88 In a single-slit diffraction experiment, what must be the ratio of the slit width to the wavelength if the second diffraction minima are to occur at an angle of 37.0° from the center of the diffraction pattern on a viewing screen?

89 A diffraction grating 3.00 cm wide produces the second order at 33.0° with light of wavelength 600 nm. What is the total number of lines on the grating?

90 A single-slit diffraction experiment is set up with light of wavelength 420 nm, incident perpendicularly on a slit of width $5.10 \mu\text{m}$. The viewing screen is 3.20 m distant. On the screen, what is the distance between the center of the diffraction pattern and the second diffraction minimum?

91 A diffraction grating has 8900 slits across 1.20 cm. If light with a wavelength of 500 nm is sent through it, how many orders (maxima) lie to one side of the central maximum?

92 In an experiment to monitor the Moon's surface with a light beam, pulsed radiation from a ruby laser ($\lambda = 0.69 \mu\text{m}$) was directed to the Moon through a reflecting telescope with a mirror radius of 1.3 m. A reflector on the Moon behaved like a circular flat mirror with radius 10 cm, reflecting the light directly back toward the telescope on Earth. The reflected light was then detected after being brought to a focus by this telescope. Approximately what fraction of the original light energy was picked up by the detector? Assume that for each direction of travel all the energy is in the central diffraction peak.

93 In June 1985, a laser beam was sent out from the Air Force Optical Station on Maui, Hawaii, and reflected back from the shuttle *Discovery* as it sped by 354 km overhead. The diameter of the central maximum of the beam at the shuttle position was said to be 9.1 m, and the beam wavelength was 500 nm. What is the effective diameter of the laser aperture at the Maui ground station? (*Hint:* A laser beam spreads only because of diffraction; assume a circular exit aperture.)

94 A diffraction grating 1.00 cm wide has 10 000 parallel slits. Monochromatic light that is incident normally is diffracted through 30° in the first order. What is the wavelength of the light?

95 SSM If you double the width of a single slit, the intensity of the central maximum of the diffraction pattern increases by a factor of 4, even though the energy passing through the slit only doubles. Explain this quantitatively.

96 When monochromatic light is incident on a slit 22.0 μm wide, the first diffraction minimum lies at 1.80° from the direction of the incident light. What is the wavelength?

97 A spy satellite orbiting at 160 km above Earth's surface has a lens with a focal length of 3.6 m and can resolve objects on the ground as small as 30 cm. For example, it can easily measure the size of an aircraft's air intake port. What is the effective diameter of the lens as determined by diffraction consideration alone? Assume $\lambda = 550 \text{ nm}$.

98 Suppose that two points are separated by 2.0 cm. If they are viewed by an eye with a pupil opening of 5.0 mm, what distance from the viewer puts them at the Rayleigh limit of resolution? Assume a light wavelength of 500 nm.

99 A diffraction grating has 200 lines/mm. Light consisting of a continuous range of wavelengths between 550 nm and 700 nm is incident perpendicularly on the grating. (a) What is the lowest order that is overlapped by another order? (b) What is the highest order for which the complete spectrum is present?

100 A diffraction grating has 200 rulings/mm, and it produces an intensity maximum at $\theta = 30.0^\circ$. (a) What are the possible wavelengths of the incident visible light? (b) To what colors do they correspond?

101 SSM Show that the dispersion of a grating is $D = (\tan \theta)/\lambda$.

102 Monochromatic light (wavelength = 450 nm) is incident perpendicularly on a single slit (width = 0.40 mm). A screen is placed parallel to the slit plane, and on it the distance between the two minima on either side of the central maximum is 1.8 mm. (a) What is the distance from the slit to the screen? (*Hint:* The angle to either minimum is small enough that $\sin \theta \approx \tan \theta$.) (b) What is the distance on the screen between the first minimum and the third minimum on the same side of the central maximum?

103 Light containing a mixture of two wavelengths, 500 and 600 nm, is incident normally on a diffraction grating. It is desired (1) that the first and second maxima for each wavelength appear at $\theta \leq 30^\circ$, (2) that the dispersion be as high as possible, and (3) that the third order for the 600 nm light be a missing order. (a) What should be the slit separation? (b) What is the smallest individual slit width that can be used? (c) For the values calculated in (a) and (b) and the light of wavelength 600 nm, what is the largest order of maxima produced by the grating?

104 A beam of x rays with wavelengths ranging from 0.120 nm to 0.0700 nm scatters from a family of reflecting planes in a crystal. The plane separation is 0.250 nm. It is observed that scattered beams are produced for 0.100 nm and 0.0750 nm. What is the angle between the incident and scattered beams?

105 Show that a grating made up of alternately transparent and opaque strips of equal width eliminates all the even orders of maxima (except $m = 0$).

106 Light of wavelength 500 nm diffracts through a slit of width 2.00 μm and onto a screen that is 2.00 m away. On the screen, what is the distance between the center of the diffraction pattern and the third diffraction minimum?

107 If, in a two-slit interference pattern, there are 8 bright fringes within the first side peak of the diffraction envelope and diffraction minima coincide with two-slit interference maxima, then what is the ratio of slit separation to slit width?

108 White light (consisting of wavelengths from 400 nm to 700 nm) is normally incident on a grating. Show that, no matter what the value of the grating spacing d , the second order and third order overlap.

109 If we make $d = a$ in Fig. 36-50, the two slits coalesce into a single slit of width $2a$. Show that Eq. 36-19 reduces to give the diffraction pattern for such a slit.

110 Derive Eq. 36-28, the expression for the half-width of the lines in a grating's diffraction pattern.

111 Prove that it is not possible to determine both wavelength of incident radiation and spacing of reflecting planes in a crystal by measuring the Bragg angles for several orders.

112 How many orders of the entire visible spectrum (400–700 nm) can be produced by a grating of 500 lines/mm?

113 An acoustic double-slit system (of slit separation d and slit width a) is driven by two loudspeakers as shown in Fig. 36-51. By use of a variable delay line, the phase of one of the speakers may be varied relative to the other speaker. Describe in detail what changes occur in the double-slit diffraction pattern at large distances as the phase difference between the speakers is varied from zero to 2π . Take both interference and diffraction effects into account.

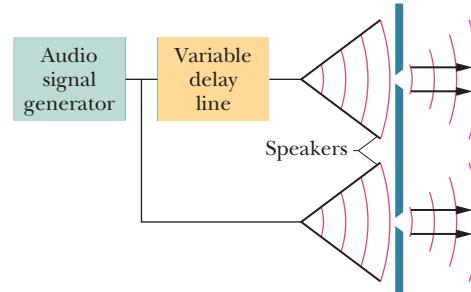


Figure 36-51 Problem 113.

114 Two emission lines have wavelengths λ and $\lambda + \Delta\lambda$, respectively, where $\Delta\lambda \ll \lambda$. Show that their angular separation $\Delta\theta$ in a grating spectrometer is given approximately by

$$\Delta\theta = \frac{\Delta\lambda}{\sqrt{(d/m)^2 - \lambda^2}},$$

where d is the slit separation and m is the order at which the lines are observed. Note that the angular separation is greater in the higher orders than the lower orders.

Electromagnetic Waves

33-1 ELECTROMAGNETIC WAVES

Learning Objectives

After reading this module, you should be able to ...

- 33.01** In the electromagnetic spectrum, identify the relative wavelengths (longer or shorter) of AM radio, FM radio, television, infrared light, visible light, ultraviolet light, x rays, and gamma rays.
- 33.02** Describe the transmission of an electromagnetic wave by an *LC* oscillator and an antenna.
- 33.03** For a transmitter with an *LC* oscillator, apply the relationships between the oscillator's inductance L , capacitance C , and angular frequency ω , and the emitted wave's frequency f and wavelength λ .
- 33.04** Identify the speed of an electromagnetic wave in vacuum (and approximately in air).
- 33.05** Identify that electromagnetic waves do not require a medium and can travel through vacuum.
- 33.06** Apply the relationship between the speed of an electromagnetic wave, the straight-line distance traveled by the wave, and the time required for the travel.
- 33.07** Apply the relationships between an electromagnetic

wave's frequency f , wavelength λ , period T , angular frequency ω , and speed c .

- 33.08** Identify that an electromagnetic wave consists of an electric component and a magnetic component that are (a) perpendicular to the direction of travel, (b) perpendicular to each other, and (c) sinusoidal waves with the same frequency and phase.
- 33.09** Apply the sinusoidal equations for the electric and magnetic components of an EM wave, written as functions of position and time.
- 33.10** Apply the relationship between the speed of light c , the permittivity constant ϵ_0 , and the permeability constant μ_0 .
- 33.11** For any instant and position, apply the relationship between the electric field magnitude E , the magnetic field magnitude B , and the speed of light c .
- 33.12** Describe the derivation of the relationship between the speed of light c and the ratio of the electric field amplitude E to the magnetic field amplitude B .

Key Ideas

- An electromagnetic wave consists of oscillating electric and magnetic fields.
- The various possible frequencies of electromagnetic waves form a spectrum, a small part of which is visible light.
- An electromagnetic wave traveling along an x axis has an electric field \vec{E} and a magnetic field \vec{B} with magnitudes that depend on x and t :

$$E = E_m \sin(kx - \omega t)$$

and

$$B = B_m \sin(kx - \omega t),$$

where E_m and B_m are the amplitudes of \vec{E} and \vec{B} . The electric field induces the magnetic field and vice versa.

- The speed of any electromagnetic wave in vacuum is c , which can be written as

$$c = \frac{E}{B} = \frac{1}{\sqrt{\mu_0 \epsilon_0}},$$

where E and B are the simultaneous magnitudes of the fields.

What Is Physics?

The information age in which we live is based almost entirely on the physics of electromagnetic waves. Like it or not, we are now globally connected by television, telephones, and the web. And like it or not, we are constantly immersed in those signals because of television, radio, and telephone transmitters.

Much of this global interconnection of information processors was not imagined by even the most visionary engineers of 40 years ago. The challenge for

today's engineers is trying to envision what the global interconnection will be like 40 years from now. The starting point in meeting that challenge is understanding the basic physics of electromagnetic waves, which come in so many different types that they are poetically said to form *Maxwell's rainbow*.

Maxwell's Rainbow

The crowning achievement of James Clerk Maxwell (see Chapter 32) was to show that a beam of light is a traveling wave of electric and magnetic fields—an **electromagnetic wave**—and thus that optics, the study of visible light, is a branch of electromagnetism. In this chapter we move from one to the other: we conclude our discussion of strictly electrical and magnetic phenomena, and we build a foundation for optics.

In Maxwell's time (the mid 1800s), the visible, infrared, and ultraviolet forms of light were the only electromagnetic waves known. Spurred on by Maxwell's work, however, Heinrich Hertz discovered what we now call radio waves and verified that they move through the laboratory at the same speed as visible light, indicating that they have the same basic nature as visible light.

As Fig. 33-1 shows, we now know a wide *spectrum* (or range) of electromagnetic waves: Maxwell's rainbow. Consider the extent to which we are immersed in electromagnetic waves throughout this spectrum. The Sun, whose radiations define the environment in which we as a species have evolved and adapted, is the dominant source. We are also crisscrossed by radio and television signals. Microwaves from radar systems and from telephone relay systems may reach us. There are electromagnetic waves from lightbulbs, from the heated engine blocks of automobiles, from x-ray machines, from lightning flashes, and from buried radioactive materials. Beyond this, radiation reaches us from stars and other objects in our galaxy and from other galaxies. Electromagnetic waves also travel in the other direction. Television signals, transmitted from Earth since about 1950, have now taken news about us (along with episodes of *I Love Lucy*, albeit very faintly) to whatever technically sophisticated inhabitants there may be on whatever planets may encircle the nearest 400 or so stars.

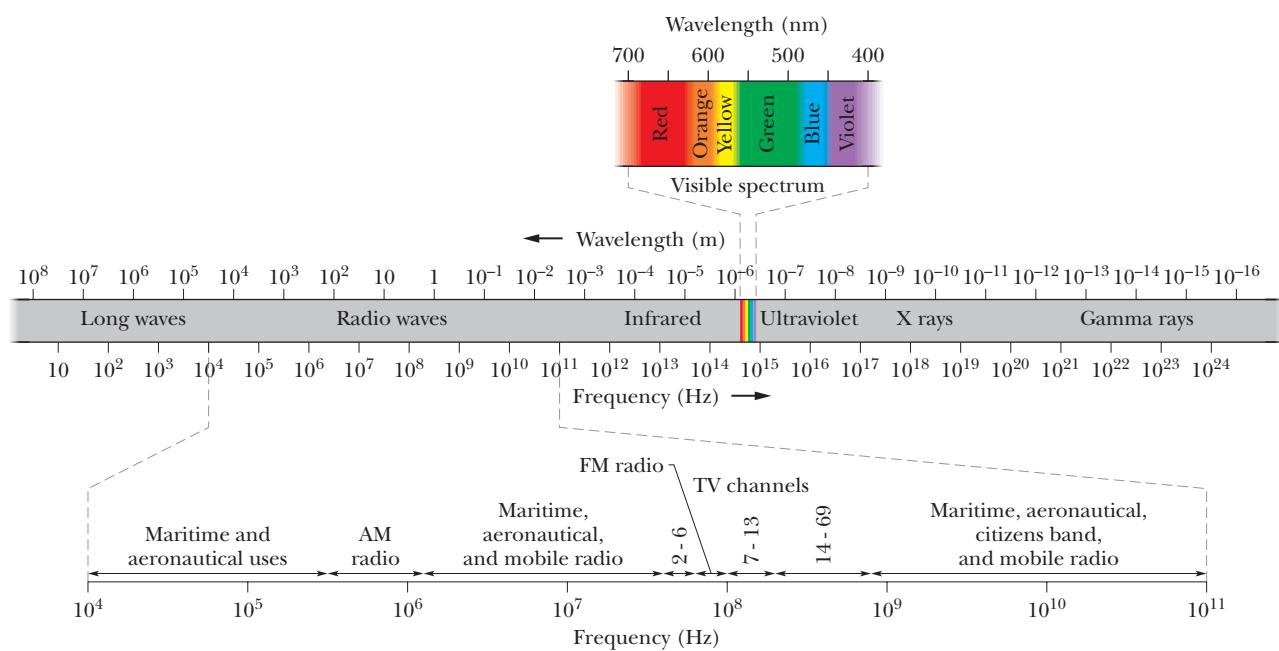


Figure 33-1 The electromagnetic spectrum.

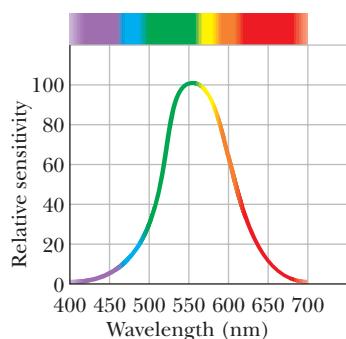


Figure 33-2 The relative sensitivity of the average human eye to electromagnetic waves at different wavelengths. This portion of the electromagnetic spectrum to which the eye is sensitive is called *visible light*.

In the wavelength scale in Fig. 33-1 (and similarly the corresponding frequency scale), each scale marker represents a change in wavelength (and correspondingly in frequency) by a factor of 10. The scale is open-ended; the wavelengths of electromagnetic waves have no inherent upper or lower bound.

Certain regions of the electromagnetic spectrum in Fig. 33-1 are identified by familiar labels, such as *x rays* and *radio waves*. These labels denote roughly defined wavelength ranges within which certain kinds of sources and detectors of electromagnetic waves are in common use. Other regions of Fig. 33-1, such as those labeled TV channels and AM radio, represent specific wavelength bands assigned by law for certain commercial or other purposes. There are no gaps in the electromagnetic spectrum—and all electromagnetic waves, no matter where they lie in the spectrum, travel through *free space* (vacuum) with the same speed c .

The visible region of the spectrum is of course of particular interest to us. Figure 33-2 shows the relative sensitivity of the human eye to light of various wavelengths. The center of the visible region is about 555 nm, which produces the sensation that we call yellow-green.

The limits of this visible spectrum are not well defined because the eye sensitivity curve approaches the zero-sensitivity line asymptotically at both long and short wavelengths. If we take the limits, arbitrarily, as the wavelengths at which eye sensitivity has dropped to 1% of its maximum value, these limits are about 430 and 690 nm; however, the eye can detect electromagnetic waves somewhat beyond these limits if they are intense enough.

The Traveling Electromagnetic Wave, Qualitatively

Some electromagnetic waves, including *x rays*, gamma rays, and visible light, are *radiated* (emitted) from sources that are of atomic or nuclear size, where quantum physics rules. Here we discuss how other electromagnetic waves are generated. To simplify matters, we restrict ourselves to that region of the spectrum (wavelength $\lambda \approx 1$ m) in which the source of the *radiation* (the emitted waves) is both macroscopic and of manageable dimensions.

Figure 33-3 shows, in broad outline, the generation of such waves. At its heart is an *LC oscillator*, which establishes an angular frequency ω ($= 1/\sqrt{LC}$). Charges and currents in this circuit vary sinusoidally at this frequency, as depicted in Fig. 31-1. An external source—possibly an ac generator—must be included to supply energy to compensate both for thermal losses in the circuit and for energy carried away by the radiated electromagnetic wave.

The *LC oscillator* of Fig. 33-3 is coupled by a transformer and a transmission line to an *antenna*, which consists essentially of two thin, solid, conducting rods. Through this coupling, the sinusoidally varying current in the oscillator causes charge to oscillate sinusoidally along the rods of the antenna at the angular frequency ω of the *LC oscillator*. The current in the rods associated with this movement of charge also varies sinusoidally, in magnitude and direction, at angular frequency ω . The antenna has the effect of an electric dipole whose electric dipole moment varies sinusoidally in magnitude and direction along the antenna.

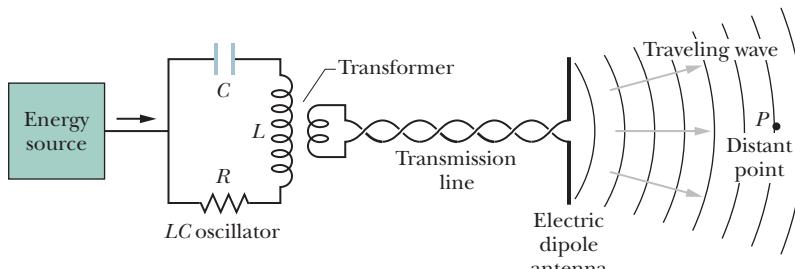


Figure 33-3 An arrangement for generating a traveling electromagnetic wave in the shortwave radio region of the spectrum: an *LC oscillator* produces a sinusoidal current in the antenna, which generates the wave. P is a distant point at which a detector can monitor the wave traveling past it.

Because the dipole moment varies in magnitude and direction, the electric field produced by the dipole varies in magnitude and direction. Also, because the current varies, the magnetic field produced by the current varies in magnitude and direction. However, the changes in the electric and magnetic fields do not happen everywhere instantaneously; rather, the changes travel outward from the antenna at the speed of light c . Together the changing fields form an electromagnetic wave that travels away from the antenna at speed c . The angular frequency of this wave is ω , the same as that of the LC oscillator.

Electromagnetic Wave. Figure 33-4 shows how the electric field \vec{E} and the magnetic field \vec{B} change with time as one wavelength of the wave sweeps past the distant point P of Fig. 33-3; in each part of Fig. 33-4, the wave is traveling directly out of the page. (We choose a distant point so that the curvature of the waves suggested in Fig. 33-3 is small enough to neglect. At such points, the wave is said to be a *plane wave*, and discussion of the wave is much simplified.) Note several key features in Fig. 33-4; they are present regardless of how the wave is created:

1. The electric and magnetic fields \vec{E} and \vec{B} are always perpendicular to the direction in which the wave is traveling. Thus, the wave is a *transverse wave*, as discussed in Chapter 16.
2. The electric field is always perpendicular to the magnetic field.
3. The cross product $\vec{E} \times \vec{B}$ always gives the direction in which the wave travels.
4. The fields always vary sinusoidally, just like the transverse waves discussed in Chapter 16. Moreover, the fields vary with the same frequency and *in phase* (in step) with each other.

In keeping with these features, we can assume that the electromagnetic wave is traveling toward P in the positive direction of an x axis, that the electric field in Fig. 33-4 is oscillating parallel to the y axis, and that the magnetic field is then oscillating parallel to the z axis (using a right-handed coordinate system, of course). Then we can write the electric and magnetic fields as sinusoidal functions of position x (along the path of the wave) and time t :

$$E = E_m \sin(kx - \omega t), \quad (33-1)$$

$$B = B_m \sin(kx - \omega t), \quad (33-2)$$

in which E_m and B_m are the amplitudes of the fields and, as in Chapter 16, ω and k are the angular frequency and angular wave number of the wave, respectively. From these equations, we note that not only do the two fields form the electromagnetic wave but each also forms its own wave. Equation 33-1 gives the *electric wave component* of the electromagnetic wave, and Eq. 33-2 gives the *magnetic wave component*. As we shall discuss below, these two wave components cannot exist independently.

Wave Speed. From Eq. 16-13, we know that the speed of the wave is ω/k . However, because this is an electromagnetic wave, its speed (in vacuum) is given the symbol c rather than v . In the next section you will see that c has the value

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \quad (\text{wave speed}), \quad (33-3)$$

which is about 3.0×10^8 m/s. In other words,



All electromagnetic waves, including visible light, have the same speed c in vacuum.

You will also see that the wave speed c and the amplitudes of the electric and

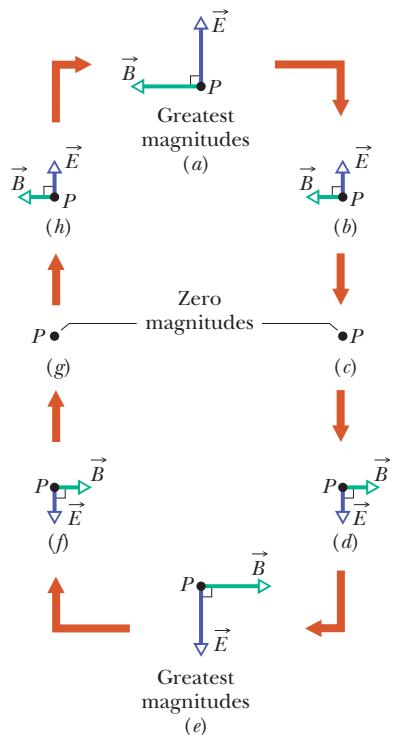


Figure 33-4 (a)–(h) The variation in the electric field \vec{E} and the magnetic field \vec{B} at the distant point P of Fig. 33-3 as one wavelength of the electromagnetic wave travels past it. In this perspective, the wave is traveling directly out of the page. The two fields vary sinusoidally in magnitude and direction. Note that they are always perpendicular to each other and to the wave's direction of travel.

magnetic fields are related by

$$\frac{E_m}{B_m} = c \quad (\text{amplitude ratio}). \quad (33-4)$$

If we divide Eq. 33-1 by Eq. 33-2 and then substitute with Eq. 33-4, we find that the magnitudes of the fields at every instant and at any point are related by

$$\frac{E}{B} = c \quad (\text{magnitude ratio}). \quad (33-5)$$

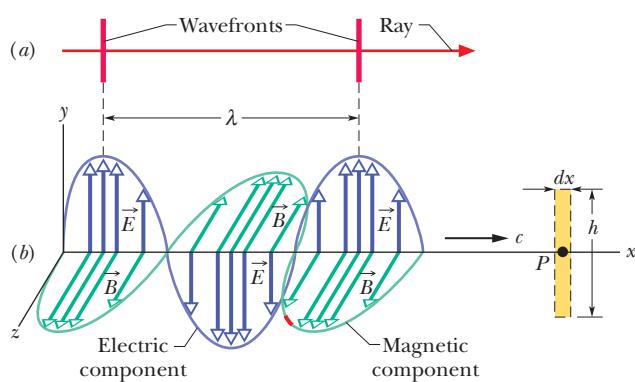
Rays and Wavefronts. We can represent the electromagnetic wave as in Fig. 33-5a, with a *ray* (a directed line showing the wave's direction of travel) or with *wavefronts* (imaginary surfaces over which the wave has the same magnitude of electric field), or both. The two wavefronts shown in Fig. 33-5a are separated by one wavelength $\lambda (= 2\pi/k)$ of the wave. (Waves traveling in approximately the same direction form a *beam*, such as a laser beam, which can also be represented with a ray.)

Drawing the Wave. We can also represent the wave as in Fig. 33-5b, which shows the electric and magnetic field vectors in a “snapshot” of the wave at a certain instant. The curves through the tips of the vectors represent the sinusoidal oscillations given by Eqs. 33-1 and 33-2; the wave components \vec{E} and \vec{B} are in phase, perpendicular to each other, and perpendicular to the wave's direction of travel.

Interpretation of Fig. 33-5b requires some care. The similar drawings for a transverse wave on a taut string that we discussed in Chapter 16 represented the up and down displacement of sections of the string as the wave passed (*something actually moved*). Figure 33-5b is more abstract. At the instant shown, the electric and magnetic fields each have a certain magnitude and direction (but always perpendicular to the x axis) at each point along the x axis. We choose to represent these vector quantities with a pair of arrows for each point, and so we must draw arrows of different lengths for different points, all directed away from the x axis, like thorns on a rose stem. However, the arrows represent field values only at points that are on the x axis. Neither the arrows nor the sinusoidal curves represent a sideways motion of anything, nor do the arrows connect points on the x axis with points off the axis.

Feedback. Drawings like Fig. 33-5 help us visualize what is actually a very complicated situation. First consider the magnetic field. Because it varies sinusoidally, it induces (via Faraday's law of induction) a perpendicular electric field that also varies sinusoidally. However, because that electric field is varying sinusoidally, it induces (via Maxwell's law of induction) a perpendicular magnetic field that also varies sinusoidally. And so on. The two fields continuously create each other via induction, and the resulting sinusoidal variations in the fields travel as a wave—the electromagnetic wave. Without this amazing result, we could not see; indeed, because we need electromagnetic waves

Figure 33-5 (a) An electromagnetic wave represented with a ray and two wavefronts; the wavefronts are separated by one wavelength λ . (b) The same wave represented in a “snapshot” of its electric field \vec{E} and magnetic field \vec{B} at points on the x axis, along which the wave travels at speed c . As it travels past point P , the fields vary as shown in Fig. 33-4. The electric component of the wave consists of only the electric fields; the magnetic component consists of only the magnetic fields. The dashed rectangle at P is used in Fig. 33-6.



from the Sun to maintain Earth's temperature, without this result we could not even exist.

A Most Curious Wave

The waves we discussed in Chapters 16 and 17 require a *medium* (some material) through which or along which to travel. We had waves traveling along a string, through Earth, and through the air. However, an electromagnetic wave (let's use the term *light wave* or *light*) is curiously different in that it requires no medium for its travel. It can, indeed, travel through a medium such as air or glass, but it can also travel through the vacuum of space between a star and us.

Once the special theory of relativity became accepted, long after Einstein published it in 1905, the speed of light waves was realized to be special. One reason is that light has the same speed regardless of the frame of reference from which it is measured. If you send a beam of light along an axis and ask several observers to measure its speed while they move at different speeds along that axis, either in the direction of the light or opposite it, they will all measure the *same speed* for the light. This result is an amazing one and quite different from what would have been found if those observers had measured the speed of any other type of wave; for other waves, the speed of the observers relative to the wave would have affected their measurements.

The meter has now been defined so that the speed of light (any electromagnetic wave) in vacuum has the exact value

$$c = 299\,792\,458 \text{ m/s},$$

which can be used as a standard. In fact, if you now measure the travel time of a pulse of light from one point to another, you are not really measuring the speed of the light but rather the distance between those two points.

The Traveling Electromagnetic Wave, Quantitatively

We shall now derive Eqs. 33-3 and 33-4 and, even more important, explore the dual induction of electric and magnetic fields that gives us light.

Equation 33-4 and the Induced Electric Field

The dashed rectangle of dimensions dx and h in Fig. 33-6 is fixed at point P on the x axis and in the xy plane (it is shown on the right in Fig. 33-5b). As the electromagnetic wave moves rightward past the rectangle, the magnetic flux Φ_B through the rectangle changes and—according to Faraday's law of induction—induced electric fields appear throughout the region of the rectangle. We take \vec{E} and $\vec{E} + d\vec{E}$ to be the induced fields along the two long sides of the rectangle. These induced electric fields are, in fact, the electrical component of the electromagnetic wave.

Note the small red portion of the magnetic field component curve far from the y axis in Fig. 33-5b. Let's consider the induced electric fields at the instant when this red portion of the magnetic component is passing through the rectangle. Just then, the magnetic field through the rectangle points in the positive z direction and is decreasing in magnitude (the magnitude was greater just before the red section arrived). Because the magnetic field is decreasing, the magnetic flux Φ_B through the rectangle is also decreasing. According to Faraday's law, this change in flux is opposed by induced electric fields, which produce a magnetic field \vec{B} in the positive z direction.

According to Lenz's law, this in turn means that if we imagine the boundary of the rectangle to be a conducting loop, a counterclockwise induced current would have to appear in it. There is, of course, no conducting loop; but this analysis shows that the induced electric field vectors \vec{E} and $\vec{E} + d\vec{E}$ are indeed

The oscillating magnetic field induces an oscillating and perpendicular electric field.

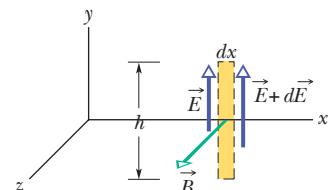


Figure 33-6 As the electromagnetic wave travels rightward past point P in Fig. 33-5b, the sinusoidal variation of the magnetic field \vec{B} through a rectangle centered at P induces electric fields along the rectangle. At the instant shown, \vec{B} is decreasing in magnitude and the induced electric field is therefore greater in magnitude on the right side of the rectangle than on the left.

oriented as shown in Fig. 33-6, with the magnitude of $\vec{E} + d\vec{E}$ greater than that of \vec{E} . Otherwise, the net induced electric field would not act counterclockwise around the rectangle.

Faraday's Law. Let us now apply Faraday's law of induction,

$$\oint \vec{E} \cdot d\vec{s} = -\frac{d\Phi_B}{dt}, \quad (33-6)$$

counterclockwise around the rectangle of Fig. 33-6. There is no contribution to the integral from the top or bottom of the rectangle because \vec{E} and $d\vec{s}$ are perpendicular to each other there. The integral then has the value

$$\oint \vec{E} \cdot d\vec{s} = (E + dE)h - Eh = h dE. \quad (33-7)$$

The flux Φ_B through this rectangle is

$$\Phi_B = (B)(h dx), \quad (33-8)$$

where B is the average magnitude of \vec{B} within the rectangle and $h dx$ is the area of the rectangle. Differentiating Eq. 33-8 with respect to t gives

$$\frac{d\Phi_B}{dt} = h dx \frac{dB}{dt}. \quad (33-9)$$

If we substitute Eqs. 33-7 and 33-9 into Eq. 33-6, we find

$$\begin{aligned} h dE &= -h dx \frac{dB}{dt} \\ \text{or} \quad \frac{dE}{dx} &= -\frac{dB}{dt}. \end{aligned} \quad (33-10)$$

Actually, both B and E are functions of *two* variables, coordinate x and time t , as Eqs. 33-1 and 33-2 show. However, in evaluating dE/dx , we must assume that t is constant because Fig. 33-6 is an "instantaneous snapshot." Also, in evaluating dB/dt we must assume that x is constant (a particular value) because we are dealing with the time rate of change of B at a particular place, the point P shown in Fig. 33-5b. The derivatives under these circumstances are *partial derivatives*, and Eq. 33-10 must be written

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t}. \quad (33-11)$$

The minus sign in this equation is appropriate and necessary because, although magnitude E is increasing with x at the site of the rectangle in Fig. 33-6, magnitude B is decreasing with t .

From Eq. 33-1 we have

$$\frac{\partial E}{\partial x} = kE_m \cos(kx - \omega t)$$

and from Eq. 33-2

$$\frac{\partial B}{\partial t} = -\omega B_m \cos(kx - \omega t).$$

Then Eq. 33-11 reduces to

$$kE_m \cos(kx - \omega t) = \omega B_m \cos(kx - \omega t). \quad (33-12)$$

The ratio ω/k for a traveling wave is its speed, which we are calling c . Equation 33-12 then becomes

$$\frac{E_m}{B_m} = c \quad (\text{amplitude ratio}), \quad (33-13)$$

which is just Eq. 33-4.

Equation 33-3 and the Induced Magnetic Field

Figure 33-7 shows another dashed rectangle at point P of Fig. 33-5b; this one is in the xz plane. As the electromagnetic wave moves rightward past this new rectangle, the electric flux Φ_E through the rectangle changes and—according to Maxwell's law of induction—induced magnetic fields appear throughout the region of the rectangle. These induced magnetic fields are, in fact, the magnetic component of the electromagnetic wave.

We see from Fig. 33-5b that at the instant chosen for the magnetic field represented in Fig. 33-6, marked in red on the magnetic component curve, the electric field through the rectangle of Fig. 33-7 is directed as shown. Recall that at the chosen instant, the magnetic field in Fig. 33-6 is decreasing. Because the two fields are in phase, the electric field in Fig. 33-7 must also be decreasing, and so must the electric flux Φ_E through the rectangle. By applying the same reasoning we applied to Fig. 33-6, we see that the changing flux Φ_E will induce a magnetic field with vectors \vec{B} and $\vec{B} + d\vec{B}$ oriented as shown in Fig. 33-7, where field $\vec{B} + d\vec{B}$ is greater than field \vec{B} .

Maxwell's Law. Let us apply Maxwell's law of induction,

$$\oint \vec{B} \cdot d\vec{s} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}, \quad (33-14)$$

by proceeding counterclockwise around the dashed rectangle of Fig. 33-7. Only the long sides of the rectangle contribute to the integral because the dot product along the short sides is zero. Thus, we can write

$$\oint \vec{B} \cdot d\vec{s} = -(B + dB)h + Bh = -h dB. \quad (33-15)$$

The flux Φ_E through the rectangle is

$$\Phi_E = (E)(h dx), \quad (33-16)$$

where E is the average magnitude of \vec{E} within the rectangle. Differentiating Eq. 33-16 with respect to t gives

$$\frac{d\Phi_E}{dt} = h dx \frac{dE}{dt}.$$

If we substitute this and Eq. 33-15 into Eq. 33-14, we find

$$-h dB = \mu_0 \epsilon_0 \left(h dx \frac{dE}{dt} \right)$$

or, changing to partial-derivative notation as we did for Eq. 33-11,

$$-\frac{\partial B}{\partial x} = \mu_0 \epsilon_0 \frac{\partial E}{\partial t}. \quad (33-17)$$

Again, the minus sign in this equation is necessary because, although B is increasing with x at point P in the rectangle in Fig. 33-7, E is decreasing with t .

Evaluating Eq. 33-17 by using Eqs. 33-1 and 33-2 leads to

$$-kB_m \cos(kx - \omega t) = -\mu_0 \epsilon_0 \omega E_m \cos(kx - \omega t),$$

which we can write as

$$\frac{E_m}{B_m} = \frac{1}{\mu_0 \epsilon_0 (\omega/k)} = \frac{1}{\mu_0 \epsilon_0 c}.$$

Combining this with Eq. 33-13 leads at once to

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \quad (\text{wave speed}), \quad (33-18)$$

which is exactly Eq. 33-3.

The oscillating electric field induces an oscillating and perpendicular magnetic field.

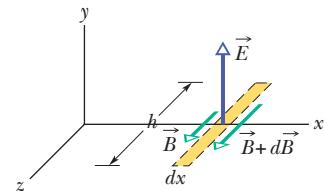
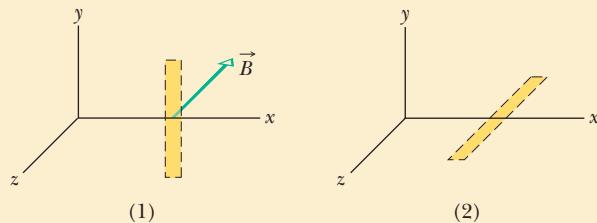


Figure 33-7 The sinusoidal variation of the electric field through this rectangle, located (but not shown) at point P in Fig. 33-5b, induces magnetic fields along the rectangle. The instant shown is that of Fig. 33-6: \vec{E} is decreasing in magnitude, and the magnitude of the induced magnetic field is greater on the right side of the rectangle than on the left.



Checkpoint 1

The magnetic field \vec{B} through the rectangle of Fig. 33-6 is shown at a different instant in part 1 of the figure here; \vec{B} is directed in the xz plane, parallel to the z axis, and its magnitude is increasing. (a) Complete part 1 by drawing the induced electric fields, indicating both directions and relative magnitudes (as in Fig. 33-6). (b) For the same instant, complete part 2 of the figure by drawing the electric field of the electromagnetic wave. Also draw the induced magnetic fields, indicating both directions and relative magnitudes (as in Fig. 33-7).



33-2 ENERGY TRANSPORT AND THE POYNTING VECTOR

Learning Objectives

After reading this module, you should be able to ...

- 33.13** Identify that an electromagnetic wave transports energy.
- 33.14** For a target, identify that an EM wave's rate of energy transport per unit area is given by the Poynting vector \vec{S} , which is related to the cross product of the electric field \vec{E} and magnetic field \vec{B} .
- 33.15** Determine the direction of travel (and thus energy transport) of an electromagnetic wave by applying the cross product for the corresponding Poynting vector.
- 33.16** Calculate the instantaneous rate S of energy flow of an EM wave in terms of the instantaneous electric field magnitude E .
- 33.17** For the electric field component of an electromagnetic wave, relate the rms value E_{rms} to the amplitude E_m .

Key Ideas

- The rate per unit area at which energy is transported via an electromagnetic wave is given by the Poynting vector \vec{S} :

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}.$$

The direction of \vec{S} (and thus of the wave's travel and the energy transport) is perpendicular to the directions of both \vec{E} and \vec{B} .

- The time-averaged rate per unit area at which energy is transported is S_{avg} , which is called the intensity I of

- 33.18** Identify an EM wave's intensity I in terms of energy transport.

- 33.19** Apply the relationships between an EM wave's intensity I and the electric field's rms value E_{rms} and amplitude E_m .

- 33.20** Apply the relationship between average power P_{avg} , energy transfer ΔE , and the time Δt taken by that transfer, and apply the relationship between the instantaneous power P and the rate of energy transfer dE/dt .

- 33.21** Identify an isotropic point source of light.

- 33.22** For an isotropic point source of light, apply the relationship between the emission power P , the distance r to a point of measurement, and the intensity I at that point.

- 33.23** In terms of energy conservation, explain why the intensity from an isotropic point source of light decreases as $1/r^2$.

the wave:

$$I = \frac{1}{c\mu_0} E_{\text{rms}}^2,$$

in which $E_{\text{rms}} = E_m/\sqrt{2}$.

- A point source of electromagnetic waves emits the waves isotropically—that is, with equal intensity in all directions. The intensity of the waves at distance r from a point source of power P_s is

$$I = \frac{P_s}{4\pi r^2}.$$

Energy Transport and the Poynting Vector

All sunbathers know that an electromagnetic wave can transport energy and deliver it to a body on which the wave falls. The rate of energy transport per unit area in such a wave is described by a vector \vec{S} , called the **Poynting vector** after physicist John Henry Poynting (1852–1914), who first discussed its properties. This vector is defined as

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B} \quad (\text{Poynting vector}). \quad (33-19)$$

Its magnitude S is related to the rate at which energy is transported by a wave across a unit area at any instant (inst):

$$S = \left(\frac{\text{energy/time}}{\text{area}} \right)_{\text{inst}} = \left(\frac{\text{power}}{\text{area}} \right)_{\text{inst}}. \quad (33-20)$$

From this we can see that the SI unit for \vec{S} is the watt per square meter (W/m^2).



The direction of the Poynting vector \vec{S} of an electromagnetic wave at any point gives the wave's direction of travel and the direction of energy transport at that point.

Because \vec{E} and \vec{B} are perpendicular to each other in an electromagnetic wave, the magnitude of $\vec{E} \times \vec{B}$ is EB . Then the magnitude of \vec{S} is

$$S = \frac{1}{\mu_0} EB, \quad (33-21)$$

in which S , E , and B are instantaneous values. The magnitudes E and B are so closely coupled to each other that we need to deal with only one of them; we choose E , largely because most instruments for detecting electromagnetic waves deal with the electric component of the wave rather than the magnetic component. Using $B = E/c$ from Eq. 33-5, we can rewrite Eq. 33-21 in terms of just the electric component as

$$S = \frac{1}{c\mu_0} E^2 \quad (\text{instantaneous energy flow rate}). \quad (33-22)$$

Intensity. By substituting $E = E_m \sin(kx - \omega t)$ into Eq. 33-22, we could obtain an equation for the energy transport rate as a function of time. More useful in practice, however, is the average energy transported over time; for that, we need to find the time-averaged value of S , written S_{avg} and also called the **intensity** I of the wave. Thus from Eq. 33-20, the intensity I is

$$I = S_{\text{avg}} = \left(\frac{\text{energy/time}}{\text{area}} \right)_{\text{avg}} = \left(\frac{\text{power}}{\text{area}} \right)_{\text{avg}}. \quad (33-23)$$

From Eq. 33-22, we find

$$I = S_{\text{avg}} = \frac{1}{c\mu_0} [E^2]_{\text{avg}} = \frac{1}{c\mu_0} [E_m^2 \sin^2(kx - \omega t)]_{\text{avg}}. \quad (33-24)$$

Over a full cycle, the average value of $\sin^2 \theta$, for any angular variable θ , is $\frac{1}{2}$ (see Fig. 31-17). In addition, we define a new quantity E_{rms} , the *root-mean-square* value of the electric field, as

$$E_{\text{rms}} = \frac{E_m}{\sqrt{2}}. \quad (33-25)$$

The energy emitted by light source S must pass through the sphere of radius r .

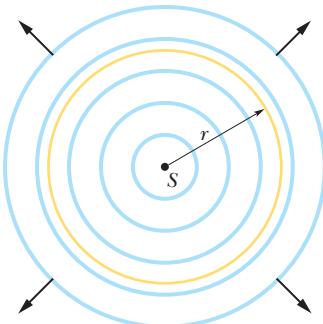


Figure 33-8 A point source S emits electromagnetic waves uniformly in all directions. The spherical wavefronts pass through an imaginary sphere of radius r that is centered on S .

We can then rewrite Eq. 33-24 as

$$I = \frac{1}{c\mu_0} E_{\text{rms}}^2. \quad (33-26)$$

Because $E = cB$ and c is such a very large number, you might conclude that the energy associated with the electric field is much greater than that associated with the magnetic field. That conclusion is incorrect; the two energies are exactly equal. To show this, we start with Eq. 25-25, which gives the energy density u ($= \frac{1}{2}\epsilon_0 E^2$) within an electric field, and substitute cB for E ; then we can write

$$u_E = \frac{1}{2}\epsilon_0 E^2 = \frac{1}{2}\epsilon_0 (cB)^2.$$

If we now substitute for c with Eq. 33-3, we get

$$u_E = \frac{1}{2}\epsilon_0 \frac{1}{\mu_0 \epsilon_0} B^2 = \frac{B^2}{2\mu_0}.$$

However, Eq. 30-55 tells us that $B^2/2\mu_0$ is the energy density u_B of a magnetic field \vec{B} ; so we see that $u_E = u_B$ everywhere along an electromagnetic wave.

Variation of Intensity with Distance

How intensity varies with distance from a real source of electromagnetic radiation is often complex—especially when the source (like a searchlight at a movie premier) beams the radiation in a particular direction. However, in some situations we can assume that the source is a *point source* that emits the light *isotropically*—that is, with equal intensity in all directions. The spherical wavefronts spreading from such an isotropic point source S at a particular instant are shown in cross section in Fig. 33-8.

Let us assume that the energy of the waves is conserved as they spread from this source. Let us also center an imaginary sphere of radius r on the source, as shown in Fig. 33-8. All the energy emitted by the source must pass through the sphere. Thus, the rate at which energy passes through the sphere via the radiation must equal the rate at which energy is emitted by the source—that is, the source power P_s . The intensity I (power per unit area) measured at the sphere must then be, from Eq. 33-23,

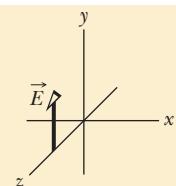
$$I = \frac{\text{power}}{\text{area}} = \frac{P_s}{4\pi r^2}, \quad (33-27)$$

where $4\pi r^2$ is the area of the sphere. Equation 33-27 tells us that the intensity of the electromagnetic radiation from an isotropic point source decreases with the square of the distance r from the source.



Checkpoint 2

The figure here gives the electric field of an electromagnetic wave at a certain point and a certain instant. The wave is transporting energy in the negative z direction. What is the direction of the magnetic field of the wave at that point and instant?



Sample Problem 33.01 Light wave: rms values of the electric and magnetic fields

When you look at the North Star (Polaris), you intercept light from a star at a distance of 431 ly and emitting energy at a rate of 2.2×10^3 times that of our Sun ($P_{\text{sun}} = 3.90 \times$

10^{26} W). Neglecting any atmospheric absorption, find the rms values of the electric and magnetic fields when the starlight reaches you.

KEY IDEAS

- The rms value E_{rms} of the electric field in light is related to the intensity I of the light via Eq. 33-26 ($I = E_{\text{rms}}^2/c\mu_0$).
- Because the source is so far away and emits light with equal intensity in all directions, the intensity I at any distance r from the source is related to the source's power P_s via Eq. 33-27 ($I = P_s/4\pi r^2$).
- The magnitudes of the electric field and magnetic field of an electromagnetic wave at any instant and at any point in the wave are related by the speed of light c according to Eq. 33-5 ($E/B = c$). Thus, the rms values of those fields are also related by Eq. 33-5.

Electric field: Putting the first two ideas together gives us

$$I = \frac{P_s}{4\pi r^2} = \frac{E_{\text{rms}}^2}{c\mu_0}$$

and

$$E_{\text{rms}} = \sqrt{\frac{P_s c \mu_0}{4\pi r^2}}.$$

By substituting $P_s = (2.2 \times 10^3)(3.90 \times 10^{26} \text{ W})$, $r = 431 \text{ ly} = 4.08 \times 10^{18} \text{ m}$, and values for the constants, we find

$$E_{\text{rms}} = 1.24 \times 10^{-3} \text{ V/m} \approx 1.2 \text{ mV/m}. \quad (\text{Answer})$$

Magnetic field: From Eq. 33-5, we write

$$\begin{aligned} B_{\text{rms}} &= \frac{E_{\text{rms}}}{c} = \frac{1.24 \times 10^{-3} \text{ V/m}}{3.00 \times 10^8 \text{ m/s}} \\ &= 4.1 \times 10^{-12} \text{ T} = 4.1 \text{ pT}. \end{aligned}$$

Cannot compare the fields: Note that E_{rms} ($= 1.2 \text{ mV/m}$) is small as judged by ordinary laboratory standards, but B_{rms} ($= 4.1 \text{ pT}$) is quite small. This difference helps to explain why most instruments used for the detection and measurement of electromagnetic waves are designed to respond to the electric component. It is wrong, however, to say that the electric component of an electromagnetic wave is "stronger" than the magnetic component. You cannot compare quantities that are measured in different units. However, these electric and magnetic components are on an equal basis because their average energies, which *can* be compared, are equal.



Additional examples, video, and practice available at WileyPLUS

33-3 RADIATION PRESSURE

Learning Objectives

After reading this module, you should be able to . . .

- 33.24** Distinguish between force and pressure.
- 33.25** Identify that an electromagnetic wave transports momentum and can exert a force and a pressure on a target.
- 33.26** For a uniform electromagnetic beam that is perpendicular to a target area, apply the relationships between that

Key Ideas

- When a surface intercepts electromagnetic radiation, a force and a pressure are exerted on the surface.
- If the radiation is totally absorbed by the surface, the force is

$$F = \frac{IA}{c} \quad (\text{total absorption}),$$

in which I is the intensity of the radiation and A is the area of the surface perpendicular to the path of the radiation.

- If the radiation is totally reflected back along its original

area, the wave's intensity, and the force on the target, for both total absorption and total backward reflection.

- 33.27** For a uniform electromagnetic beam that is perpendicular to a target area, apply the relationships between the wave's intensity and the pressure on the target, for both total absorption and total backward reflection.

path, the force is

$$F = \frac{2IA}{c} \quad (\text{total reflection back along path}).$$

- The radiation pressure p_r is the force per unit area:

$$p_r = \frac{I}{c} \quad (\text{total absorption})$$

and

$$p_r = \frac{2I}{c} \quad (\text{total reflection back along path}).$$

Radiation Pressure

Electromagnetic waves have linear momentum and thus can exert a pressure on an object when shining on it. However, the pressure must be very small because, for example, you do not feel a punch during a camera flash.

To find an expression for the pressure, let us shine a beam of electromagnetic radiation—light, for example—on an object for a time interval Δt . Further, let us assume that the object is free to move and that the radiation is entirely **absorbed** (taken up) by the object. This means that during the interval Δt , the object gains an energy ΔU from the radiation. Maxwell showed that the object also gains linear momentum. The magnitude Δp of the momentum change of the object is related to the energy change ΔU by

$$\Delta p = \frac{\Delta U}{c} \quad (\text{total absorption}), \quad (33-28)$$

where c is the speed of light. The direction of the momentum change of the object is the direction of the *incident* (incoming) beam that the object absorbs.

Instead of being absorbed, the radiation can be **reflected** by the object; that is, the radiation can be sent off in a new direction as if it bounced off the object. If the radiation is entirely reflected back along its original path, the magnitude of the momentum change of the object is twice that given above, or

$$\Delta p = \frac{2 \Delta U}{c} \quad (\text{total reflection back along path}). \quad (33-29)$$

In the same way, an object undergoes twice as much momentum change when a perfectly elastic tennis ball is bounced from it as when it is struck by a perfectly inelastic ball (a lump of wet putty, say) of the same mass and velocity. If the incident radiation is partly absorbed and partly reflected, the momentum change of the object is between $\Delta U/c$ and $2 \Delta U/c$.

Force. From Newton's second law in its linear momentum form (Module 9-3), we know that a change in momentum is related to a force by

$$F = \frac{\Delta p}{\Delta t}. \quad (33-30)$$

To find expressions for the force exerted by radiation in terms of the intensity I of the radiation, we first note that intensity is

$$I = \frac{\text{power}}{\text{area}} = \frac{\text{energy/time}}{\text{area}}.$$

Next, suppose that a flat surface of area A , perpendicular to the path of the radiation, intercepts the radiation. In time interval Δt , the energy intercepted by area A is

$$\Delta U = IA \Delta t. \quad (33-31)$$

If the energy is completely absorbed, then Eq. 33-28 tells us that $\Delta p = IA \Delta t/c$, and, from Eq. 33-30, the magnitude of the force on the area A is

$$F = \frac{IA}{c} \quad (\text{total absorption}). \quad (33-32)$$

Similarly, if the radiation is totally reflected back along its original path, Eq. 33-29 tells us that $\Delta p = 2IA \Delta t/c$ and, from Eq. 33-30,

$$F = \frac{2IA}{c} \quad (\text{total reflection back along path}). \quad (33-33)$$

If the radiation is partly absorbed and partly reflected, the magnitude of the force on area A is between the values of IA/c and $2IA/c$.

Pressure. The force per unit area on an object due to radiation is the radiation pressure p_r . We can find it for the situations of Eqs. 33-32 and 33-33 by dividing both sides of each equation by A . We obtain

$$p_r = \frac{I}{c} \quad (\text{total absorption}) \quad (33-34)$$

and $p_r = \frac{2I}{c}$ (total reflection back along path). (33-35)

Be careful not to confuse the symbol p_r for radiation pressure with the symbol p for momentum. Just as with fluid pressure in Chapter 14, the SI unit of radiation pressure is the newton per square meter (N/m^2), which is called the pascal (Pa).

The development of laser technology has permitted researchers to achieve radiation pressures much greater than, say, that due to a camera flashlamp. This comes about because a beam of laser light—unlike a beam of light from a small lamp filament—can be focused to a tiny spot. This permits the delivery of great amounts of energy to small objects placed at that spot.



Checkpoint 3

Light of uniform intensity shines perpendicularly on a totally absorbing surface, fully illuminating the surface. If the area of the surface is decreased, do (a) the radiation pressure and (b) the radiation force on the surface increase, decrease, or stay the same?

33-4 POLARIZATION

Learning Objectives

After reading this module, you should be able to . . .

33.28 Distinguish between polarized light and unpolarized light.

33.29 For a light beam headed toward you, sketch representations of polarized light and unpolarized light.

33.30 When a beam is sent into a polarizing sheet, explain the function of the sheet in terms of its polarizing direction (or axis) and the electric field component that is absorbed and the component that is transmitted.

33.31 For light that emerges from a polarizing sheet, identify its polarization relative to the sheet's polarizing direction.

33.32 For a light beam incident perpendicularly on a polarizing sheet, apply the one-half rule and the cosine-squared rule, distinguishing their uses.

33.33 Distinguish between a polarizer and an analyzer.

33.34 Explain what is meant if two sheets are crossed.

33.35 When a beam is sent into a system of polarizing sheets, work through the sheets one by one, finding the transmitted intensity and polarization.

Key Ideas

- Electromagnetic waves are polarized if their electric field vectors are all in a single plane, called the plane of oscillation. Light waves from common sources are not polarized; that is, they are unpolarized, or polarized randomly.

- When a polarizing sheet is placed in the path of light, only electric field components of the light parallel to the sheet's polarizing direction are transmitted by the sheet; components perpendicular to the polarizing direction are absorbed. The light that emerges from a polarizing sheet is polarized parallel to the polarizing direction of the sheet.

- If the original light is initially unpolarized, the transmitted intensity I is half the original intensity I_0 :

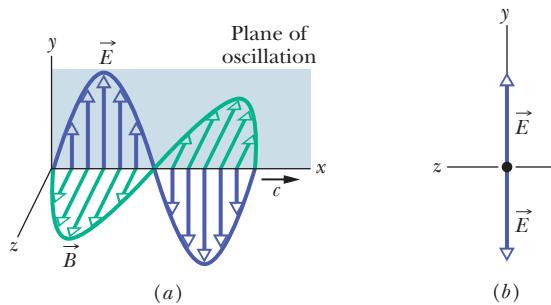
$$I = \frac{1}{2}I_0.$$

- If the original light is initially polarized, the transmitted intensity depends on the angle θ between the polarization direction of the original light and the polarizing direction of the sheet:

$$I = I_0 \cos^2 \theta.$$

Polarization

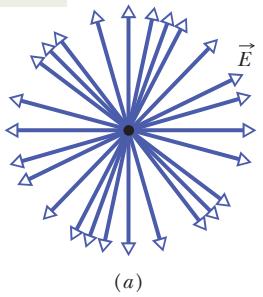
VHF (very high frequency) television antennas in England are oriented vertically, but those in North America are horizontal. The difference is due to the direction of oscillation of the electromagnetic waves carrying the TV signal. In England, the transmitting equipment is designed to produce waves that are **polarized** vertically; that is, their electric field oscillates vertically. Thus, for the



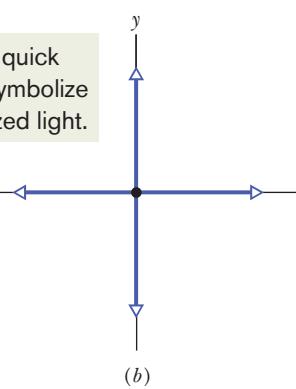
Vertically polarized light headed toward you—the electric fields are all vertical.

Figure 33-9 (a) The plane of oscillation of a polarized electromagnetic wave. (b) To represent the polarization, we view the plane of oscillation head-on and indicate the directions of the oscillating electric field with a double arrow.

Unpolarized light headed toward you—the electric fields are in all directions in the plane.



(a)



(b)

This is a quick way to symbolize unpolarized light.

electric field of the incident television waves to drive a current along an antenna (and provide a signal to a television set), the antenna must be vertical. In North America, the waves are polarized horizontally.

Figure 33-9a shows an electromagnetic wave with its electric field oscillating parallel to the vertical y axis. The plane containing the \vec{E} vectors is called the **plane of oscillation** of the wave (hence, the wave is said to be *plane-polarized* in the y direction). We can represent the wave's *polarization* (state of being polarized) by showing the directions of the electric field oscillations in a head-on view of the plane of oscillation, as in Fig. 33-9b. The vertical double arrow in that figure indicates that as the wave travels past us, its electric field oscillates vertically—it continuously changes between being directed up and down the y axis.

Polarized Light

The electromagnetic waves emitted by a television station all have the same polarization, but the electromagnetic waves emitted by any common source of light (such as the Sun or a bulb) are **polarized randomly**, or **unpolarized** (the two terms mean the same thing). That is, the electric field at any given point is always perpendicular to the direction of travel of the waves but changes directions randomly. Thus, if we try to represent a head-on view of the oscillations over some time period, we do not have a simple drawing with a single double arrow like that of Fig. 33-9b; instead we have a mess of double arrows like that in Fig. 33-10a.

In principle, we can simplify the mess by resolving each electric field of Fig. 33-10a into y and z components. Then as the wave travels past us, the net y component oscillates parallel to the y axis and the net z component oscillates parallel to the z axis. We can then represent the unpolarized light with a pair of double arrows as shown in Fig. 33-10b. The double arrow along the y axis represents the oscillations of the net y component of the electric field. The double arrow along the z axis represents the oscillations of the net z component of the electric field. In doing all this, we effectively change unpolarized light into the superposition of two polarized waves whose planes of oscillation are perpendicular to each other—one plane contains the y axis and the other contains the z axis. One reason to make this change is that drawing Fig. 33-10b is a lot easier than drawing Fig. 33-10a.

We can draw similar figures to represent light that is **partially polarized** (its field oscillations are not completely random as in Fig. 33-10a, nor are they parallel to a single axis as in Fig. 33-9b). For this situation, we draw one of the double arrows in a perpendicular pair of double arrows longer than the other one.

Polarizing Direction. We can transform unpolarized visible light into polarized light by sending it through a *polarizing sheet*, as is shown in Fig. 33-11. Such sheets, commercially known as Polaroids or Polaroid filters, were invented in 1932 by Edwin Land while he was an undergraduate student. A polarizing sheet consists of certain long molecules embedded in plastic. When the sheet is manu-

Figure 33-10 (a) Unpolarized light consists of waves with randomly directed electric fields. Here the waves are all traveling along the same axis, directly out of the page, and all have the same amplitude E . (b) A second way of representing unpolarized light—the light is the superposition of two polarized waves whose planes of oscillation are perpendicular to each other.

factured, it is stretched to align the molecules in parallel rows, like rows in a plowed field. When light is then sent through the sheet, electric field components along one direction pass through the sheet, while components perpendicular to that direction are absorbed by the molecules and disappear.

We shall not dwell on the molecules but, instead, shall assign to the sheet a *polarizing direction*, along which electric field components are passed:



An electric field component parallel to the polarizing direction is passed (*transmitted*) by a polarizing sheet; a component perpendicular to it is absorbed.

Thus, the electric field of the light emerging from the sheet consists of only the components that are parallel to the polarizing direction of the sheet; hence the light is polarized in that direction. In Fig. 33-11, the vertical electric field components are transmitted by the sheet; the horizontal components are absorbed. The transmitted waves are then vertically polarized.

Intensity of Transmitted Polarized Light

We now consider the intensity of light transmitted by a polarizing sheet. We start with unpolarized light, whose electric field oscillations we can resolve into y and z components as represented in Fig. 33-10b. Further, we can arrange for the y axis to be parallel to the polarizing direction of the sheet. Then only the y components of the light's electric field are passed by the sheet; the z components are absorbed. As suggested by Fig. 33-10b, if the original waves are randomly oriented, the sum of the y components and the sum of the z components are equal. When the z components are absorbed, half the intensity I_0 of the original light is lost. The intensity I of the emerging polarized light is then

$$I = \frac{1}{2}I_0 \quad (\text{one-half rule}). \quad (33-36)$$

Let us call this the *one-half rule*; we can use it *only* when the light reaching a polarizing sheet is unpolarized.

Suppose now that the light reaching a polarizing sheet is already polarized. Figure 33-12 shows a polarizing sheet in the plane of the page and the electric field \vec{E} of such a polarized light wave traveling toward the sheet (and thus prior to any absorption). We can resolve \vec{E} into two components relative to the polarizing direction of the sheet: parallel component E_y is transmitted by the sheet, and perpendicular component E_z is absorbed. Since θ is the angle between \vec{E} and the polarizing direction of the sheet, the transmitted parallel component is

$$E_y = E \cos \theta. \quad (33-37)$$

Recall that the intensity of an electromagnetic wave (such as our light wave) is proportional to the square of the electric field's magnitude (Eq. 33-26, $I = E_{\text{rms}}^2/c\mu_0$). In our present case then, the intensity I of the emerging wave is proportional to E_y^2 and the intensity I_0 of the original wave is proportional to E^2 . Hence, from Eq. 33-37 we can write $I/I_0 = \cos^2 \theta$, or

$$I = I_0 \cos^2 \theta \quad (\text{cosine-squared rule}). \quad (33-38)$$

Let us call this the *cosine-squared rule*; we can use it *only* when the light reaching a polarizing sheet is already polarized. Then the transmitted intensity I is a maximum and is equal to the original intensity I_0 when the original wave is polarized parallel to the polarizing direction of the sheet (when θ in Eq. 33-38 is 0° or 180°). The transmitted intensity is zero when the original wave is polarized perpendicular to the polarizing direction of the sheet (when θ is 90°).

The sheet's polarizing axis is vertical, so only vertically polarized light emerges.

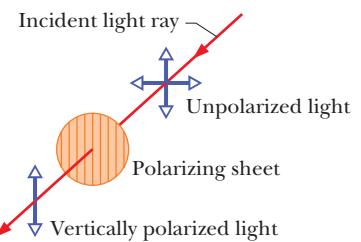


Figure 33-11 Unpolarized light becomes polarized when it is sent through a polarizing sheet. Its direction of polarization is then parallel to the polarizing direction of the sheet, which is represented here by the vertical lines drawn in the sheet.

The sheet's polarizing axis is vertical, so only vertical components of the electric fields pass.

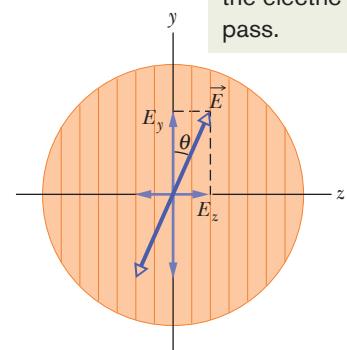


Figure 33-12 Polarized light approaching a polarizing sheet. The electric field \vec{E} of the light can be resolved into components E_y (parallel to the polarizing direction of the sheet) and E_z (perpendicular to that direction). Component E_y will be transmitted by the sheet; component E_z will be absorbed.

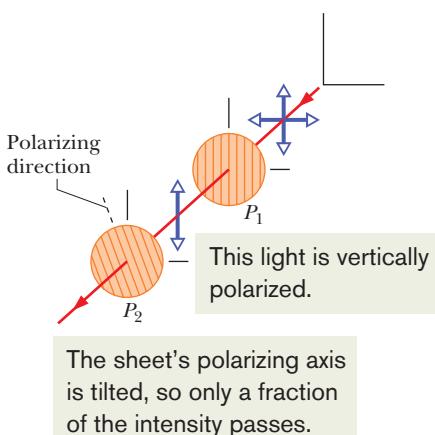


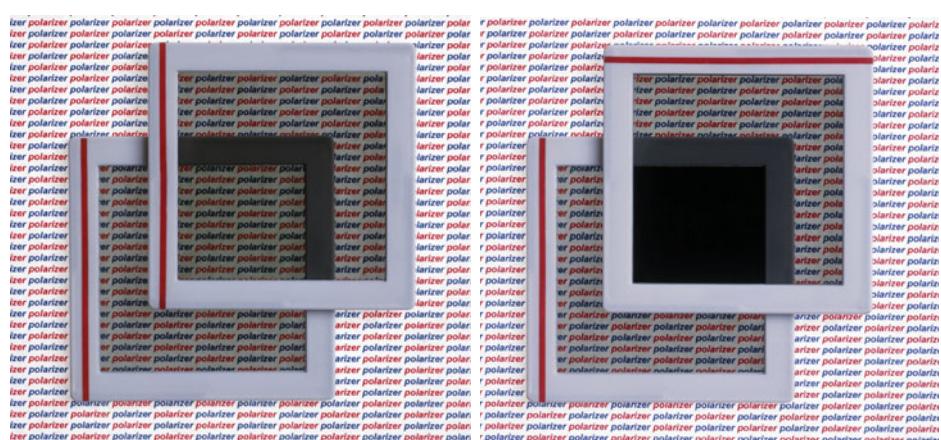
Figure 33-13 The light transmitted by polarizing sheet P_1 is vertically polarized, as represented by the vertical double arrow. The amount of that light that is then transmitted by polarizing sheet P_2 depends on the angle between the polarization direction of that light and the polarizing direction of P_2 (indicated by the lines drawn in the sheet and by the dashed line).

Two Polarizing Sheets. Figure 33-13 shows an arrangement in which initially unpolarized light is sent through two polarizing sheets P_1 and P_2 . (Often, the first sheet is called the *polarizer*, and the second the *analyzer*.) Because the polarizing direction of P_1 is vertical, the light transmitted by P_1 to P_2 is polarized vertically. If the polarizing direction of P_2 is also vertical, then all the light transmitted by P_1 is transmitted by P_2 . If the polarizing direction of P_2 is horizontal, none of the light transmitted by P_1 is transmitted by P_2 . We reach the same conclusions by considering only the *relative* orientations of the two sheets: If their polarizing directions are parallel, all the light passed by the first sheet is passed by the second sheet (Fig. 33-14a). If those directions are perpendicular (the sheets are said to be *crossed*), no light is passed by the second sheet (Fig. 33-14b). Finally, if the two polarizing directions of Fig. 33-13 make an angle between 0° and 90° , some of the light transmitted by P_1 will be transmitted by P_2 , as set by Eq. 33-38.

Other Means. Light can be polarized by means other than polarizing sheets, such as by reflection (discussed in Module 33-7) and by scattering from atoms or molecules. In *scattering*, light that is intercepted by an object, such as a molecule, is sent off in many, perhaps random, directions. An example is the scattering of sunlight by molecules in the atmosphere, which gives the sky its general glow.

Although direct sunlight is unpolarized, light from much of the sky is at least partially polarized by such scattering. Bees use the polarization of sky light in navigating to and from their hives. Similarly, the Vikings used it to navigate across the North Sea when the daytime Sun was below the horizon (because of the high latitude of the North Sea). These early seafarers had discovered certain crystals (now called cordierite) that changed color when rotated in polarized light. By looking at the sky through such a crystal while rotating it about their line of sight, they could locate the hidden Sun and thus determine which way was south.

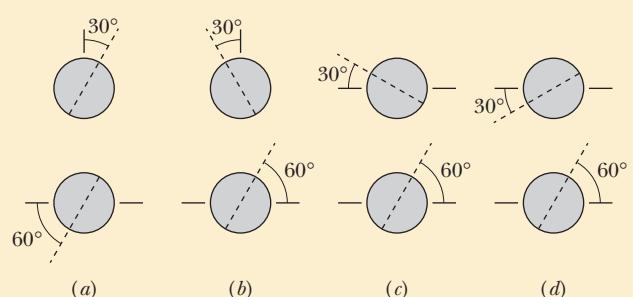
Figure 33-14 (a) Overlapping polarizing sheets transmit light fairly well when their polarizing directions have the same orientation, but (b) they block most of the light when they are crossed.



Richard Megna/Fundamental Photographs

Checkpoint 4

The figure shows four pairs of polarizing sheets, seen face-on. Each pair is mounted in the path of initially unpolarized light. The polarizing direction of each sheet (indicated by the dashed line) is referenced to either a horizontal x axis or a vertical y axis. Rank the pairs according to the fraction of the initial intensity that they pass, greatest first.



Sample Problem 33.02 Polarization and intensity with three polarizing sheets

Figure 33-15a, drawn in perspective, shows a system of three polarizing sheets in the path of initially unpolarized light. The polarizing direction of the first sheet is parallel to the y axis, that of the second sheet is at an angle of 60° counterclockwise from the y axis, and that of the third sheet is parallel to the x axis. What fraction of the initial intensity I_0 of the light emerges from the three-sheet system, and in which direction is that emerging light polarized?

KEY IDEAS

1. We work through the system sheet by sheet, from the first one encountered by the light to the last one.

2. To find the intensity transmitted by any sheet, we apply either the one-half rule or the cosine-squared rule, depending on whether the light reaching the sheet is unpolarized or already polarized.
3. The light that is transmitted by a polarizing sheet is always polarized parallel to the polarizing direction of the sheet.

First sheet: The original light wave is represented in Fig. 33-15b, using the head-on, double-arrow representation of Fig. 33-10b. Because the light is initially unpolarized, the intensity I_1 of the light transmitted by the first sheet is given by the one-half rule (Eq. 33-36):

$$I_1 = \frac{1}{2} I_0.$$

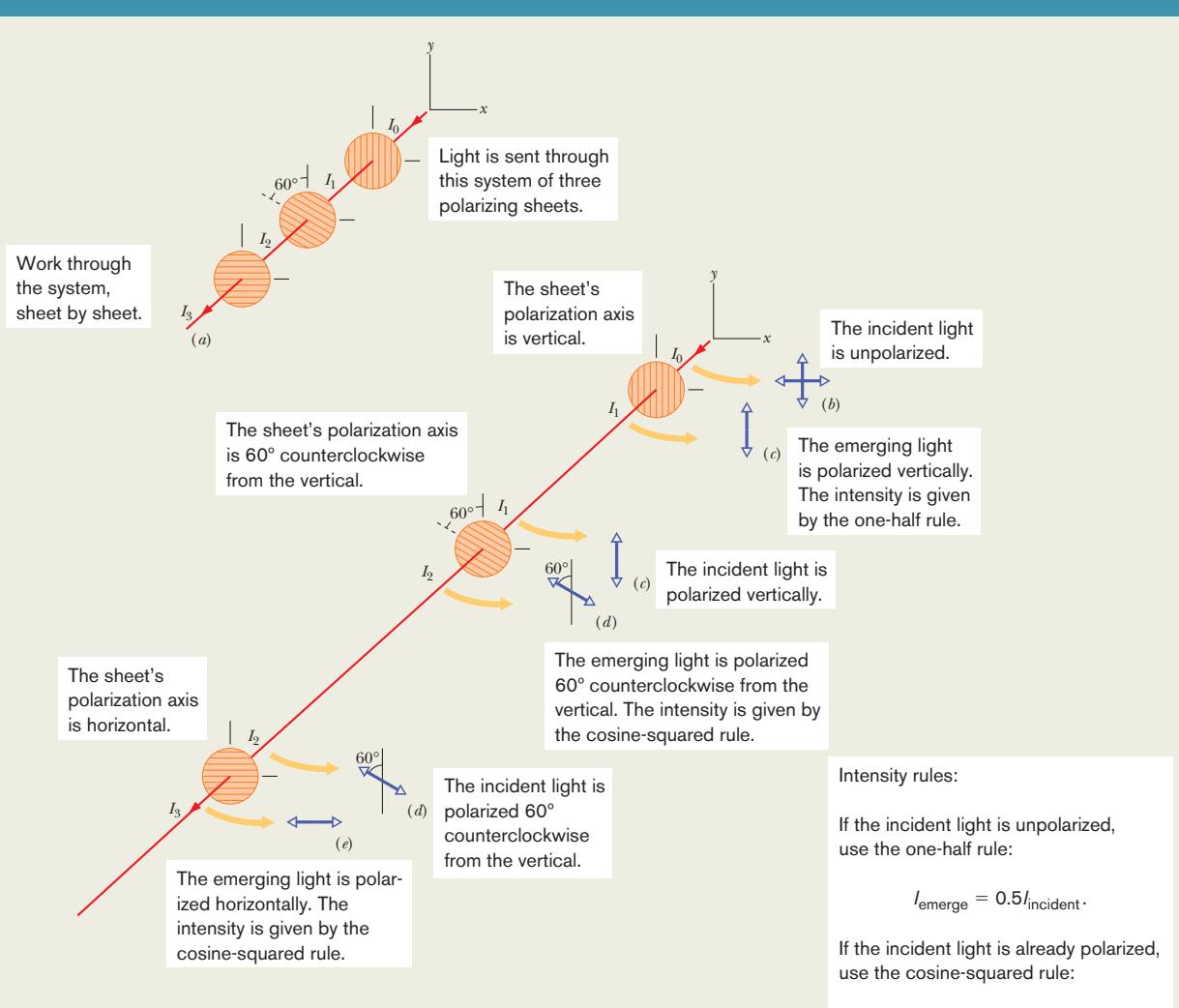


Figure 33-15 (a) Initially unpolarized light of intensity I_0 is sent into a system of three polarizing sheets. The intensities I_1 , I_2 , and I_3 of the light transmitted by the sheets are labeled. Shown also are the polarizations, from head-on views, of (b) the initial light and the light transmitted by (c) the first sheet, (d) the second sheet, and (e) the third sheet.

but be sure to insert the angle between the polarization of the incident light and the polarization axis of the sheet.

Because the polarizing direction of the first sheet is parallel to the y axis, the polarization of the light transmitted by it is also, as shown in the head-on view of Fig. 33-15c.

Second sheet: Because the light reaching the second sheet is polarized, the intensity I_2 of the light transmitted by that sheet is given by the cosine-squared rule (Eq. 33-38). The angle θ in the rule is the angle between the polarization direction of the entering light (parallel to the y axis) and the polarizing direction of the second sheet (60° counterclockwise from the y axis), and so θ is 60° . (The larger angle between the two directions, namely 120° , can also be used.) We have

$$I_2 = I_1 \cos^2 60^\circ.$$

The polarization of this transmitted light is parallel to the polarizing direction of the sheet transmitting it—that is, 60° counterclockwise from the y axis, as shown in the head-on view of Fig. 33-15d.

Third sheet: Because the light reaching the third sheet is

polarized, the intensity I_3 of the light transmitted by that sheet is given by the cosine-squared rule. The angle θ is now the angle between the polarization direction of the entering light (Fig. 33-15d) and the polarizing direction of the third sheet (parallel to the x axis), and so $\theta = 30^\circ$. Thus,

$$I_3 = I_2 \cos^2 30^\circ.$$

This final transmitted light is polarized parallel to the x axis (Fig. 33-15e). We find its intensity by substituting first for I_2 and then for I_1 in the equation above:

$$\begin{aligned} I_3 &= I_2 \cos^2 30^\circ = (I_1 \cos^2 60^\circ) \cos^2 30^\circ \\ &= (\frac{1}{2} I_0) \cos^2 60^\circ \cos^2 30^\circ = 0.094 I_0. \end{aligned}$$

Thus, $\frac{I_3}{I_0} = 0.094$. (Answer)

That is to say, 9.4% of the initial intensity emerges from the three-sheet system. (If we now remove the second sheet, what fraction of the initial intensity emerges from the system?)



Additional examples, video, and practice available at WileyPLUS

33-5 REFLECTION AND REFRACTION

Learning Objectives

After reading this module, you should be able to . . .

33.36 With a sketch, show the reflection of a light ray from an interface and identify the incident ray, the reflected ray, the normal, the angle of incidence, and the angle of reflection.

33.37 For a reflection, relate the angle of incidence and the angle of reflection.

33.38 With a sketch, show the refraction of a light ray at an interface and identify the incident ray, the refracted ray, the normal on each side of the interface, the angle of incidence, and the angle of refraction.

33.39 For refraction of light, apply Snell's law to relate the index of refraction and the angle of the ray on one side of the interface to those quantities on the other side.

33.40 In a sketch and using a line along the undeflected direction, show the refraction of light from one material into

a second material that has a greater index, a smaller index, and the same index, and, for each situation, describe the refraction in terms of the ray being bent toward the normal, away from the normal, or not at all.

33.41 Identify that refraction occurs only at an interface and not in the interior of a material.

33.42 Identify chromatic dispersion.

33.43 For a beam of red and blue light (or other colors) refracting at an interface, identify which color has the greater bending and which has the greater angle of refraction when they enter a material with a lower index than the initial material and a greater index.

33.44 Describe how the primary and secondary rainbows are formed and explain why they are circular arcs.

Key Ideas

- Geometrical optics is an approximate treatment of light in which light waves are represented as straight-line rays.

- When a light ray encounters a boundary between two transparent media, a reflected ray and a refracted ray generally appear. Both rays remain in the plane of incidence. The angle of reflection is equal to the angle of incidence, and

the angle of refraction is related to the angle of incidence by Snell's law,

$$n_2 \sin \theta_2 = n_1 \sin \theta_1 \quad (\text{refraction}),$$

where n_1 and n_2 are the indexes of refraction of the media in which the incident and refracted rays travel.

Reflection and Refraction

Although a light wave spreads as it moves away from its source, we can often approximate its travel as being in a straight line; we did so for the light wave in Fig. 33-5a. The study of the properties of light waves under that approximation is called *geometrical optics*. For the rest of this chapter and all of Chapter 34, we shall discuss the geometrical optics of visible light.

The photograph in Fig. 33-16a shows an example of light waves traveling in approximately straight lines. A narrow beam of light (the *incident beam*), angled downward from the left and traveling through air, encounters a *plane* (flat) water surface. Part of the light is **reflected** by the surface, forming a beam directed upward toward the right, traveling as if the original beam had bounced from the surface. The rest of the light travels through the surface and into the water, forming a beam directed downward to the right. Because light can travel through it, the water is said to be *transparent*; that is, we can see through it. (In this chapter we shall consider only transparent materials and not opaque materials, through which light cannot travel.)

The travel of light through a surface (or *interface*) that separates two media is called **refraction**, and the light is said to be *refracted*. Unless an incident beam of light is perpendicular to the surface, refraction changes the light's direction of travel. For this reason, the beam is said to be "bent" by the refraction. Note in Fig. 33-16a that the bending occurs only at the surface; within the water, the light travels in a straight line.

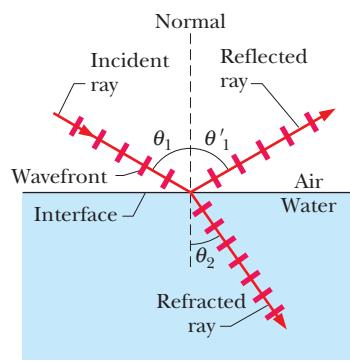
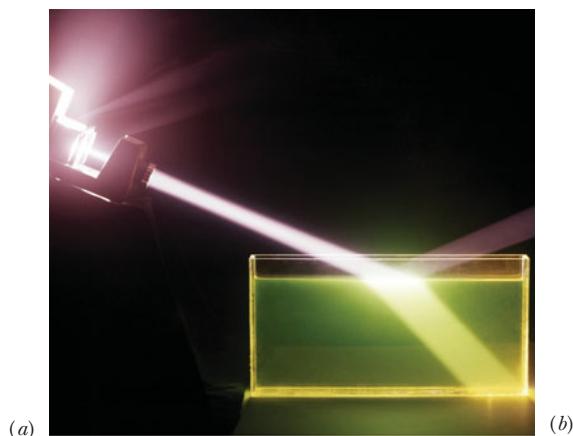
In Figure 33-16b, the beams of light in the photograph are represented with an *incident ray*, a *reflected ray*, and a *refracted ray* (and wavefronts). Each ray is oriented with respect to a line, called the *normal*, that is perpendicular to the surface at the point of reflection and refraction. In Fig. 33-16b, the **angle of incidence** is θ_1 , the **angle of reflection** is θ'_1 , and the **angle of refraction** is θ_2 , all measured relative to the normal. The plane containing the incident ray and the normal is the *plane of incidence*, which is in the plane of the page in Fig. 33-16b.

Experiment shows that reflection and refraction are governed by two laws:

Law of reflection: A reflected ray lies in the plane of incidence and has an angle of reflection equal to the angle of incidence (both relative to the normal). In Fig. 33-16b, this means that

$$\theta'_1 = \theta_1 \quad (\text{reflection}). \quad (33-39)$$

(We shall now usually drop the prime on the angle of reflection.)



©1974 FP/Fundamental Photographs

Figure 33-16 (a) A photograph showing an incident beam of light reflected and refracted by a horizontal water surface. (b) A ray representation of (a). The angles of incidence (θ_1), reflection (θ'_1), and refraction (θ_2) are marked.

Table 33-1 Some Indexes of Refraction^a

Medium	Index	Medium	Index
Vacuum	Exactly 1	Typical crown glass	1.52
Air (STP) ^b	1.00029	Sodium chloride	1.54
Water (20°C)	1.33	Polystyrene	1.55
Acetone	1.36	Carbon disulfide	1.63
Ethyl alcohol	1.36	Heavy flint glass	1.65
Sugar solution (30%)	1.38	Sapphire	1.77
Fused quartz	1.46	Heaviest flint glass	1.89
Sugar solution (80%)	1.49	Diamond	2.42

^aFor a wavelength of 589 nm (yellow sodium light).

^bSTP means “standard temperature (0°C) and pressure (1 atm).”

Law of refraction: A refracted ray lies in the plane of incidence and has an angle of refraction θ_2 that is related to the angle of incidence θ_1 by

$$n_2 \sin \theta_2 = n_1 \sin \theta_1 \quad (\text{refraction}). \quad (33-40)$$

Here each of the symbols n_1 and n_2 is a dimensionless constant, called the **index of refraction**, that is associated with a medium involved in the refraction. We derive this equation, called **Snell's law**, in Chapter 35. As we shall discuss there, the index of refraction of a medium is equal to c/v , where v is the speed of light in that medium and c is its speed in vacuum.

Table 33-1 gives the indexes of refraction of vacuum and some common substances. For vacuum, n is defined to be exactly 1; for air, n is very close to 1.0 (an approximation we shall often make). Nothing has an index of refraction below 1.

We can rearrange Eq. 33-40 as

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1 \quad (33-41)$$

to compare the angle of refraction θ_2 with the angle of incidence θ_1 . We can then see that the relative value of θ_2 depends on the relative values of n_2 and n_1 :

1. If n_2 is equal to n_1 , then θ_2 is equal to θ_1 and refraction does not bend the light beam, which continues in the *undeflected direction*, as in Fig. 33-17a.

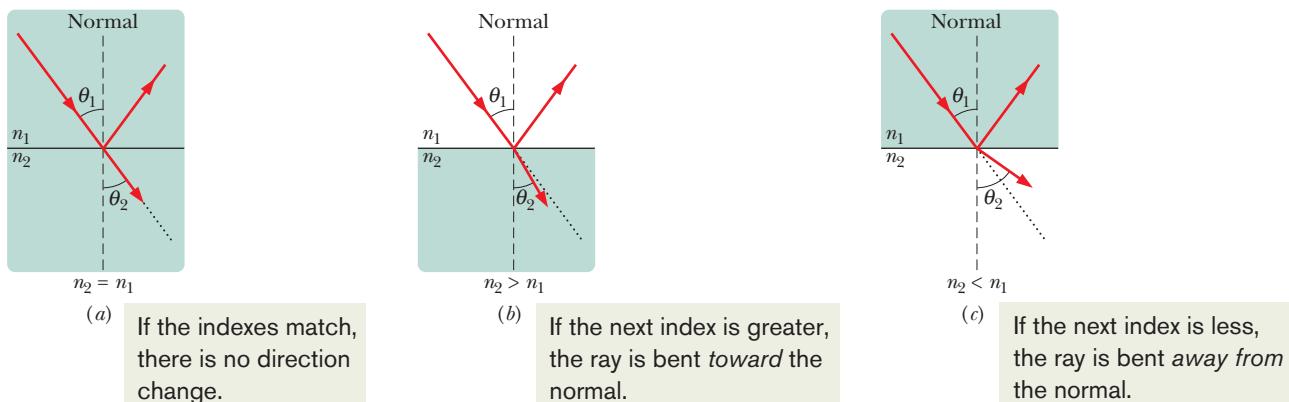


Figure 33-17 Refraction of light traveling from a medium with an index of refraction n_1 into a medium with an index of refraction n_2 . (a) The beam does not bend when $n_2 = n_1$; the refracted light then travels in the *undeflected direction* (the dotted line), which is the same as the direction of the incident beam. The beam bends (b) toward the normal when $n_2 > n_1$ and (c) away from the normal when $n_2 < n_1$.

2. If n_2 is greater than n_1 , then θ_2 is less than θ_1 . In this case, refraction bends the light beam away from the undeflected direction and toward the normal, as in Fig. 33-17b.
3. If n_2 is less than n_1 , then θ_2 is greater than θ_1 . In this case, refraction bends the light beam away from the undeflected direction and away from the normal, as in Fig. 33-17c.

Refraction *cannot* bend a beam so much that the refracted ray is on the same side of the normal as the incident ray.

Chromatic Dispersion

The index of refraction n encountered by light in any medium except vacuum depends on the wavelength of the light. The dependence of n on wavelength implies that when a light beam consists of rays of different wavelengths, the rays will be refracted at different angles by a surface; that is, the light will be spread out by the refraction. This spreading of light is called **chromatic dispersion**, in which “chromatic” refers to the colors associated with the individual wavelengths and “dispersion” refers to the spreading of the light according to its wavelengths or colors. The refractions of Figs. 33-16 and 33-17 do not show chromatic dispersion because the beams are *monochromatic* (of a single wavelength or color).

Generally, the index of refraction of a given medium is *greater* for a shorter wavelength (corresponding to, say, blue light) than for a longer wavelength (say, red light). As an example, Fig. 33-18 shows how the index of refraction of fused quartz depends on the wavelength of light. Such dependence means that when a beam made up of waves of both blue and red light is refracted through a surface, such as from air into quartz or vice versa, the blue *component* (the ray corresponding to the wave of blue light) bends more than the red component.

A beam of *white light* consists of components of all (or nearly all) the colors in the visible spectrum with approximately uniform intensities. When you see such a beam, you perceive white rather than the individual colors. In Fig. 33-19a, a beam of white light in air is incident on a glass surface. (Because the pages of this book are white, a beam of white light is represented with a gray ray here. Also, a beam of monochromatic light is generally represented with a red ray.) Of the refracted light in Fig. 33-19a, only the red and blue components are shown. Because the blue component is bent more than the red component, the angle of refraction θ_{2b} for the blue component is *smaller* than the angle of refraction θ_{2r} for the red component. (Remember, angles are measured relative to the normal.) In Fig. 33-19b, a ray of white light in glass is incident on a glass–air interface. Again, the blue component is bent more than the red component, but now θ_{2b} is greater than θ_{2r} .

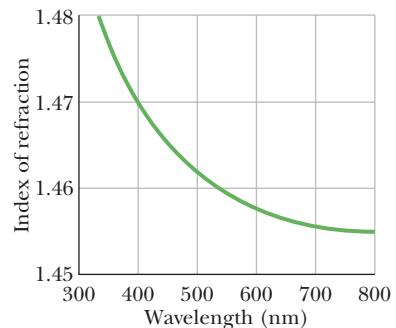


Figure 33-18 The index of refraction as a function of wavelength for fused quartz. The graph indicates that a beam of short-wavelength light, for which the index of refraction is higher, is bent more upon entering or leaving quartz than a beam of long-wavelength light.

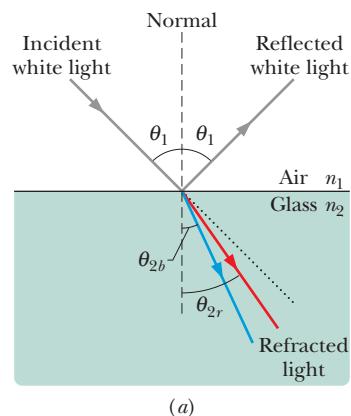
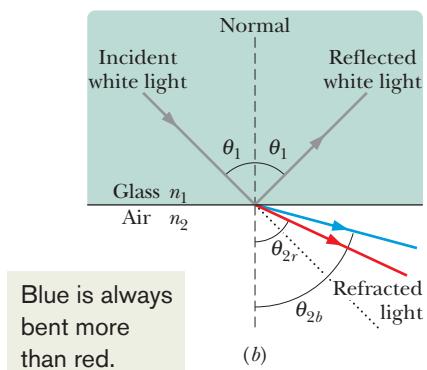
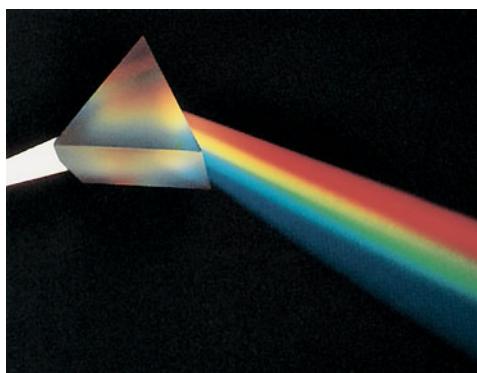


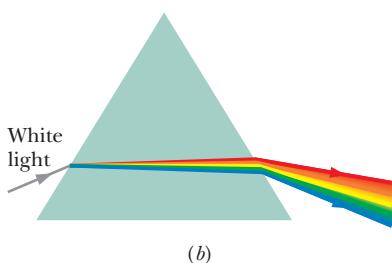
Figure 33-19 Chromatic dispersion of white light. The blue component is bent more than the red component. (a) Passing from air to glass, the blue component ends up with the smaller angle of refraction. (b) Passing from glass to air, the blue component ends up with the greater angle of refraction. Each dotted line represents the direction in which the light would continue to travel if it were not bent by the refraction.





Courtesy Bausch & Lomb

(a)



(b)

Figure 33-20 (a) A triangular prism separating white light into its component colors. (b) Chromatic dispersion occurs at the first surface and is increased at the second surface.

To increase the color separation, we can use a solid glass prism with a triangular cross section, as in Fig. 33-20a. The dispersion at the first surface (on the left in Figs. 33-20a, b) is then enhanced by the dispersion at the second surface.

Rainbows

The most charming example of chromatic dispersion is a rainbow. When sunlight (which consists of all visible colors) is intercepted by a falling raindrop, some of the light refracts into the drop, reflects once from the drop's inner surface, and then refracts out of the drop. Figure 33-21a shows the situation when the Sun is on the horizon at the left (and thus when the rays of sunlight are horizontal). The first refraction separates the sunlight into its component colors, and the second refraction increases the separation. (Only the red and blue rays are shown in the figure.) If many falling drops are brightly illuminated, you can see the separated colors they produce when the drops are at an angle of 42° from the direction of the *antisolar point* A , the point directly opposite the Sun in your view.

To locate the drops, face away from the Sun and point both arms directly away from the Sun, toward the shadow of your head. Then move your right arm directly up, directly rightward, or in any intermediate direction until the angle between your arms is 42° . If illuminated drops happen to be in the direction of your right arm, you see color in that direction.

Because any drop at an angle of 42° in any direction from A can contribute to the rainbow, the rainbow is always a 42° circular arc around A (Fig. 33-21b) and the top of a rainbow is never more than 42° above the horizon. When the Sun is above the horizon, the direction of A is below the horizon, and only a shorter, lower rainbow arc is possible (Fig. 33-21c).

Because rainbows formed in this way involve one reflection of light inside each drop, they are often called *primary rainbows*. A *secondary rainbow* involves two reflections inside a drop, as shown in Fig. 33-21d. Colors appear in the secondary rainbow at an angle of 52° from the direction of A . A secondary rainbow

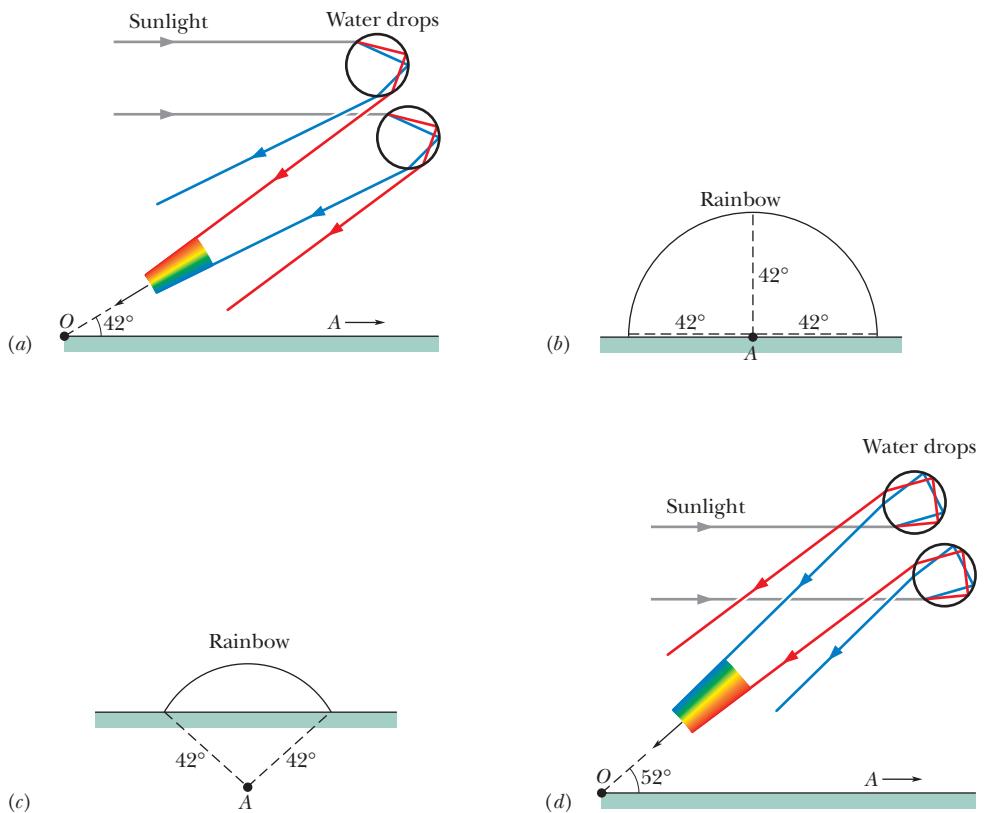


Figure 33-21 (a) The separation of colors when sunlight refracts into and out of falling raindrops leads to a primary rainbow. The antisolar point A is on the horizon at the right. The rainbow colors appear at an angle of 42° from the direction of A . (b) Drops at 42° from A in any direction can contribute to the rainbow. (c) The rainbow arc when the Sun is higher (and thus A is lower). (d) The separation of colors leading to a secondary rainbow.

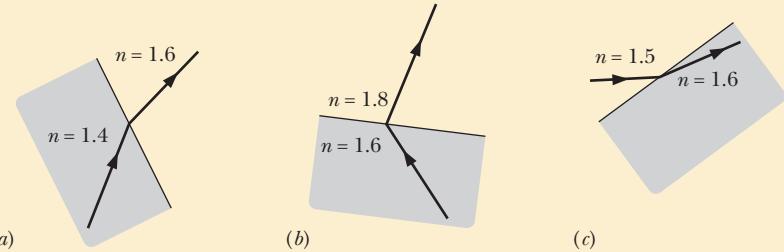
is wider and dimmer than a primary rainbow and thus is more difficult to see. Also, the order of colors in a secondary rainbow is reversed from the order in a primary rainbow, as you can see by comparing parts *a* and *d* of Fig. 33-21.

Rainbows involving three or four reflections occur in the direction of the Sun and cannot be seen against the glare of sunshine in that part of the sky but have been photographed with special techniques.



Checkpoint 5

Which of the three drawings here (if any) show physically possible refraction?



Sample Problem 33.03 Reflection and refraction of a monochromatic beam

(a) In Fig. 33-22*a*, a beam of monochromatic light reflects and refracts at point *A* on the interface between material 1 with index of refraction $n_1 = 1.33$ and material 2 with index of refraction $n_2 = 1.77$. The incident beam makes an angle of 50° with the interface. What is the angle of reflection at point *A*? What is the angle of refraction there?

KEY IDEAS

(1) The angle of reflection is equal to the angle of incidence, and both angles are measured relative to the normal to the surface at the point of reflection. (2) When light reaches the interface between two materials with different indexes of refraction (call them n_1 and n_2), part of the light can be refracted by the interface according to Snell's law, Eq. 33-40:

$$n_2 \sin \theta_2 = n_1 \sin \theta_1, \quad (33-42)$$

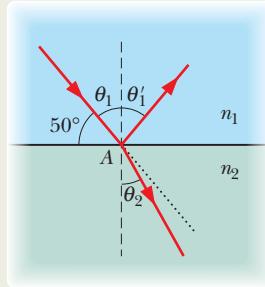
where both angles are measured relative to the normal at the point of refraction.

Calculations: In Fig. 33-22*a*, the normal at point *A* is drawn as a dashed line through the point. Note that the angle of incidence θ_1 is not the given 50° but is $90^\circ - 50^\circ = 40^\circ$. Thus, the angle of reflection is

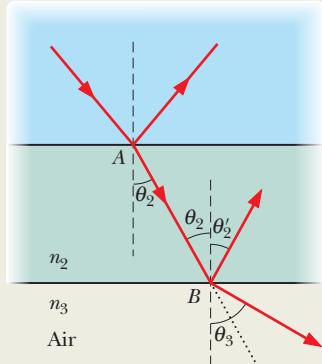
$$\theta'_1 = \theta_1 = 40^\circ. \quad (\text{Answer})$$

The light that passes from material 1 into material 2 undergoes refraction at point *A* on the interface between the two materials. Again we measure angles between light rays and a normal, here at the point of refraction. Thus, in Fig. 33-22*a*, the angle of refraction is the angle marked θ_2 . Solving Eq. 33-42 for θ_2 gives us

$$\theta_2 = \sin^{-1} \left(\frac{n_1}{n_2} \sin \theta_1 \right) = \sin^{-1} \left(\frac{1.33}{1.77} \sin 40^\circ \right) \\ = 28.88^\circ \approx 29^\circ. \quad (\text{Answer})$$



(a)



(b)

Figure 33-22 (a) Light reflects and refracts at point *A* on the interface between materials 1 and 2. (b) The light that passes through material 2 reflects and refracts at point *B* on the interface between materials 2 and 3 (air). Each dashed line is a normal. Each dotted line gives the incident direction of travel.

This result means that the beam swings toward the normal (it was at 40° to the normal and is now at 29°). The reason is that when the light travels across the interface, it moves into a material with a greater index of refraction. **Caution:** Note that the beam does *not* swing through the normal so that it appears on the left side of Fig. 33-22*a*.

(b) The light that enters material 2 at point *A* then reaches point *B* on the interface between material 2 and material 3, which is air, as shown in Fig. 33-22*b*. The interface through *B* is parallel to that through *A*. At *B*, some of the light reflects and the rest enters the air. What is the angle of reflection? What is the angle of refraction into the air?

Calculations: We first need to relate one of the angles at

point *B* with a known angle at point *A*. Because the interface through point *B* is parallel to that through point *A*, the incident angle at *B* must be equal to the angle of refraction θ_2 , as shown in Fig. 33-22*b*. Then for reflection, we again use the law of reflection. Thus, the angle of reflection at *B* is

$$\theta'_2 = \theta_2 = 28.88^\circ \approx 29^\circ. \quad (\text{Answer})$$

Next, the light that passes from material 2 into the air undergoes refraction at point *B*, with refraction angle θ_3 . Thus, we again apply Snell's law of refraction, but this time

we write Eq. 33-40 as

$$n_3 \sin \theta_3 = n_2 \sin \theta_2. \quad (33-43)$$

Solving for θ_3 then leads to

$$\begin{aligned} \theta_3 &= \sin^{-1} \left(\frac{n_2}{n_3} \sin \theta_2 \right) = \sin^{-1} \left(\frac{1.77}{1.00} \sin 28.88^\circ \right) \\ &= 58.75^\circ \approx 59^\circ. \end{aligned} \quad (\text{Answer})$$

Thus, the beam swings away from the normal (it was at 29° to the normal and is now at 59°) because it moves into a material (air) with a lower index of refraction.



Additional examples, video, and practice available at WileyPLUS

33-6 TOTAL INTERNAL REFLECTION

Learning Objectives

After reading this module, you should be able to ...

33.45 With sketches, explain total internal reflection and include the angle of incidence, the critical angle, and the relative values of the indexes of refraction on the two sides of the interface.

33.46 Identify the angle of refraction for incidence at a critical angle.

33.47 For a given pair of indexes of refraction, calculate the critical angle.

Key Idea

• A wave encountering a boundary across which the index of refraction decreases will experience total internal reflection if the angle of incidence exceeds a critical angle θ_c , where

$$\theta_c = \sin^{-1} \frac{n_2}{n_1} \quad (\text{critical angle}).$$

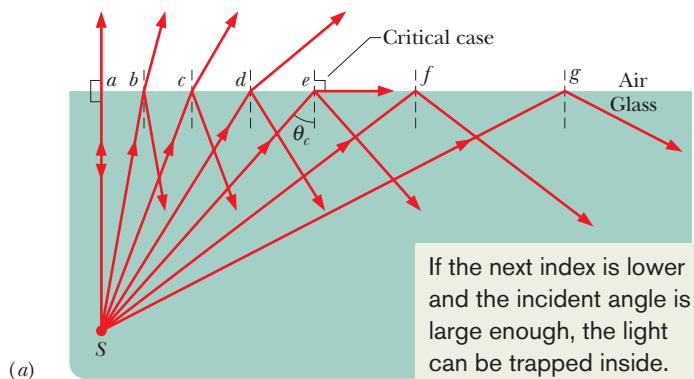
Total Internal Reflection

Figure 33-23*a* shows rays of monochromatic light from a point source *S* in glass incident on the interface between the glass and air. For ray *a*, which is perpendicular to the interface, part of the light reflects at the interface and the rest travels through it with no change in direction.

For rays *b* through *e*, which have progressively larger angles of incidence at the interface, there are also both reflection and refraction at the interface. As the angle of incidence increases, the angle of refraction increases; for ray *e* it is 90° , which means that the refracted ray points directly along the interface. The angle of incidence giving this situation is called the **critical angle** θ_c . For angles of incidence larger than θ_c , such as for rays *f* and *g*, there is no refracted ray and *all* the light is reflected; this effect is called **total internal reflection** because all the light remains inside the glass.

To find θ_c , we use Eq. 33-40; we arbitrarily associate subscript 1 with the glass and subscript 2 with the air, and then we substitute θ_c for θ_1 and 90° for θ_2 , which leads to

$$n_1 \sin \theta_c = n_2 \sin 90^\circ, \quad (33-44)$$



If the next index is lower and the incident angle is large enough, the light can be trapped inside.



Ken Kay/Fundamental Photographs

Figure 33-23 (a) Total internal reflection of light from a point source S in glass occurs for all angles of incidence greater than the critical angle θ_c . At the critical angle, the refracted ray points along the air–glass interface. (b) A source in a tank of water.

which gives us

$$\theta_c = \sin^{-1} \frac{n_2}{n_1} \quad (\text{critical angle}). \quad (33-45)$$

Because the sine of an angle cannot exceed unity, n_2 cannot exceed n_1 in this equation. This restriction tells us that total internal reflection cannot occur when the incident light is in the medium of lower index of refraction. If source S were in the air in Fig. 33-23a, all its rays that are incident on the air–glass interface (including f and g) would be both reflected *and* refracted at the interface.

Total internal reflection has found many applications in medical technology. For example, a physician can view the interior of an artery of a patient by running two thin bundles of *optical fibers* through the chest wall and into an artery (Fig. 33-24). Light introduced at the outer end of one bundle undergoes repeated total internal reflection within the fibers so that, even though the bundle provides a curved path, most of the light ends up exiting the other end and illuminating the interior of the artery. Some of the light reflected from the interior then comes back up the second bundle in a similar way, to be detected and converted to an image on a monitor's screen for the physician to view. The physician can then perform a surgical procedure, such as the placement of a stent.



©Laurent/Phototake

Figure 33-24 An endoscope used to inspect an artery.

33-7 POLARIZATION BY REFLECTION

Learning Objectives

After reading this module, you should be able to . . .

33.48 With sketches, explain how unpolarized light can be converted to polarized light by reflection from an interface.

33.49 Identify Brewster's angle.

33.50 Apply the relationship between Brewster's angle and the indexes of refraction on the two sides of an interface.

33.51 Explain the function of polarizing sunglasses.

Key Idea

- A reflected wave will be fully polarized, with its \vec{E} vectors perpendicular to the plane of incidence, if it strikes a boundary at the Brewster angle θ_B , where

$$\theta_B = \tan^{-1} \frac{n_2}{n_1} \quad (\text{Brewster angle}).$$

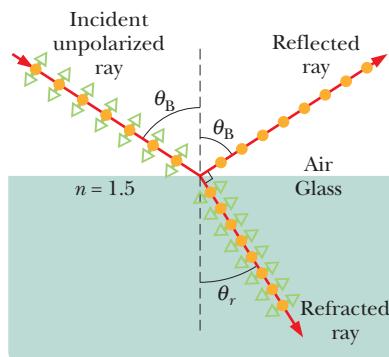


Figure 33-25 A ray of unpolarized light in air is incident on a glass surface at the Brewster angle θ_B . The electric fields along that ray have been resolved into components perpendicular to the page (the plane of incidence, reflection, and refraction) and components parallel to the page. The reflected light consists only of components perpendicular to the page and is thus polarized in that direction. The refracted light consists of the original components parallel to the page and weaker components perpendicular to the page; this light is partially polarized.

Polarization by Reflection

You can vary the glare you see in sunlight that has been reflected from, say, water by looking through a polarizing sheet (such as a polarizing sunglass lens) and then rotating the sheet's polarizing axis around your line of sight. You can do so because any light that is reflected from a surface is either fully or partially polarized by the reflection.

Figure 33-25 shows a ray of unpolarized light incident on a glass surface. Let us resolve the electric field vectors of the light into two components. The *perpendicular components* are perpendicular to the plane of incidence and thus also to the page in Fig. 33-25; these components are represented with dots (as if we see the tips of the vectors). The *parallel components* are parallel to the plane of incidence and the page; they are represented with double-headed arrows. Because the light is unpolarized, these two components are of equal magnitude.

In general, the reflected light also has both components but with unequal magnitudes. This means that the reflected light is partially polarized—the electric fields oscillating along one direction have greater amplitudes than those oscillating along other directions. However, when the light is incident at a particular incident angle, called the *Brewster angle* θ_B , the reflected light has only perpendicular components, as shown in Fig. 33-25. The reflected light is then fully polarized perpendicular to the plane of incidence. The parallel components of the incident light do not disappear but (along with perpendicular components) refract into the glass.

Polarizing Sunglasses. Glass, water, and the other dielectric materials discussed in Module 25-5 can partially and fully polarize light by reflection. When you intercept sunlight reflected from such a surface, you see a bright spot (the glare) on the surface where the reflection takes place. If the surface is horizontal as in Fig. 33-25, the reflected light is partially or fully polarized horizontally. To eliminate such glare from horizontal surfaces, the lenses in polarizing sunglasses are mounted with their polarizing direction vertical.



Brewster's Law

For light incident at the Brewster angle θ_B , we find experimentally that the reflected and refracted rays are perpendicular to each other. Because the reflected ray is reflected at the angle θ_B in Fig. 33-25 and the refracted ray is at an angle θ_r , we have

$$\theta_B + \theta_r = 90^\circ. \quad (33-46)$$

These two angles can also be related with Eq. 33-40. Arbitrarily assigning subscript 1 in Eq. 33-40 to the material through which the incident and reflected rays travel, we have, from that equation,

$$n_1 \sin \theta_B = n_2 \sin \theta_r. \quad (33-47)$$

Combining these equations leads to

$$n_1 \sin \theta_B = n_2 \sin(90^\circ - \theta_B) = n_2 \cos \theta_B, \quad (33-48)$$

which gives us

$$\theta_B = \tan^{-1} \frac{n_2}{n_1} \quad (\text{Brewster angle}). \quad (33-49)$$

(Note carefully that the subscripts in Eq. 33-49 are *not* arbitrary because of our decision as to their meanings.) If the incident and reflected rays travel *in air*, we can approximate n_1 as unity and let n represent n_2 in order to write Eq. 33-49 as

$$\theta_B = \tan^{-1} n \quad (\text{Brewster's law}). \quad (33-50)$$

This simplified version of Eq. 33-49 is known as **Brewster's law**. Like θ_B , it is named after Sir David Brewster, who found both experimentally in 1812.

Review & Summary

Electromagnetic Waves An electromagnetic wave consists of oscillating electric and magnetic fields. The various possible frequencies of electromagnetic waves form a *spectrum*, a small part of which is visible light. An electromagnetic wave traveling along an x axis has an electric field \vec{E} and a magnetic field \vec{B} with magnitudes that depend on x and t :

$$E = E_m \sin(kx - \omega t)$$

and $B = B_m \sin(kx - \omega t)$, (33-1, 33-2)

where E_m and B_m are the amplitudes of \vec{E} and \vec{B} . The oscillating electric field induces the magnetic field, and the oscillating magnetic field induces the electric field. The speed of any electromagnetic wave in vacuum is c , which can be written as

$$c = \frac{E}{B} = \frac{1}{\sqrt{\mu_0 \epsilon_0}}, \quad (33-5, 33-3)$$

where E and B are the simultaneous (but nonzero) magnitudes of the two fields.

Energy Flow The rate per unit area at which energy is transported via an electromagnetic wave is given by the Poynting vector \vec{S} :

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}. \quad (33-19)$$

The direction of \vec{S} (and thus of the wave's travel and the energy transport) is perpendicular to the directions of both \vec{E} and \vec{B} . The time-averaged rate per unit area at which energy is transported is S_{avg} , which is called the *intensity* I of the wave:

$$I = \frac{1}{c\mu_0} E_{\text{rms}}^2, \quad (33-26)$$

in which $E_{\text{rms}} = E_m/\sqrt{2}$. A *point source* of electromagnetic waves emits the waves *isotropically*—that is, with equal intensity in all directions. The intensity of the waves at distance r from a point source of power P_s is

$$I = \frac{P_s}{4\pi r^2}. \quad (33-27)$$

Radiation Pressure When a surface intercepts electromagnetic radiation, a force and a pressure are exerted on the surface. If the radiation is totally absorbed by the surface, the force is

$$F = \frac{IA}{c} \quad (\text{total absorption}), \quad (33-32)$$

in which I is the intensity of the radiation and A is the area of the surface perpendicular to the path of the radiation. If the radiation is totally reflected back along its original path, the force is

$$F = \frac{2IA}{c} \quad (\text{total reflection back along path}). \quad (33-33)$$

The radiation pressure p_r is the force per unit area:

$$p_r = \frac{I}{c} \quad (\text{total absorption}) \quad (33-34)$$

and $p_r = \frac{2I}{c} \quad (\text{total reflection back along path}). \quad (33-35)$

Polarization Electromagnetic waves are **polarized** if their electric field vectors are all in a single plane, called the *plane of oscillation*. From a head-on view, the field vectors oscillate parallel to a single axis perpendicular to the path taken by the waves. Light waves from common sources are not polarized; that is, they are **unpolarized**, or **polarized randomly**. From a head-on view, the vectors oscillate parallel to every possible axis that is perpendicular to the path taken by the waves.

Polarizing Sheets When a polarizing sheet is placed in the path of light, only electric field components of the light parallel to the sheet's **polarizing direction** are *transmitted* by the sheet; components perpendicular to the polarizing direction are absorbed. The light that emerges from a polarizing sheet is polarized parallel to the polarizing direction of the sheet.

If the original light is initially unpolarized, the transmitted intensity I is half the original intensity I_0 :

$$I = \frac{1}{2} I_0. \quad (33-36)$$

If the original light is initially polarized, the transmitted intensity depends on the angle θ between the polarization direction of the original light (the axis along which the fields oscillate) and the polarizing direction of the sheet:

$$I = I_0 \cos^2 \theta. \quad (33-38)$$

Geometrical Optics *Geometrical optics* is an approximate treatment of light in which light waves are represented as straight-line rays.

Reflection and Refraction When a light ray encounters a boundary between two transparent media, a **reflected** ray and a **refracted** ray generally appear. Both rays remain in the plane of incidence. The **angle of reflection** is equal to the angle of incidence, and the **angle of refraction** is related to the angle of incidence by Snell's law,

$$n_2 \sin \theta_2 = n_1 \sin \theta_1 \quad (\text{refraction}), \quad (33-40)$$

where n_1 and n_2 are the indexes of refraction of the media in which the incident and refracted rays travel.

Total Internal Reflection A wave encountering a boundary across which the index of refraction decreases will experience **total internal reflection** if the angle of incidence exceeds a **critical angle** θ_c , where

$$\theta_c = \sin^{-1} \frac{n_2}{n_1} \quad (\text{critical angle}). \quad (33-45)$$

Polarization by Reflection A reflected wave will be fully **polarized**, with its \vec{E} vectors perpendicular to the plane of incidence, if the incident, unpolarized wave strikes a boundary at the **Brewster angle** θ_B , where

$$\theta_B = \tan^{-1} \frac{n_2}{n_1} \quad (\text{Brewster angle}). \quad (33-49)$$

Questions

- 1** If the magnetic field of a light wave oscillates parallel to a y axis and is given by $B_y = B_m \sin(kz - \omega t)$, (a) in what direction does the wave travel and (b) parallel to which axis does the associated electric field oscillate?

- 2** Suppose we rotate the second sheet in Fig. 33-15a, starting with the polarization direction aligned with the y axis ($\theta = 0^\circ$) and ending with it aligned with the x axis ($\theta = 90^\circ$). Which of the four curves in Fig. 33-26 best shows the intensity of the light through the three-sheet system during this 90° rotation?

- 3** Figure 33-27 shows light reaching a polarizing sheet whose polarizing direction is parallel to a y axis. We shall rotate the sheet 40° clockwise about the light's indicated line of travel. During this rotation, does the fraction of the initial light intensity passed by the sheet increase, decrease, or remain the same if the light is (a) initially unpolarized, (b) initially polarized parallel to the x axis, and (c) initially polarized parallel to the y axis?

- 4** Figure 33-28 shows the electric and magnetic fields of an electromagnetic wave at a certain instant. Is the wave traveling into the page or out of the page?

- 5** In the arrangement of Fig. 33-15a, start with light that is initially polarized parallel to the x axis, and write the ratio of its final intensity I_3 to its initial intensity I_0 as $I_3/I_0 = A \cos^n \theta$. What are A , n , and θ if we rotate the polarizing direction of the first sheet (a) 60° counterclockwise and (b) 90° clockwise from what is shown?

- 6** In Fig. 33-29, unpolarized light is sent into a system of five polarizing sheets. Their polarizing directions, measured counterclockwise from the positive direction of the y axis, are the following: sheet 1, 35° ; sheet 2, 0° ; sheet 3, 0° ; sheet 4, 110° ; sheet 5, 45° . Sheet 3 is then rotated 180° counterclockwise about the light ray. During that rotation, at what angles (measured counterclockwise from the y axis) is the transmission of light through the system eliminated?

- 7** Figure 33-30 shows rays of monochromatic light propagating through three materials a , b , and c . Rank the materials according to the index of refraction, greatest first.

- 8** Figure 33-31 shows the multiple reflections of a light ray along a glass corridor where the walls are either parallel or perpendicular to one another. If the angle of incidence at point a is 30° , what are

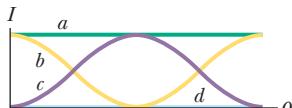


Figure 33-26 Question 2.

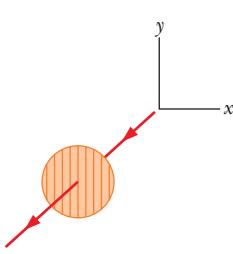


Figure 33-27 Question 3.

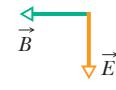


Figure 33-28
Question 4.

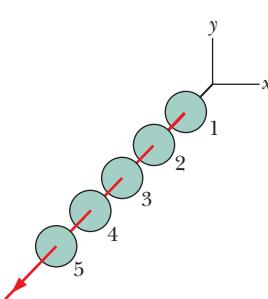


Figure 33-29 Question 6.

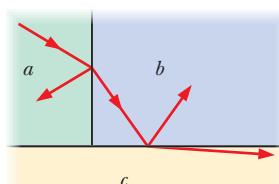


Figure 33-30 Question 7.

the angles of reflection of the light ray at points b , c , d , e , and f ?

- 9** Figure 33-32 shows four long horizontal layers A – D of different materials, with air above and below them. The index of refraction of each material is given. Rays of light are sent into the left end of each layer as shown. In which layer is there the possibility of totally trapping the light in that layer so that, after many reflections, all the light reaches the right end of the layer?

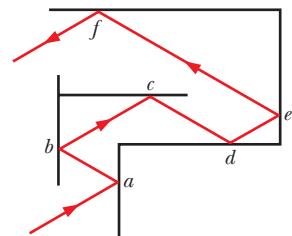


Figure 33-31 Question 8.

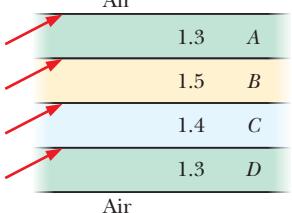


Figure 33-32 Question 9.

- 10** The leftmost block in Fig. 33-33 depicts total internal reflection for light inside a material with an index of refraction n_1 when air is outside the material. A light ray reaching point A from anywhere within the shaded region at the left (such as the ray shown) fully reflects at that point and ends up in the shaded region at the right. The other blocks show similar situations for two other materials. Rank the indexes of refraction of the three materials, greatest first.

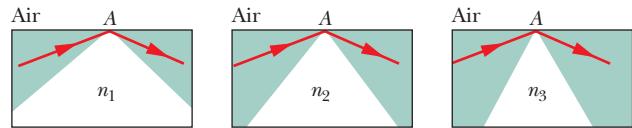


Figure 33-33 Question 10.

- 11** Each part of Fig. 33-34 shows light that refracts through an interface between two materials. The incident ray (shown gray in the figure) consists of red and blue light. The approximate index of refraction for visible light is indicated for each material. Which of the three parts show physically possible refraction? (Hint: First consider the refraction in general, regardless of the color, and then consider how red and blue light refract differently.)

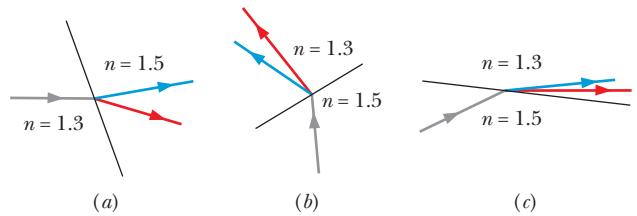


Figure 33-34 Question 11.

- 12** In Fig. 33-35, light travels from material a , through three layers of other materials with surfaces parallel to one another, and then back into another layer of material a . The refractions (but not the associated reflections) at the surfaces are shown. Rank the materials according to index of refraction, greatest first. (Hint: The parallel arrangement of the surfaces allows comparison.)

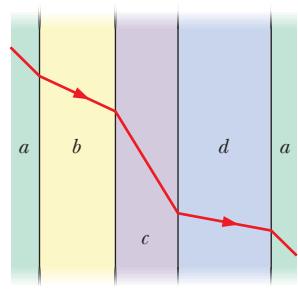


Figure 33-35 Question 12.

Problems



Tutoring problem available (at instructor's discretion) in WileyPLUS and WebAssign



Worked-out solution available in Student Solutions Manual



Number of dots indicates level of problem difficulty



Additional information available in *The Flying Circus of Physics* and at flyingcircusofphysics.com



Worked-out solution is at



Interactive solution is at

<http://www.wiley.com/college/halliday>

Module 33-1 Electromagnetic Waves

•1 A certain helium–neon laser emits red light in a narrow band of wavelengths centered at 632.8 nm and with a “wavelength width” (such as on the scale of Fig. 33-1) of 0.0100 nm. What is the corresponding “frequency width” for the emission?

•2 Project Seafarer was an ambitious program to construct an enormous antenna, buried underground on a site about 10 000 km² in area. Its purpose was to transmit signals to submarines while they were deeply submerged. If the effective wavelength were 1.0×10^4 Earth radii, what would be the (a) frequency and (b) period of the radiations emitted? Ordinarily, electromagnetic radiations do not penetrate very far into conductors such as seawater, and so normal signals cannot reach the submarines.

•3 From Fig. 33-2, approximate the (a) smaller and (b) larger wavelength at which the eye of a standard observer has half the eye’s maximum sensitivity. What are the (c) wavelength, (d) frequency, and (e) period of the light at which the eye is the most sensitive?

•4 About how far apart must you hold your hands for them to be separated by 1.0 nano-light-second (the distance light travels in 1.0 ns)?

•5 **SSM** What inductance must be connected to a 17 pF capacitor in an oscillator capable of generating 550 nm (i.e., visible) electromagnetic waves? Comment on your answer.

•6 What is the wavelength of the electromagnetic wave emitted by the oscillator–antenna system of Fig. 33-3 if $L = 0.253 \mu\text{H}$ and $C = 25.0 \text{ pF}$?

Module 33-2 Energy Transport and the Poynting Vector

•7 What is the intensity of a traveling plane electromagnetic wave if B_m is $1.0 \times 10^{-4} \text{ T}$?

•8 Assume (unrealistically) that a TV station acts as a point source broadcasting isotropically at 1.0 MW. What is the intensity of the transmitted signal reaching Proxima Centauri, the star nearest our solar system, 4.3 ly away? (An alien civilization at that distance might be able to watch *X Files*.) A light-year (ly) is the distance light travels in one year.

•9 **ILW** Some neodymium–glass lasers can provide 100 TW of power in 1.0 ns pulses at a wavelength of 0.26 μm. How much energy is contained in a single pulse?

•10 A plane electromagnetic wave has a maximum electric field magnitude of $3.20 \times 10^{-4} \text{ V/m}$. Find the magnetic field amplitude.

•11 **ILW** A plane electromagnetic wave traveling in the positive direction of an x axis in vacuum has components $E_x = E_y = 0$ and $E_z = (2.0 \text{ V/m}) \cos[(\pi \times 10^{15} \text{ s}^{-1})(t - x/c)]$. (a) What is the amplitude of the magnetic field component? (b) Parallel to which axis does the magnetic field oscillate? (c) When the electric field component is in the positive direction of the z axis at a certain point P , what is the direction of the magnetic field component there?

•12 In a plane radio wave the maximum value of the electric field component is 5.00 V/m. Calculate (a) the maximum value of the magnetic field component and (b) the wave intensity.

•13 Sunlight just outside Earth’s atmosphere has an intensity of 1.40 kW/m^2 . Calculate (a) E_m and (b) B_m for sunlight there, assuming it to be a plane wave.

•14 **GO** An isotropic point source emits light at wavelength 500 nm, at the rate of 200 W. A light detector is positioned 400 m from the source. What is the maximum rate $\partial B/\partial t$ at which the magnetic component of the light changes with time at the detector’s location?

•15 An airplane flying at a distance of 10 km from a radio transmitter receives a signal of intensity $10 \mu\text{W/m}^2$. What is the amplitude of the (a) electric and (b) magnetic component of the signal at the airplane? (c) If the transmitter radiates uniformly over a hemisphere, what is the transmission power?

•16 Frank D. Drake, an investigator in the SETI (Search for Extra-Terrestrial Intelligence) program, once said that the large radio telescope in Arecibo, Puerto Rico (Fig. 33-36), “can detect a signal which lays down on the entire surface of the earth a power of only one picowatt.” (a) What is the power that would be received by the Arecibo antenna for such a signal? The antenna diameter is 300 m. (b) What would be the power of an isotropic source at the center of our galaxy that could provide such a signal? The galactic center is 2.2×10^4 ly away. A light-year is the distance light travels in one year.



Courtesy SRI International, USRA, UMET

Figure 33-36 Problem 16. Radio telescope at Arecibo.

•17 The maximum electric field 10 m from an isotropic point source of light is 2.0 V/m. What are (a) the maximum value of the magnetic field and (b) the average intensity of the light there? (c) What is the power of the source?

•18 The intensity I of light from an isotropic point source is determined as a function of distance r from the source. Figure 33-37 gives

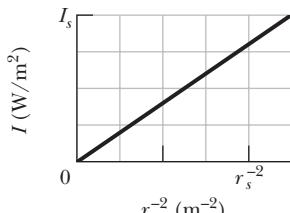


Figure 33-37 Problem 18.

intensity I versus the inverse square r^{-2} of that distance. The vertical axis scale is set by $I_s = 200 \text{ W/m}^2$, and the horizontal axis scale is set by $r_s^{-2} = 8.0 \text{ m}^{-2}$. What is the power of the source?

Module 33-3 Radiation Pressure

•19 SSM High-power lasers are used to compress a plasma (a gas of charged particles) by radiation pressure. A laser generating radiation pulses with peak power $1.5 \times 10^3 \text{ MW}$ is focused onto 1.0 mm^2 of high-electron-density plasma. Find the pressure exerted on the plasma if the plasma reflects all the light beams directly back along their paths.

•20 Radiation from the Sun reaching Earth (just outside the atmosphere) has an intensity of 1.4 kW/m^2 . (a) Assuming that Earth (and its atmosphere) behaves like a flat disk perpendicular to the Sun's rays and that all the incident energy is absorbed, calculate the force on Earth due to radiation pressure. (b) For comparison, calculate the force due to the Sun's gravitational attraction.

•21 ILW What is the radiation pressure 1.5 m away from a 500 W lightbulb? Assume that the surface on which the pressure is exerted faces the bulb and is perfectly absorbing and that the bulb radiates uniformly in all directions.

•22 A black, totally absorbing piece of cardboard of area $A = 2.0 \text{ cm}^2$ intercepts light with an intensity of 10 W/m^2 from a camera strobe light. What radiation pressure is produced on the cardboard by the light?

•23 Someone plans to float a small, totally absorbing sphere 0.500 m above an isotropic point source of light, so that the upward radiation force from the light matches the downward gravitational force on the sphere. The sphere's density is 19.0 g/cm^3 , and its radius is 2.00 mm . (a) What power would be required of the light source? (b) Even if such a source were made, why would the support of the sphere be unstable?

•24 GO It has been proposed that a spaceship might be propelled in the solar system by radiation pressure, using a large sail made of foil. How large must the surface area of the sail be if the radiation force is to be equal in magnitude to the Sun's gravitational attraction? Assume that the mass of the ship + sail is 1500 kg , that the sail is perfectly reflecting, and that the sail is oriented perpendicular to the Sun's rays. See Appendix C for needed data. (With a larger sail, the ship is continuously driven away from the Sun.)

•25 SSM WWW Prove, for a plane electromagnetic wave that is normally incident on a flat surface, that the radiation pressure on the surface is equal to the energy density in the incident beam. (This relation between pressure and energy density holds no matter what fraction of the incident energy is reflected.)

•26 In Fig. 33-38, a laser beam of power 4.60 W and diameter $D = 2.60 \text{ mm}$ is directed upward at one circular face (of diameter $d < 2.60 \text{ mm}$) of a perfectly reflecting cylinder. The cylinder is levitated because the upward radiation force matches the downward gravitational force. If the cylinder's density is 1.20 g/cm^3 , what is its height H ?

•27 SSM WWW A plane electromagnetic wave, with wavelength 3.0 m , travels in vacuum in the positive direction of an x axis. The electric field, of amplitude 300 V/m , oscillates parallel to

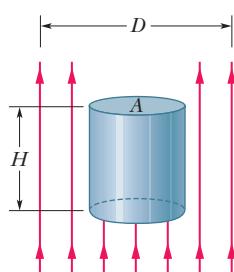


Figure 33-38
Problem 26.

the y axis. What are the (a) frequency, (b) angular frequency, and (c) angular wave number of the wave? (d) What is the amplitude of the magnetic field component? (e) Parallel to which axis does the magnetic field oscillate? (f) What is the time-averaged rate of energy flow in watts per square meter associated with this wave? The wave uniformly illuminates a surface of area 2.0 m^2 . If the surface totally absorbs the wave, what are (g) the rate at which momentum is transferred to the surface and (h) the radiation pressure on the surface?

•28 The average intensity of the solar radiation that strikes normally on a surface just outside Earth's atmosphere is 1.4 kW/m^2 . (a) What radiation pressure p_r is exerted on this surface, assuming complete absorption? (b) For comparison, find the ratio of p_r to Earth's sea-level atmospheric pressure, which is $1.0 \times 10^5 \text{ Pa}$.

•29 SSM A small spaceship with a mass of only $1.5 \times 10^3 \text{ kg}$ (including an astronaut) is drifting in outer space with negligible gravitational forces acting on it. If the astronaut turns on a 10 kW laser beam, what speed will the ship attain in 1.0 day because of the momentum carried away by the beam?

•30 A small laser emits light at power 5.00 mW and wavelength 633 nm . The laser beam is focused (narrowed) until its diameter matches the 1266 nm diameter of a sphere placed in its path. The sphere is perfectly absorbing and has density $5.00 \times 10^3 \text{ kg/m}^3$. What are (a) the beam intensity at the sphere's location, (b) the radiation pressure on the sphere, (c) the magnitude of the corresponding force, and (d) the magnitude of the acceleration that force alone would give the sphere?

•31 GO As a comet swings around the Sun, ice on the comet's surface vaporizes, releasing trapped dust particles and ions. The ions, because they are electrically charged, are forced by the electrically charged *solar wind* into a straight *ion tail* that points radially away from the Sun (Fig. 33-39). The (electrically neutral) dust particles are pushed radially outward from the Sun by the radiation force on them from sunlight. Assume that the dust particles are spherical, have density $3.5 \times 10^3 \text{ kg/m}^3$, and are totally absorbing. (a) What radius must a particle have in order to follow a straight path, like path 2 in the figure? (b) If its radius is larger, does its path curve away from the Sun (like path 1) or toward the Sun (like path 3)?

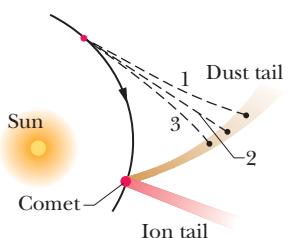


Figure 33-39 Problem 31.

Module 33-4 Polarization

•32 In Fig. 33-40, initially unpolarized light is sent into a system of three polarizing sheets whose polarizing directions make angles

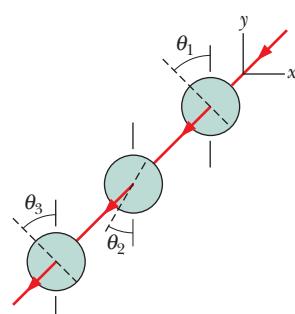


Figure 33-40 Problems 32 and 33.

of $\theta_1 = \theta_2 = \theta_3 = 50^\circ$ with the direction of the y axis. What percentage of the initial intensity is transmitted by the system? (Hint: Be careful with the angles.)

•33 SSM In Fig. 33-40, initially unpolarized light is sent into a system of three polarizing sheets whose polarizing directions make angles of $\theta_1 = 40^\circ$, $\theta_2 = 20^\circ$, and $\theta_3 = 40^\circ$ with the direction of the y axis. What percentage of the light's initial intensity is transmitted by the system? (Hint: Be careful with the angles.)

•34 GO In Fig. 33-41, a beam of unpolarized light, with intensity 43 W/m^2 , is sent into a system of two polarizing sheets with polarizing directions at angles $\theta_1 = 70^\circ$ and $\theta_2 = 90^\circ$ to the y axis. What is the intensity of the light transmitted by the system?

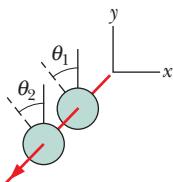


Figure 33-41
Problems 34, 35,
and 42.

•35 ILW In Fig. 33-41, a beam of light, with intensity 43 W/m^2 and polarization parallel to a y axis, is sent into a system of two polarizing sheets with polarizing directions at angles $\theta_1 = 70^\circ$ and $\theta_2 = 90^\circ$ to the y axis. What is the intensity of the light transmitted by the two-sheet system?

•36 At a beach the light is generally partially polarized due to reflections off sand and water. At a particular beach on a particular day near sundown, the horizontal component of the electric field vector is 2.3 times the vertical component. A standing sunbather puts on polarizing sunglasses; the glasses eliminate the horizontal field component. (a) What fraction of the light intensity received before the glasses were put on now reaches the sunbather's eyes? (b) The sunbather, still wearing the glasses, lies on his side. What fraction of the light intensity received before the glasses were put on now reaches his eyes?

•37 SSM WWW We want to rotate the direction of polarization of a beam of polarized light through 90° by sending the beam through one or more polarizing sheets. (a) What is the minimum number of sheets required? (b) What is the minimum number of sheets required if the transmitted intensity is to be more than 60% of the original intensity?

•38 GO In Fig. 33-42, unpolarized light is sent into a system of three polarizing sheets. The angles θ_1 , θ_2 , and θ_3 of the polarizing directions are measured counterclockwise from the positive direction of the y axis (they are not drawn to scale). Angles θ_1 and θ_3 are fixed, but angle θ_2 can be varied. Figure 33-43 gives the intensity of the light emerging from sheet 3 as a function of θ_2 . (The scale of the intensity axis is not indicated.) What percentage of the light's initial intensity is transmitted by the system when $\theta_2 = 30^\circ$?

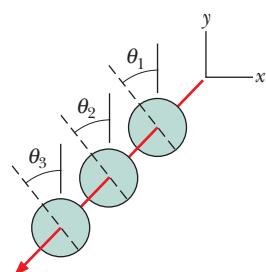


Figure 33-42
Problems 38, 40,
and 44.

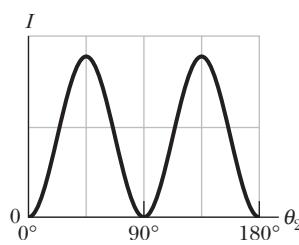


Figure 33-43 Problem 38.

•39 Unpolarized light of intensity 10 mW/m^2 is sent into a polarizing sheet as in Fig. 33-11. What are (a) the amplitude of the electric field component of the transmitted light and (b) the radiation pressure on the sheet due to its absorbing some of the light?

•40 GO In Fig. 33-42, unpolarized light is sent into a system of three polarizing sheets. The angles θ_1 , θ_2 , and θ_3 of the polarizing directions are measured counterclockwise from the positive direction of the y axis (they are not drawn to scale). Angles θ_1 and θ_3 are fixed, but angle θ_2 can be varied. Figure 33-44 gives the intensity of the light emerging from sheet 3 as a function of θ_2 . (The scale of the intensity axis is not indicated.) What percentage of the light's initial intensity is transmitted by the three-sheet system when $\theta_2 = 90^\circ$?

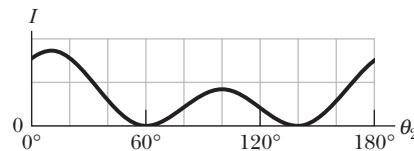


Figure 33-44 Problem 40.

•41 A beam of polarized light is sent into a system of two polarizing sheets. Relative to the polarization direction of that incident light, the polarizing directions of the sheets are at angles θ for the first sheet and 90° for the second sheet. If 0.10 of the incident intensity is transmitted by the two sheets, what is θ ?

•42 GO In Fig. 33-41, unpolarized light is sent into a system of two polarizing sheets. The angles θ_1 and θ_2 of the polarizing directions of the sheets are measured counterclockwise from the positive direction of the y axis (they are not drawn to scale in the figure). Angle θ_1 is fixed but angle θ_2 can be varied. Figure 33-45 gives the intensity of the light emerging from sheet 2 as a function of θ_2 . (The scale of the intensity axis is not indicated.) What percentage of the light's initial intensity is transmitted by the two-sheet system when $\theta_2 = 90^\circ$?

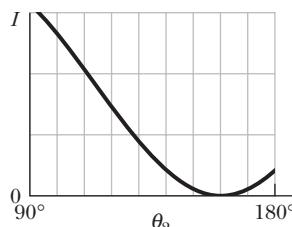


Figure 33-45 Problem 42.

•43 A beam of partially polarized light can be considered to be a mixture of polarized and unpolarized light. Suppose we send such a beam through a polarizing filter and then rotate the filter through 360° while keeping it perpendicular to the beam. If the transmitted intensity varies by a factor of 5.0 during the rotation, what fraction of the intensity of the original beam is associated with the beam's polarized light?

•44 In Fig. 33-42, unpolarized light is sent into a system of three polarizing sheets, which transmits 0.0500 of the initial light intensity. The polarizing directions of the first and third sheets are at angles $\theta_1 = 0^\circ$ and $\theta_3 = 90^\circ$. What are the (a) smaller and (b) larger possible values of angle θ_2 ($< 90^\circ$) for the polarizing direction of sheet 2?

Module 33-5 Reflection and Refraction

•45 When the rectangular metal tank in Fig. 33-46 is filled to the top with an unknown liquid, observer O , with eyes level with the top of the tank, can just see corner E . A ray that refracts toward O at the top surface of the liquid is shown. If $D = 85.0\text{ cm}$ and $L = 1.10\text{ m}$, what is the index of refraction of the liquid?

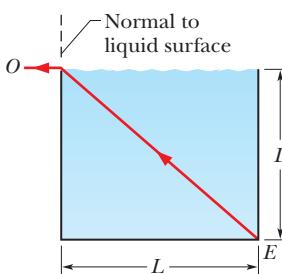


Figure 33-46 Problem 45.

•46 In Fig. 33-47a, a light ray in an underlying material is incident at angle θ_1 on a boundary with water, and some of the light refracts into the water. There are two choices of underlying material. For each, the angle of refraction θ_2 versus the incident angle θ_1 is given in Fig. 33-47b. The horizontal axis scale is set by $\theta_{1s} = 90^\circ$. Without calculation, determine whether the index of refraction of (a) material 1 and (b) material 2 is greater or less than the index of water ($n = 1.33$). What is the index of refraction of (c) material 1 and (d) material 2?

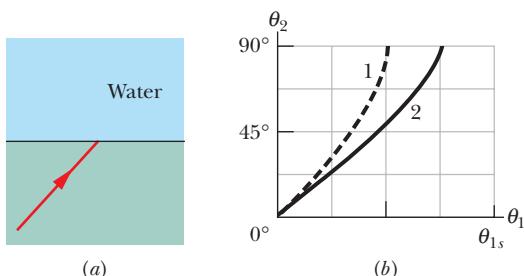


Figure 33-47 Problem 46.

•47 Light in vacuum is incident on the surface of a glass slab. In the vacuum the beam makes an angle of 32.0° with the normal to the surface, while in the glass it makes an angle of 21.0° with the normal. What is the index of refraction of the glass?

•48 In Fig. 33-48a, a light ray in water is incident at angle θ_1 on a boundary with an underlying material, into which some of the light refracts. There are two choices of underlying material. For each, the angle of refraction θ_2 versus the incident angle θ_1 is given in Fig. 33-48b. The vertical axis scale is set by $\theta_{2s} = 90^\circ$. Without calculation, determine whether the index of refraction of (a) material 1 and (b) material 2 is greater or less than the index of water ($n = 1.33$). What is the index of refraction of (c) material 1 and (d) material 2?

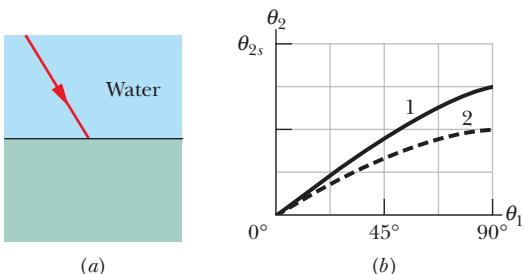


Figure 33-48 Problem 48.

•49 Figure 33-49 shows light reflecting from two perpendicular reflecting surfaces A and B . Find the angle between the incoming ray i and the outgoing ray r' .

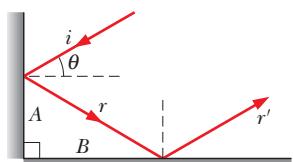


Figure 33-49 Problem 49.

•50 In Fig. 33-50a, a beam of light in material 1 is incident on a boundary at an angle $\theta_1 = 40^\circ$. Some of the light travels through material 2, and then some of it emerges into material 3. The two boundaries between the three materials are parallel. The final direction of the beam depends, in part, on the index of refraction n_3 of the third material. Figure 33-50b gives the angle of refraction θ_3 in that material versus n_3 for a range of possible n_3 values. The vertical axis scale is set by $\theta_{3a} = 30.0^\circ$ and $\theta_{3b} = 50.0^\circ$. (a) What is the index of refraction of material 1, or is the index impossible to calculate without more information? (b) What is the index of refraction of material 2, or is the index impossible to calculate without more information? (c) If θ_1 is changed to 70° and the index of refraction of material 3 is 2.4, what is θ_3 ?

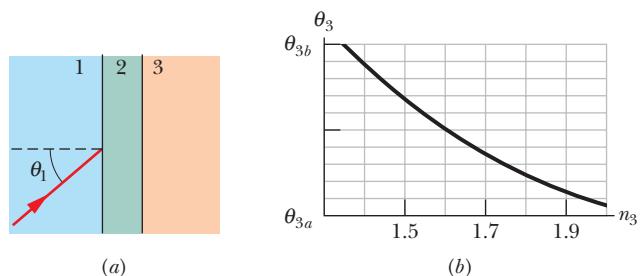


Figure 33-50 Problem 50.

•51 In Fig. 33-51, light is incident at angle $\theta_1 = 40.1^\circ$ on a boundary between two transparent materials. Some of the light travels down through the next three layers of transparent materials, while some of it reflects upward and then escapes into the air. If $n_1 = 1.30$, $n_2 = 1.40$, $n_3 = 1.32$, and $n_4 = 1.45$, what is the value of (a) θ_5 in the air and (b) θ_4 in the bottom material?

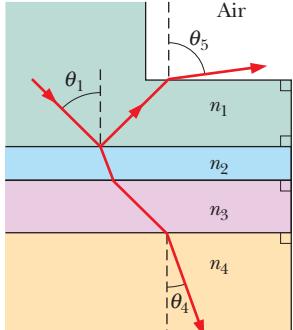


Figure 33-51 Problem 51.

•52 In Fig. 33-52a, a beam of light in material 1 is incident on a boundary at an angle of $\theta_1 = 30^\circ$. The extent of refraction of the light into material 2 depends, in part, on the index of refraction n_2 of material 2. Figure 33-52b gives the angle of refraction θ_2 versus n_2 for a range of possible n_2 values. The vertical axis scale is set by $\theta_{2a} = 20.0^\circ$ and $\theta_{2b} = 40.0^\circ$. (a) What is the index of refraction of

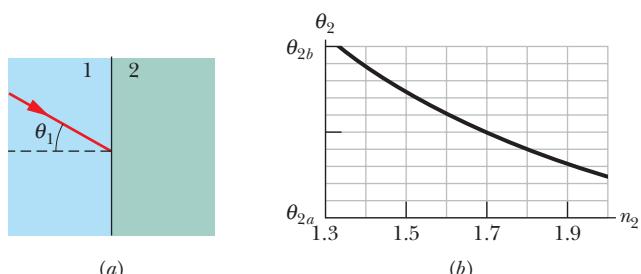


Figure 33-52 Problem 52.

material 1? (b) If the incident angle is changed to 60° and material 2 has $n_2 = 2.4$, then what is angle θ_2 ?

••53 SSM WWW ILW In Fig. 33-53, a ray is incident on one face of a triangular glass prism in air. The angle of incidence θ is chosen so that the emerging ray also makes the same angle θ with the normal to the other face. Show that the index of refraction n of the glass prism is given by

$$n = \frac{\sin \frac{1}{2}(\psi + \phi)}{\sin \frac{1}{2}\phi},$$

where ϕ is the vertex angle of the prism and ψ is the *deviation angle*, the total angle through which the beam is turned in passing through the prism. (Under these conditions the deviation angle ψ has the smallest possible value, which is called the *angle of minimum deviation*.)

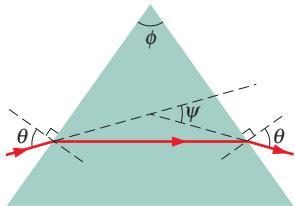


Figure 33-53 Problems 53 and 64.

••54 *Dispersion in a window pane.* In Fig. 33-54, a beam of white light is incident at angle $\theta = 50^\circ$ on a common window pane (shown in cross section). For the pane's type of glass, the index of refraction for visible light ranges from 1.524 at the blue end of the spectrum to 1.509 at the red end. The two sides of the pane are parallel. What is the angular spread of the colors in the beam (a) when the light enters the pane and (b) when it emerges from the opposite side? (Hint: When you look at an object through a window pane, are the colors in the light from the object dispersed as shown in, say, Fig. 33-20?)

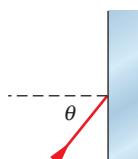


Figure 33-54
Problem 54.

••55 GO SSM In Fig. 33-55, a 2.00-m-long vertical pole extends from the bottom of a swimming pool to a point 50.0 cm above the water. Sunlight is incident at angle $\theta = 55.0^\circ$. What is the length of the shadow of the pole on the level bottom of the pool?

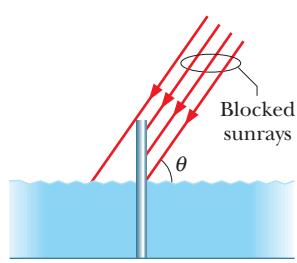


Figure 33-55 Problem 55.

••56 *Rainbows from square drops.* Suppose that, on some surreal world, raindrops had a square cross section and always fell with one face horizontal. Figure 33-56 shows such a falling drop, with a white beam of sunlight incident at $\theta = 70.0^\circ$ at point P. The part of the light that enters the drop then travels to point A, where some of it refracts out into the air and the rest reflects. That reflected light then travels to point B, where again some of the light refracts out into the air and the rest reflects. What is the difference in the angles of the red light ($n = 1.331$) and the blue light ($n = 1.343$) that emerge at

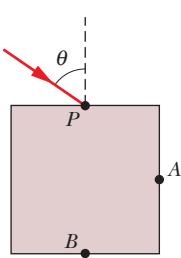


Figure 33-56
Problem 56.

(a) point A and (b) point B? (This angular difference in the light emerging at, say, point A would be the rainbow's angular width.)

Module 33-6 Total Internal Reflection

••57 A point source of light is 80.0 cm below the surface of a body of water. Find the diameter of the circle at the surface through which light emerges from the water.

••58 The index of refraction of benzene is 1.8. What is the critical angle for a light ray traveling in benzene toward a flat layer of air above the benzene?

••59 SSM ILW In Fig. 33-57, a ray of light is perpendicular to the face ab of a glass prism ($n = 1.52$). Find the largest value for the angle ϕ so that the ray is totally reflected at face ac if the prism is immersed (a) in air and (b) in water.



Figure 33-57 Problem 59.

••60 In Fig. 33-58, light from ray A refracts from material 1 ($n_1 = 1.60$) into a thin layer of material 2 ($n_2 = 1.80$), crosses that layer, and is then incident at the critical angle on the interface between materials 2 and 3 ($n_3 = 1.30$). (a) What is the value of incident angle θ_A ? (b) If θ_A is decreased, does part of the light refract into material 3?

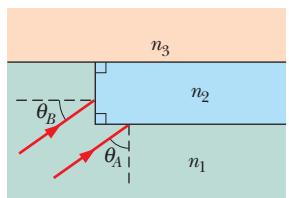


Figure 33-58 Problem 60.

Light from ray B refracts from material 1 into the thin layer, crosses that layer, and is then incident at the critical angle on the interface between materials 2 and 3. (c) What is the value of incident angle θ_B ? (d) If θ_B is decreased, does part of the light refract into material 3?

••61 GO In Fig. 33-59, light initially in material 1 refracts into material 2, crosses that material, and is then incident at the critical angle on the interface between materials 2 and 3. The indexes of refraction are $n_1 = 1.60$, $n_2 = 1.40$, and $n_3 = 1.20$. (a) What is angle θ ? (b) If θ is increased, is there refraction of light into material 3?

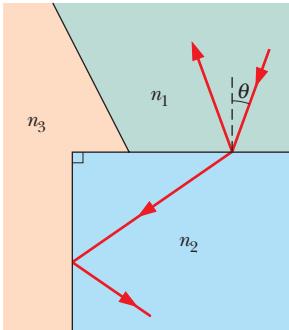


Figure 33-59 Problem 61.

••62 GO A catfish is 2.00 m below the surface of a smooth lake. (a) What is the diameter of the circle on the surface through which the fish can see the world outside the water? (b) If the fish descends, does the diameter of the circle increase, decrease, or remain the same?

••63 In Fig. 33-60, light enters a 90° triangular prism at point P with incident angle θ , and then some of it refracts at point Q with an angle of refraction of 90° . (a) What is the index of refraction of the prism in terms of θ ? (b) What, numerically, is the maximum value that the index of refraction can have? Does light emerge at Q if the incident angle at P is (c) increased slightly and (d) decreased slightly?

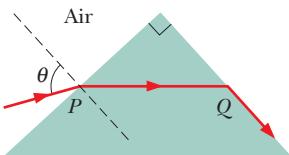


Figure 33-60 Problem 63.

••64 Suppose the prism of Fig. 33-53 has apex angle $\phi = 60.0^\circ$ and index of refraction $n = 1.60$. (a) What is the smallest angle of incidence θ for which a ray can enter the left face of the prism and exit the right face? (b) What angle of incidence θ is required for the ray to exit the prism with an identical angle θ for its refraction, as it does in Fig. 33-53?

••65 GO Figure 33-61 depicts a simplistic optical fiber: a plastic core ($n_1 = 1.58$) is surrounded by a plastic sheath ($n_2 = 1.53$). A light ray is incident on one end of the fiber at angle θ . The ray is to undergo total internal reflection at point A , where it encounters the core–sheath boundary. (Thus there is no loss of light through that boundary.) What is the maximum value of θ that allows total internal reflection at A ?

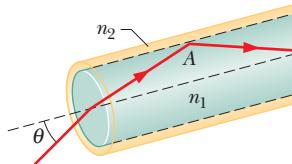


Figure 33-61 Problem 65.

••66 GO In Fig. 33-62, a light ray in air is incident at angle θ_1 on a block of transparent plastic with an index of refraction of 1.56. The dimensions indicated are $H = 2.00\text{ cm}$ and $W = 3.00\text{ cm}$. The light passes through the block to one of its sides and there undergoes reflection (inside the block) and possibly refraction (out into the air). This is the point of *first reflection*. The reflected light then passes through the block to another of its sides—a point of *second reflection*. If $\theta_1 = 40^\circ$, on which side is the point of (a) first reflection and (b) second reflection? If there is refraction at the point of (c) first reflection and (d) second reflection, give the angle of refraction; if not, answer “none.” If $\theta_1 = 70^\circ$, on which side is the point of (e) first reflection and (f) second reflection? If there is refraction at the point of (g) first reflection and (h) second reflection, give the angle of refraction; if not, answer “none.”

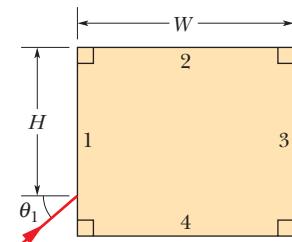


Figure 33-62 Problem 66.

••67 GO In the ray diagram of Fig. 33-63, where the angles are not drawn to scale, the ray is incident at the critical angle on the interface between materials 2 and 3. Angle $\phi = 60.0^\circ$, and two of the indexes of refraction are $n_1 = 1.70$ and $n_2 = 1.60$. Find (a) index of refraction n_3 and (b) angle θ . (c) If θ is decreased, does light refract into material 3?

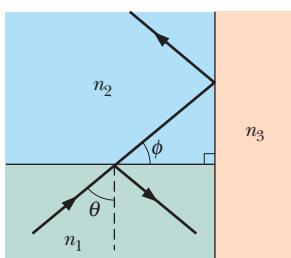


Figure 33-63 Problem 67.

Module 33-7 Polarization by Reflection

•68 (a) At what angle of incidence will the light reflected from water be completely polarized? (b) Does this angle depend on the wavelength of the light?

•69 SSM Light that is traveling in water (with an index of refraction of 1.33) is incident on a plate of glass (with index of refraction 1.53). At what angle of incidence does the reflected light end up fully polarized?

•70 In Fig. 33-64, a light ray in air is incident on a flat layer of material 2 that has an index of refraction $n_2 = 1.5$. Beneath material 2 is material 3 with an index of refraction n_3 . The ray is incident on the air–material 2 interface at the Brewster angle for that interface. The ray of light refracted into material 3 happens to be incident on the material 2–material 3 interface at the Brewster angle for that interface. What is the value of n_3 ?

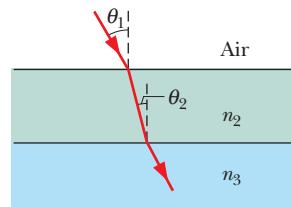


Figure 33-64 Problem 70.

Additional Problems

71 SSM (a) How long does it take a radio signal to travel 150 km from a transmitter to a receiving antenna? (b) We see a full Moon by reflected sunlight. How much earlier did the light that enters our eye leave the Sun? The Earth–Moon and Earth–Sun distances are $3.8 \times 10^5\text{ km}$ and $1.5 \times 10^8\text{ km}$, respectively. (c) What is the round-trip travel time for light between Earth and a spaceship orbiting Saturn, $1.3 \times 10^9\text{ km}$ distant? (d) The Crab nebula, which is about 6500 light-years (ly) distant, is thought to be the result of a supernova explosion recorded by Chinese astronomers in A.D. 1054. In approximately what year did the explosion actually occur? (When we look into the night sky, we are effectively looking back in time.)

72 An electromagnetic wave with frequency $4.00 \times 10^{14}\text{ Hz}$ travels through vacuum in the positive direction of an x axis. The wave has its electric field oscillating parallel to the y axis, with an amplitude E_m . At time $t = 0$, the electric field at point P on the x axis has a value of $+E_m/4$ and is decreasing with time. What is the distance along the x axis from point P to the first point with $E = 0$ if we search in (a) the negative direction and (b) the positive direction of the x axis?

73 SSM The electric component of a beam of polarized light is

$$E_y = (5.00 \text{ V/m}) \sin[(1.00 \times 10^6 \text{ m}^{-1})z + \omega t].$$

(a) Write an expression for the magnetic field component of the wave, including a value for ω . What are the (b) wavelength, (c) period, and (d) intensity of this light? (e) Parallel to which axis does the magnetic field oscillate? (f) In which region of the electromagnetic spectrum is this wave?

74 A particle in the solar system is under the combined influence of the Sun’s gravitational attraction and the radiation force due to the Sun’s rays. Assume that the particle is a sphere of density $1.0 \times 10^3 \text{ kg/m}^3$ and that all the incident light is absorbed. (a) Show that, if its radius is less than some critical radius R , the particle will be blown out of the solar system. (b) Calculate the critical radius.

75 SSM In Fig. 33-65, a light ray enters a glass slab at point A at incident angle $\theta_1 = 45.0^\circ$ and then undergoes total internal reflection at point B. (The reflection at A is not shown.) What minimum value for the index of refraction of the glass can be inferred from this information?

76 GO In Fig. 33-66, unpolarized light with an intensity of 25 W/m^2 is sent into a system of four polarizing sheets with polarizing directions at angles $\theta_1 = 40^\circ$, $\theta_2 = 20^\circ$, $\theta_3 = 20^\circ$, and $\theta_4 = 30^\circ$. What is the intensity of the light that emerges from the system?

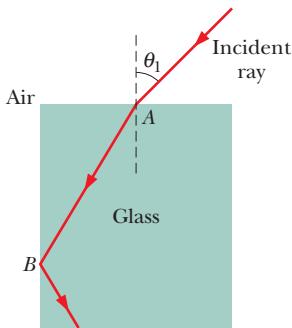


Figure 33-65 Problem 75.

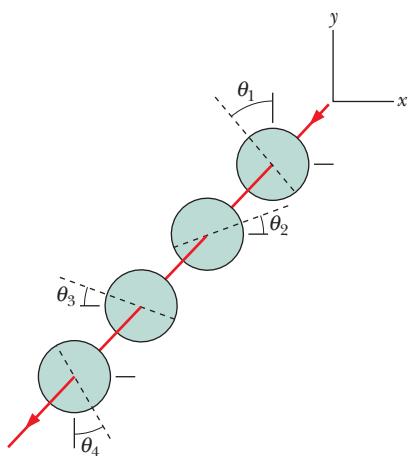


Figure 33-66 Problem 76.

77 **Rainbow.** Figure 33-67 shows a light ray entering and then leaving a falling, spherical raindrop after one internal reflection (see Fig. 33-21a). The final direction of travel is deviated (turned) from the initial direction of travel by angular deviation θ_{dev} . (a) Show that θ_{dev} is

$$\theta_{\text{dev}} = 180^\circ + 2\theta_i - 4\theta_r,$$

where θ_i is the angle of incidence of the ray on the drop and θ_r is the angle of refraction of the ray within the drop. (b) Using Snell's law, substitute for θ_r in terms of θ_i and the index of refraction n of the water. Then, on a graphing calculator or with a computer graphing package, graph θ_{dev} versus θ_i for the range of possible θ_i values and for $n = 1.331$ for red light (at one end of the visible spectrum) and $n = 1.333$ for blue light (at the other end).

The red-light curve and the blue-light curve have different minima, which means that there is a different *angle of minimum deviation* for each color. The light of any given color that leaves the drop at that color's angle of minimum deviation is especially bright because rays bunch up at that angle. Thus, the bright red light leaves the drop at one angle and the bright blue light leaves it at another angle.

Determine the angle of minimum deviation from the θ_{dev} curve

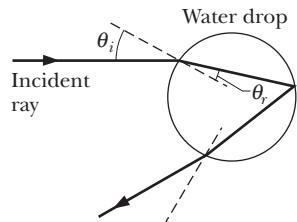


Figure 33-67 Problem 77.

for (c) red light and (d) blue light. (e) If these colors form the inner and outer edges of a rainbow (Fig. 33-21a), what is the angular width of the rainbow?

78 The *primary rainbow* described in Problem 77 is the type commonly seen in regions where rainbows appear. It is produced by light reflecting once inside the drops. Rarer is the *secondary rainbow* described in Module 33-5, produced by light reflecting twice inside the drops (Fig. 33-68a). (a) Show that the angular deviation of light entering and then leaving a spherical water drop is

$$\theta_{\text{dev}} = (180^\circ)k + 2\theta_i - 2(k+1)\theta_r,$$

where k is the number of internal reflections. Using the procedure of Problem 77, find the angle of minimum deviation for (b) red light and (c) blue light in a secondary rainbow. (d) What is the angular width of that rainbow (Fig. 33-21d)?

The *tertiary rainbow* depends on three internal reflections (Fig. 33-68b). It probably occurs but, as noted in Module 33-5, cannot be seen with the eye because it is very faint and lies in the bright sky surrounding the Sun. What is the angle of minimum deviation for (e) the red light and (f) the blue light in this rainbow? (g) What is the rainbow's angular width?

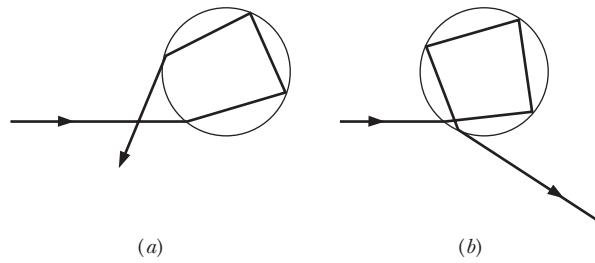


Figure 33-68 Problem 78.

79 SSM (a) Prove that a ray of light incident on the surface of a sheet of plate glass of thickness t emerges from the opposite face parallel to its initial direction but displaced sideways, as in Fig. 33-69. (b) Show that, for small angles of incidence θ , this displacement is given by

$$x = t\theta \frac{n-1}{n},$$

where n is the index of refraction of the glass and θ is measured in radians.

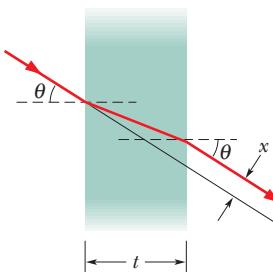


Figure 33-69 Problem 79.

80 An electromagnetic wave is traveling in the negative direction of a y axis. At a particular position and time, the electric field is directed along the positive direction of the z axis and has a magnitude of 100 V/m . What are the (a) magnitude and (b) direction of the corresponding magnetic field?

81 The magnetic component of a polarized wave of light is

$$B_x = (4.0 \times 10^{-6} \text{ T}) \sin[(1.57 \times 10^7 \text{ m}^{-1})y + \omega t].$$

- (a) Parallel to which axis is the light polarized? What are the (b) frequency and (c) intensity of the light?

82 In Fig. 33-70, unpolarized light is sent into the system of three polarizing sheets, where the polarizing directions of the first and third sheets are at angles $\theta_1 = 30^\circ$ (counterclockwise) and $\theta_3 = 30^\circ$ (clockwise). What fraction of the initial light intensity emerges from the system?

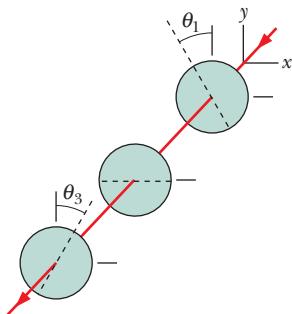


Figure 33-70 Problem 82.

83 SSM A ray of white light traveling through fused quartz is incident at a quartz-air interface at angle θ_1 . Assume that the index of refraction of quartz is $n = 1.456$ at the red end of the visible range and $n = 1.470$ at the blue end. If θ_1 is (a) 42.00° , (b) 43.10° , and (c) 44.00° , is the refracted light white, white dominated by the red end of the visible range, or white dominated by the blue end of the visible range, or is there no refracted light?

84 Three polarizing sheets are stacked. The first and third are crossed; the one between has its polarizing direction at 45.0° to the polarizing directions of the other two. What fraction of the intensity of an originally unpolarized beam is transmitted by the stack?

85 In a region of space where gravitational forces can be neglected, a sphere is accelerated by a uniform light beam of intensity 6.0 mW/m^2 . The sphere is totally absorbing and has a radius of $2.0 \mu\text{m}$ and a uniform density of $5.0 \times 10^3 \text{ kg/m}^3$. What is the magnitude of the sphere's acceleration due to the light?

86 An unpolarized beam of light is sent into a stack of four polarizing sheets, oriented so that the angle between the polarizing directions of adjacent sheets is 30° . What fraction of the incident intensity is transmitted by the system?

87 SSM During a test, a NATO surveillance radar system, operating at 12 GHz at 180 kW of power, attempts to detect an incoming stealth aircraft at 90 km. Assume that the radar beam is emitted uniformly over a hemisphere. (a) What is the intensity of the beam when the beam reaches the aircraft's location? The aircraft reflects radar waves as though it has a cross-sectional area of only 0.22 m^2 . (b) What is the power of the aircraft's reflection? Assume that the beam is reflected uniformly over a hemisphere. Back at the radar site, what are (c) the intensity, (d) the maximum value of the electric field vector, and (e) the rms value of the magnetic field of the reflected radar beam?

88 The magnetic component of an electromagnetic wave in vacuum has an amplitude of 85.8 nT and an angular wave number of 4.00 m^{-1} . What are (a) the frequency of the wave, (b) the rms value of the electric component, and (c) the intensity of the light?

89 Calculate the (a) upper and (b) lower limit of the Brewster angle for white light incident on fused quartz. Assume that the wavelength limits of the light are 400 and 700 nm.

90 In Fig. 33-71, two light rays pass from air through five layers of transparent plastic and then back into air. The layers have parallel interfaces and unknown thicknesses; their indexes of refraction are $n_1 = 1.7$, $n_2 = 1.6$, $n_3 = 1.5$, $n_4 = 1.4$, and $n_5 = 1.6$. Ray b is incident

at angle $\theta_b = 20^\circ$. Relative to a normal at the last interface, at what angle do (a) ray a and (b) ray b emerge? (Hint: Solving the problem algebraically can save time.) If the air at the left and right sides in the figure were, instead, glass with index of refraction 1.5, at what angle would (c) ray a and (d) ray b emerge?

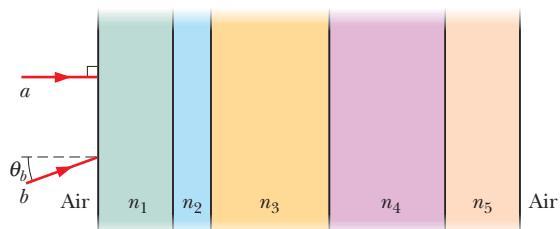


Figure 33-71 Problem 90.

91 A helium-neon laser, radiating at 632.8 nm , has a power output of 3.0 mW . The beam diverges (spreads) at angle $\theta = 0.17 \text{ mrad}$ (Fig. 33-72). (a) What is the intensity of the beam 40 m from the laser? (b)

What is the power of a point source providing that intensity at that distance?

92 In about A.D. 150, Claudius Ptolemy gave the following measured values for the angle of incidence θ_1 and the angle of refraction θ_2 for a light beam passing from air to water:

θ_1	θ_2	θ_1	θ_2
10°	8°	50°	35°
20°	$15^\circ 30'$	60°	$40^\circ 30'$
30°	$22^\circ 30'$	70°	$45^\circ 30'$
40°	29°	80°	50°

Assuming these data are consistent with the law of refraction, use them to find the index of refraction of water. These data are interesting as perhaps the oldest recorded physical measurements.

93 A beam of initially unpolarized light is sent through two polarizing sheets placed one on top of the other. What must be the angle between the polarizing directions of the sheets if the intensity of the transmitted light is to be one-third the incident intensity?

94 In Fig. 33-73, a long, straight copper wire (diameter 2.50 mm and resistance 1.00Ω per 300 m) carries a uniform current of 25.0 A in the positive x direction. For point P on the wire's surface, calculate the magnitudes of (a) the electric field \vec{E} , (b) the magnetic field \vec{B} , and (c) the Poynting vector \vec{S} , and (d) determine the direction of \vec{S} .

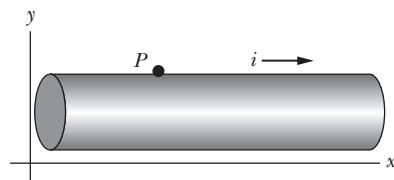


Figure 33-73 Problem 94.

95 Figure 33-74 shows a cylindrical resistor of length l , radius a , and resistivity ρ , carrying current i . (a) Show that the Poynting vector \vec{S} at the surface of the resistor is everywhere directed normal to the surface, as shown. (b) Show that the rate P at which energy flows into the resistor through its cylindrical surface, calculated by integrating the Poynting vector over this surface, is equal to the rate at which thermal energy is produced:

$$\int \vec{S} \cdot d\vec{A} = i^2 R,$$

where $d\vec{A}$ is an element of area on the cylindrical surface and R is the resistance.

96 A thin, totally absorbing sheet of mass m , face area A , and specific heat c_s is fully illuminated by a perpendicular beam of a plane electromagnetic wave. The magnitude of the maximum electric field of the wave is E_m . What is the rate dT/dt at which the sheet's temperature increases due to the absorption of the wave?

97 Two polarizing sheets, one directly above the other, transmit $p\%$ of the initially unpolarized light that is perpendicularly incident on the top sheet. What is the angle between the polarizing directions of the two sheets?

98 A laser beam of intensity I reflects from a flat, totally reflecting surface of area A , with a normal at angle θ with the beam. Write an expression for the beam's radiation pressure $p_r(\theta)$ on the surface in terms of the beam's pressure $p_{r\perp}$ when $\theta = 0^\circ$.

99 A beam of intensity I reflects from a long, totally reflecting cylinder of radius R ; the beam is perpendicular to the central axis of the cylinder and has a diameter larger than $2R$. What is the beam's force per unit length on the cylinder?

100 In Fig. 33-75, unpolarized light is sent into a system of three polarizing sheets, where the polarizing directions of the first and second sheets are at angles $\theta_1 = 20^\circ$ and $\theta_2 = 40^\circ$. What fraction of the initial light intensity emerges from the system?

101 In Fig. 33-76, unpolarized light is sent into a system of three polarizing sheets with polarizing directions at angles $\theta_1 = 20^\circ$, $\theta_2 = 60^\circ$, and $\theta_3 = 40^\circ$. What fraction of the initial light intensity emerges from the system?

102 A square, perfectly reflecting surface is oriented in space to be perpendicular to the light rays from the Sun. The surface has an

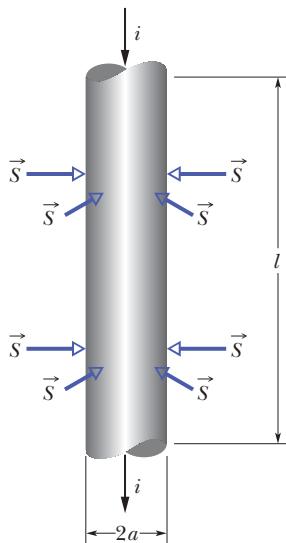


Figure 33-74 Problem 95.

edge length of 2.0 m and is located 3.0×10^{11} m from the Sun's center. What is the radiation force on the surface from the light rays?

103 The rms value of the electric field in a certain light wave is 0.200 V/m. What is the amplitude of the associated magnetic field?

104 In Fig. 33-77, an albatross glides at a constant 15 m/s horizontally above level ground, moving in a vertical plane that contains the Sun. It glides toward a wall of height $h = 2.0$ m, which it will just barely clear. At that time of day, the angle of the Sun relative to the ground is $\theta = 30^\circ$. At what speed does the shadow of the albatross move (a) across the level ground and then (b) up the wall? Suppose that later a hawk happens to glide along the same path, also at 15 m/s. You see that when its shadow reaches the wall, the speed of the shadow noticeably increases. (c) Is the Sun now higher or lower in the sky than when the albatross flew by earlier? (d) If the speed of the hawk's shadow on the wall is 45 m/s, what is the angle θ of the Sun just then?

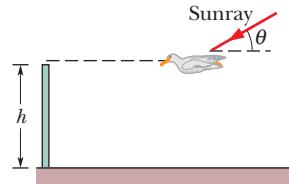


Figure 33-77 Problem 104.

105 The magnetic component of a polarized wave of light is given by $B_x = (4.00 \mu T) \sin [ky + (2.00 \times 10^{15} s^{-1})t]$. (a) In which direction does the wave travel, (b) parallel to which axis is it polarized, and (c) what is its intensity? (d) Write an expression for the electric field of the wave, including a value for the angular wave number. (e) What is the wavelength? (f) In which region of the electromagnetic spectrum is this electromagnetic wave?

106 In Fig. 33-78, where $n_1 = 1.70$, $n_2 = 1.50$, and $n_3 = 1.30$, light refracts from material 1 into material 2. If it is incident at point A at the critical angle for the interface between materials 2 and 3, what are (a) the angle of refraction at point B and (b) the initial angle θ ? If, instead, light is incident at B at the critical angle for the interface between materials 2 and 3, what are (c) the angle of refraction at point A and (d) the initial angle θ ? If, instead of all that, light is incident at point A at Brewster's angle for the interface between materials 2 and 3, what are (e) the angle of refraction at point B and (f) the initial angle θ ?

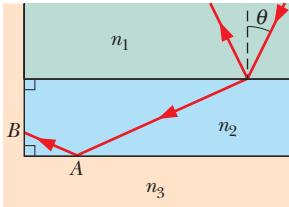


Figure 33-78 Problem 106.

107 When red light in vacuum is incident at the Brewster angle on a certain glass slab, the angle of refraction is 32.0° . What are (a) the index of refraction of the glass and (b) the Brewster angle?

108 Start from Eqs. 33-11 and 33-17 and show that $E(x, t)$ and $B(x, t)$, the electric and magnetic field components of a plane traveling electromagnetic wave, must satisfy the "wave equations"

$$\frac{\partial^2 E}{\partial t^2} = c^2 \frac{\partial^2 E}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 B}{\partial t^2} = c^2 \frac{\partial^2 B}{\partial x^2}.$$

109 **SSM** (a) Show that Eqs. 33-1 and 33-2 satisfy the wave equations displayed in Problem 108. (b) Show that any expressions of the form $E = E_m f(kx \pm \omega t)$ and $B = B_m f(kx \pm \omega t)$, where $f(kx \pm \omega t)$ denotes an arbitrary function, also satisfy these wave equations.

110 A point source of light emits isotropically with a power of 200 W. What is the force due to the light on a totally absorbing sphere of radius 2.0 cm at a distance of 20 m from the source?

Figure 33-75 Problem 100.

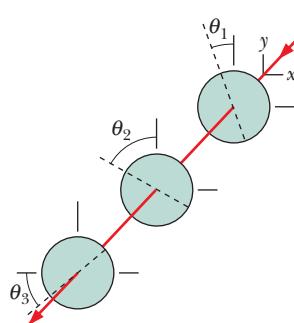


Figure 33-76 Problem 101.

Interference in thin films

February 21, 2020

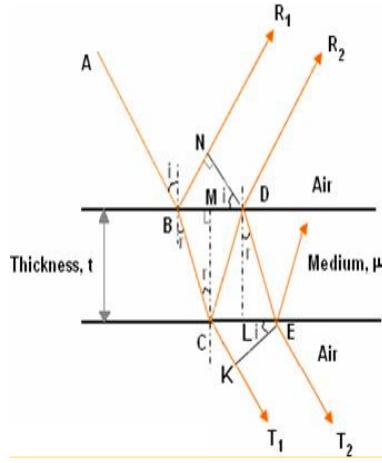
1 Introduction

Waves reflecting from two surfaces can interfere constructively and destructively. In this case it is light waves that are being reflected from the front and rear surfaces of thin soap or oil films. The interference produces a pattern of beautiful colors in white light, or dark and light bands in monochromatic light. Consider the case of a thin film of oil floating on water. Thin-film interference can take place if these two light waves interfere constructively:

1. The light from the air reflecting off the top surface
2. The light traveling from the air, through the oil, reflecting off the bottom surface, traveling back through the oil and out into the air again.

2 Theory

Consider a thin film of uniform thickness t and refractive index μ . Let us consider a monochromatic ray AB which is made to incident on the film, at B part of the ray is reflected (R_1) and a part is refracted along BC. At C The beam BC again suffers partial reflection and partial refraction, the reflected beam CD suffers partial reflection and partial refraction at D. The refracted beam R_2 moves in air. These two reflected rays R_1 and R_2 interfere to produce interference pattern.



The optical path difference needs to consider the refractive index of the medium as well. Therefore the path difference between the two rays R_1 and R_2 is given by

$$\delta = \mu(BC + CD) - BN$$

In ΔBDN

$$\sin i = \frac{BN}{BD}, BC = CD$$

$$\text{as } \Delta BDN = \Delta MCD$$

$$\text{Therefore in } \Delta BMC \cos r = \frac{t}{BC}$$

$$\text{Therefore } \delta = 2\mu BC - 2i$$

$$\text{In } \Delta BMC \cos r = \frac{t}{BC} \text{ therefore}$$

$$\delta = \frac{2\mu t}{\cos r} - BD \sin i$$

$$\delta = \frac{2\mu t}{\cos r} - 2BM \sin i$$

$$\text{In } \Delta BMC \tan r = \frac{BM}{t} \text{ therefore}$$

$$\delta = \frac{2\mu t}{\cos r} - 2t(\tan r) \sin i$$

$$\text{According to Snell's law } \mu = \frac{\sin i}{\sin r} \text{ therefore}$$

$$\delta = \frac{2\mu t}{\cos r} - 2\mu t(\tan r) \sin r$$

$$\delta = \frac{2\mu t}{\cos r} - \frac{2\mu t \sin^2 r}{\cos r}$$

$$\delta = \frac{2\mu t}{\cos r} (1 - \sin^2 r)$$

$$\delta = 2\mu t \cos r$$

Correction on account of phase change at reflection: when a beam is reflected from a denser medium (ray R_1 at B), a path change of $\frac{\lambda}{2}$ occurs.

Therefore the total path difference would be

$$\delta = 2\mu t \cos r + \frac{\lambda}{2}$$

Condition for Maxima

Maxima occurs when the path difference is

$$\delta = n\lambda$$

Therefore

$$2\mu t \cos r + \frac{\lambda}{2} = n\lambda$$

$$2\mu t \cos r = \frac{(2n-1)\lambda}{2}$$

Condition for Minima

Minima occurs when the path difference is

$$\delta = \frac{(2n+1)\lambda}{2}$$

Therefore

$$2\mu t \cos r = n\lambda$$

3. Newton's Rings

Background

Coherent light
Phase relationship
Path difference
Interference in thin film
Newton's ring apparatus

Aim of the experiment

To study the formation of Newton's rings in the air-film in between a plano-convex lens and a glass plate using nearly monochromatic light from a sodium-source and hence to determine the radius of curvature of the plano-convex lens.

Apparatus required

A nearly monochromatic source of light (source of sodium light)

A plano-convex lens

An optically flat glass plates

A convex lens

A traveling microscope

Theory

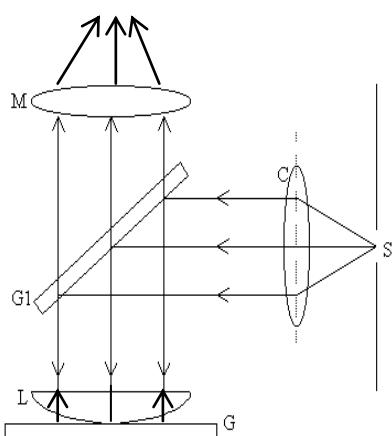


Fig. 1 Experimental set-up to observe Newton's ring

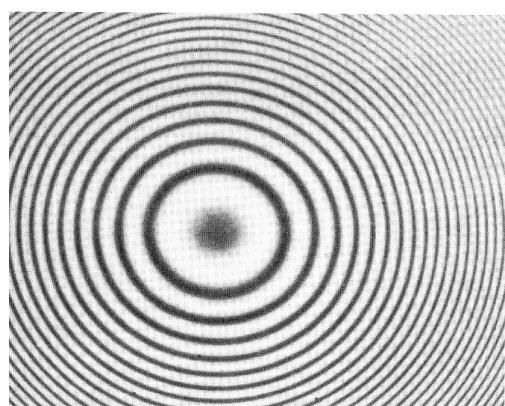


Fig. 2. Newton's rings

When a parallel beam of monochromatic light is incident normally on a combination of a plano-convex lens L and a glass plate G, as shown in Fig.1, a part of each incident ray is reflected from the lower surface of the lens, and a part, after refraction through the air film between the lens and the plate, is reflected back from the plate surface. These two reflected rays are coherent,

hence they will interfere and produce a system of alternate dark and bright rings with the point of contact between the lens and the plate as the center. These rings are known as Newton's ring.

For a normal incidence of monochromatic light, the path difference between the reflected rays (see Fig.1) is very nearly equal to $2\mu t$ where μ and t are the refractive index and thickness of the air-film respectively. The fact that the wave is reflected from air to glass surface introduces a phase shift of π . Therefore, for bright fringe

$$2\mu t = \left(n + \frac{1}{2}\right)\lambda ; n = 0,1,2,3 \quad (1)$$

and for dark fringe

$$2\mu t = n\lambda ; n = 0,1,2,3 \quad (2)$$

For n -th (bright or dark) ring (see Fig. 2), we also have

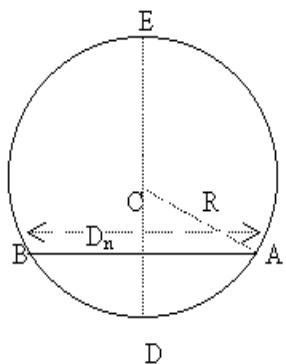


Fig. 2 Geometry used to determine the thickness of the air-film

$$\frac{D_n^2}{4} + (R - t)^2 = R^2 \quad (3)$$

where D_n = the diameter of the n -th ring and R = the radius of curvature of the lower surface of the plano-convex lens.

On neglecting t^2 , equation (3) reduces to

$$D_n^2 = 8tR \quad (4)$$

From equations (1) and (4), we get,

$$D_n^2 = 4\left(n + \frac{1}{2}\right)\frac{\lambda R}{\mu}, \text{ for } n\text{-th bright ring} \quad (5)$$

$$D_{n+m}^2 = 4\left(n + m + \frac{1}{2}\right)\frac{\lambda R}{\mu}, \\ \text{for } (n+m)\text{-th bright ring} \quad (6)$$

Similarly, from equations (2) and (4), we obtain

$$D_n^2 = \frac{4n\lambda R}{\mu}, \text{ for } n\text{-th dark ring} \quad (7)$$

$$D_{n+m}^2 = \frac{4(n+m)\lambda R}{\mu}, \text{ for } (n+m)\text{-th dark ring} \quad (8)$$

Thus for bright as well as dark rings, we obtain

$$R = \frac{\mu \left(D_{n+m}^2 - D_n^2 \right)}{4m\lambda} \quad \dots \quad (9)$$

Since $\mu=1$ for air-film, above equation gives

$$R = \frac{\left(D_{n+m}^2 - D_n^2 \right)}{4m\lambda} \quad \dots \quad (10)$$

Procedure

1. Level the traveling microscope with its axis vertical. Arrange the set-up as shown in Fig.1 and focus the microscope on the air-film. Newton's Rings will be clearly seen.
2. Adjust the glass plate G1 for maximum visibility of the point of contact of lens L with the glass plate G and hence for maximum visibility of Newton's Rings. In this orientation, G1 is at 45° to the incident beam of light.
3. Move the microscope to the right of the central dark spot (say order 'n', this is because the central ring is often broad and may not necessarily will be zero order) and set it on the extreme tenth ($n+10^{\text{th}}$ order) distinct bright ring so that the cross-wire perpendicular to the direction of movement of the microscope passes through the bright ring and is tangential to it. Record the microscope position from the horizontal scale along with its number with bright ring around the central dark spot as the first bright ring. Move the microscope to left and record the position of the next bright ring. Repeat it till you reach to the tenth bright ring on the left. From these measurements, evaluate the diameters of different rings. Repeat these measurements for microscope movement from left to right and evaluate the diameters of different rings. Determine the average diameters of different rings.

Observations

Vernier constant for the horizontal scale of the microscope (Least Count) :

Table 1
Measurements of the diameter of the ring

Ring No. (n)	Microscope readings (cm) on the						Diameter $D_{n+m} =$ $R_1 - R_2$ (cm)	D_{n+m}^2 (cm ²)	$m_1 - m_2$	$D_{n+m_1}^2 - D_{n+m_2}^2$ (cm ²)				
	Left (R_1)			Right (R_2)										
	Main Scale	Vernier	Total	Main scale	Vernier	Total								
n+10														
n+9														
n+8														
n+7														
n+6														
n+5														
n+4														
n+3														
n+2														
n+1														

Calculation and Results

Plot a graph between $D_{n+m_1}^2 - D_{n+m_2}^2$ vs $m_1 - m_2$

Table 2
Calculation of radius of curvature, R, from the graph

$D_{n+m_1}^2 - D_{n+m_2}^2$ (cm ²) from graph	$m_1 - m_2$	λ (cm) (5893×10^{-8})	$R = \frac{D_{n+m_1}^2 - D_{n+m_2}^2}{4(m_1 - m_2)\lambda}$ (cm)

Estimate error in R

The radius of curvature is calculated from Equation (3), viz.

$$R = \frac{D_{n+m_1}^2 - D_{n+m_2}^2}{4(m_1 - m_2)\lambda}$$

Since D_{n+m_1} and D_{n+m_2} are only measured, the maximum proportional error in R is given by

$$\frac{\delta R}{R} = \frac{\delta(D_{n+m_1}^2 - D_{n+m_2}^2)}{D_{n+m_1}^2 - D_{n+m_2}^2} = \frac{2(\delta D_{n+m_1})D_{n+m_1} + 2(\delta D_{n+m_2})D_{n+m_2}}{D_{n+m_1}^2 - D_{n+m_2}^2}$$

Since D_{n+m_1} or D_{n+m_2} is measured by taking the difference between the two readings of a scale provided with a vernier, the maximum error in measuring each of these quantities is twice the vernier constant i.e. $2v.c$.

Therefore, $\delta D_n = 2v.c$

$$\text{Hence, } \frac{\delta R}{R} = 4v.c \frac{(D_{n+m_1} + D_{n+m_2})}{D_{n+m_1}^2 - D_{n+m_2}^2} = \frac{4v.c}{(D_{n+m_1} - D_{n+m_2})}$$

Discussion

- (i) The Newton's ring experiment can be also used to find the wavelength of a monochromatic light. In this case, the radius of curvature of the convex surface of the given lens is supplied or is determined otherwise. By employing sodium light whose mean wavelength is 5893\AA , R can be determined from Eqn.(3), as in the present experiment. Then the same equation can be used to find the wavelength λ of any other given monochromatic light.

- (ii) R is calculated from Eq. (10). An error in the actual ring number $n+m_1$ does not affect the result.
- (iii) Since the first few rings near the center are deformed, they must be avoided while taking readings for the rings.
- (iv) Care must be taken not to disturb the lens and glass plate combination in any way during the experiment.

Questions

1. In the Newton's ring experiment, how does interference occur?
2. Where have the fringes formed?
3. Why are the fringes circular?
4. Are all rings equispaced?
5. Why is an extended source used in this experiment?
6. What will happen if a point source or an illuminated slit is used instead of the extended source?
7. In place of lens, if a wedge shaped film formed by two glass plates is supplied to you, will you be able to observe Newton's ring? Why?
8. How is the central spot in your experiment, bright or dark? Why?
9. Instead of reflected rays, if you look at transmitted rays, what do you expect to observe?
10. What happens with the central spot when a liquid of refractive index μ greater than that of the lens and less than that of the glass plate is introduced between the lens and the glass plate?
11. Is it possible to determine the refractive index of the liquid by this experiment?
12. What would happen to the ring if the space between lens and the plate is filled with a liquid of refractive index μ ?
13. What do you expect to see in the microscope if you use a white light source?
14. What is the difference between biprism fringes and Newton's ring fringes?
15. On which factors does the diameter of a ring depend?
16. What would happen if a glass plate is replaced by a plane mirror?
17. Why should a lens of large radius of curvature be used in this experiment?
18. Is it desirable to measure the radius of curvature of the given lens by a spherometer in the usual way?
19. What do you understand by (a) fringes of equal thickness (b) fringes of equal inclination and (c) fringes of equal chromatic order.
20. How does the sodium source, which you are using in your experiment work?

References

1. Fundamental of Optics by F. Jenkins and H. White 535 JEN/F
2. Optics by A.Ghatak 535 GHA/O
3. Optics by E. Hecht 535 HEC/O

Graph : Newton's Rings

Chapter 4

Einstein Coefficients and Light Amplification

4.1 Introduction

In this chapter we discuss interaction of radiation and atoms and obtain the relationship between absorption and emission processes. We show that for light amplification a state of population inversion should be created in the atomic system. We also obtain an expression for the gain coefficient of the system. This is followed by a discussion of two-level, three-level, and four-level systems using the rate equation approach. Finally a discussion of various mechanisms leading to broadening of spectral lines is discussed.

4.2 The Einstein Coefficients

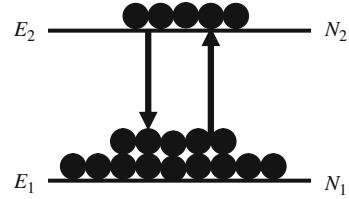
We consider two levels of an atomic system as shown in Fig. 4.1 and let N_1 and N_2 be the number of atoms per unit volume present in the energy levels E_1 and E_2 , respectively. The atomic system can interact with electromagnetic radiation in three distinct ways:

- An atom in the lower energy level E_1 can absorb the incident radiation at a frequency $\omega = (E_2 - E_1) / \hbar$ and be excited to E_2 ; this excitation process requires the presence of radiation. The rate at which absorption takes place from level 1 to level 2 will be proportional to the number of atoms present in the level E_1 and also to the energy density of the radiation at the frequency $\omega = (E_2 - E_1) / \hbar$. Thus if $u(\omega)d\omega$ represents the radiation energy per unit volume between ω and $\omega + d\omega$ then we may write the number of atoms undergoing absorptions per unit time per unit volume from level 1 to level 2 as

$$\Gamma_{12} = B_{12}u(\omega)N_1 \quad (4.1)$$

where B_{12} is a constant of proportionality and depends on the energy levels E_1 and E_2 . Notice here that $u(\omega)$ has the units of energy density per frequency interval.

Fig. 4.1 Two states of an atom with energies E_1 and E_2 with corresponding population densities of N_1 and N_2 , respectively



- (b) For the reverse process, namely the deexcitation of the atom from E_2 to E_1 , Einstein postulated that an atom can make a transition from E_2 to E_1 through two distinct processes, namely *stimulated emission* and *spontaneous emission*. In the case of stimulated emission, the radiation which is incident on the atom stimulates it to emit radiation and the rate of transition to the lower energy level is proportional to the energy density of radiation at the frequency ω . Thus, the number of stimulated emissions per unit time per unit volume will be

$$\Gamma_{21} = B_{21}u(\omega)N_2 \quad (4.2)$$

where B_{21} is the coefficient of proportionality and depends on the energy levels.

- (c) An atom which is in the upper energy level E_2 can also make a spontaneous emission; this rate will be proportional to N_2 only and thus we have for the number atoms making spontaneous emissions per unit time per unit volume

$$U_{21} = A_{21}N_2 \quad (4.3)$$

At thermal equilibrium between the atomic system and the radiation field, the number of upward transitions must be equal to the number of downward transitions. Hence, at thermal equilibrium

$$N_1B_{12}u(\omega) = N_2A_{21} + N_2B_{21}u(\omega)$$

or

$$u(\omega) = \frac{A_{21}}{(N_1/N_2)B_{12} - B_{21}} \quad (4.4)$$

Using Boltzmann's law, the ratio of the equilibrium populations of levels 1 and 2 at temperature T is

$$\frac{N_1}{N_2} = e^{(E_2-E_1)/k_B T} = e^{\hbar\omega/k_B T} \quad (4.5)$$

where $k_B (= 1.38 \times 10^{-23} \text{ J/K})$ is the Boltzmann's constant. Hence

$$u(\omega) = \frac{A_{21}}{B_{12}e^{\hbar\omega/k_B T} - B_{21}} \quad (4.6)$$

Now according to Planck's law, the radiation energy density per unit frequency interval is given by (see Appendix F)

$$u(\omega) = \frac{\hbar\omega^3 n_0^3}{\pi^2 c^3} \frac{1}{e^{\hbar\omega/k_B T} - 1} \quad (4.7)$$

where c is the velocity of light in free space and n_0 is the refractive index of the medium.

Comparing Eqs. (4.6) and (4.7), we obtain

$$B_{12} = B_{21} = B \quad (4.8)$$

and

$$\frac{A_{21}}{B_{21}} = \frac{\hbar\omega^3 n_0^3}{\pi^2 c^3} \quad (4.9)$$

Thus the stimulated emission rate per atom is the same as the absorption rate per atom and the ratio of spontaneous to stimulated emission coefficients is given by Eq. (4.9). The coefficients A and B are referred to as the Einstein A and B coefficients.

At thermal equilibrium, the ratio of the number of spontaneous to stimulated emissions is given by

$$R = \frac{A_{21}N_2}{B_{21}N_2 u(\omega)} = e^{\hbar\omega/k_B T} - 1 \quad (4.10)$$

Thus at thermal equilibrium at a temperature T , for frequencies, $\omega \gg k_B T/\hbar$, the number of spontaneous emissions far exceeds the number of stimulated emissions.

Example 4.1 Let us consider an optical source at $T = 1000$ K. At this temperature

$$\frac{k_B T}{\hbar} = \frac{1.38 \times 10^{-23} (\text{J/K}) \times 10^3 (\text{K})}{1.054 \times 10^{-34} (\text{Js})} \approx 1.3 \times 10^{14} \text{ s}^{-1}$$

Thus for $\omega \gg 1.3 \times 10^{14} \text{ s}^{-1}$, the radiation would be mostly due to spontaneous emission. For $\lambda \approx 500 \text{ nm}$, $\omega \approx 3.8 \times 10^{15} \text{ s}^{-1}$ and

$$R \approx e^{29.2} \approx 5.0 \times 10^{12}$$

Thus at optical frequencies the emission from a hot body is predominantly due to spontaneous transitions and hence the light from usual light sources is incoherent.

We shall now obtain the relationship between the Einstein A coefficient and the spontaneous lifetime of level 2. Let us assume that an atom in level 2 can make a spontaneous transition only to level 1. Then since the number of atoms making spontaneous transitions per unit time per unit volume is $A_{21}N_2$, we may write the rate of change of population of level 2 with time due to spontaneous emission as

$$\frac{dN_2}{dt} = -A_{21}N_2 \quad (4.11)$$

the solution of which is

$$N_2(t) = N_2(0)e^{-A_{21}t} \quad (4.12)$$

Thus the population of level 2 reduces by $1/e$ in a time $t_{\text{sp}} = 1/A_{21}$ which is called the spontaneous lifetime associated with the transition $2 \rightarrow 1$.

Example 4.2 In the $2P \rightarrow 1S$ transition in the hydrogen atom, the lifetime of the $2P$ state for spontaneous emission is given by

$$t_{\text{sp}} = \frac{1}{A_{21}} \approx 1.6 \times 10^{-9} \text{ s}$$

Thus

$$A_{21} \approx 6 \times 10^8 \text{ s}^{-1}$$

The frequency of the transition is given by

$$\omega \approx 1.55 \times 10^{16} \text{ s}^{-1} \quad (\hbar\omega \approx 10.2 \text{ eV})$$

Thus

$$B_{21} = \frac{\pi^2 c^3}{\hbar \omega^3 n_0^3} A_{21} \approx 4.1 \times 10^{20} \text{ m}^3/\text{Js}^2$$

where we have assumed $n_0 \approx 1$. (Note the unit for B_{21} .)

Now, if one observes the spectrum of the radiation due to the spontaneous emission from a collection of atoms, one finds that the radiation is not strictly monochromatic but is spread over a certain frequency range. Similarly, if one measures the absorption by a collection of atoms as a function of frequency, one again finds that the atoms are capable of absorbing not just a single frequency but radiation over a band of frequencies. This implies that energy levels have widths and the atoms can interact with radiation over a range of frequencies but the strength of interaction is a function of frequency (see Fig. 4.2). This function that describes the frequency dependence is called the lineshape function and is represented by $g(\omega)$. The function is usually normalized according to

$$\int g(\omega) d\omega = 1 \quad (4.13)$$

Explicit expressions for $g(\omega)$ will be obtained in Section 4.5.

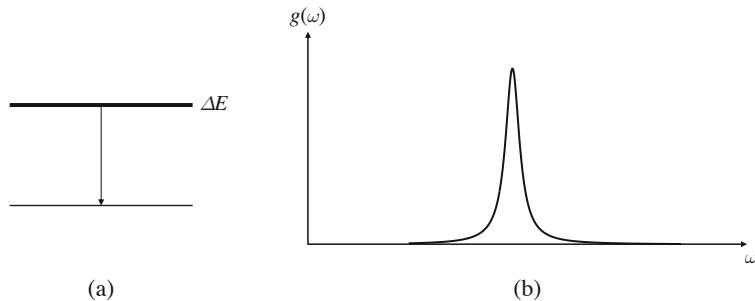


Fig. 4.2 (a) Because of the finite lifetime of a state each state has a certain width so that the atom can absorb or emit radiation over a range of frequencies. The corresponding lineshape is shown in (b)

From the above we may say that out of the total N_2 and N_1 atoms per unit volume, only $N_2 g(\omega)d\omega$ and $N_1 g(\omega)d\omega$ atoms per unit volume will be capable of interacting with radiation of frequency lying between ω and $\omega + d\omega$. Hence the total number of stimulated emissions per unit time per unit volume will now be given by

$$\begin{aligned}\Gamma_{21} &= \int B_{21} u(\omega) N_2 g(\omega) d\omega \\ &= N_2 \frac{\pi^2 c^3}{\hbar n_0^3 t_{sp}} \int \frac{u(\omega) g(\omega)}{\omega^3} d\omega\end{aligned}\quad (4.14)$$

where we have used Eq. (4.9) and $A_{21} = 1/t_{sp}$.

We now consider two specific cases.

- (1) If the atoms are interacting with radiation whose spectrum is very broad compared to that of $g(\omega)$ (see Fig. 4.3a), then one may assume that over the region of integration where $g(\omega)$ is appreciable $u(\omega)/\omega^3$ is essentially constant and thus may be taken out of the integral in Eq. (4.14). Using the normalization integral, Eq. (4.14) becomes

$$\Gamma_{21} = N_2 \frac{\pi^2 c^3}{\hbar \omega^3 n_0^3 t_{sp}} u(\omega) \quad (4.15)$$

where ω now represents the transition frequency. Equation (4.15) is consistent with Eq. (4.2) if we use Eq. (4.9) for B_{21} . Thus Eq. (4.15) represents the rate of stimulated emission per unit volume when the atom interacts with broadband radiation.

- (2) We now consider the other extreme case in which the atom is interacting with near-monochromatic radiation. If the frequency of the incident radiation is ω' , then the $u(\omega)$ curve will be extremely sharply peaked at $\omega = \omega'$ as compared to $g(\omega)$ (see Fig. 4.3b) and thus $g(\omega)/\omega^3$ can be taken out of the integral to obtain

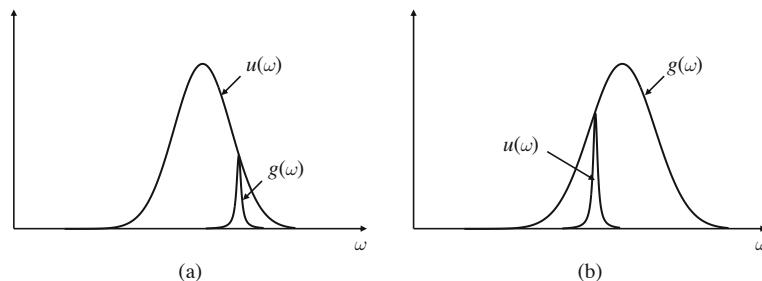


Fig. 4.3 (a) Atoms characterized by the lineshape function $g(\omega)$ interacting with broadband radiation. (b) Atoms interacting with near-monochromatic radiation

$$\begin{aligned}\Gamma_{21} &= N_2 \frac{\pi^2 c^3}{\hbar \omega'^3 n_0^3 t_{sp}} g(\omega') \int u(\omega) d\omega \\ &= N_2 \frac{\pi^2 c^3}{\hbar \omega'^3 n_0^3 t_{sp}} g(\omega') u\end{aligned}\quad (4.16)$$

where

$$u = \int u(\omega) d\omega \quad (4.17)$$

is the energy density of the incident near-monochromatic radiation. It may be noted that u has dimensions of energy per unit volume unlike $u(\omega)$ which has the dimensions of energy per unit volume per unit frequency interval. Thus when the atom described by a lineshape function $g(\omega)$ interacts with near-monochromatic radiation at frequency ω' , the stimulated emission rate per unit volume is given by Eq. (4.16).

In a similar manner, the number of stimulated absorptions per unit time per unit volume will be

$$\Gamma_{12} = N_1 \frac{\pi^2 c^3}{\hbar \omega'^3 n_0^3 t_{sp}} g(\omega') u \quad (4.18)$$

4.2.1 Absorption and Emission Cross Sections

The rates of absorption and stimulated emission can also be characterized in terms of the parameters referred to as absorption and emission cross sections. To do this, we first notice that the energy density u and the intensity I of the propagating electromagnetic wave are related through the following equation (see Section 2.2):

$$u = \frac{I}{c/n_0} = \frac{n_0 I}{c} \quad (4.19)$$

The number of photons crossing a unit area per unit time also referred to as the photon flux ϕ is related to the intensity I through the following equation:

$$\phi = \frac{I}{\hbar \omega} \quad (4.20)$$

Thus Eq. (4.18) can be written as

$$\begin{aligned}\Gamma_{12} &= N_1 \frac{\pi^2 c^2}{\omega'^2 n_0^2 t_{sp}} g(\omega') \phi \\ &= \sigma_a N_1 \phi\end{aligned}\quad (4.21)$$

where σ_a represents the absorption cross section (with dimensions of area) for this transition and is given by

$$\sigma_a = \frac{\pi^2 c^2}{\omega^2 n_0^2 t_{sp}} g(\omega) \quad (4.22)$$

Similarly we can define the emission cross section σ_e through the rate Γ_{21} . Since Γ_{12} and Γ_{21} are equal, the absorption and emission cross sections are equal.

Note that the absorption and emission cross sections are functions of frequency and are related to the line broadening function $g(\omega)$ and the lifetime t_{sp} .

The peak emission cross sections for some of the important laser transitions are given in Table 4.1.

Table 4.1 Table giving transition cross section for some important laser lines

Laser transition	Wavelength (nm)	Cross section (m^2)	Lifetime (μs)
He-Ne laser	632.8	5.8×10^{-17}	30×10^{-3}
Argon ion	514.5	2.5×10^{-17}	6×10^{-3}
Nd:YAG	1064	2.8×10^{-23}	230

Example 4.3 Consider the transition in neon atom at the wavelength of 1150 nm. This transition is Doppler broadened with a linewidth of 900 MHz and the upper state spontaneous lifetime is 100 ns. Using Eq. (4.22) we can calculate the peak absorption cross section. If we assume $g(\omega_0) \sim 1/\Delta\omega$, we obtain $\sigma_a \sim 5.8 \times 10^{-16} m^2$

4.3 Light Amplification

We next consider a collection of atoms and let a near-monochromatic radiation of energy density u at frequency ω' pass through it. We shall now obtain the rate of change of intensity of the radiation as it passes through the medium.

Let us consider two planes P_1 and P_2 of area S situated at z and $z + dz$, z being the direction of propagation of the radiation (see Fig. 4.4). If $I(z)$ and $I(z+dz)$ represent the intensity of the radiation at z and $z + dz$, respectively, then the net amount of energy entering the volume Sdz between P_1 and P_2 will be

$$\begin{aligned} [I(z) - I(z + dz)]S &= [I(z) - I(z) - \frac{dI}{dz}dz]S \\ &= -\frac{dI}{dz}Sdz \end{aligned} \quad (4.23)$$

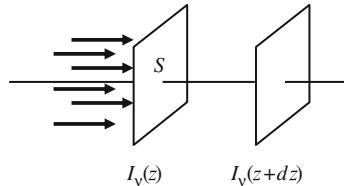


Fig. 4.4 Propagation of radiation at frequency ω' through a medium leading to a change of intensity with propagation

This must be equal to the net energy absorbed by the atoms in the volume Sdz . The energy absorbed by the atoms in going from level 1 to level 2 will be $\Gamma_{12}Sdz\hbar\omega'$ where $\hbar\omega'$ is the energy absorbed when an atom goes from level 1 to level 2. Similarly the energy released through stimulated emissions from level 2 to level 1 will be $\Gamma_{21}Sdz\hbar\omega'$. We shall neglect the energy arising from spontaneous emission since it appears over a broad frequency range and is also emitted in all directions. Thus the fraction of the spontaneous emission which would be at the radiation frequency ω' and which would be traveling along the z -direction will be very small. Thus the net energy absorbed per unit time in the volume Sdz will be

$$\begin{aligned} (\Gamma_{12} - \Gamma_{21})\hbar\omega' Sdz &= \frac{\pi^2 c^3}{\hbar\omega'^3 n_0^3 t_{sp}} \frac{1}{ug(\omega')}(N_1 - N_2)\hbar\omega' Sdz \\ &= \frac{\pi^2 c^3}{\omega'^2 n_0^3 t_{sp}} ug(\omega')(N_1 - N_2) Sdz \end{aligned} \quad (4.24)$$

Now, the energy density u and the intensity of radiation I are related through Eq. (4.19). Thus using Eqs. (4.23) and (4.24) we obtain

$$\frac{dI}{dz} = -\alpha I \quad (4.25)$$

where

$$\alpha = \frac{\pi^2 c^2}{\omega^2 n_0^2 t_{sp}} g(\omega)(N_1 - N_2) = -\gamma \quad (4.26)$$

and we have removed the prime on ω with the understanding that ω represents the frequency of the incident radiation. Hence if $N_1 > N_2$, α is positive (and γ is negative) and the intensity decreases with z leading to an attenuation of the beam. On the other hand, if $N_2 > N_1$ then α is negative (and γ is positive) the beam is amplified with z . Figure 4.5 shows typical plots of $\alpha(\omega)$ versus ω for $N_1 > N_2$ and $N_2 > N_1$. Obviously the frequency dependence of α will be almost the same as that of the lineshape function $g(\omega)$. The condition $N_2 > N_1$ is called population inversion and it is under this condition that one can obtain optical amplification.

In Eq. (4.26) if $(N_1 - N_2)$ is independent of I , then we have from Eq. (4.25)

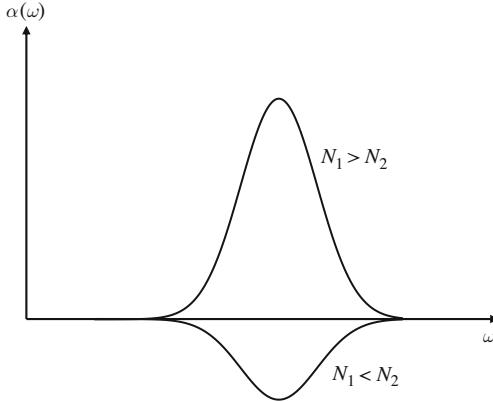
$$I(z) = I(0)e^{-\alpha z} \quad (4.27)$$

i.e., an exponential attenuation when $N_1 > N_2$ and an exponential amplification when $N_2 > N_1$. We should mention that such an exponential decrease or increase of intensity is obtained for low intensities; for large intensities saturation sets in and $(N_1 - N_2)$ is no longer independent of I (see Chapter 5).

Example 4.3 We consider a ruby laser (see Chapter 11) with the following characteristics:

$$\begin{aligned} n_0 &= 1.76, t_{sp} = 3 \times 10^{-3} \text{s}, \lambda_0 = 6943 \text{\AA} \\ g(\omega_0) &\approx 1/\Delta\omega \approx 1.1 \times 10^{-12} \text{s} \end{aligned}$$

Fig. 4.5 A typical variation of $\alpha(\omega)$ with ω for an amplifying medium corresponding to $N_2 > N_1$ (lower curve) and for an attenuating medium with $N_2 < N_1$ (upper curve)



where we have assumed that for the normalized lineshape function¹

$$g(\omega_0) \approx \frac{1}{\Delta\omega} \quad (4.28)$$

where $\Delta\omega$ represents the full width at half maximum of the lineshape function and ω_0 represents the frequency at the centre of the line. At thermal equilibrium at 300 K,

$$\frac{N_2}{N_1} = e^{-hv/k_B T} \approx 10^{-30} \approx 0$$

A typical chromium ion density in a ruby laser is about $1.6 \times 10^{19} \text{ cm}^{-3}$ and since at 300 K most atoms are in the ground level, the absorption coefficient at the centre of the line would be

$$\begin{aligned} \alpha &= 1.4 \times 10^{-19} (N_1 - N_2) \approx 1.4 \times 10^{-19} \times 1.6 \times 10^{19} \\ &\approx 2.2 \text{ cm}^{-1} \end{aligned}$$

If a population inversion density of $5 \times 10^{16} \text{ cm}^{-3}$ is generated (which represents a typical value) then the gain coefficient will be

$$\begin{aligned} -\alpha &\approx 1.4 \times 10^{-19} \times 5 \times 10^{16} \\ &\approx 7 \times 10^{-3} \text{ cm}^{-1} \end{aligned}$$

Example 4.4 As another example we consider the Nd:YAG laser (see Chapter 11) for which

$$n_0 = 1.82, \quad t_{sp} = 0.23 \times 10^{-3} \text{ s}, \quad \lambda_0 = 1.06 \mu\text{m}$$

$$\Delta v = \frac{\Delta\omega}{2\pi} \approx \frac{1}{2\pi g(\omega_0)} \approx 1.95 \times 10^{11} \text{ Hz}$$

If we want a gain of 1 m^{-1} , the inversion required can be calculated from Eq. (4.26) as

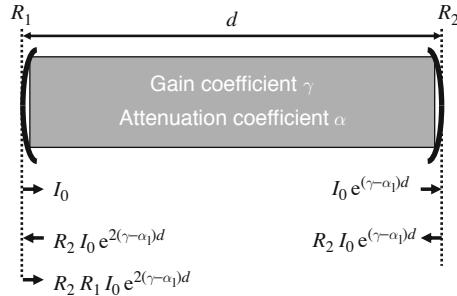
$$\begin{aligned} (N_2 - N_1) &= \frac{4v^2 n_0^2 t_{sp} \alpha}{c^2 g(\omega)} \\ &\approx 3.3 \times 10^{15} \text{ cm}^{-3} \end{aligned}$$

¹We will show in Section 4.5 that $g(\omega_0)\Delta\omega$ equals $(2/\pi)$ and $(4 \ln 2/\pi)^{1/2}$ for Lorentzian and Gaussian lineshape functions, respectively.

4.4 The Threshold Condition

In the last section we saw that in order that a medium be capable of amplifying incident radiation, one must create a state of population inversion in the medium. Such a medium will behave as an amplifier for those frequencies which fall within its linewidth. In order to generate radiation, this amplifying medium is placed in an optical resonator which consists of a pair of mirrors facing each other much like in a Fabry–Perot etalon (see Fig. 4.6). Radiation which bounces back and forth between the mirrors is amplified by the amplifying medium and also suffers losses due to the finite reflectivity of the mirrors and other scattering and diffraction losses. If the oscillations have to be sustained in the cavity then the losses must be exactly compensated by the gain. Thus a minimum population inversion density is required to overcome the losses and this is called the threshold population inversion.

Fig. 4.6 A typical optical resonator consisting of a pair of mirrors facing each other. The active medium is placed inside the cavity



In order to obtain an expression for the threshold population inversion, let d represent the length of the resonator and let R_1 and R_2 represent the reflectivities of the mirrors (see Fig. 4.6). Let α_1 represent the average loss per unit length due to all loss mechanisms (other than the finite reflectivity) such as scattering loss and diffraction loss due to finite mirror sizes. Let us consider a radiation with intensity I_0 leaving mirror M_1 . As it propagates through the medium and reaches the second mirror, it is amplified by $e^{\gamma d}$ and also suffers a loss of $e^{-\alpha_1 d}$; for an amplifying medium γ is positive and $e^{\gamma d} > 1$. The intensity of the reflected beam at the second mirror will be $I_0 R_2 e^{(\gamma - \alpha_1)d}$. A second passage through the resonator and a reflection at the first mirror leads to an intensity for the radiation after one complete round trip of $I_0 R_1 R_2 e^{2(\gamma - \alpha_1)d}$. Hence for laser oscillation to begin

$$R_1 R_2 e^{2(\gamma - \alpha_1)d} \geq 1 \quad (4.29)$$

the equality sign giving the threshold value for α (i.e., for population inversion). Indeed, when the laser is oscillating in a steady state with a continuous wave oscillation, then the equality sign in Eq. (4.29) must be satisfied. If the inversion is increased then the LHS becomes greater than unity; this implies that the round trip gain is greater than the round trip loss. This would result in an increasing intensity inside the laser till saturation effects take over, which would result in a decrease

in the inversion (we shall explicitly show saturation effects in [Chapter 5](#)). Thus the gain is brought back to its value at threshold.

Equation (4.29) can be written as

$$\gamma \geq \alpha_1 - \frac{1}{2d} \ln R_1 R_2 \quad (4.30)$$

This RHS of Eq. (4.30) depends on the passive cavity parameters only. This can be related to the passive cavity lifetime t_c which is the time in which energy in the cavity reduces by a factor 1/e. In the absence of amplification by the medium, the intensity at a point reduces by a factor $R_1 R_2 e^{-2\alpha_1 d} = e^{-(2\alpha_1 d - \ln R_1 R_2)}$ in a time corresponding to one round-trip time. One round-trip time corresponds to $t = 2d(c/n_0) = 2dn_0/c$. Hence if the intensity reduces as e^{-t/t_c} , then in a time $t = 2dn_0/c$, the factor by which the intensity will be reduced is $e^{-2dn_0/c t_c}$. Thus

$$e^{-(2\alpha_1 d - \ln R_1 R_2)} = e^{-2dn_0/c t_c}$$

or

$$\frac{1}{t_c} = \frac{c}{2dn_0} (2\alpha_1 d - \ln R_1 R_2) \quad (4.31)$$

Using Eqs. (4.26) and (4.31), Eq. (4.30) becomes

$$(N_2 - N_1) \geq \frac{4v^2 n_0^3}{c^3} \frac{t_{sp}}{t_c} \frac{1}{g(\omega)} \quad (4.32)$$

Corresponding to the equality sign, we have the threshold population inversion density required for the oscillation of the laser.

According to Eq. (4.32), in order to have a low threshold value of the population inversion, the following conditions must hold:

- (a) The value of t_c should be large, i.e., the cavity losses must be small.
- (b) Since $g(\omega)$ is normalized according to Eq. (4.13) the peak value of $g(\omega)$ will be inversely proportional to the width $\Delta\omega$ of the $g(\omega)$ function [see Eq. (4.28)]. Thus smaller widths give larger values of $g(\omega)$ which implies lower threshold values of $(N_2 - N_1)$. Also since the largest $g(\omega)$ appears at the line centre, the resonator mode which lies closest to the line centre will reach threshold first and begin to oscillate.
- (c) Smaller values of t_{sp} (i.e., strongly allowed transitions) also lead to smaller values of threshold inversion. At the same time for smaller relaxation times (t_{sp}), larger pumping power will be required to maintain a given population inversion. In general, population inversion is more easily obtained on transitions which have longer relaxation times.
- (d) The value of $g(\omega)$ at the centre of the line is inversely proportional to $\Delta\omega$ which, for example, in the case of Doppler broadening is proportional to ω (see [Section 4.5](#)). Thus the threshold population inversion increases approximately in proportion to ω^3 . Thus it is much easier to obtain laser action at infrared wavelengths than in the ultraviolet region.

Example 4.5 We first consider a ruby laser² which has the following typical parameters:

$$\begin{aligned}\lambda_0 &= 6943\text{\AA}, t_{\text{sp}} \approx 3 \times 10^{-3}\text{s}, n_0 = 1.76, d = 5\text{ cm} \\ R_1 &= R_2 = 0.9, \alpha_1 \approx 0 \\ g(\omega_0) &= \frac{1}{\Delta\omega} = \frac{1}{2\pi\Delta\nu} \approx 1.1 \times 10^{-12}\text{s}\end{aligned}$$

Thus for the above values

$$t_c \approx 2.8 \times 10^{-9}\text{s}$$

and

$$(N_2 - N_1)_{\text{th}} \approx 1.5 \times 10^{17}\text{ cm}^{-3}$$

Typical Cr⁺³ ion densities are about $1.6 \times 10^{19}\text{ cm}^{-3}$. Thus the fractional excess population is very small. The above population inversion corresponds to a gain of about 0.02 cm^{-1} or to 0.09 dB/cm .

Example 4.6 As another example, we consider a He–Ne laser with the following typical characteristics:

$$\begin{aligned}\lambda_0 &= 6328\text{\AA}, t_{\text{sp}} = 10^{-7}\text{s}, n_0 \approx 1, d = 20\text{ cm} \\ R_1 &= R_2 = 0.98, \alpha_1 \approx 0 \\ \Delta\nu &\approx 10^9\text{Hz} \\ g(\omega_0) &\approx \frac{1}{2\pi\Delta\nu} \approx 0.16 \times 10^{-9}\text{s}\end{aligned}$$

for the above values

$$t_c \approx 3.3 \times 10^{-8}\text{s}$$

and

$$(N_2 - N_1)_{\text{th}} \approx 6.24 \times 10^8\text{ cm}^{-3}$$

4.5 Line Broadening Mechanisms

As we mentioned in Section 4.2 the radiation coming out of a collection of atoms making transitions between two energy levels is never perfectly monochromatic. This line broadening is described in terms of the lineshape function $g(\omega)$ that was introduced in Section 4.2. In this section, we shall discuss some important line broadening mechanisms and obtain the corresponding $g(\omega)$. A study of line broadening is extremely important since it determines the operation characteristics of the laser such as the threshold population inversion and the number of oscillating modes.

The various broadening mechanisms can be broadly classified as homogeneous or inhomogeneous broadening. In the case of homogenous broadening (like natural or collision broadening) the mechanisms act to broaden the response of each atom in an identical fashion, and for such a case the probability of absorption or emission of radiation of a certain frequency is the same for all atoms in the collection. Thus there

²Ruby laser active medium consists of Cr⁺³-doped ion Al₂O₃ and is an example of a three level laser. More details regarding the ruby laser are given in Section 10.2.

is nothing which distinguishes one group of atoms from another in the collection. In the case of inhomogeneous broadening, different groups of atoms are distinguished by different frequency responses. Thus, for example, in Doppler broadening groups of atoms having different velocity components are distinguishable and they have different spectral responses. Similarly broadening caused by local inhomogeneities of a crystal lattice acts to shift the central frequency of the response of individual atoms by different amounts, thereby leading to inhomogeneous broadening. In the following, we shall discuss natural, collision, and Doppler broadening.

4.5.1 Natural Broadening

We have seen earlier that an excited atom can emit its energy in the form of spontaneous emission. In order to investigate the spectral distribution of this spontaneous radiation, we recall that the rate of decrease of the number of atoms in level 2 due to transitions from level 2 to level 1 is [see Eq. (4.11)]

$$\frac{dN_2}{dt} = -A_{21}N_2 \quad (4.33)$$

For every transition an energy $\hbar\omega_0 = E_2 - E_1$ is released. Thus the energy emitted per unit time per unit volume will be

$$\begin{aligned} W(t) &= \left| \frac{dN_2}{dt} \right| \hbar\omega_0 \\ &= N_{20}A_{21}\hbar\omega_0 e^{-A_{21}t} \end{aligned} \quad (4.34)$$

where we have used Eqs. (4.33) and (4.12). Since Eq. (4.34) describes the variation of the intensity of the spontaneously emitted radiation, we may write the electric field associated with the spontaneous radiation as

$$E(t) = E_0 e^{i\omega_0 t} e^{-t/2t_{sp}} \quad (4.35)$$

where $t_{sp} = 1/A_{21}$ and we have used the fact that intensity is proportional to the square of the electric field. Thus the electric field associated with spontaneous emission decreases exponentially.

In order to calculate the spectrum associated with the wave described by the Eq. (4.35), we first take the Fourier transform:

$$\begin{aligned} \tilde{E}(\omega) &= \int_{-\infty}^{\infty} E(t) e^{-i\omega t} dt \\ &= E_0 \int_0^{\infty} \exp [i(\omega_0 - \omega)t - t/2t_{sp}] dt \\ &= E_0 \frac{1}{\frac{1}{2t_{sp}} + i(\omega - \omega_0)} \end{aligned} \quad (4.36)$$

where $t=0$ is the time at which the atoms start emitting radiation. The power spectrum associated with the radiation will be proportional to $|E_0(\omega)|^2$. Hence we may write the lineshape function associated with the spontaneously emitted radiation as

$$g(\omega) = K \frac{1}{(\omega - \omega_0)^2 + 1/4t_{\text{sp}}^2}$$

where K is a constant of proportionality which is determined such that $g(\omega)$ satisfies the normalization condition given by Eq. (4.13). Substituting for $g(\omega)$ in Eq. (4.13) and integrating, one can show that

$$K = \frac{1}{2\pi t_{\text{sp}}}$$

Thus the normalized lineshape function is

$$g(\omega) = \frac{2t_{\text{sp}}}{\pi} \frac{1}{1 + 4(\omega - \omega_0)^2 t_{\text{sp}}^2} \quad (4.37)$$

The above functional form is referred to as a Lorentzian and is plotted in Fig. 4.7. The full width at half maximum (FWHM) of the Lorentzian is

$$\Delta\omega_N = \frac{1}{t_{\text{sp}}} \quad (4.38)$$

Thus, Eq. (4.37) can also be written as

$$g(\omega) = \frac{2}{\pi \Delta\omega_N} \frac{1}{1 + 4(\omega - \omega_0)^2 / (\Delta\omega_N)^2} \quad (4.39)$$

A more precise derivation of Eq. (4.39) is given in Appendix G.

Example 4.7 The spontaneous lifetime of the sodium level leading to a D₁ line ($\lambda = 589.1\text{nm}$) is 16 ns. Thus the natural linewidth (FWHM) will be

$$\Delta\nu_N = \frac{1}{2\pi t_{\text{sp}}} \approx 10 \text{ MHz} \quad (4.40)$$

which corresponds to $\Delta\lambda \approx 0.001 \text{ nm}$.

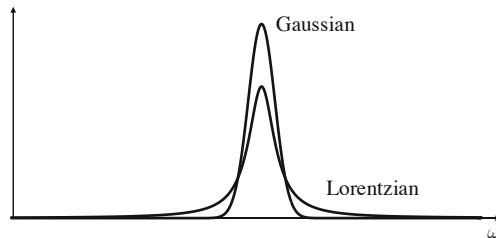


Fig. 4.7 The Lorentzian and Gaussian lineshape functions having the same FWHM

4.5.2 Collision Broadening

In a gas, random collisions occur between the atoms. In such a collision process, the energy levels of the atoms change when the atoms are very close due to their mutual interaction. Let us consider an atom which is emitting radiation and which collides with another atom. When the colliding atoms are far apart, their energy levels are unperturbed and the radiation emitted is purely sinusoidal (if we neglect the decay in the amplitude due to spontaneous emission). As the atoms come close together their energy levels are perturbed and thus the frequency of emission changes during the collision time. After the collision the emission frequency returns to its original value.

If τ_c represents the time between collisions and $\Delta\tau_c$ the collision time then one can obtain order of magnitude expressions as follows:

$$\begin{aligned}\Delta\tau_c &\approx \frac{\text{interatomic distance}}{\text{average thermal velocity}} \\ &\approx \frac{1 \text{ \AA}^\circ}{500 \text{ m/s}} \approx 2 \times 10^{-13} \text{ s}\end{aligned}$$

$$\begin{aligned}\tau_c &\approx \frac{\text{mean free path}}{\text{average thermal velocity}} \approx \frac{5 \times 10^{-4} \text{ m}}{500 \text{ m/s}} \\ &\approx 10^{-6} \text{ s}\end{aligned}$$

Thus the collision time is very small compared to the time between collisions and hence the collision may be taken to be almost instantaneous. Since the collision time $\Delta\tau_c$ is random, the phase of the wave after the collision is arbitrary with respect to the phase before the collision. Thus each collision may be assumed to lead to random phase changes as shown in Fig. 4.8. The wave shown in Fig. 4.8 is no longer monochromatic and this broadening is referred to as *collision broadening*.

In order to obtain the lineshape function for collision broadening, we note that the field associated with the wave shown in Fig. 4.8 can be represented by

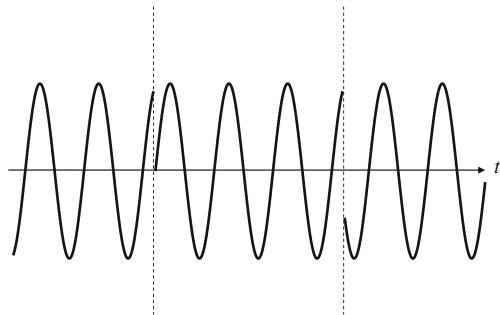


Fig. 4.8 The wave coming out of an atom undergoing random collisions at which there are abrupt phase changes

$$E(t) = E_0 e^{i(\omega_0 t + \phi)} \quad (4.41)$$

where the phase ϕ remains constant for $t_0 \leq t \leq t_0 + \tau_c$ and at each collision the phase ϕ changes randomly.

Since the wave is sinusoidal between two collisions, the spectrum of such a wave will be given by

$$\begin{aligned} \tilde{E}(\omega) &= \frac{1}{2\pi} \int_{t_0}^{t_0 + \tau_c} E_0 e^{i(\omega_0 t + \phi)} e^{-i\omega t} dt \\ &= \frac{1}{2\pi} E_0 e^{i[(\omega_0 - \omega)t_0 + \phi]} \frac{e^{i(\omega_0 - \omega)\tau_c} - 1}{i(\omega_0 - \omega)} \end{aligned} \quad (4.42)$$

The power spectrum of such a wave will be

$$I(\omega) \propto |\tilde{E}(\omega)|^2 = \left(\frac{E_0}{\pi}\right)^2 \frac{\sin^2[(\omega - \omega_0)\tau_c/2]}{(\omega - \omega_0)^2} \quad (4.43)$$

Now, at any instant, the radiation coming out of the atomic collection would be from atoms with different values of τ_c . In order to obtain the power spectrum we must multiply $I(\omega)$ by the probability $P(\tau_c)d\tau_c$ that the atom suffers a collision in the time interval between τ_c and $\tau_c + d\tau_c$ and integrate over τ_c from 0 to ∞ . It can be shown from kinetic theory that (see, e.g., Gopal (1974))

$$P(\tau_c)d\tau_c = \left(\frac{1}{\tau_0}\right) e^{-\tau_c/\tau_0} d\tau_c \quad (4.44)$$

where τ_0 represents the mean time between two collisions. Notice that

$$\int_0^\infty P(\tau_c)d\tau_c = 1, \quad \int_0^\infty \tau_c P(\tau_c)d\tau_c = \tau_0 \quad (4.45)$$

Hence the lineshape function for collision broadening will be

$$\begin{aligned} g(\omega) &\propto \int_0^\infty I(\omega) P(\tau_c)d\tau_c \\ &= \left(\frac{E_0}{\pi}\right)^2 \frac{1}{2} \frac{1}{(\omega - \omega_0)^2 + 1/\tau_0^2} \end{aligned}$$

which is again a Lorentzian. The normalized lineshape function will thus be

$$g(\omega)d\omega = \frac{\tau_0}{\pi} \frac{1}{1 + (\omega - \omega_0)^2 \tau_0^2} d\omega \quad (4.46)$$

and the FWHM will be

$$\Delta\omega_c = 2/\tau_0 \quad (4.47)$$

Thus a mean collision time of $\sim 10^{-6}$ s corresponds to a $\Delta\nu$ of about 0.3 MHz.

The mean time between collisions depends on the mean free path and the average speed of the atoms in the gas which in turn would depend on the pressure and temperature of the gas as well as the mass of the atom. An approximate expression for the average collision time is

$$\tau_0 = \frac{1}{8\pi} \left(\frac{2}{3}\right)^{1/2} \frac{(Mk_B T)^{1/2}}{pa^2}$$

where M is the atomic mass, a is the radius of the atom (assumed to be a hard sphere), and p is the pressure of the gas.

Example 4.8 In a He–Ne laser the pressure of gas is typically 0.5 torr. (Torr is a unit of pressure and 1 Torr = 1 mm of Hg). If we assume $a \sim 0.1$ nm, $T = 300$ K, $M = 20 \times 1.67 \times 10^{-27}$ kg, we obtain $\tau_0 \sim 580$ ns.

Problem 4.1 In the presence of both natural and collision broadening, in addition to the sudden phase changes at every collision, there will also be an exponential decay of the field as represented by Eq. (4.35). Show that in such a case, the FWHM is given by

$$\Delta\omega = \frac{1}{t_{sp}} + \frac{2}{t_0} \quad (4.48)$$

4.5.3 Doppler Broadening

In a gas, atoms move randomly and when a moving atom interacts with electromagnetic radiation, the apparent frequency of the wave is different from that seen from a stationary atom; this is called the Doppler effect and the broadening caused by this is termed Doppler broadening.

In order to obtain $g(\omega)$ for Doppler broadening, we consider radiation of frequency ω passing through a collection of atoms which have a resonant frequency ω_0 and which move randomly (we neglect natural and collision broadening in this discussion). In order that an atom may interact with the incident radiation, it is necessary that the apparent frequency seen by the atom in its frame of reference be ω_0 . If the radiation is assumed to propagate along the z -direction, then the apparent frequency seen by the atom having a z -component of velocity v_z will be

$$\tilde{\omega} = \omega \left(1 - \frac{v_z}{c}\right) \quad (4.49)$$

Hence for a strong interaction, the frequency of the incident radiation must be such that $\tilde{\omega} = \omega_0$. Thus

$$\omega = \omega_0 \left(1 - \frac{v_z}{c}\right)^{-1} \approx \omega_0 \left(1 + \frac{v_z}{c}\right) \quad (4.50)$$

where we have assumed $v_z \ll c$. Thus the effect of the motion is to change the resonant frequency of the atom.

In order to obtain the $g(\omega)$ due to Doppler broadening, we note that the probability that an atom has a z component of velocity lying between v_z and $v_z + dv_z$ is given by the Maxwell distribution

$$P(v_z) dv_z = \left(\frac{M}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left(-\frac{Mv_z^2}{2k_B T} \right) dv_z \quad (4.51)$$

where M is the mass of the atom and T the absolute temperature of the gas. Hence the probability $g(\omega)d\omega$ that the transition frequency lies between ω and $\omega + d\omega$ is equal to the probability that the z component of the velocity of the atom lies between v_z and $v_z + dv_z$ where

$$v_z = \frac{(\omega - \omega_0)}{\omega_0} c$$

Thus

$$g(\omega)d\omega = \frac{c}{\omega_0} \left(\frac{M}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left[-\frac{Mc^2}{2k_B T} \frac{(\omega - \omega_0)^2}{\omega_0^2} \right] d\omega \quad (4.52)$$

which corresponds to a Gaussian distribution. The lineshape function is peaked at ω_0 , and the FWHM is given by

$$\Delta\omega_D = 2\omega_0 \left(\frac{2k_B T}{Mc^2} \ln 2 \right)^{\frac{1}{2}} \quad (4.53)$$

In terms of $\Delta\omega_D$ Eq. (4.52) can be written as

$$g(\omega)d\omega = \frac{2}{\Delta\omega_D} \left(\frac{\ln 2}{\pi} \right)^{\frac{1}{2}} \exp \left[-4 \ln 2 \frac{(\omega - \omega_0)^2}{(\Delta\omega_D)^2} \right] d\omega \quad (4.54)$$

Figure 4.7 shows a comparative plot of a Lorentzian and a Gaussian line having the same FWHM. It can be seen that the peak value of the Gaussian is more and that the Lorentzian has a much longer tail. As an example, for the D₁ line of sodium $\lambda \approx 589.1$ nm at $T = 500$ K, $\Delta\nu_D = 1.7 \times 10^9$ Hz which corresponds to $\Delta\lambda_D \approx 0.02$ Å. For neon atoms corresponding to $\lambda = 6328$ Å (the red line of the He–Ne laser) at 300 K, we have $\Delta\nu_D \approx 1600$ MHz where we have used $M_{Ne} \approx 20 \times 1.67 \times 10^{-27}$ kg. For the vibrational transition of the carbon dioxide molecule leading to the 10.6 μm radiation, at $T = 300$ K, we have

$$\Delta\nu_D \approx 5.6 \times 10^7 \text{ Hz} \Rightarrow \Delta\lambda_D \approx 0.19 \text{ Å}$$

where we have used $M_{CO_2} \approx 44 \times 1.67 \times 10^{-27}$ kg

In all the above discussions we have considered a single broadening mechanisms at a time. In general, all broadening mechanisms will be present simultaneously and

the resultant lineshape function has to be evaluated by performing a convolution of the different lineshape functions.

Problem 4.2 Obtain the lineshape function in the presence of both natural and Doppler broadening

Solution From Maxwell's velocity distribution, the fraction of atoms with their center frequency lying between ω' and $\omega' + d\omega'$ is given by

$$f(\omega')d\omega' = \left(\frac{M}{2\pi k_B T} \right)^{\frac{1}{2}} \frac{c}{\omega_0} \exp \left[-\frac{Mc^2}{2k_B T} \frac{(\omega' - \omega_0)^2}{\omega_0^2} \right] d\omega' \quad (4.55)$$

These atoms are characterized by a naturally broadened lineshape function described by

$$h(\omega - \omega') = \frac{2t_{sp}}{\pi} \frac{1}{1 + (\omega - \omega')^2 4t_{sp}^2} \quad (4.56)$$

Thus the resultant lineshape function will be given by

$$g(\omega) = \int f(\omega') h(\omega - \omega') d\omega' \quad (4.57)$$

which is nothing but the convolution of $f(\omega')$ with $h(\omega)$

Example 4.9 Neodymium doped in YAG and in glass are two very important lasers. The host YAG is crystalline while glass is amorphous. Thus the broadening in YAG host is expected to be much smaller than in glass host. In fact the linewidth at 300 K for Nd:YAG is about 120 GHz while that for Nd:glass is about 5400 GHz.

4.6 Saturation Behavior of Homogeneously and Inhomogeneously Broadened Transitions

In Section 4.5 we discussed the various line broadening mechanisms belonging to both homogeneous and inhomogeneous broadenings. In this section, we briefly discuss the difference in saturation behavior between the two kinds of broadenings.

Let us first consider a homogeneously broadened laser medium placed inside a resonator and let us assume that there is a resonator mode coinciding exactly with the center of the line. Initially as the pumping rate is below threshold, the gain in the resonator is less than the losses and the laser does not oscillate. As the pumping rate is increased, first to reach threshold is the mode at the center as it has the minimum threshold. We have seen earlier that when the laser is oscillating in steady state, the gain is exactly equal to the loss at the oscillating frequency. Thus at steady state even when the pumping power is increased beyond threshold, the gain at the oscillating frequency does not increase beyond the threshold value; this is because of the fact that the losses remain constant. In fact, increasing the pumping power will be accompanied by an increase in the power in the mode which in turn would be accompanied by a stronger saturation of the laser transition, thus reducing the gain

at the oscillation frequency again to the value at threshold. It may be mentioned that the gain could exceed the threshold value on a transient basis but not under steady state operation.

Now in a homogeneously broadened transition all the atoms have identical line-shapes peaked at the same frequency. Thus all atoms interact with the same oscillating mode and the increase in pumping power cannot increase the gain at other frequencies and thus the laser will oscillate only in a single longitudinal mode (see Fig. 4.9). This observation has been verified experimentally on some homogeneously broadened transitions such as Nd:YAG laser. The fact that a laser with homogeneously broadened transition can oscillate in many modes is due to spatial hole burning. This can be understood from the fact that each mode is a standing wave pattern between the resonator mirrors. Thus there are regions of high population inversion (at the nodes of the field where the field amplitude is very small) and regions of saturated population inversion (at the antinodes of the field where the field has maximum value). If one considers another mode which has (at least over some portions) antinodes at the nodes corresponding to the central oscillating mode, then this mode can draw energy from the atoms and, if the loss can be compensated by gain, this mode can also oscillate.

In contrast to the case of homogeneous broadening, if the laser medium is inhomogeneously broadened then a given mode at a central frequency can interact with only a group of atoms whose response curve contains the mode frequency (see Fig. 4.10). Thus if the pumping is increased beyond threshold, the gain at the oscillating frequency remains fixed but the gain at other frequencies can go on increasing (see Fig. 4.10). Thus, in an inhomogeneously broadened line one can have multimode oscillation and as one can see from Fig. 4.10. Each oscillating mode “burns holes in the frequency space” of the gain profile. These general conclusions regarding homogeneously and inhomogeneously broadened lines have been verified experimentally.

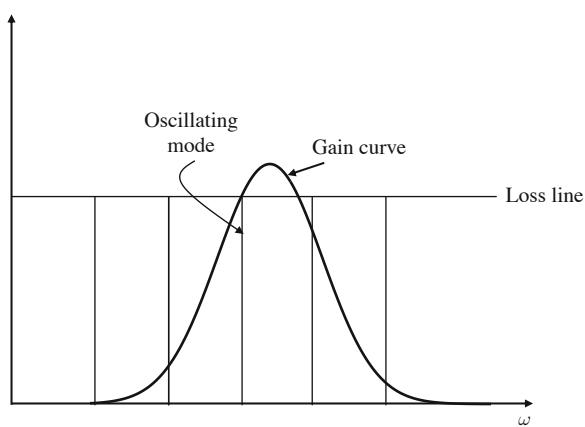
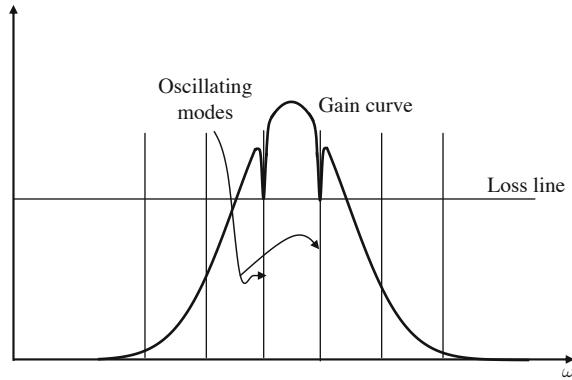


Fig. 4.9 In a homogeneously broadened transition, gain can compensate loss at only one oscillating mode leading to single longitudinal mode operation

Fig. 4.10 As the pumping is increased beyond threshold, under steady-state operation the gain at the various oscillating frequencies cannot increase beyond the threshold value but the gain at other frequencies may be much above the threshold value. The various frequencies are said to burn holes in the gain curve



Various techniques for single longitudinal mode oscillation of inhomogeneously broadened lasers are discussed in [Chapter 7](#).

Let us now consider an inhomogeneously broadened laser medium and let us assume that only a single mode exists within the entire gain profile. Let us also assume to begin with that the frequency of the mode does not coincide with the line center and that we slowly change the frequency of the mode so that it passes through the center of the profile to the other side of the peak in the gain profile. In order to determine the variation of the power output as the frequency is scanned through the line center, we observe that a mode of the laser is actually made up of two traveling waves traveling along opposite directions along the resonator axis. Thus when the mode frequency does not coincide with the line center, the wave travelling from left to right in the resonator will interact with those atoms whose z -directed velocities are near to [see Eq. (4.49)]:

$$v_z = \frac{\omega - \omega_{21}}{\omega_{21}} c \quad (4.58)$$

while the wave moving from right to left would interact with those atoms whose z -directed velocity would be

$$v_z = -\frac{\omega - \omega_{21}}{\omega_{21}} c \quad (4.59)$$

Thus there are two groups of atoms with equal and opposite z -directed velocities which are strongly interacting with the mode. As the frequency of the mode is tuned to the center these groups of atoms change with the frequency, and at the line center, the mode can interact only with the groups of atoms having a zero value of z -directed velocity. Thus the power output must decrease slightly when the mode frequency is tuned through the line center. In fact, this has been observed experimentally and is referred to as the Lamb dip – the presence of a Lamb dip in a He–Ne laser was shown by McFarlane, Bennet, and Lamb ([1963](#)).

4.7 Quantum Theory for the Evaluation of the Transition Rates and Einstein Coefficients

For the calculation of transition rates we consider the atom to be in the presence of an oscillating electric field given by

$$\mathbf{E}(t) = \hat{\mathbf{e}} E_0 \cos \omega t \quad (4.60)$$

which is switched on at $t = 0$; $\hat{\mathbf{e}}$ represents the unit vector along the direction of the electric field. The frequency ω is assumed to be very close to the resonant frequency $[(E_2 - E_1)/\hbar]$ corresponding to the transition from state 1 to 2 (see Fig. 4.1). We will show that the presence of the higher excited states can be neglected because of the corresponding transition frequencies are far away from ω . In the presence of the electric field, the time-dependent Schrödinger equation becomes

$$i\hbar \frac{\partial \Psi}{\partial t} = (H_0 + H') \Psi \quad (4.61)$$

where

$$H' = -e\mathbf{E} \cdot \mathbf{r} = -eE_0(\hat{\mathbf{e}} \cdot \mathbf{r}) \cos \omega t \quad (4.62)$$

represents the interaction energy of the electron with the electric field and H_0 (which is independent of time) represents the Hamiltonian of the atom; $e (< 0)$ represents the charge of the electron.³ Since H_0 is independent of time, the solution of the Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = H_0 \Psi \quad (4.63)$$

is of the form

$$\Psi = \sum \psi_n(\mathbf{r}) e^{-iE_n t/\hbar} \quad (4.64)$$

where $\psi_n(\mathbf{r})$ and E_n are the eigenfunctions and eigenvalues of H_0 :

$$H_0 \psi_n(\mathbf{r}) = E_n \psi_n(\mathbf{r}) \quad (4.65)$$

The functions $\psi_n(\mathbf{r})$ are known as the atomic wave functions and satisfy the orthonormality condition

$$\int \psi_n^*(\mathbf{r}) \psi_m(\mathbf{r}) d\tau = \delta_{mn} = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} \quad (4.66)$$

³We are considering here a single electron atom with r representing the position of the electron with respect to the nucleus. Thus the electric dipole moment of the atom is given by $p = e r$ because the direction of the dipole moment is from negative to the positive charge. The interaction energy of a dipole placed in an electric field E is $-\vec{p} \cdot \vec{E}$ which leads to Eq. (4.62).

The solution of Eq. (4.63) can be written as a linear combination of the atomic wave functions:

$$\Psi(\mathbf{r}, t) = \sum_n C_n(t) \psi_n(\mathbf{r}) e^{-i\omega_n t} \quad (4.67)$$

where

$$\omega_n = \frac{E_n}{\hbar} \quad (4.68)$$

and the coefficients are now time dependent to account for transitions among the various energy levels due to the perturbation. Substituting from Eq. (4.67) in Eq. (4.63) we obtain

$$\begin{aligned} i\hbar \sum_n \left(\frac{dC_n}{dt} - i\omega_n C_n \right) e^{-i\omega_n t} \psi_n(\mathbf{r}) &= \sum_n E_n C_n(t) \psi_n(\mathbf{r}) e^{-i\omega_n t} \\ &- eE_0 (\hat{\mathbf{e}} \cdot \mathbf{r}) \sum_n C_n(t) \psi_n(\mathbf{r}) e^{-i\omega_n t} \cos \omega t \end{aligned}$$

where we have used Eq. (4.65). It is immediately seen that the second term on the left-hand side exactly cancels with the first term on the right-hand side. If we multiply by ψ_m^* and integrate we would get

$$i\hbar \frac{dC_m}{dt} = \frac{1}{2} E_0 \sum_n D_{mn} C_n(t) \left(e^{i(\omega_{mn} + \omega)t} + e^{i(\omega_{mn} - \omega)t} \right) \quad (4.69)$$

where use has been made of the orthogonality relation [Eq. (4.66)] and

$$\omega_{mn} = \omega_m - \omega_n = \frac{E_m - E_n}{\hbar} \quad (4.70)$$

$$D_{mn} = \hat{\mathbf{e}} \cdot \mathbf{P}_{mn} \quad (4.71)$$

$$\mathbf{P}_{mn} = -e \int \psi_m^*(\mathbf{r}) \mathbf{r} \psi_n(\mathbf{r}) d\tau = |e| \int \psi_m^*(\mathbf{r}) \mathbf{r} \psi_n(\mathbf{r}) d\tau \quad (4.72)$$

We wish to solve Eq. (4.69) subject to the boundary condition

$$\begin{aligned} C_k(t = 0) &= 1 \\ C_n(t = 0) &= 0 \quad \text{for } n \neq k \end{aligned} \quad (4.73)$$

i.e., at $t = 0$, the atom is assumed to be in the state characterized by the wave function ψ_k . Equation (4.69) represents an infinite set of coupled equations, and as a first approximation, one may replace $C_n(t)$ by $C_n(0)$ on the right-hand side of Eq. (4.69). Thus

$$i\hbar \frac{dC_m}{dt} = \frac{1}{2} E_0 D_{mk} \left(e^{i(\omega_{mk} + \omega)t} + e^{i(\omega_{mk} - \omega)t} \right) \quad (4.74)$$

Integrating, one obtains

$$C_m(t) - C_m(0) \approx -\frac{E_0}{2\hbar} D_{mk} \left[\frac{e^{i(\omega_{mk}+\omega)t} - 1}{(\omega_{mk} + \omega)} + \frac{e^{i(\omega_{mk}-\omega)t} - 1}{(\omega_{mk} - \omega)} \right] \quad (4.75)$$

or, for $m \neq k$

$$C_m(t) \approx -i \frac{E_0}{\hbar} D_{mk} \left[e^{i(\omega_{mk}+\omega)t/2} \frac{\sin(\omega_{mk}+\omega)t/2}{(\omega_{mk} + \omega)} + e^{i(\omega_{mk}-\omega)t/2} \frac{\sin(\omega_{mk}-\omega)t/2}{(\omega_{mk} - \omega)} \right] \quad (4.76)$$

It can be easily seen that for large values of t , the function

$$\frac{\sin(\omega_{mk}-\omega)t/2}{(\omega_{mk} - \omega)}$$

is very sharply peaked around $\omega \approx \omega_{mk}$ and negligible everywhere else (see Fig. 4.11). Thus for states for which ω_{mk} is significantly different from ω , $C_m(t)$ would be negligible and transitions between such states will not be stimulated by the incident field. This justifies our earlier statement that the presence of only those excited states be considered which are close to the resonance frequency.

In an emission process, $\omega_k > \omega_m$ and hence ω_{mk} is negative; thus it is the first term on the right-hand side of Eq. (4.76) which contributes. On the other hand, in an absorption process, $\omega_{mk} > 0$ and the second term in Eq. (4.76) contributes.

We consider absorption of radiation and assume that at $t = 0$ the atom is in state 1, the corresponding wave function being $\psi_1(\vec{r})$. We also assume ω to be close to $\omega_{21} [= (E_2 - E_1)/\hbar]$ – see Fig. 4.1. The probability for the transition to occur to state 2 is given by

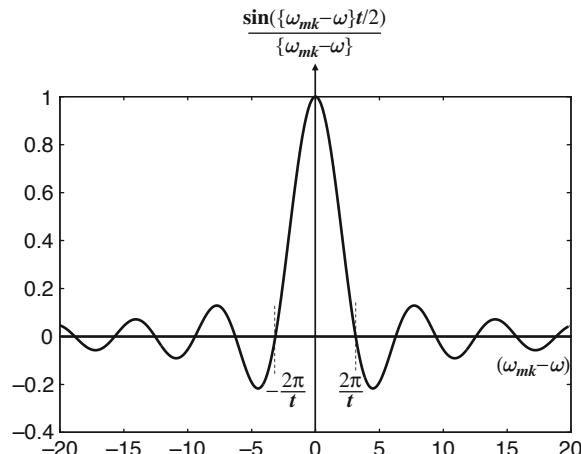


Fig. 4.11 For large values of t , the function $\frac{\sin(\omega_{mk}-\omega)t/2}{(\omega_{mk}-\omega)}$ is very sharply peaked around $\omega = \omega_{mk}$

$$|C_2(t)|^2 = \frac{1}{4} \frac{D_{21}^2 E_0^2}{\hbar^2} \left\{ \frac{\sin[(\omega_{21} - \omega)/2]t}{(\omega_{21} - \omega)/2} \right\}^2 \quad (4.77)$$

The above expression represents the probability for stimulated absorption of radiation. In deriving Eq. (4.77) we have assumed that $|C_2(t)|^2 \ll 1$; thus the result is accurate when

$$\frac{D_{21}^2 E_0^2 t^2}{\hbar^2} \ll 1 \quad \text{or} \quad \frac{\left(\frac{D_{21}^2 E_0^2}{\hbar^2} \right)}{(\omega_{21} - \omega)^2} \ll 1 \quad (4.78)$$

A more exact result for a two-state system will be discussed in Section 4.8.

We next assume that the quantity $(\omega_{21} - \omega)$ has a range of values either on account of the field having a continuous spectrum or the atom is capable of interaction with radiation having a range of frequencies.

4.7.1 Interaction with Radiation Having a Broad Spectrum

We first consider the field having a continuous spectrum characterized by $u(\omega)$ which is defined such that $u(\omega) d\omega$ represents the energy associated with the field per unit volume within the frequency interval ω and $\omega + d\omega$. Since the average energy density associated with an electromagnetic wave is $(1/2)\epsilon_0 E_0^2$ where ϵ_0 is the permittivity of free space, we replace E_0^2 in Eq. (4.77) by $(2/\epsilon_0) u(\omega) d\omega$ and integrate over all frequencies, which gives us the following expression for the transition probability:

$$\Gamma_{12} = \frac{1}{2\epsilon_0} \frac{D_{21}^2}{\hbar^2} \int u(\omega) \left\{ \frac{\sin[(\omega_{21} - \omega)/2]t}{(\omega_{21} - \omega)/2} \right\}^2 d\omega \quad (4.79)$$

Assuming that $u(\omega)$ varies very slowly in comparison to the quantity inside the square brackets, we replace $u(\omega)$ by its value at $\omega = \omega_{21}$ and take it out of the integral to obtain

$$\begin{aligned} \Gamma_{12} &\approx \frac{1}{2\epsilon_0} \frac{D_{21}^2}{\hbar^2} u(\omega_{21}) \left(\int \frac{\sin^2 \xi}{\xi^2} d\xi \right) 2t \\ &= \frac{\pi}{\epsilon_0} \frac{D_{21}^2}{\hbar^2} u(\omega_{21}) t \end{aligned} \quad (4.80)$$

where $\xi = \frac{\omega_{21} - \omega}{2}t$. The above expression shows that⁴ the probability of transition is proportional to time; thus the probability per unit time (which we denote by w_{12})

⁴It may be noted that Eq. (4.80) predicts an indefinite increase in the transition probability with time; however, the first-order perturbation theory itself breaks down when Γ_{21} is not appreciably less than unity. Thus Eq. (4.80) gives correct results as long as $\Gamma_{21} \ll 1$.

would be given by

$$w_{12} \approx \frac{\pi}{\epsilon_0} \frac{D_{21}^2}{\hbar^2} u(\omega_{21}) \quad (4.81)$$

Now (omitting the subscripts) we have

$$\mathbf{D} = \hat{\mathbf{e}} \cdot \mathbf{P} = P \cos \theta \quad (4.82)$$

where θ is the angle that $\hat{\mathbf{e}}$ (i.e., the electric field) makes with the dipole moment vector \mathbf{P} . Assuming that the dipole moment vector is randomly oriented, the average value of D^2 is given by

$$\overline{D^2} = P^2 \langle \cos^2 \theta \rangle = \frac{1}{3} P^2 \quad (4.83)$$

where use has been made of the following relation:

$$\langle \cos^2 \theta \rangle = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \cos^2 \theta \sin \theta \, d\theta \, d\phi = \frac{1}{3} \quad (4.84)$$

Thus

$$w_{12} = \frac{\pi}{3\epsilon_0} \frac{P^2}{\hbar^2} u(\omega_{21}) \quad (4.85)$$

If there are N_1 atoms per unit volume in state 1 then the total number of absorptions per unit time per unit volume would be $N_1 w_{12}$, which would be equal to

$$N_1 \frac{\pi}{3\epsilon_0} \frac{P^2}{\hbar^2} u(\omega_{21}) \quad (4.86)$$

Comparing Eqs. (4.86) and (4.1), we obtain

$$B_{12} = \frac{\pi}{3\epsilon_0} \frac{P^2}{\hbar^2} = \frac{4\pi^2}{3\hbar^2} \left(\frac{e^2}{4\pi\epsilon_0} \right) \left| \int \psi_2^* \mathbf{r} \psi_1 d\tau \right|^2 \quad (4.87)$$

The corresponding expression for stimulated emission is obtained by starting with the first term on the right-hand side of Eq. (4.76) and proceeding in a similar fashion. The final expression is identical to Eq. (4.87) except for an interchange of indices 1 and 2.

Using Eq. (4.87) we get the following expression for the A coefficient

$$A = \frac{4}{3} \left(\frac{e^2}{4\pi\epsilon_0} \frac{1}{\hbar c} \right) \frac{\omega^3}{c^2} \left| \int \psi_2^* \mathbf{r} \psi_1 d\tau \right|^2 \quad (4.88)$$

It may be of interest to note that

$$\left(\frac{e^2}{4\pi\epsilon_0} \frac{1}{hc} \right) \approx \frac{1}{137} \quad (4.89)$$

Using this value, we obtain

$$A = \frac{4}{3} \frac{1}{137} \frac{\omega^3}{c^2} \left| \int \psi_2^* \mathbf{r} \psi_1 d\tau \right|^2 \quad (4.90)$$

As an example we calculate the A coefficient for the $2P \rightarrow 1S$ transition in the hydrogen atom, i.e., the transition from the $(n = 2, l = 1, m = 0)$ state to the $(n = 1, l = 0, m = 0)$ state. For these states (see, e.g., Ghatak and Lokanathan (2004))

$$\psi_1 = \frac{1}{(4\pi)^{1/2}} \frac{2}{a_0^{3/2}} \exp\left(-\frac{r}{a_0}\right) \quad (4.91)$$

and

$$\psi_2 = \frac{1}{(2a_0)^{3/2}} \frac{r}{a_0 \sqrt{3}} \exp\left(-\frac{r}{2a_0}\right) \left[\left(\frac{3}{4\pi}\right)^{1/2} \cos\theta \right] \quad (4.92)$$

where $a_0 = (\hbar^2/m)(4\pi\epsilon_0/e^2) \approx 0.5 \times 10^{-10}$ m. In order to evaluate the matrix element, we write

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \sin\theta \sin\phi \\ z &= r \cos\theta \end{aligned} \quad (4.93)$$

Now,

$$\begin{aligned} \int \psi_1^* x \psi_2 d\tau &= \frac{1}{4\pi\sqrt{2}} \frac{1}{a_0^4} \left(\int_0^\infty r^2 dr e^{-3r/2a_0} r^2 \right) \times \left(\int_0^\pi \cos\theta \sin^2\theta d\theta \right) \\ &\times \left(\int_0^{2\pi} \cos\phi d\phi \right) = 0 \end{aligned}$$

because the integral over ϕ vanishes. Similarly

$$\int \psi_1^* y \psi_2 d\tau = 0 \quad (4.94)$$

The only non-vanishing integral is

$$\begin{aligned} \int \psi_1^* z \psi_2 d\tau &= \frac{1}{4\pi\sqrt{2}} \frac{1}{a_0^4} \left(\int_0^\infty r^2 dr e^{-3r/2a_0} r^2 \right) \\ &\times \left(\int_0^\pi \cos^2 \theta \sin \theta d\theta \right) \times \left(\int_0^{2\pi} d\phi \right) = 4\sqrt{2} \left(\frac{2}{3} \right)^5 a_0 \end{aligned}$$

Thus⁵

$$\left| \int \psi_1^* \mathbf{r} \psi_2 d\tau \right|^2 = 2^5 \left(\frac{2}{3} \right)^{10} a_0^2 \quad (4.95)$$

Further for the 2P → 1S transition

$$\omega = \frac{1}{\hbar} \frac{3}{8a_0} \left(\frac{e^2}{4\pi\epsilon_0} \right) = \frac{3c}{8a_0} \left(\frac{e^2}{4\pi\epsilon_0} \frac{1}{\hbar c} \right) \approx \frac{3 \times 3 \times 10^8}{8 \times 0.51 \times 10^{-10}} \frac{1}{137} \approx 1.5 \times 10^{16} \text{s}^{-1} \quad (4.96)$$

Substituting in Eq. (4.90), we obtain

$$\begin{aligned} A &\approx \frac{4}{3} \frac{1}{137} \frac{(1.5 \times 10^{16})^3}{(3 \times 10^8)^2} 2^5 \left(\frac{2}{3} \right)^{10} (0.5 \times 10^{-10})^2 \\ &= 6 \times 10^8 \text{s}^{-1} \end{aligned} \quad (4.97)$$

The mean lifetime of the state, τ , is the inverse of A giving

$$\tau \approx 1.6 \times 10^{-9} \text{s}$$

Thus the lifetime of the hydrogen atom in the upper level corresponding to the 2P → 1S transition is about 1.6 ns. Transitions having such small lifetimes are referred to as strongly allowed transitions.

In contrast, the levels used in laser transitions are such that the upper laser level has a very long lifetime ($\sim 10^{-3}$ – 10^{-6} s). A level having such a long lifetime is referred to as a metastable level, and such transitions come under the class of weakly allowed or nearly forbidden transitions. The strength of an atomic transition is usually expressed in terms of the f -value defined by the following equation:

$$f_{21} = \frac{2}{3} \frac{m\omega_{21}}{\hbar} |D_{21}|^2 \quad (4.98)$$

⁵ It can be shown that $|\int \psi_1^* \mathbf{r} \psi_2 d\tau|^2$ has the same value for transition from anyone of the states $(n = 2, l = 1, m = 0)$ or $(n = 2, l = 1, m = -1)$ or $(n = 2, l = 1, m = -1)$ to $(n = 2, l = 0, m = 0)$ state. However, the matrix element for the transition from $(n = 2, l = 0, m = 0)$ state to the $(n = 1, l = 0, m = 0)$ state is zero. This implies that the corresponding dipole transition is forbidden.

For strongly allowed transitions, f is of the order of unity, for example, for the $2P \rightarrow 1S$ transition in the hydrogen atom, $f = 0.416$. On the other hand, for the transitions from the upper laser level, $f \sim 10^{-3}$ – 10^{-6} .

4.7.2 Interaction of a Near-Monochromatic Wave with an Atom Having a Broad Frequency Response

We next consider a nearly monochromatic field interacting with atoms characterized by the lineshape function $g(\omega)$. For such a case the probability for the atom being in the upper state would be given by

$$\begin{aligned}\Gamma_{12} &= \frac{1}{4} \frac{D_{21}^2 E_0^2}{\hbar^2} \int g(\omega') \left\{ \frac{\sin [(\omega' - \omega)/2] t}{(\omega' - \omega)/2} \right\}^2 d\omega' \\ &= \frac{1}{4} \frac{D_{21}^2 E_0^2}{\hbar^2} g(\omega) 2\pi t \\ &= \frac{\pi P^2}{3\hbar^2 \varepsilon_0} g(\omega) u_\omega t\end{aligned}\quad (4.99)$$

where in the last step we have replaced D_{21}^2 and E_0^2 by $\frac{1}{3}P^2$ and $2u_\omega/\varepsilon_0$, respectively. Since

$$B_{12} = B_{21} = \frac{\pi}{3\varepsilon_0} \frac{P^2}{\hbar^2} = \frac{\pi^2 c^3}{\hbar \omega^3 t_{sp}} \quad (4.100)$$

we obtain the following expression for the transition rate (per unit time) per unit volume:

$$W_{12} = N_1 \frac{\pi^2 c^3}{\hbar \omega^3 t_{sp}} u_\omega g(\omega) \quad (4.101)$$

which is consistent with Eq. (4.18).

4.8 More Accurate Solution for the Two-Level System

A more accurate solution of the time-dependent Schrödinger equation can be obtained if we assume that the atom can exist in only two possible states characterized by $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$. Thus Eq. (4.67) gets replaced by

$$\Psi(\mathbf{r}, t) = C_1(t)\psi_1(\mathbf{r})e^{-i\omega_1 t} + C_2(t)\psi_2(\mathbf{r})e^{-i\omega_2 t} \quad (4.102)$$

If we substitute from Eq. (4.102) into Eq. (4.61), multiply by ψ_1^* and integrate, we would get [cf. Eq. (4.74)]

$$i\hbar \frac{dC_1}{dt} = \frac{1}{2} E_0 D_{12} C_2(t) \left(e^{-i(\omega' - \omega)t} + e^{-i(\omega' + \omega)t} \right) \quad (4.103)$$

Similarly

$$i\hbar \frac{dC_2}{dt} = \frac{1}{2} E_0 D_{12} C_1(t) \left(e^{i(\omega'+\omega)t} + e^{i(\omega'-\omega)t} \right) \quad (4.104)$$

where use has been made of the fact that

$$\int \psi_1^* \mathbf{r} \psi_1 d\tau = \int \psi_2^* \mathbf{r} \psi_2 d\tau = 0$$

and

$$\omega' = \frac{(E_2 - E_1)}{\hbar} = \omega_{21} = -\omega_{12} \quad (4.105)$$

In the rotating wave approximation, considering absorption we neglect the terms $e^{-i(\omega'+\omega)t}$ and $e^{i(\omega'+\omega)t}$ in Eqs. (4.103) and (4.104) and obtain

$$\frac{dC_1}{dt} = -\frac{i}{2\hbar} E_0 D_{12} C_2(t) e^{i(\omega-\omega')t} \quad (4.106)$$

$$\frac{dC_2}{dt} = -\frac{i}{2\hbar} E_0 D_{12} C_1(t) e^{-i(\omega-\omega')t} \quad (4.107)$$

If we assume a solution of the form

$$C_1(t) = e^{i\Omega t} \quad (4.108)$$

then from Eq. (4.107),

$$C_2(t) = -\frac{2\hbar\Omega}{E_0 D_{12}} e^{i(\Omega-\omega+\omega')t} \quad (4.109)$$

Substituting in Eq. (4.106), we get

$$-\frac{2\hbar\Omega}{E_0 D_{12}} (\Omega - \omega + \omega') = -\frac{i}{2\hbar} E_0 D_{21}$$

or

$$\Omega (\Omega - \omega + \omega') - \frac{\Omega_0^2}{4} = 0 \quad (4.110)$$

where

$$\Omega_0^2 = \frac{E_0^2 D_{12} D_{21}}{\hbar^2} = \frac{E_0^2 D^2}{\hbar^2} \quad (4.111)$$

and

$$D = D_{12} = D_{21} \quad (4.112)$$

Equation (4.110) gives

$$\Omega_{1,2} = \frac{1}{2} \left\{ -(\omega' - \omega) \pm \left[(\omega' - \omega)^2 + \Omega_0^2 \right]^{1/2} \right\} \quad (4.113)$$

Thus the general solution will be

$$C_1(t) = A_1 e^{i\Omega_1 t} + A_2 e^{i\Omega_2 t} \quad (4.114)$$

$$C_2(t) = -\frac{2}{\Omega_0} e^{i(\omega' - \omega)t} (A_1 \Omega_1 e^{i\Omega_1 t} + A_2 \Omega_2 e^{i\Omega_2 t}) \quad (4.115)$$

If we now assume that the atom is initially in the ground state, i.e.,

$$C_1(0) = 1, \quad C_2(0) = 0 \quad (4.116)$$

then

$$A_1 = -\frac{\Omega_2}{\Omega_1} A_2 \quad (4.117)$$

and

$$1 = A_1 + A_2 = A_2 \frac{\Omega_1 - \Omega_2}{\Omega_1} = \left[(\omega' - \omega)^2 + \Omega_0^2 \right]^{1/2} \frac{A_2}{\Omega_1}$$

or

$$A_2 = \frac{\Omega_1}{\Omega'} \quad (4.118)$$

where

$$\Omega' = \left[(\omega' - \omega)^2 + \Omega_0^2 \right]^{1/2} \quad (4.119)$$

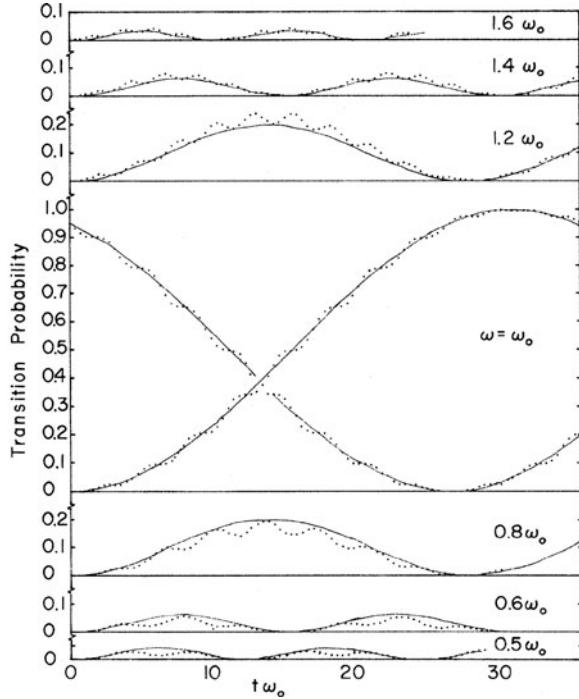
On substitution we finally obtain

$$C_2(t) = -i \frac{\Omega_0}{\Omega'} e^{i(\omega' - \omega)t/2} \sin \left(\frac{\Omega' t}{2} \right) \quad (4.120)$$

Thus the transition probability for absorption is given by

$$|C_2(t)|^2 = \left(\frac{\sin(\Omega' t/2)}{\Omega'/2} \right)^2 \left(\frac{\Omega_0}{2} \right)^2 \quad (4.121)$$

Fig. 4.12 Variation of the transition probability with time for a two-level system for different frequencies of the electromagnetic field. The curves correspond to the function $DE_0/\hbar = 0.1\omega'$. The solid line corresponds to Eq. (4.121) and the dotted curve corresponds to an accurate numerical computation (Reprinted with permission from Salzman (1971). © 1971 American Institute of Physics)



which has been plotted in Fig. 4.12. Also shown in the figure are the results of the exact numerical calculations without resorting to the rotating wave approximation. At resonance $\omega = \omega'$ and one obtains

$$|C_2(t)|^2 = \sin^2\left(\frac{\Omega_0 t}{2}\right) \quad (4.122)$$

which shows that the system flip flops between states 1 and 2. A comparison of Eqs. (4.122) and (4.77) shows that the perturbation theory result is valid if

$$\left(\frac{D_{21}E_0}{\hbar}t\right)^2 \ll 1 \quad \text{or} \quad \left(\frac{D_{21}E_0}{\hbar}\right)^2 \frac{1}{(\omega' - \omega)^2} \ll 1 \quad (4.123)$$

It may be of interest to note that the solutions obtained in this section are exact when $\omega = 0$ (i.e., a constant electric field) and if D_{21} is replaced by $2D_{21}$ in the solution given by Eq. (4.121). This follows from the fact that for $\omega = 0$, the exact equations [Eqs. (4.103) and (4.104)] are the same as Eqs. (4.106) and (4.107) with D_{21} replaced by $2D_{21}$.

Problems

Problem 4.3 Consider the two-level system shown in Fig. 4.1 with $E_1 = -13.6 \text{ eV}$ and $E_2 = -3.4 \text{ eV}$. Assume $A_{21} \approx 6 \times 10^8 \text{ s}^{-1}$. (a) What is the frequency of light emitted due to transitions from E_2 and E_1 ? Assuming the emission to have only natural broadening, what is the FWHM of the emission? What is the population ratio N_2/N_1 at $T = 300 \text{ K}$?

$$[\text{Answer : (a)} v \approx 2.5 \times 10^{15} \text{ Hz}, \Delta v = A_{21}/2\pi \cong 10^8 \text{ Hz}, N_2/N_1 \approx e^{-394}]$$

Problem 4.4 Given that the gain coefficient in a Doppler-broadened line is

$$\alpha(v) = \alpha(v_0) \exp \left[-4 \ln 2 (v - v_0)^2 / (\Delta v_0)^2 \right]$$

where v_0 is the centre frequency and Δv_0 is the FWHM and that the gain coefficient at the line centre is twice the loss averaged per unit length, calculate the bandwidth over which oscillation can take place. [Answer: Δv_0].

Problem 4.5 Consider an atomic system as shown below:

$$\begin{array}{ll} 3 & E_3 = 3 \text{ eV} \\ 2 & E_2 = 1 \text{ eV} \\ 1 & E_1 = 0 \text{ eV} \end{array}$$

The A coefficient of the various transitions are given by

$$A_{32} = 7 \times 10^7 \text{ s}^{-1}, \quad A_{31} = 10^7 \text{ s}^{-1}, \quad A_{21} = 10^8 \text{ s}^{-1}$$

- (a) What is the spontaneous lifetime of level 3?
- (b) If the steady-state population of level 3 is 10^{15} atoms/cm³, what is the power emitted spontaneously in the $3 \rightarrow 2$ transition? [Answer: (a) $t_{sp} = 1.2 \times 10^{-8} \text{ s}$ (b) $2.2 \times 10^{10} \text{ W/m}^3$]

Problem 4.6 Consider the transition in neon that emits 632.8 nm in the He–Ne laser and assume a temperature of 300 K. For a collision time of 500 ns, and a lifetime of 30 ns, obtain the broadening due to collisions, lifetime, and Doppler and show that the Doppler broadening is the dominant mechanism.

Problem 4.7 Consider an atomic system under thermal equilibrium at $T = 1000 \text{ K}$. The number of absorptions per unit time corresponding to a wavelength of 1 μm is found to be 10^{22} s^{-1} . What would be the number of stimulated emissions per unit time between the two energy levels? [Ans: 10^{22} s^{-1}]

Problem 4.8 Consider a laser with plane mirrors having reflectivities of 0.9 each and of length 50 cm filled with the gain medium. Neglecting scattering and other cavity losses, estimate the threshold gain coefficient (in m^{-1}) required to start laser oscillation. [Ans: 0.21 m^{-1}]

Problem 4.9 An atomic transition has a linewidth of $\Delta v = 10^8 \text{ Hz}$. Estimate the approximate value of $g(\omega)$ at the center of the line. [Ans: $\sim 1.6 \times 10^{-95} \text{ s}^{-1}$]

Problem 4.10 There is a 10% loss per round trip in a ruby laser resonator having a 10 cm long ruby crystal as the active medium. Calculate the cavity lifetime, assuming that the mirrors are coated on the ends of the ruby crystal. Given: Refractive index of ruby at the laser wavelength is 1.78 [Ans: 11.3 ns]

Problem 4.11 In a ruby crystal, a population inversion density of $(N_2 - N_1) = 5 \times 10^{17} \text{ cm}^{-3}$ is generated by pumping. Assuming $g(v_0) = 5 \times 10^{-12} \text{ s}$, $t_{sp} = 3 \times 10^{-3} \text{ s}$, wavelength of 694.3 nm and a refractive index of 1.78, obtain the gain coefficient $\gamma(v_0)$. By what factor will a beam get amplified if it passes through 5 cm of such a crystal? [Ans: $5 \times 10^{-2} \text{ cm}^{-1}$, 1.28]

Problem 4.12 An optical amplifier of length 10 cm amplifies an input power of 1 to 1.1 W. Calculate the gain coefficient in m^{-1} . [Ans: 0.95 m^{-1}]

Problem 4.13 Doppler broadening leads to a linewidth given by

$$\Delta\nu_D = 2v_0 \sqrt{\frac{2k_B T}{Mc^2} \ln 2}$$

Estimate the broadening for the 632.8 nm transition of Ne (used in the He–Ne laser) assuming $T = 300 \text{ K}$ and atomic mass of Ne to be 20. What would be the corresponding linewidth of the $10.6 \mu\text{m}$ transition of the CO₂ molecule? [Ans: $1.6 \times 10^9 \text{ Hz}$, $6 \times 10^7 \text{ Hz}$]

Problem 4.14 In a typical He–Ne laser the threshold population inversion density is 10^9 cm^{-3} . What is the value of the population inversion density when the laser is oscillating in steady state with an output power of 2 mW?

Problem 4.15 Given that the gain coefficient in a Doppler-broadened line is

$$\gamma(\nu) = \gamma_0 \exp \left[-\frac{4 \ln 2(\nu - \nu_0)^2}{(\Delta\nu_0)^2} \right]$$

and that the gain coefficient at the center of the line is four times the loss averaged per unit length, obtain the bandwidth over which oscillation will take place. [Ans: $\sqrt{2}\Delta\nu_0$]

Problem 4.16 A laser resonator 1 m long is filled with a medium having a gain coefficient of 0.02 m^{-1} . If one of the mirrors is 100 % reflecting, what should be the minimum reflectivity of the other mirror so that the laser may oscillate? [Ans: $\sim 96\%$]

Chapter 5

Laser Rate Equations

5.1 Introduction

In [Chapter 4](#) we studied the interaction of radiation with matter and found that under the action of radiation of proper frequencies, the atomic populations of various energy levels change. In this chapter, we will be studying the rate equations which govern the rate at which populations of various energy levels change under the action of the pump and in the presence of laser radiation. The rate equations approach provides a convenient means of studying the time dependence of the atomic populations of various levels in the presence of radiation at frequencies corresponding to the different transitions of the atom. It also gives the steady-state population difference between the actual levels involved in the laser transition and allows one to study whether an inversion of population is achievable in a transition and, if so, what would be the minimum pumping rate required to maintain a steady population inversion between two levels, the gain that such a medium would provide at and near the transition frequency, and the phase shift effects that such a medium would introduce are discussed in detail in [Chapter 6](#). Thus [Chapter 6](#) discusses the behavior of a system having two levels when there is a population inversion between the two levels, and this chapter deals with the means of obtaining an inversion between two levels of an atomic system by making use of other energy levels. The rate equations can also be solved to obtain the transient behavior of the laser, which gives rise to phenomena like Q-switching and spiking.

The atomic rate equations along with the rate equation for the photon number in the cavity form a set of coupled nonlinear equations. These equations can be solved under the steady-state regime and one can study the evolution of the photon number as one passes through the threshold pumping region.

In Section [5.2](#) we discuss a two-level system and show that it is not possible to achieve population inversion in steady state in a two-level system. Sections [5.3](#) and [5.4](#) discuss three-level and four-level laser systems and obtain the dependence of inversion on the pump power. In Section [5.5](#) we obtain the variation of laser power around threshold showing the sudden increase in the output power as a function of pumping. This is a very characteristic behavior of a laser. Finally in Section [5.6](#) we discuss the optimum output coupling for maximizing the output power of a laser.

5.2 The Two-Level System

We first consider a two-level system consisting of energy levels E_1 and E_2 with N_1 and N_2 atoms per unit volume, respectively [see (Fig. 5.1)]. Let radiation at frequency ω with energy density u be incident on the system. The number of atoms per unit volume which absorbs the radiation and is excited to the upper level will be [see Eq. (4.18)]

$$\Gamma_{12} = \frac{\pi^2 c^3}{\hbar \omega^3 t_{sp} n_0^3} u g(\omega) N_1 = W_{12} N_1 \quad (5.1)$$

where

$$W_{12} = \frac{\pi^2 c^3}{\hbar \omega^3 t_{sp} n_0^3} u g(\omega) \quad (5.2)$$

The number of atoms undergoing stimulated emissions from E_2 to E_1 per unit volume per unit time will be [see Eqs. (4.16) and (4.18)]

$$\Gamma_{21} = W_{21} N_2 = W_{12} N_1 \quad (5.3)$$

where we have used the fact that the absorption probability is the same as the stimulated emission probability. In addition to the above two transitions, atoms in the level E_2 would also undergo spontaneous transitions from E_2 to E_1 . If A_{21} and S_{21} represent the radiative and non-radiative transition¹ rates from E_2 to E_1 , then the number of atoms undergoing spontaneous transitions per unit time per unit volume from E_2 to E_1 will be $T_{21} N_2$ where

$$T_{21} = A_{21} + S_{21} \quad (5.4)$$

Thus we may write the rate of change of population of energy levels E_2 and E_1 as

$$\frac{dN_2}{dt} = W_{12}(N_1 - N_2) - T_{21}N_2 \quad (5.5)$$

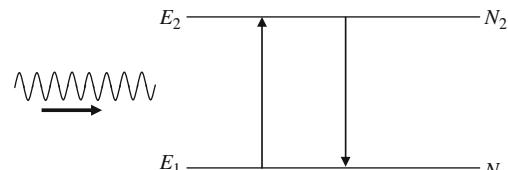


Fig. 5.1 A two-level system

¹In a non-radiative transitions when the atom de-excites, the energy is transferred to the translational, vibrational or rotational energies of the surrounding atoms or molecules.

$$\frac{dN_1}{dt} = -W_{12}(N_1 - N_2) + T_{21}N_2 \quad (5.6)$$

As can be seen from Eqs. (5.5) and (5.6)

$$\begin{aligned} \frac{d}{dt}(N_1 + N_2) &= 0 \\ \Rightarrow N_1 + N_2 &= \text{a constant} = N \quad (\text{say}) \end{aligned} \quad (5.7)$$

which is nothing but the fact that the total number of atoms N per unit volume is constant. At steady state

$$\frac{dN_1}{dt} = 0 = \frac{dN_2}{dt} \quad (5.8)$$

which gives us

$$\frac{N_2}{N_1} = \frac{W_{12}}{W_{12} + T_{21}} \quad (5.9)$$

Since both W_{12} and T_{21} are positive quantities, Eq. (5.9) shows us that we can never obtain a steady-state population inversion by optical pumping between just two levels.

Let us now have a look at the population difference between the two levels. From Eq. (5.9) we have

$$\frac{N_2 - N_1}{N_2 + N_1} = -\frac{T_{21}}{2W_{12} + T_{21}}$$

or if we write $\Delta N = N_2 - N_1$, we have

$$\frac{\Delta N}{N} = -\frac{1}{1 + 2W_{12}/T_{21}} \quad (5.10)$$

In order to put Eq. (5.10) in a slightly different form, we first assume that the transition from 2 to 1 is mostly radiative, i.e., $A_{21} \gg S_{21}$ and $T_{21} \approx A_{21}$. We also introduce a lineshape function $\tilde{g}(\omega)$ which is normalized to have unit value at $\omega = \omega_0$, the center of the line, i.e.,

$$\tilde{g}(\omega) = \frac{g(\omega)}{g(\omega_0)} \quad (5.11)$$

Since $g(\omega) \leq g(\omega_0)$ for all ω , we have $0 < \tilde{g}(\omega) < 1$. Substituting the value of W_{12} in terms of u from Eq. (5.2) and observing that $u = n_0 I/c$, where I is the intensity of the incident radiation at ω , we have

$$\begin{aligned} \frac{W_{12}}{T_{21}} &= \frac{\pi^2 c^3}{\hbar \omega^3 t_{sp} n_0^3} I \frac{n_0}{c} \tilde{g}(\omega) g(\omega_0) \frac{1}{A_{21}} \\ &= \frac{\pi^2 c^3}{\hbar \omega^3 n_0^2} g(\omega_0) \tilde{g}(\omega) I \end{aligned} \quad (5.12)$$

where we have used the fact that $A_{21}t_{\text{sp}} = 1$. Hence Eq. (5.10) becomes

$$\frac{\Delta N}{N} = -\frac{1}{1 + (I/I_s)\tilde{g}(\omega)} \quad (5.13)$$

where

$$I_s \equiv \frac{\hbar\omega^3 n_0^2}{2\pi^2 c^2 g(\omega_0)} \quad (5.14)$$

is called the saturation intensity. In order to see what I_s represents let us consider a monochromatic wave at frequency ω_0 interacting with a two-level system. Since $\tilde{g}(\omega_0) = 1$, we see from Eq. (5.13) that for $I \ll I_s$, the density of population difference between the two levels ΔN is almost independent of the intensity of the incident radiation. On the other hand for I comparable to I_s , ΔN becomes a function of I and indeed for $I = I_s$, the value of ΔN is half the value at low incident intensities.

We showed in Section 4.3 that the loss/gain coefficient for a population difference $\Delta N = N_2 - N_1$ between two levels is given by [see Eq. (4.26)]

$$\begin{aligned} \alpha &= -\frac{\pi^2 c^2}{\omega^2 t_{\text{sp}} n_0^2} g(\omega) \Delta N \\ &= \frac{\alpha_0}{1 + (I/I_s)\tilde{g}(\omega)} \end{aligned} \quad (5.15)$$

where

$$\alpha_0 = \frac{\pi^2 c^2}{\omega^2 t_{\text{sp}} n_0^2} g(\omega) N \quad (5.16)$$

corresponds to the small signal loss, i.e., the loss coefficient when $I \ll I_s$. We can see from Problem 5.1 that with α given by Eq. (5.15), the loss is exponential for $I \ll I_s$ while it becomes linear for $I \gg I_s$. Thus we see that the attenuation caused by a medium decreases as the incident intensity increases to values comparable to the saturation intensity. Organic dyes having reasonably low values of I_s ($\sim 5 \text{ MW/cm}^2$) are used as saturable absorbers in mode locking and Q-switching of lasers (see Section 7.7.1).

Problem 5.1 Using Eq. (5.15) in Eq. (4.25) obtain the variation of I with z .

[Answer:

$$\ln \frac{I}{I_0} + \frac{\tilde{g}(\omega)}{I_s} (I - I_0) = -\alpha_0 z$$

where I_0 is the intensity at $z=0$.]

5.3 The Three-Level Laser System

In the last section we saw that one cannot create a steady-state population inversion between two levels just by using pumping between these levels. Thus in order to produce a steady-state population inversion, one makes use of either a three-level or a four-level system. In this section we shall discuss a three-level system.

We consider a three-level system consisting of energy levels E_1 , E_2 , and E_3 all of which are assumed to be nondegenerate. Let N_1 , N_2 , and N_3 represent the population densities of the three levels [see (Fig. 5.2)]. The pump is assumed to lift atoms from level 1 to level 3 from which they decay rapidly to level 2 through some nonradiative process. Thus the pump effectively transfers atoms from the ground level 1 to the excited level 2 which is now the upper laser level; the lower laser level being the ground state 1. If the relaxation from level 3 to level 2 is very fast, then the atoms will relax down to level 2 rather than to level 1. Since the upper level 3 is not a laser level, it can be a broad level (or a group of broad levels) so that a broadband light source may be efficiently used as a pump source (see, e.g., the ruby laser discussed in Chapter 11).

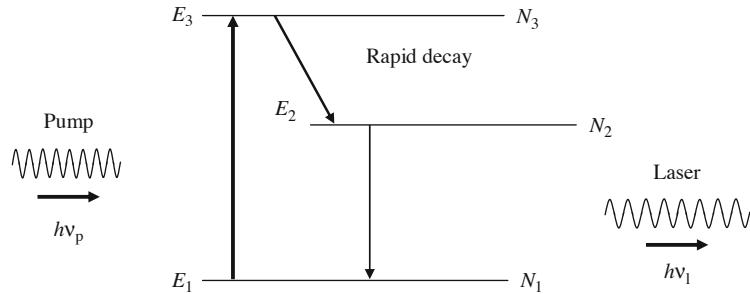


Fig. 5.2 A three-level system. The pump excites the atoms from level E_1 to level E_3 from where the atoms undergo a fast decay to level E_2 . The laser action takes place between levels E_2 and E_1

If we assume that transitions take place only between these three levels then we may write

$$N = N_1 + N_2 + N_3 \quad (5.17)$$

where N represents the total number of atoms per unit volume.

We may now write the rate equations describing the rate of change of N_1 , N_2 and N_3 . For example, the rate of change of N_3 may be written as

$$\frac{dN_3}{dt} = W_p(N_1 - N_3) - T_{32}N_3 \quad (5.18)$$

where W_p is the rate of pumping per atom from level 1 to level 3 which depends on the pump intensity. The first term in Eq. (5.18) represents stimulated transitions

between levels 1 and 3 and $T_{32}N_3$ represents the spontaneous transition from level 3 to level 2:

$$T_{32} = A_{32} + S_{32} \quad (5.19)$$

A_{32} and S_{32} correspond, respectively, to the radiative and nonradiative transition rates between levels 3 and 2. In writing Eq. (5.18) we have neglected $T_{31}N_3$ which corresponds to spontaneous transitions between levels 3 and 1 since most atoms raised to level 3 are assumed to make transitions to level 2 rather than to level 1.

In a similar manner, we may write

$$\frac{dN_2}{dt} = W_1(N_1 - N_2) + N_3T_{32} - N_2T_{21} \quad (5.20)$$

and

$$\frac{dN_1}{dt} = W_p(N_3 - N_1) + W_1(N_2 - N_1) + N_2T_{21} \quad (5.21)$$

where

$$W_1 = \frac{\pi^2 c^2}{\hbar \omega^3 n_0^2} A_{21} g(\omega) I_1 \quad (5.22)$$

represents the stimulated transition rate per atom between levels 1 and 2, I_1 is the intensity of the radiation in the $2 \rightarrow 1$ transition and $g(\omega)$ represents the lineshape function describing the transitions between levels 1 and 2. Further,

$$T_{21} = A_{21} + S_{21} \quad (5.23)$$

with A_{21} and S_{21} representing the radiative and nonradiative relaxation rates between levels 1 and 2. For efficient laser action since the transition must be mostly radiative, we shall assume $A_{21} \gg S_{21}$.

At steady state we must have

$$\frac{dN_1}{dt} = 0 = \frac{dN_2}{dt} = \frac{dN_3}{dt} \quad (5.24)$$

From Eq. (5.18) we obtain

$$N_3 = \frac{W_p}{W_p + T_{32}} N_1 \quad (5.25)$$

Using Eqs. (5.20), (5.21), and (5.25) we get

$$N_2 = \frac{W_1(T_{32} + W_p) + W_p T_{32}}{(W_p + T_{32})(W_1 + T_{21})} N_1 \quad (5.26)$$

Thus from Eqs. (5.17), (5.25), and (5.26) we get

$$\frac{N_2 - N_1}{N} = \frac{[W_p(T_{32} - T_{21}) - T_{32}T_{21}]}{[3W_pW_1 + 2W_pT_{21} + 2T_{32}W_1 + T_{32}W_p + T_{32}T_{21}]} \quad (5.27)$$

From the above equation, one may see that in order to obtain population inversion between levels 2 and 1, i.e., for $(N_2 - N_1)$ to be positive, a necessary (but not sufficient) condition is that

$$T_{32} > T_{21} \quad (5.28)$$

Since the lifetimes of levels 3 and 2 are inversely proportional to the relaxation rates, according to Eq. (5.28), the lifetime of level 3 must be smaller than that of level 2 for attainment of population inversion between levels 1 and 2. If this condition is satisfied then according to Eq. (5.27), there is a minimum pumping rate required to achieve population inversion which is given by

$$W_{pt} = \frac{T_{32}T_{21}}{T_{32} - T_{21}} \quad (5.29)$$

If $T_{32} \gg T_{21}$,

$$W_{pt} \approx T_{21} \quad (5.30)$$

and under the same approximation, Eq. (5.27) becomes

$$\frac{N_2 - N_1}{N} = \frac{(W_p - T_{21})/(W_p + T_{21})}{\left[1 + \frac{3W_p + 2T_{32}}{T_{32}(W_p + T_{21})} W_1\right]} \quad (5.31)$$

Below the threshold for laser oscillation, W_1 is very small and hence we may write

$$\frac{N_2 - N_1}{N} = \frac{(W_p - T_{21})}{(W_p + T_{21})} \quad (5.32)$$

Thus when W_1 is small, i.e., when the intensity of the radiation corresponding to the laser transition is small [see Eq. (5.22)], then the population inversion is independent of I_1 and there is an exponential amplification of the beam. As the laser starts oscillating, W_1 becomes large and from Eq. (5.31) we see that this reduces the inversion $N_2 - N_1$ which in turn reduces the amplification. When the laser oscillates under steady-state conditions, the intensity of the radiation at the laser transition increases to such a value that the value of $N_2 - N_1$ is the same as the threshold value.

Recalling Eq. (5.31), we see that for a population inversion $N_2 - N_1$, the gain coefficient of the laser medium is

$$\begin{aligned} \gamma &= \frac{\pi^2 c^2}{\omega^2 t_{sp} n_0^2} g(\omega)(N_2 - N_1) \\ &= \frac{\gamma_0}{1 + \frac{3W_p + 2T_{32}}{T_{32}(W_p + T_{21})} W_1} \end{aligned} \quad (5.33)$$

where

$$\gamma_0 = \frac{\pi^2 c^2}{\omega^2 t_{\text{sp}} n_0^2} g(\omega) N \frac{W_p - T_{21}}{W_p + T_{21}} \quad (5.34)$$

is the small signal gain coefficient. If we now carry out a similar analysis to that in Section 5.2, we may write

$$\gamma = \frac{\gamma_0}{1 + (I/I_s)\tilde{g}(\omega)} \quad (5.35)$$

where

$$\begin{aligned} \tilde{g}(\omega) &= g(\omega)/g(\omega_0) \\ I_s &= \frac{\hbar\omega^3 n_0^2}{\pi^2 c^2 A_{21} g(\omega_0)} \frac{T_{32}(W_p + T_{21})}{(3W_p + 2T_{32})} \end{aligned} \quad (5.36)$$

I_s being the saturation intensity [see the discussion following Eq. (5.16)].

If T_{32} is very large then there will be very few atoms residing in level 3. Consequently, we may write

$$N = N_1 + N_2 + N_3 \approx N_1 + N_2 \quad (5.37)$$

Substituting in Eq. (5.32), we get

$$\frac{N_2 - N_1}{N_2 + N_1} = \frac{W_p - T_{21}}{W_p + T_{21}}$$

or

$$W_p N_1 = T_{21} N_2 \quad (5.38)$$

The left-hand side of the above equation represents the number of atoms being lifted (by the pump) per unit volume per unit time from level 1 to level 2 via level 3 and the right-hand side corresponds to the spontaneous emission rate per unit volume from level 2 to level 1. These rates must be equal under steady-state conditions for $W_1 \approx 0$, i.e., below the threshold.

We shall now estimate the threshold pumping power required to start laser oscillation. In order to do this, we first observe that the threshold inversion required is usually very small compared to N (i.e., $N_2 - N_1 \ll N$ – see the example of the ruby laser discussed in Chapter 11). Thus from Eq. (5.38), we see that the threshold value of W_p required to start laser oscillation is also approximately equal to T_{21} . Now the number of atoms being pumped per unit time per unit volume from level 1 to level 3 is $W_p N_1$. If v_p represents the average pump frequency corresponding to excitation to E_3 from E_1 , then the power required per unit volume will be

$$P = W_p N_1 h v_p \quad (5.39)$$

Thus the threshold pump power for laser oscillation is given by

$$P_t = T_{21} N_1 h v_p \quad (5.40)$$

Since $N_2 - N_1 \ll N$ and $N_3 \approx 0$, $N_1 \approx N_2 \approx N/2$. Also assuming the transition from level 2 to level 1 to be mainly radiative (i.e., $A_{21} \gg S_{21}$), we have

$$P_t \approx Nh\nu_p/2t_{sp} \quad (5.41)$$

where we have used $A_{21} = 1/t_{sp}$.

As an example, we consider the ruby laser for which we have the following values of the various parameters:

$$N \approx 1.6 \times 10^{19} \text{ cm}^{-3} \quad t_{sp} \approx 3 \times 10^{-3} \text{ s} \quad \nu_p \approx 6.25 \times 10^{14} \text{ Hz} \quad (5.42)$$

Substitution in Eq. (5.41) gives us

$$P_t \approx 1100 \text{ W/cm}^3 \quad (5.43)$$

If we assume that the efficiency of the pumping source to be 25% and also that only 25% of the pump light is absorbed on passage through the ruby rod, then the electrical threshold power comes out to be about 18 kW/cm^3 of the active medium. This is consistent with the threshold powers obtained experimentally.

Under pulsed operation if we assume that the pumping pulse is much shorter than the lifetime of level 2, then the atoms excited to the upper laser level do not appreciably decay during the duration of the pulse and the threshold pump energy would be

$$U_{pt} = \frac{N}{2}h\nu_p$$

per unit volume of the active medium. For the case of ruby laser, with the above efficiencies of pumping and absorption, one obtains

$$U_{pt} \approx 54 \text{ J/cm}^3$$

It may be noted here that even though ruby laser is a three-level laser system, because of various other factors mentioned below it does operate with not too large a pumping power. Thus, for example, the absorption band of ruby crystal is very well matched to the emission spectrum of available pump lamps so that the pumping efficiency is quite high. Also most of the atoms pumped to level 3 drop down to level 2 which has a very long lifetime which is nearly radiative. In addition the line width of laser transition is also very narrow.

5.4 The Four-Level Laser System

In the last section we found that since the lower laser was the ground level, one has to lift more than 50% of the atoms in the ground level in order to obtain population inversion. This problem can be overcome by using another level of the atomic system and having the lower laser level also as an excited level. The four-level laser

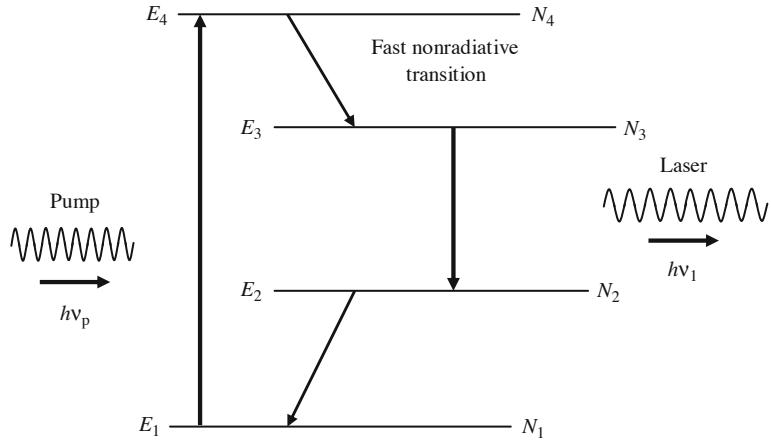


Fig. 5.3 A four-level system; the pump lifts atoms from level E_1 to level E_4 from where they decay rapidly to level E_3 and laser emission takes place between levels E_3 and E_2 . Atoms drop down from level E_2 to level E_1

system is shown in Fig 5.3. Level 1 is the ground level and levels 2, 3, and 4 are excited levels of the system. Atoms from level 1 are pumped to level 4 from where they make a fast nonradiative relaxation to level 3. Level 3 which corresponds to the upper laser level is usually a metastable level having a long lifetime. The transition from level 3 to level 2 forms the laser transition. In order that atoms do not accumulate in level 2 and hence destroy the population inversion between levels 3 and 2, level 2 must have a very small lifetime so that atoms from level 2 are quickly removed to level 1 ready for pumping to level 4. If the relaxation rate of atoms from level 2 to level 1 is faster than the rate of arrival of atoms to level 2 then one can obtain population inversion between levels 3 and 2 even for very small pump powers. Level 4 can be a collection of a large number of levels or a broad level. In such a case an optical pump source emitting over a broad range of frequencies can be used to pump atoms from level 1 to level 4 effectively. In addition, level 2 is required to be sufficiently above the ground level so that, at ordinary temperatures, level 2 is almost unpopulated. The population of level 2 can also be reduced by lowering the temperature of the system.

We shall now write the rate equations corresponding to the populations of the four levels. Let N_1, N_2, N_3 , and N_4 be the population densities of levels 1, 2, 3, and 4, respectively. The rate of change of N_4 can be written as

$$\frac{dN_4}{dt} = W_p(N_1 - N_4) - T_{43}N_4 \quad (5.44)$$

where, as before, W_pN_1 is the number of atoms being pumped per unit time per unit volume, W_pN_4 is the stimulated emission rate per unit volume,

$$T_{43} = A_{43} + S_{43} \quad (5.45)$$

is the relaxation rate from level 4 to level 3 and is the sum of the radiative (A_{43}) and nonradiative (S_{43}) rates. In writing Eq. (5.44) we have neglected (T_{42}) and (T_{41}) in comparison to (T_{43}), i.e., we have assumed that the atoms in level 4 relax to level 3 rather than to levels 2 and 1.

Similarly, the rate equation for level 3 may be written as

$$\frac{dN_3}{dt} = W_1(N_2 - N_3) + T_{43}N_4 - T_{32}N_3 \quad (5.46)$$

where

$$W_1 = \frac{\pi^2 c^2}{\hbar \omega^3 n_0^2} A_{32} g_1(\omega) I_1 \quad (5.47)$$

represents the stimulated transition rate per atom between levels 3 and 2 and the subscript 1 stands for laser transition; $g_1(\omega)$ is the lineshape function describing the $3 \leftrightarrow 2$ transition and I_1 is the intensity of the radiation at the frequency $\omega = (E_3 - E_2)/\hbar$. Also

$$T_{32} = A_{32} + S_{32} \quad (5.48)$$

is the net spontaneous relaxation rate from level 3 to level 2 and consists of the radiative (A_{32}) and the nonradiative (S_{32}) contributions. Again we have neglected any spontaneous transition from level 3 to level 1. In a similar manner, we can write

$$\frac{dN_2}{dt} = -W_1(N_2 - N_3) + T_{32}N_3 - T_{21}N_2 \quad (5.49)$$

$$\frac{dN_1}{dt} = -W_p(N_1 - N_4) + T_{21}N_2 \quad (5.50)$$

where

$$T_{21} = A_{21} + S_{21} \quad (5.51)$$

is the spontaneous relaxation rate from $2 \rightarrow 1$.

Under steady-state conditions

$$\frac{dN_1}{dt} = \frac{dN_2}{dt} = \frac{dN_3}{dt} = \frac{dN_4}{dt} = 0 \quad (5.52)$$

We will thus get four simultaneous equations in N_1, N_2, N_3 , and N_4 and in addition we have

$$N = N_1 + N_2 + N_3 + N_4 \quad (5.53)$$

for the total number of atoms per unit volume in the system.

From Eq. (5.44) we obtain, setting $dN_4/dt = 0$

$$\frac{N_4}{N_1} = \frac{W_p}{(W_p + T_{43})} \quad (5.54)$$

If the relaxation from level 4 to level 3 is very rapid then $T_{43} \gg W_p$ and hence $N_4 \ll N_1$. Using this approximation in the remaining three equations we can obtain for the population difference,

$$\frac{N_3 - N_2}{N} \approx \frac{W_p(T_{21} - T_{32})}{W_p(T_{21} + T_{32}) + T_{32}T_{21} + W_l(2W_p + T_{21})} \quad (5.55)$$

Thus in order to be able to obtain population inversion between levels 3 and 2, we must have

$$T_{21} > T_{32} \quad (5.56)$$

i.e., the spontaneous rate of deexcitation of level 2 to level 1 must be larger than the spontaneous rate of deexcitation of level 3 to level 2.

If we now assume $T_{21} \gg T_{32}$, then from Eq. (5.55) we obtain

$$\frac{N_3 - N_2}{N} \approx \frac{W_p}{W_p + T_{32}} \frac{1}{1 + W_l(T_{21} + 2W_p)/T_{21}(W_p + T_{32})} \quad (5.57)$$

From the above equation we see that even for very small pump rates one can obtain population inversion between levels 3 and 2. This is contrary to what we found in a three-level system, where there was a minimum pump rate, W_{pt} , required to achieve inversion. The first factor in Eq. (5.57) which is independent of W_l [i.e., independent of the intensity of radiation corresponding to the laser transition – see Eq. (5.47)] – gives the small signal gain coefficient whereas the second factor in Eq. (5.57) gives the saturation behavior.

Just below threshold for laser oscillation, $W_l \approx 0$, and hence from Eq. (5.57) we obtain

$$\frac{\Delta N}{N} \approx \frac{W_p}{(W_p + T_{32})} \quad (5.58)$$

where $\Delta N = N_3 - N_2$ is the population inversion density. We shall now consider two examples of four-level systems.

Example 5.1 The Nd:YAG laser corresponds to a four-level laser system (see Chapter 11). For such a laser, typical values of various parameters are

$$\begin{aligned} \lambda_0 &= 1.06 \mu\text{m} (\nu = 2.83 \times 10^{14} \text{ Hz}), & \Delta\nu &= 1.95 \times 10^{11} \text{ Hz}, \\ t_{sp} &= 2.3 \times 10^{-4} \text{ s}, & N &= 6 \times 10^{19} \text{ cm}^{-3}, & n_0 &= 1.82 \end{aligned} \quad (5.59)$$

If we consider a resonator cavity of length 7 cm and $R_1 = 1.00$, $R_2 = 0.90$, neglecting other loss factors (i.e., $\alpha_1 = 0$)

$$t_c = -\frac{2n_0d}{c \ln R_1 R_2} \approx 8 \times 10^{-9} \text{ s}$$

We now use Eq. (4.32) to estimate the population inversion density to start laser oscillation corresponding to the center of the laser transition:

$$\begin{aligned} (\Delta N)_t &= \frac{4v^2 n_0^3}{c^3} \frac{1}{g(\omega)} \frac{t_{sp}}{t_c} \\ &= \frac{4v^2 n_0^3}{c^3} \pi^2 \Delta v \frac{t_{sp}}{t_c} \end{aligned} \quad (5.60)$$

where for a homogenous transition (see Section 4.5)

$$g(\omega_0) = 2/\pi \Delta\omega = 1/\pi^2 \Delta v \quad (5.61)$$

Thus substituting various values, we obtain

$$(\Delta N)_t \approx 4 \times 10^{15} \text{ cm}^{-3} \quad (5.62)$$

Since $(\Delta N)_t \ll N$, we may assume in Eq. (5.58) $T_{32} \gg W_p$ and hence we obtain for the threshold pumping rate required to start laser oscillation

$$\begin{aligned} W_{pt} &\approx \frac{(\Delta N)_t}{N} T_{32} \approx \frac{(\Delta N)_t}{N} \frac{1}{t_{sp}} \\ &= \frac{4 \times 10^{15}}{6 \times 10^{19}} \times \frac{1}{2.3 \times 10^{-4}} \approx 0.3 \text{ s}^{-1} \end{aligned}$$

At this pumping rate the number of atoms being pumped from level 1 to level 4 is $W_{pt}N_1$ and since N_2, N_3 and N_4 are all very small compared to N_1 , we have $N_1 \approx N$. For every atom lifted from level 1 to level 4 an energy $h\nu_p$ has to be given to the atom where ν_p is the average pump frequency corresponding to the $1 \rightarrow 4$ transition. Assuming $\nu_p \approx 4 \times 10^{14} \text{ Hz}$ we obtain for the threshold pump power required per unit volume of the laser medium

$$\begin{aligned} P_{th} &= W_{pt}N_1 h\nu_p \approx W_{pt}N h\nu_p \\ &= 0.3 \times 6 \times 10^{19} \times 6.6 \times 10^{-34} \times 4 \times 10^{14} \\ &\approx 4.8 \text{ W/cm}^3 \end{aligned}$$

which is about three orders of magnitude smaller than that obtained for ruby.

Example 5.2 As a second example of a four-level laser system, we consider the He–Ne laser (see Chapter 11). We use the following data:

$$\begin{aligned} \lambda_0 &= 0.6328 \times 10^{-4} \text{ cm} (\nu = 4.74 \times 10^{14} \text{ Hz}), \\ t_{sp} &= 10^{-7} \text{ s}, \quad \Delta v = 10^9 \text{ Hz}, \quad n_0 \approx 1 \end{aligned} \quad (5.63)$$

If we consider the resonator to be of length 10 cm and having mirrors of reflectivities $R_1 = R_2 = 0.98$, then assuming the absence of other loss mechanisms ($\alpha_l = 0$),

$$\begin{aligned} t_c &= -2n_0 d/c \ln R_1 R_2 \\ &\approx 1.6 \times 10^{-8} \text{ s} \end{aligned} \quad (5.64)$$

For an inhomogeneously broadened transition (see Section 4.5)

$$\begin{aligned} g(\omega_0) &= \frac{2}{\Delta\omega} \left(\frac{\ln 2}{\pi} \right)^{\frac{1}{2}} \\ &\approx 1.5 \times 10^{-10} \text{s} \end{aligned} \quad (5.65)$$

Thus the threshold population inversion required is

$$(\Delta N)_t \approx 1.4 \times 10^9 \text{ cm}^{-3} \quad (5.66)$$

Hence the threshold pump power required to start laser oscillation is

$$\begin{aligned} P_{th} &= W_{pt} N_1 (E_4 - E_1) \\ &\approx \frac{(\Delta N)_t}{t_{sp}} h v_p \end{aligned} \quad (5.67)$$

where again we assume $(\Delta N)_t \ll N$ and $T_{32} \approx A_{32} = 1/t_{sp}$. Assuming $v_p \approx 5 \times 10^{15}$ Hz, we obtain

$$\begin{aligned} P_{th} &= \frac{1.4 \times 10^9 \times 6.6 \times 10^{-34} \times 5 \times 10^{15}}{10^{-7}} \\ &\approx 50 \text{ mW/cm}^3 \end{aligned} \quad (5.68)$$

which again is very small compared to the threshold powers required for ruby laser.

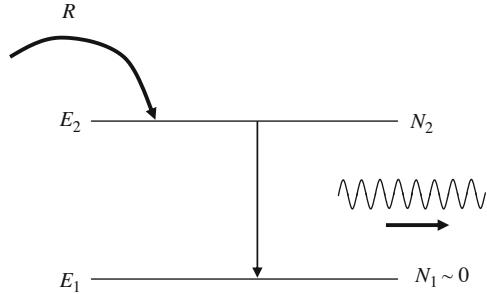
5.5 Variation of Laser Power Around Threshold

In the earlier sections we considered the three-level and four-level laser systems and obtained conditions for the attainment of population inversion. In this section we shall discuss the variation of the power in the laser transition as the pumping rate passes through threshold.

We consider the two levels involved in the laser transition in a four-level laser² and assume that the lower laser level has a very fast relaxation rate to lower levels so that it is essentially unpopulated. We will assume that only one mode has sufficient gain to oscillate and that the line is homogeneously broadened so that the same induced rate applies to all atoms (see Section 4.5). Let R represent the number of atoms that are being pumped into the upper level per unit time per unit volume [see

²A similar analysis can also be performed for a three-level laser system but the general conclusions of this simple analysis remain valid.

Fig. 5.4 The upper level is pumped at a rate R per unit volume and the lower level is assumed to be unpopulated due to rapid relaxation to other lower levels



(Fig. 5.4)]. If the population density of the upper level is N_2 , then the number of atoms undergoing stimulated emissions from level 2 to level 1 per unit time will be [see Eq. (4.16)]

$$F_{21} = \Gamma_{21}V = \frac{\pi^2 c^3}{\hbar \omega^3 n_0^3} A_{21} u g(\omega) N_2 V \quad (5.69)$$

where u is the density of radiation at the oscillating mode frequency ω , V represents the volume of the active medium, and n_0 is the refractive index of the medium.

Instead of working with the energy density u , we introduce the number of photons n in the oscillating cavity mode. Since each photon carries an energy $\hbar\omega$, the number of photons n in the cavity mode will be given by

$$n = uV/\hbar\omega \quad (5.70)$$

Thus

$$F_{21} = \frac{\pi^2 c^3}{\omega^2 n_0^3} A_{21} g(\omega) N_2 n = K n N_2 \quad (5.71)$$

where

$$K \equiv (\pi^2 c^3 / \omega^2 n_0^3) A_{21} g(\omega) \quad (5.72)$$

The spontaneous relaxation rate from level 2 to level 1 in the whole volume will be $T_{21}N_2V$ where

$$T_{21} = A_{21} + S_{21} \quad (5.73)$$

is the total relaxation rate consisting of the radiative (A_{21}) and the nonradiative (S_{21}) components. Hence we have for the net rate of the change of population of level 2

$$\frac{d}{dt}(N_2 V) = -K n N_2 - T_{21} N_2 V + R V$$

or

$$\frac{dN_2}{dt} = -\frac{K n N_2}{V} - T_{21} N_2 + R \quad (5.74)$$

In order to write a rate equation describing the variation of photon number n in the oscillating mode in the cavity, we note that n change due to

- a) All stimulated emissions caused by the n photons existing in the cavity mode which results in a rate of increase of n of KnN_2 since every stimulated emission from level 2 to level 1 caused by radiation in that mode will result in the addition of a photon in that mode. There is no absorption since we have assumed the lower level to be unpopulated.
- b) In order to estimate the increase in the number of photons in the cavity mode due to spontaneous emission, we must note that not all spontaneous emission occurring from the $2 \rightarrow 1$ transition will contribute to a photon in the oscillating mode. As we will show in Section 7.2 for an optical resonator which has dimensions which are large compared to the wavelength of light, there are an extremely large number of modes ($\sim 10^8$) that have their frequencies within the atomic linewidth. Thus when an atom deexcites from level 2 to level 1 by spontaneous emission it may appear in any one of these modes. Since we are only interested in the number of photons in the oscillating cavity mode, we must first obtain the rate of spontaneous emission into a mode of oscillation of the cavity. In order to obtain this we recall from Section 4.2 that the number of spontaneous emissions occurring between ω and $\omega + d\omega$ will be

$$G_{21}d\omega = A_{21}N_2g(\omega)d\omega V \quad (5.75)$$

We shall show in Appendix E that the number of oscillating modes lying in a frequency interval between ω and $\omega + d\omega$ is

$$N(\omega)d\omega = n_0^3 \frac{\omega^2}{\pi^2 c^3} V d\omega \quad (5.76)$$

where n_0 is the refractive index of the medium. Thus the spontaneous emission rate per mode of oscillation at frequency ω is

$$\begin{aligned} S_{21} &= \frac{G_{21}d\omega}{N(\omega)d\omega} = \frac{\pi^2 c^3}{n_0^3 \omega^2} g(\omega) A_{21} N_2 \\ &= K N_2 \end{aligned} \quad (5.77)$$

i.e., the rate of spontaneous emission into a particular cavity mode is the same as the rate of stimulated emission into the same mode when there is just one photon in that mode. This result can indeed be obtained by rigorous quantum mechanical derivation (see [Chapter 9](#)).

- c) The photons in the cavity mode are also lost due to the finite cavity lifetime. Since the energy in the cavity reduces with time as e^{-t/t_c} (see Section 4.4) the rate of decrease of photon number in the cavity will also be n/t_c .

Thus we can write for the total rate of change of n

$$\frac{dn}{dt} = KnN_2 + KN_2 - \frac{n}{t_c} \quad (5.78)$$

Eqs. (5.74) and (5.78) represent the pair of coupled rate equations describing the variation of N_2 and n with time.

Under steady-state conditions both time derivatives are zero. Thus we obtain from Eq. (5.78),

$$N_2 = \frac{n}{n+1} \frac{1}{Kt_c} \quad (5.79)$$

The above equation implies that under steady-state conditions $N_2 \leq 1/Kt_c$. When the laser is oscillating under steady-state conditions $n \gg 1$ and $N_2 \approx 1/Kt_c$. If we substitute the value of K from Eq. (5.72) we find that (for $n \gg 1$)

$$N_2 \approx \frac{\omega^2 n_0^3}{\pi^2 c^3} \frac{t_{sp}}{t_c} \frac{1}{g(\omega)} \quad (5.80)$$

which is nothing but the threshold population inversion density required for laser oscillation (cf. Eq. (4.32)). Thus Eq. (5.79) implies that when the laser oscillates under steady-state conditions, the population inversion density is almost equal to and can never exceed the threshold value. This is also obvious since if the inversion density exceeds the threshold value, the gain in the cavity will exceed the loss and thus the laser power will start increasing. This increase will continue till saturation effects take over and reduce N_2 to the threshold value.

Substituting from Eq. (5.79) into Eq. (5.74) and putting $dN_2/dt = 0$, we get

$$\frac{K}{VT_{21}} n^2 + n \left(1 - \frac{R}{R_t} \right) - \frac{R}{R_t} = 0 \quad (5.81)$$

where

$$R_t = \frac{T_{21}}{Kt_c} \quad (5.82)$$

The solution of the above equation which gives a positive value of n is

$$n = \frac{VT_{21}}{2K} \left\{ \left(\frac{R}{R_t} - 1 \right) + \left[\left(1 - \frac{R}{R_t} \right)^2 + \frac{4K}{VT_{21}} \frac{R}{R_t} \right]^{\frac{1}{2}} \right\} \quad (5.83)$$

The above equation gives the photon number in the cavity under steady-state conditions for a pump rate R .

For a typical laser system, for example an Nd:glass laser (see Chapter 11),

$$\begin{aligned} V &\approx 10 \text{ cm}^3, & n_0 &\approx 1.5 \\ \lambda &\approx 1.06 \mu\text{m}, & \Delta\nu &\approx 3 \times 10^{12} \text{ Hz} \end{aligned}$$

so that

$$\frac{K}{VT_{21}} = \frac{c^3}{8v^2 n_0^3} \frac{1}{V\pi \Delta v} \approx 1.3 \times 10^{-13} \quad (5.84)$$

where we have used $T_{21} \approx A_{21}$. For such small values of K/VT_{21} , unless R/R_t is extremely close to unity, we can make a binomial expansion in Eq. (5.83) to get

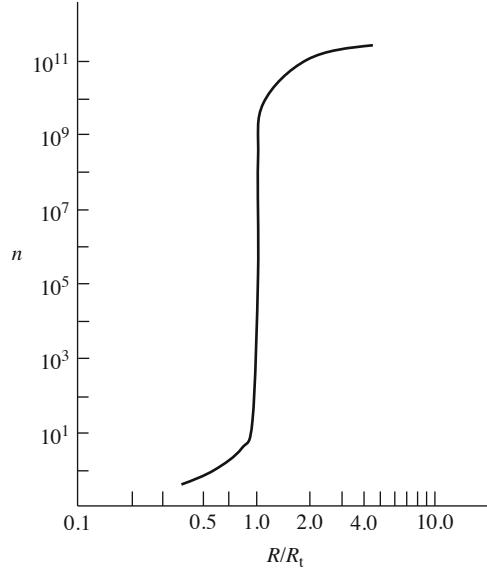
$$\begin{aligned} n &\approx \frac{R/R_t}{1-R/R_t} & \text{for } \frac{R}{R_t} < 1 - \Delta \\ n &\approx \frac{VT_{21}}{K} \left(\frac{R}{R_t} - 1 \right) & \text{for } \frac{R}{R_t} > 1 + \Delta \end{aligned} \quad (5.85)$$

where $\Delta \gg (2K/VT_{21})^{1/2}$. Further

$$n \approx \left(\frac{VT_{21}}{K} \right)^{1/2} \quad \text{for } \frac{R}{R_t} = 1 \quad (5.86)$$

Figure 5.5 shows a typical variation of n with R/R_t . As is evident $n \approx 1$ for $R < R_t$ and approaches 10^{12} for $R > R_t$. Thus R_t as given by Eq. (5.82) gives the threshold pump rate for laser oscillation.

Fig. 5.5 Variation of photon number n in the cavity mode as a function of pumping rate R ; R_t corresponds to the threshold pumping rate. Note the steep rise in the photon number as one crosses the threshold for laser oscillation



Problem 5.2 Show that the threshold pump rate R_t given by Eq. (5.82) is consistent with that obtained in Section 4.4.

From the above analysis it follows that when the pumping rate is below threshold ($R < R_t$) then the number of photons in the cavity mode is very small (~ 1). As one approaches the threshold, the number of photons in the preferred cavity mode

(having higher gain and lower cavity losses) increases at a tremendous rate and as one passes the threshold, the number of photons in the oscillating cavity mode becomes extremely large. At the same time the number of photons in other cavity modes which are below threshold remains orders of magnitude smaller.

In addition to the sudden increase in the number of photons in the cavity mode and hence laser output power, the output also changes from an incoherent to a coherent emission. The output becomes an almost pure sinusoidal wave with a well-defined wave front, apart from small amplitude and phase fluctuations caused by the ever-present spontaneous emission.³ It is this spontaneous emission which determines the ultimate linewidth of the laser.

If the only mechanism in the cavity is that arising from output coupling due to the finite reflectivity of one of the mirrors, then the output laser power will be

$$P_{\text{out}} = \frac{n\hbar v}{t_c} \quad (5.87)$$

where n/t_c is the number of photons escaping from the cavity per unit time and $\hbar v$ is the energy of each photon. Taking K/VT_{21} as given by Eq. (5.84) and $t_c \approx 10^{-8}\text{s}$, for $R/R_t = 2$ we obtain

$$P_{\text{out}} = 144 \text{ W}$$

Example 5.3 It is interesting to compare the number of photons per cavity mode in an oscillating laser and in a black body at a temperature T . The number of photons/mode in a black body is (see Appendix D)

$$n = \frac{1}{e^{\hbar\omega/k_B T} - 1} \quad (5.88)$$

Hence for $\lambda = 1.06 \mu\text{m}$, $T = 1000 \text{ K}$, we obtain

$$n \approx \frac{1}{e^{13.5} - 1} \approx 1.4 \times 10^{-6}$$

which is orders of magnitude smaller than in an oscillating laser [see (Fig 5.5)].

From Eq. (5.85) we may write for the change in number of photons dn for a change dR in the pump rate as

$$\frac{dn}{dR} = \frac{VT_{21}}{K} \frac{1}{R_t} = Vt_c$$

or

$$VdR = \frac{dn}{t_c} \quad (5.89)$$

where we have used Eq. (5.82). The LHS of Eq. (5.89) represents the additional number of atoms that are being pumped per unit time into the upper laser level and

³In an actual laser system, the ultimate purity of the output beam is restricted due to mechanical vibrations of the laser, mirrors, temperature fluctuations, etc.

the RHS represents the additional number of photons that is being lost from the cavity. Thus above threshold all the increase in pump rate goes toward the increase in the laser power.

Example 5.4 Let us consider an Nd:glass laser (see Chapter 11) with the parameters given in page 119 and having

$$\begin{aligned} d &= 10 \text{ cm} \\ R_1 &= 0.95, \quad R_2 = 1.00 \end{aligned}$$

For these values of the parameters, using Eq. (4.31) we have

$$t_c \approx -\frac{2n_0d}{c \ln R_1 R_2} \approx 1.96 \times 10^{-8} \text{ s}$$

and

$$\frac{VT_{21}}{K} \approx \frac{V}{Kt_{sp}} = \frac{4v^2 V n_0^3}{c^3 g(\omega)} \quad (5.90)$$

Thus for $R/R_t = 2$, i.e., for a pumping rate twice the threshold value (see Eq. (5.85))

$$n = VT_{21}/K \approx 7.7 \times 10^{12}$$

Hence the energy inside the cavity is

$$\begin{aligned} E &= nhv \\ &\approx 1.4 \times 10^{-6} \text{ J} \end{aligned} \quad (5.91)$$

If the only loss mechanism is the finite reflectivity of one of the mirrors, then the output power will be

$$P_{\text{out}} = \frac{n h v}{t_c} \approx 74 \text{ W}$$

Problem 5.3 In the above example, if it is required that there be 1 W of power from the mirror at the left and 73 W of power from the right mirror, what should the reflectivities of the two mirrors be? Assume the absence of all other loss mechanisms in the cavity.

[Answer: $R_1 = 0.9993$, $R_2 = 0.9507$]

Example 5.5 In this example, we will obtain the relationship between the output power of the laser and the energy present inside the cavity by considering radiation to be making to and fro oscillations in the cavity. Figure 5.6a shows the cavity of length l bounded by mirrors of reflectivities 1 and R and filled by a medium characterized by the gain coefficient α . Let us for simplicity assume absence of all other loss mechanisms. Figure 5.6b shows schematically the variation of intensity along the length of the resonator when the laser oscillates under steady-state conditions. For such a case, the intensity after one round trip I_4 must be equal to the intensity at the same point at the start of the round trip. Hence

$$R e^{2\alpha l} = 1 \quad (5.92)$$

Also, recalling the definition of cavity lifetime (see Eq. (4.31) with $\alpha_1 = 0$), we have

$$t_c = -\frac{2l}{c} \ln R = \frac{1}{\alpha c} \quad (5.93)$$

Now let us consider a plane P inside the resonator. Let the distance of the plane from mirror M_1 be x . Thus if I_1 is the intensity of the beam at mirror M_1 , then assuming exponential amplification, the intensity of the beam going from left to right at P is

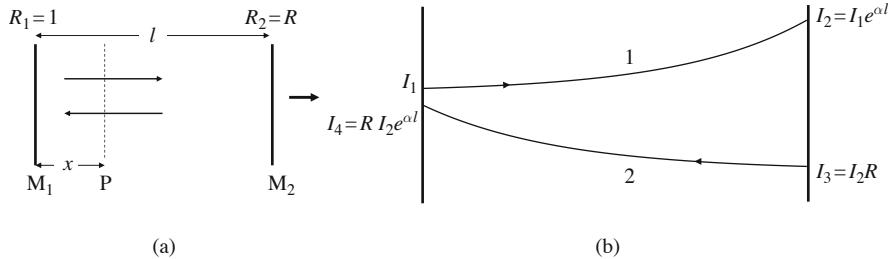


Fig. 5.6 (a) A resonator of length l bound by mirrors of reflectivities 1 and R and filled by a medium of gain coefficient α . (b) Curves 1 and 2 represent the qualitative variation of intensity associated with the waves propagating in the forward and backward directions within the cavity. The sudden drop in intensity from I_2 to I_3 is due to the finite reflectivity of the mirror M_2

$$I_+ = I_1 e^{\alpha x} \quad (5.94)$$

Similarly, the intensity of the beam going from right to left at P is

$$\begin{aligned} I_- &= I_1 e^{\alpha l} R e^{\alpha(l-x)} = I_1 R e^{2\alpha l} e^{-\alpha x} \\ &= I_1 e^{-\alpha x} \end{aligned} \quad (5.95)$$

Hence the energy density at x is

$$u(x) = \frac{I_+ + I_-}{c} = \frac{I_1}{c} (e^{\alpha x} + e^{-\alpha x}) \quad (5.96)$$

If A is the area of cross section, then the total energy in the cavity is

$$\begin{aligned} W &= \int \int u dA dx = A \int_0^l u dx \\ &= \frac{AI_1}{\alpha c} e^{\alpha l} (1 - R) \\ &= AI_1 t_c e^{\alpha l} (1 - R) \end{aligned} \quad (5.97)$$

where we have used Eqs. (5.92) and (5.93) and have assumed, for the sake of simplicity, uniform intensity distribution in the transverse direction. Now the power emerging from mirror M_2 is

$$\begin{aligned} P_{\text{out}} &= I_2 A (1 - R) \\ &= I_1 A e^{\alpha l} (1 - R) \\ &= W/t_c \end{aligned} \quad (5.98)$$

which is consistent with Eq. (5.87)

5.6 Optimum Output Coupling

In the last section we obtained the steady-state energy inside the resonator cavity as a function of the pump rate. In order to get an output laser beam, one of the mirrors is made partially transparent so that a part of the energy is coupled out. In

this section we shall obtain the optimum reflectivity of the mirror so as to have a maximum output power.

The fact that an optimum output coupling exists can be understood as follows. If one has an almost zero output coupling (i.e., if both mirrors are almost 100% reflecting) then even though the laser may be oscillating, the output power will be almost zero. As one starts to increase the output coupling, the energy inside the cavity will start to decrease since the cavity loss is being increased but, since one is taking out a larger fraction of power the output power starts increasing. The output power will start decreasing again if the reflectivity of the mirror is continuously reduced since if it is made too small, then for that pumping rate, the losses will exceed the gain and the laser will stop oscillating. Thus for a given pumping rate, there must be an optimum output coupling which gives the maximum output power.

In Section 4.4 we showed that the cavity lifetime of a passive resonator is

$$\begin{aligned}\frac{1}{t_c} &= \frac{c}{2dn_0} (2\alpha_1 d - \ln R_1 R_2) \\ &= \frac{1}{t_i} + \frac{1}{t_e}\end{aligned}\quad (5.99)$$

where

$$\frac{1}{t_i} = \frac{c\alpha_1}{n_0}, \quad \frac{1}{t_e} = -\frac{c}{2dn_0} \ln R_1 R_2 \quad (5.100)$$

t_i accounts for all loss mechanisms except for the output coupling due to the finite mirror reflectivities and t_e for the loss due to output coupling only. Thus, the number photons escaping the cavity due to finite mirror reflectivity will be n/t_e and hence the output power will be

$$\begin{aligned}P_{\text{out}} &= \frac{n\hbar v}{t_e} \\ &= \frac{\hbar v}{t_e} \frac{V T_{21}}{K} \left[\frac{R K}{T_{21}} \left(\frac{1}{t_i} + \frac{1}{t_e} \right)^{-1} - 1 \right]\end{aligned}\quad (5.101)$$

where we have used Eqs. (5.82), (5.85), and (5.99). The optimum output power will correspond to the value of t_e satisfying $\partial P_{\text{out}}/\partial t_e = 0$ which gives

$$\frac{1}{t_e} = \left(\frac{R K}{T_{21} t_i} \right)^{\frac{1}{2}} - \frac{1}{t_i} \quad (5.102)$$

Using Eqs. (5.99) and (5.82), the above equation can be simplified to

$$\frac{1}{t_e} = \frac{1}{t_i} \left(\frac{R}{R_t} - 1 \right) \quad (5.103)$$

Substituting for t_e from Eq. (5.103) in Eq. (5.101) we obtain the maximum output power as

$$P_{\max} = h\nu RV \left[1 - \left(\frac{T_{21}}{KRt_i} \right)^{1/2} \right]^2 \quad (5.104)$$

It is interesting to note that the optimum t_e and hence the optimum reflectivity is a function of the pump rate R .

Even though the output power passes through a maximum as the transmittivity $T = (1-R)$ of the mirror is increased, the energy inside the cavity monotonically reduces from a maximum value as T is increased. This may be seen from the fact that the energy in the cavity is

$$E = nhv = \frac{VT_{21}}{K} \left(\frac{KRt_c}{T_{21}} - 1 \right) hv \quad (5.105)$$

Thus as T is increased, t_c reduces and hence E reduces monotonically finally becoming zero when

$$t_c = \frac{T_{21}}{KR} \quad (5.106)$$

beyond which the losses become more than the gain.

Problems

Problem 5.4 Using Eq. (5.103) calculate the optimum reflectivity of one of the mirrors of the resonator (assuming the other mirror to have 100% reflectivity) for $R = 2R_t$. Assume the length of the resonator to be 100 cm, $n_0 = 1$ and the intrinsic loss per unit length to be $3 \times 10^{-5} \text{ cm}^{-1}$.

Problem 5.5 Consider an atomic system as shown below:

- 3 ————— $E_3 = 3 \text{ eV}$
- 2 ————— $E_2 = 1 \text{ eV}$
- 1 ————— $E_1 = 0 \text{ eV}$

The A coefficient of the various transitions are given by

$$A_{32} = 7 \times 10^7 \text{ s}^{-1}, \quad A_{31} = 10^7 \text{ s}^{-1}, \quad A_{21} = 10^8 \text{ s}^{-1}$$

- Show that this system cannot be used for continuous wave laser oscillation between levels 2 and 1.
- Suppose at $t = 0$, N_0 atoms are lifted to level 3 by some external mechanism describe the change of populations in levels 1, 2, and 3.

Problem 5.6 Using Eq. (5.103) calculate the optimum reflectivity of one of the mirrors of the resonator (assuming the other mirror to be 100% reflecting) for $R = 2R_t$. Assume the length of the resonator to

be 50 cm, $n_0 = 1$ and the intrinsic loss per unit length to be $3 \times 10^{-4} \text{ m}^{-1}$. If the power output at the optimum coupling is 10 mW, what is the corresponding energy inside the cavity?

[Answer: $R \approx 0.9997$, energy $\cong 1.1 \times 10^{-7} \text{ J}$]

Problem 5.7 Consider a laser with plane mirrors having reflectivities of 0.9 each and of length 50 cm filled with the gain medium. Neglecting scattering and other cavity losses, estimate the threshold gain coefficient (in m^{-1}) required to start laser oscillation.

Problem 5.8 The cavity of a 6328 Å He–Ne laser is 1 m long and has mirror of reflectivities 100 and 98%; the internal cavity losses are negligible. (a) If the steady-state power output is 10 mW, what is the energy stored in the resonator? (b) What is the linewidth of the above passive cavity? (c) If the oscillating linewidth is 1500 MHz, how many longitudinal modes would oscillate?

Problem 5.9 Consider a two-level system shown below:

$$\begin{array}{ll} \text{---} & E_2 = 2 \text{ eV} \\ A_{21} = 10^7 \text{ s}^{-1} & \\ \text{---} & E_1 = 0 \end{array}$$

- a) What is the frequency of light emitted due to transitions from E_2 to E_1 ?
- b) Assuming the emission to have only natural broadening, what is the FWHM of the emission?
- c) What is the population ratio N_2/N_1 at 300 K?
- d) An atomic system containing N_0 atoms/cm³ of the above atoms is radiated by a beam of intensity I_0 at the line center. Write down the rate equation and obtain the population difference between E_2 and E_1 under steady-state condition. Calculate the incident intensity required to produce a population ratio $N_1 = 2 N_2$.

Problem 5.10 The active medium of a three-level atomic system is characterized by the following spontaneous emission rates: $A_{21} = 10^8 \text{ s}^{-1}$, $A_{31} = 10^6 \text{ s}^{-1}$, $A_{32} = 10^5 \text{ s}^{-1}$. (Neglect non-radiative transitions.) Can we use the atomic system to realize a laser? (YES or NO). Justify your answer.

Problem 5.11 Consider a three-level laser system with lasing between levels E_2 and E_1 . The level E_2 has a lift time of 1 μs. Assuming the transition $E_3 \rightarrow E_2$ to be very rapid, estimate the number of atoms that needs to be pumped per unit time per unit volume from level E_1 to reach threshold for achieving population inversion. Given that the total population density of the atoms is 10^{19} cm^{-3} .

Chapter 10

Properties of Lasers

10.1 Introduction

So far we have discussed the physics behind laser operation. Basically the light from both a laser and any ordinary source of light is electromagnetic in nature, but laser light can be extremely monochromatic, highly directional, and very intense. Apart from these laser light also differs from light produced by thermal emission in the basic quantum properties.

In this chapter we shall look at the different properties of laser light and in Part II of the book we shall discuss some of the most interesting applications of lasers.

10.2 Laser Beam Characteristics

Light from the laser arises primarily from stimulated emission and the resonator cavity within which the amplifying medium is kept leads to the following special properties:

- Directionality
- Spectral purity
- High power
- Extremely short pulse durations

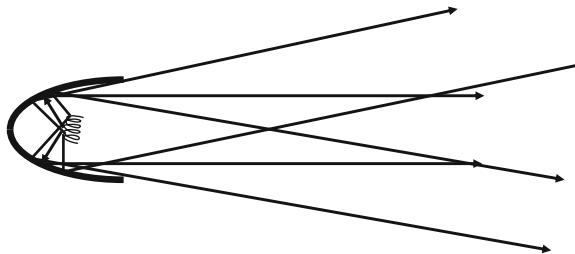
Table 10.1 gives the achievable laser characteristics; these special properties lead to many applications of lasers.

Directionality: Light from a source of light such as a torchlight diverges significantly as it propagates (see Fig. 10.1). But the beam coming from a laser is in the form of a pencil of rays and seems to propagate without any divergence. The laser beam also diverges but by a much smaller magnitude. The wave nature of light imparts an intrinsic divergence to the beam due to the phenomenon of diffraction (see Chapter 2). Thus unlike a torchlight where the divergence is due to the finite size of the filament, the divergence of the laser beam is limited by diffraction depending on the laser types and can be less than 10^{-5} radians (~ 2 s of arc). This extremely small divergence leads to the many application of the laser in surveying, remote sensing, lidar, etc.

Table 10.1 Some of the special properties possessed by laser beams from different types of lasers

• Directionality	(Divergence $\sim 10^{-7}$ rad)
• Spectral purity	$(\Delta\lambda \sim 10^{-9} \mu\text{m})$
• High power	$(P \sim 10^{18} \text{ W/cm}^2)$
• Ultra short pulses	$(\Delta t \sim 10^{-15} \text{ s})$
• High electric fields	$(E \sim 10^{12} \text{ V/m})$
• Small focused areas	$(\sim 10^{-12} \text{ m}^2)$

Fig. 10.1 Light from a torch has a divergence primarily due to the fact that light emanating from different points on the filament propagates along different directions after reflection from the parabolic mirror

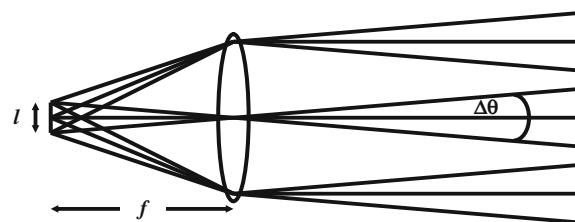


As an example consider a tiny filament lamp placed at the focus of a convex lens as shown in Fig. 10.2. The filament can be considered to be made up of a number of point sources, and thus the light emanating from different points on the filament will travel along different directions after passing through the lens and the exiting beam will diverge. If the linear dimension of the filament is about 2 mm and if the focal length of the convex lens is 10 cm then the angular divergence of the beam (due to the finite size of the filament) is approximately 1° ($= 0.02$ radians). This divergence could be reduced provided we reduce the dimension of the filament, but then the amount of light will also get correspondingly reduced.

Compared to the filament, the divergence of a laser beam is primarily due to diffraction. For most laser beams, the *spot size* (the radius of the cross section of the laser beam) of the beam is about a few millimeters. As discussed in Chapter 2, if the laser beam has a free space wavelength of λ_0 and a spot size w_0 , then the divergence angle of the beam is given by

$$\theta \approx \frac{\lambda_0}{\pi w_0} \quad (10.1)$$

Fig. 10.2 Light from the filament of a bulb placed at the focus of a convex lens diverges after passing through it



For a typical spot size of 1 mm and a wavelength of $0.6 \mu\text{m}$, the divergence angle is given approximately by 0.01° . A beam is said to be *diffraction limited* if it diverges only due to diffraction and usually laser beams are diffraction limited. We may mention that the laser beam from a laser diode has a significant divergence due to the small spot size of the beam. At the same time, unlike the case of the torch, the divergence of the beam can be reduced by simply using a lens in front of the laser diode.

Tight Focusing: Because of highly directional properties of the laser beams, they can be focused to very small areas of a few $(\mu\text{m})^2$. The limits to focusing are again determined by diffraction effects. Smaller the wavelength, smaller the size of the focused spot. This property leads to applications in surgery, material processing, compact discs, etc.

When a convex lens images a point object, the size of the image point is directly proportional to the wavelength of the light wave and also to the ratio of focal length to the diameter. The ratio of focal length to the diameter of a lens is also called the *f-number*. This parameter is used in specifying the quality of camera lenses. Thus an $f/2$ lens implies that the focal length to diameter ratio is 2. If the focal length of this camera lens is 50 mm then its diameter is 25 mm. Smaller the *f-number* for a given focal length larger is the diameter of the lens. Smaller the wavelength, smaller the spot size, and similarly smaller the *f-number*, smaller the image size. Since an object can be considered to be made up of points, if we consider the imaging by the convex lens, the resolution provided by the lens will depend on the *f-number* and the wavelength. For a given wavelength, for better resolution we must have a smaller *f-number*. Smaller the *f-number* of a camera, better will be the resolution of the camera.

Thus when a laser beam is allowed to fall on a convex lens then the radius of the focused spot is directly proportional to the wavelength and to the *f-number*, provided the laser beam fills the entire area of the lens (see Fig. 10.3). If we take a lens having an *f-number* of 2 (i.e., focal length is twice the lens diameter), then for a laser wavelength of 600 nm the radius of the focused spot will be about $1.5 \mu\text{m}$. Thus the area of such a focused spot would be about $7 \mu\text{m}^2$. If the laser beam has a power of 1 MW ($= 10^6 \text{ W}$) then the intensity at the focused spot would be approximately 14 TW/cm^2 . Such intensities of light lead to electric fields of 10^9 V/m . Such high

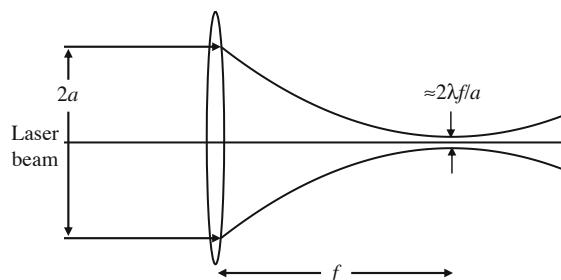


Fig. 10.3 If a truncated plane wave (of diameter $2a$) is incident on a lens without any aberration of focal length f , then the wave emerging from the lens will get focused to spot of radius $\approx \lambda f/a$

Fig. 10.4 Focusing of a 3 MW peak power-pulsed ruby laser beam. At the focus, the electric field strengths are of the order of a billion Volts per meter which results in the creation of a spark in the air. (Photograph courtesy Dr. R. W. Terhune)

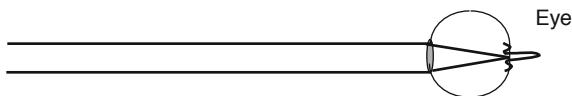
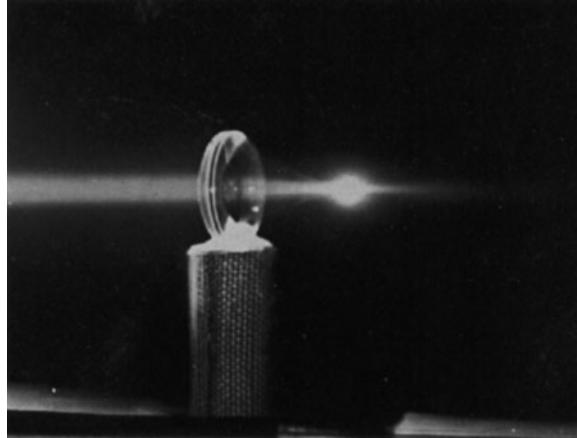


Fig. 10.5 When a laser beam falls on the eye, then it gets focused to a very small diffraction-limited size producing very high intensities even for small powers as 2 mW

electric fields can create a spark in air – see Fig. 10.4. This shows that laser beams (because of their high directionality) can be focused to extremely small regions producing very high intensities and electric fields.

When the electric fields are very high inside a medium, the light beam can change the properties of the medium. Such effects are termed non-linear effects; some effects of this non-linearity are discussed in Chapters 14 and 18.

We may mention here that a low-power (≈ 2 mW) diffraction-limited laser beam incident on the eye gets focused to a very small spot (see Fig. 10.5) and can produce an intensity of about 100 W/cm^2 on the retina – this could indeed damage the retina. On the other hand, when we look at a 20 W bulb at a distance of about 5 m from the eye, the eye produces an image of the bulb on the retina and this would produce an intensity of only about 10 W/m^2 on the retina of the eye (see Fig. 10.6). Thus, whereas it is quite safe to look at a 20 W bulb, it is very dangerous to look directly into a 2 mW laser beam. Indeed, because a laser beam can be focused to very narrow areas, it has found important applications in areas like eye surgery and laser cutting.

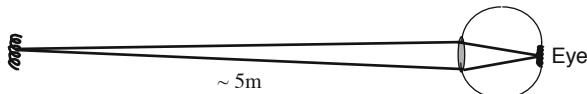


Fig. 10.6 Looking at a bulb produces an image of the bulb on the retina and even 20 W bulb does not produce very high intensities

It may be of interest to mention that, if we are directly looking at the sun, the power density in the image formed is about 30 kW/m^2 . This follows from the fact that on the earth, about 1.35 kW of solar energy is incident (normally) on an area of 1 m^2 . Thus the energy entering the eye is about 4 mW . Since the sun subtends about 0.5° on the earth, the radius of the image of the sun (on the retina) is about $2 \times 10^{-4} \text{ m}$. Therefore if we are directly looking at the sun the power density in the image formed is about 30 kW/m^2 . *Thus, never look into the sun; the retina will be damaged not only because of high intensities but also because of large ultraviolet content of the sunlight.*

A very interesting application of the extreme directionality of laser beams is in the realization of artificial stars in the sky. At a height of about 95 km above the surface of earth there is a layer containing sodium atoms. If a laser beam at a wavelength of 589 nm is sent up, then the sodium atoms absorb the radiation, get excited to a higher energy level, and then emit spontaneously when they get de-excited. Some of this radiation is traveling toward the earth and resembles a star. The position of this artificial star or guide star can be adjusted by changing the direction of the laser beam. Typically pulsed lasers emitting a power of about 20 W and pulse widths of 100 ns are used to create the star. One of the interesting applications of this guide star is in the correction of images formed by telescopes on the earth. Since the light coming from objects outside the earth has to pass via the turbulent atmosphere, the image of any extra terrestrial object will not be stable. By looking at the image of this artificial star it is possible to determine the correction to the optical system required in real time for canceling the effects of turbulence. Figure 10.7 shows the images of the application of laser guide star in imaging the dense star cluster at the center of the Milky Way.

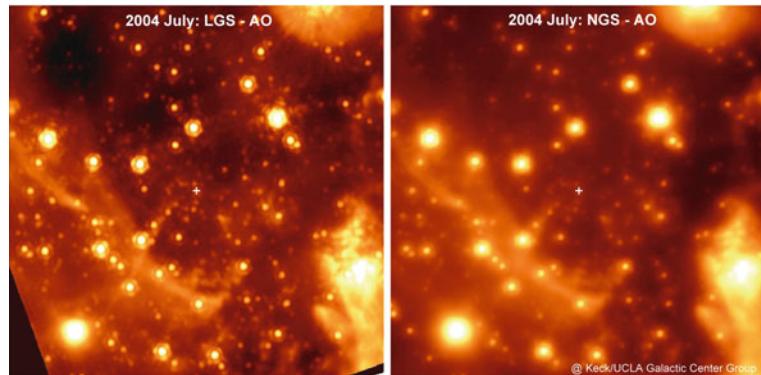


Fig. 10.7 Images of the dense star cluster at the center of the Milky Way Galaxy in infrared light at $3.6 \mu\text{m}$ wavelength. *Left:* image using laser guide star adaptive optics. *Right:* best natural guide star image. The laser guide star image has a total integration time of 8 min, while the natural guide star image has 150 min. The plus sign marks the position of the central million-solar-mass black hole, Sgr A*, in both images. (Credit: UCLA Galactic Center Group and W.M. Keck Observatory Laser Guide Star Team; photograph provided by Prof Claire Max, University of California, USA)

Problem 10.1 Consider a confocal laser resonator made of mirrors of radii of curvature 1 m each. Assuming that the laser is oscillating in the fundamental Gaussian mode, obtain the divergence of the exiting laser beam if the free space wavelength of the laser is 500 nm. [Ans: $w_0 \sim 300 \mu\text{m}$, 90 s of arc].

Problem 10.2 Referring to Problem 10.1 if the separation between the mirrors is kept 1 m while the mirror radii of curvatures are increased to 5 m, obtain the corresponding divergence. [Ans: $w_0 \sim 5 \text{ mm}$, 6 s of arc].

Spectral Purity: Laser beams can have an extremely small spectral width, of the order of 10^{-6} \AA . Compare this with a typical source such as a sodium lamp which has a spectral width of about 0.1 \AA . The stimulated emission process coupled with the optical resonator within which the amplifying medium is placed is responsible for the very small spectral widths. In general a laser may oscillate in a number of frequencies simultaneously unless special techniques are adopted (see Chapter 7). This includes using Fabry–Perot filters within the laser cavity to allow only one frequency to oscillate. Even in a laser oscillating in a single frequency, there could be random but small variations in the frequency of oscillation due to temperature fluctuations and vibrations of the mirrors of the cavity. Stabilization of frequency is achieved using various techniques, for example, the laser is coupled to another very stable cavity and the emission from the laser gets locked to this stable cavity. Frequency-stabilized lasers with frequency stability of better than 10^{-8} (i.e.,

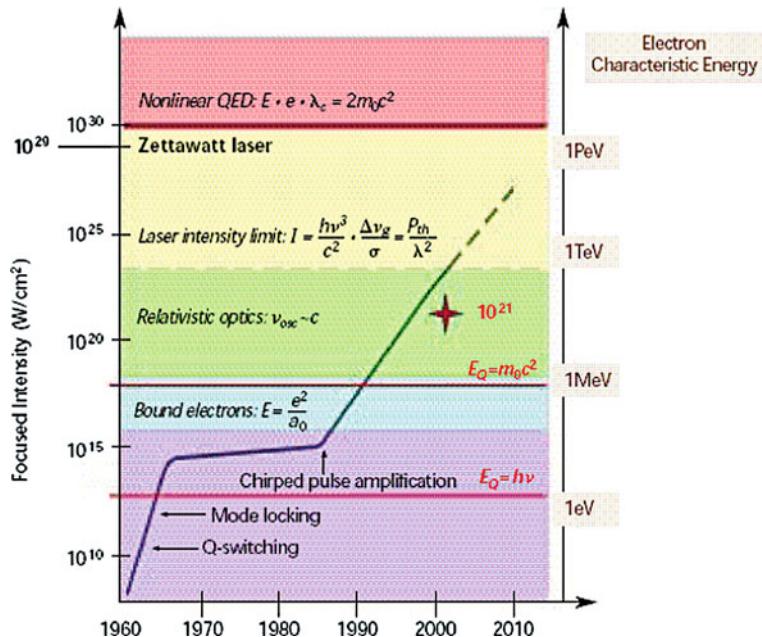


Fig. 10.8 Increase in achievable laser intensity with the year. The increase has a large slope around 1960 due to the invention of the laser and then again after 1985. (Adapted from Mourou and Yanovsky (2004) © 2004 OSA)

fractional frequency shift of less than 10 parts per billion) are commercially available. Because of high spectral purity, lasers find applications in holography, optical communications, spectroscopy, etc.

High Power: Lasers can generate extremely high powers, and since they can also be focused to very small areas, it is possible to generate extremely high-intensity values. Figure 10.8 shows how the intensity achievable using laser beams has increased every year. At intensities such as 10^{21} W/m^2 , the electric fields are so high that electrons can get accelerated to relativistic velocities (velocities approaching that of light) leading to very interesting effects. Apart from scientific investigations of extreme conditions, continuous wave lasers having power levels $\sim 10^5 \text{ W}$ and pulsed lasers having a total energy $\sim 50000 \text{ J}$ have applications in welding, cutting, laser fusion, star wars, etc.

10.3 Coherence Properties of Laser Light

In this section we will introduce the concepts of temporal and spatial coherence since these play a significant role in their applications.

10.3.1 Temporal Coherence

In order to understand the concept of temporal coherence, we consider a Michelson interferometer arrangement as shown in Fig. 10.9. S represents an extended near monochromatic source, G represents a beam splitter, and M_1 and M_2 are two plane mirrors. The mirror M_2 is fixed while the mirror M_1 can be moved either toward

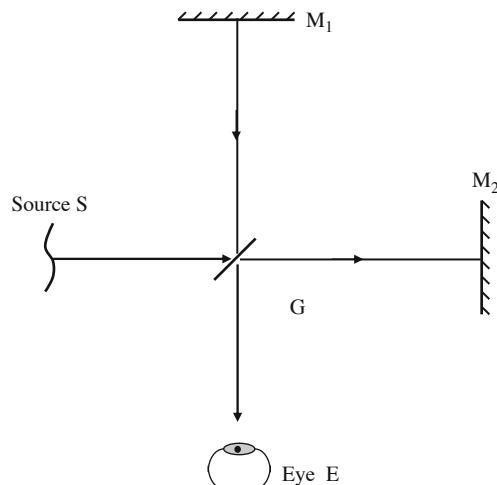


Fig. 10.9 A Michelson interferometer setup

or away from G. Light from the source S is incident on G and is divided into two equal portions; one part travels toward M_1 and is reflected back and the other part is reflected back from M_2 . The two reflected waves interfere and produce interference fringes which are visible from E. When the mirrors M_1 and M_2 are nearly equidistant from G, i.e., when the two waves traversing the two different paths take the same amount of time, then it is observed that the contrast of the interference fringes formed is good. If now the mirror M_1 is slowly moved away from G, then it is seen that for ordinary extended source of light (like a sodium lamp), the contrast in the fringes goes on decreasing and when the difference between the distances from G to M_1 and M_2 is about a few millimeters to a few centimeters, the fringes are no longer visible. This decrease in contrast of the fringes can be explained as follows: The source S is emitting small wave trains of an average duration τ_c (say) and there is no phase relationship between different wave trains (see Section 4.5). This is in contrast to an infinitely long pure sinusoidal wave train, which is also referred to as a monochromatic wave. When the difference in time taken by the wave trains to travel the paths G to M_1 and back and G to M_2 and back is much less than the average duration τ_c , then the interference is produced between two wave trains each one being derived from the same wave train. Hence even though different wave trains emanating from the source S do not have definite phase relationship, since one is superimposing two wave trains derived from the same wave train, fringes of good contrast will be seen. On the other hand, if the difference in the time taken to traverse the paths to M_1 and back and to M_2 and back is much more than τ_c , then one is superimposing two wave trains which are derived from two different wave trains, and since there is no definite phase relationship between two wave trains emanating from S, interference fringes will not be observed. Hence as the mirror M_1 is moved, the contrast in the fringes becomes poorer and poorer and for large separations no fringes would be seen. The time τ_c is referred to as the coherence time and the length of the wave train $c\tau_c$ is referred to as the longitudinal coherence length. It may be mentioned that there is no definite distance at which the interference pattern disappears; as the distance increases, the contrast in the fringes becomes gradually poorer and eventually the fringes disappear.

As an example, for the neon 632.8 nm line from a discharge lamp, the interference fringes would vanish if the path difference between the two mirrors is about a few centimeters. Thus for this source, $\tau_c \sim 100$ ps. On the other hand, for the red cadmium line at 643.8 nm, the coherence length is about 30 cm, which gives $\tau_c \sim 1$ ns.

The decrease in contrast of the fringes can also be interpreted as being due to the fact that the source S is not emitting a single frequency but emits over a band of frequencies (see Section 4.5). When the path difference is zero or very small, the different wavelength components produce fringe patterns superimposed on one another and the fringe contrast is good. On the other hand, when the path difference is increased, different wavelength components produce fringe patterns which are slightly displaced with respect to one another and the fringe contrast becomes poor. Thus the non-monochromaticity of the light source can equally well be interpreted as the reason for poor fringe visibility for large optical path differences.

The equivalence of the above two approaches can also be seen using Fourier analysis (see Section 4.5). One can indeed show that a wave having a coherence time $\sim \tau_c$ is essentially a superposition of harmonic waves having frequencies in the range $v_0 - \Delta\nu/2 \leq v \leq v_0 + \Delta\nu/2$ where

$$\Delta\nu \sim \frac{1}{2\pi\tau_c} \quad (10.2)$$

Thus the longer the coherence time, the smaller the frequency width. For ordinary sources $\tau_c \sim 100$ ps and

$$\Delta\nu \sim 10^{10} \text{ Hz}$$

For $\lambda = 600$ nm, $v = 5 \times 10^{14}$ Hz and

$$\frac{\Delta\nu}{v_0} \sim 0.00002$$

The quantity $\Delta\nu/v_0$ represents the monochromaticity and one can see that even for ordinary light sources it is quite small. In Section 4.5 we have discussed some of the mechanisms leading to the broadening of spectral lines emitted by atoms.

In a laser, in contrast to an ordinary source of light, the optical resonant cavity is excited in different longitudinal modes of the cavity which are specified by discrete frequencies of oscillation. In an optical resonator without the amplifying medium, the finite loss of the resonator leads to an exponentially decaying output amplitude which leads to a finite linewidth of the output. On the other hand, in an actual laser oscillating in steady state, the loss is exactly compensated by the gain provided by the laser medium, and when the laser is oscillating in a single mode, the output is essentially a pure sinusoidal wave. Superposed on this are the random emissions arising out of spontaneous emission and it is this spontaneous emission which limits the ultimate monochromaticity of the laser (see Section 7.5).

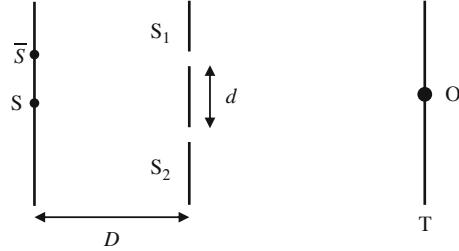
In contrast to $\Delta\nu \sim 10^{10}$ Hz, for an ordinary source of light, for a well-controlled laser one can obtain $\Delta\nu \sim 500$ Hz, which gives $\tau_c \sim 2$ ms. The corresponding coherence length is about 600 km. Such long coherence lengths imply that the laser could be used for performing interference experiments with very large path differences.

For a laser oscillating in many modes, the monochromaticity depends obviously on the number of oscillating modes (see Problem 7.3). Also for a pulsed laser, the minimum linewidth is limited by the duration of the pulse. Thus for a 1 ps pulse, the coherence time is 1 ps and the spectral width would be about 10^{12} Hz.

10.3.2 Spatial Coherence

In order to understand the concept of spatial coherence, we consider the Young's double-hole experiment as shown in Fig. 10.10. S represents a source placed in front of a screen with two holes S_1 and S_2 and the interference pattern between the

Fig. 10.10 Young's double-slit experimental arrangement



waves emanating from S_1 and S_2 is observed on screen T . We restrict ourselves to the region near O for which the optical path lengths S_1O and S_2O are equal. If S represents a point source then it illuminates the pinholes S_1 and S_2 with spherical waves. Since the holes S_1 and S_2 are being illuminated coherently, the interference fringes formed near O will be of good contrast. Consider now another point source \bar{S} placed near S and assume that the waves from S and \bar{S} have no phase relationship. In such a case the interference pattern observed on the screen T will be a superposition of the *intensity* distributions of the interference patterns formed due to S and \bar{S} . If \bar{S} is moved slowly away from S , the contrast in the interference pattern on T will become poorer because of the fact that the interference pattern produced by \bar{S} is slightly shifted in relation to that produced by S . For a particular separation, the interference maximum produced by S falls on the interference minimum produced by \bar{S} and the minimum produced by S falls on the maximum produced by \bar{S} . For such a position the interference fringe pattern on the screen T is washed away.

In order to obtain an approximate expression for the separation $S\bar{S}$ for disappearance of fringes, we assume that S and O are equidistant from S_1 and S_2 . If the position of \bar{S} is such that the path difference between $\bar{S}S_2$ and $\bar{S}S_1$ is $\lambda/2$ (where λ is the wavelength of light used), then the source \bar{S} produces an interference minimum at O and the two fringe patterns would be out of step. If we assume $S\bar{S} = l$, $S_1S_2 = d$, and the distance between S and the plane of the pinholes is D , we obtain

$$\bar{S}S_2 = \left[D^2 + \left(\frac{d}{2} + l \right)^2 \right]^{1/2} \approx D + \frac{1}{2D} \left(\frac{d}{2} + l \right)^2 \quad (10.3)$$

$$\bar{S}S_1 = \left[D^2 + \left(\frac{d}{2} - l \right)^2 \right]^{1/2} \approx D + \frac{1}{2D} \left(\frac{d}{2} - l \right)^2 \quad (10.4)$$

where we have assumed that $D \gg d, l$. Thus for disappearance of fringes,

$$\bar{S}S_2 - \bar{S}S_1 = \frac{\lambda}{2} \approx \frac{lD}{D}$$

or

$$l \approx \frac{\lambda D}{2d} \quad (10.5)$$

For an extended source made up of independent point sources, one may say that good interference fringes will be observed as long as

$$l \ll \frac{\lambda D}{d} \quad (10.6)$$

Equivalently for a given source of width l , interference fringes of good contrast will be formed by interference of light from two point sources S_1 and S_2 separated by a distance d as long as

$$d \ll \frac{\lambda D}{l}$$

Since l/d is the angle (say θ) subtended by the source at the slits above equation can also be written as

$$d \ll \frac{\lambda}{\theta} \quad (10.7)$$

The distance l_w ($=\lambda/\theta$) is referred to as the lateral coherence width. It can be seen from Eq. (10.7) that l_w depends inversely on θ .

Example 10.1 The angle subtended by sun on the earth is 32 s of arc which is approximately 0.01 radians. Thus assuming a wavelength of 500 nm, the lateral coherence width of the sun would be 50 μm . Thus if we have a pair of pinholes separated by a distance much less than 50 μm , and illuminated by the sun, interference pattern of good contrast will be obtained on the screen.

Using ordinary extended sources, one must pass the light through a pinhole in order to produce a spatially coherent light beam. In contrast, the laser beam is highly spatially coherent. For example, Fig. 10.11 shows the interference pattern obtained by Nelson and Collins (1961) by placing a pair of slits of width 7.5 μm separated by a distance of 54 μm on the end of a ruby rod of a ruby laser. The interference pattern agrees with the theoretical prediction to within 20%. To show that the spatial

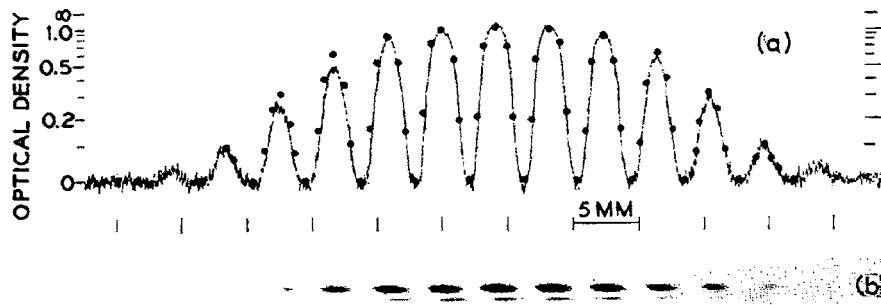


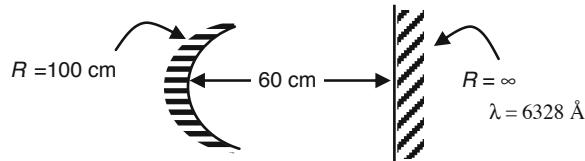
Fig. 10.11 Interference fringes observed by placing a pair of slits in front of a ruby laser showing the spatial coherence of the laser beam (Reprinted with permission from D.F. Nelson and R.J. Collins, Spatial coherence in the output of a maser, J. Appl. Phys. 32 (1961) 739. © 1961 American Institute of Physics)

coherence is indeed due to laser action, they showed that below threshold no interference pattern was observed; only a uniform darkening of the photographic plate was obtained.

Problem 10.3 A 1550 nm semiconductor laser emits an elliptical Gaussian beam (i.e., the spot sizes in the x - and y -directions are not equal) with minimum spot sizes of 100 and 10 μm in two orthogonal directions at the output facet of the laser. At what distance from the facet does the beam become circular and what is the spot size at this point?

Problem 10.4 Consider a symmetric spherical resonator consisting of two concave mirrors of radii of curvature 1 m and separated by 20 cm operating at 1 mm. What will be the angle of divergence of the laser beam emanating from such a laser?

Problem 10.5 Consider an optical resonator shown below:



- Calculate the intensity distribution of the fundamental Gaussian mode at the plane mirror and at the spherical mirror
- If both mirrors are partially reflecting, which of the beams coming out (one from the plane mirror and the other from the spherical mirror) would have a larger diffraction divergence?
- If the same resonator is used for oscillation at 1.5 μm by what approximate factor would the diffraction divergence of the beam at 1.5 μm increase or decrease compared to 1 μm ?

Problem 10.6 What is the approximate angular divergence of the output beam from a He–Ne laser (operating at 6328 Å) having a 1 m long confocal cavity?

Problem 10.7 A parallel laser beam with a diameter of 2 mm and a power of 10 W falls on a convex lens of diameter 25 mm and focal length 10 mm. If the wavelength of the laser beam is 500 nm, estimate the intensity at the focused spot?

Problem 10.8 Estimate the intensity levels produced on the retina if someone accidentally happens to look straight into a laser beam of 10 mW power and a diameter of 2 mm. Assume the pupil diameter to be 5 mm and the wavelength of light as 600 nm.

Problem 10.9 In continuation, estimate the intensity levels produced on the retina if someone accidentally looks at the sun. Assume the pupil diameter to be 2 mm and the wavelength of light as 600 nm. Also calculate the intensity levels produced on the retina when someone looks straight at a 60 W bulb. Take reasonable values of various parameters.

Problem 10.10 An interference experiment is to be conducted using a He–Ne laser. We have two lasers one oscillating in a single longitudinal mode with a linewidth of 10 MHz and the other with two modes with linewidths of 10 MHz and separated by a frequency of 600 MHz. Estimate the minimum path difference for which the interference pattern will disappear when either of the lasers is used.

Problem 10.11 A He–Ne laser of length 20 cm oscillates in two longitudinal modes. If the output of the laser is incident on a photodetector (a device that converts light into electrical current) whose output current is proportional to the incident intensity, what will be the time variation of the output current?

Problem 10.12 Consider a laser beam of circular cross section of diameter 10 cm and of wavelength 800 nm pointed toward the moon which is at a distance of 3.76×10^5 km. What will be the approximate diameter of the laser spot on the moon? Neglect effects due to atmospheric turbulence, etc. [Ans: ~ 6 km].

Problem 10.13 Consider a lens having a focal length of 5 mm and an f-number of 2. Obtain the area of the focused spot using a laser at 650 nm and a laser at 400 nm.

Problem 10.14 Consider two sources, one having a spectral width of 0.1 nm and the other with a spectral width of 0.001 nm. Obtain the coherence length of the two sources.

Problem 10.15 Consider a laser emitting pulses of duration 100 fs. What would be the approximate spectral width of the pulse. What will be the physical length of the optical pulse in free space?

Problem 10.16 The breakdown electric field of air is $30 \text{ V}/\mu\text{m}$. To what intensity level does this correspond?

Chapter 11

Some Laser Systems

11.1 Introduction

In this chapter we shall discuss some specific laser systems and their important operating characteristics. The systems that we shall consider are some of the more important lasers that are in widespread use today for different applications. The lasers considered are

- (a) solid-state lasers: ruby, Nd:YAG, Nd:glass;
- (b) gas lasers: He-Ne, argon ion, and CO₂;
- (c) liquid lasers: dyes;
- (d) excimer lasers;

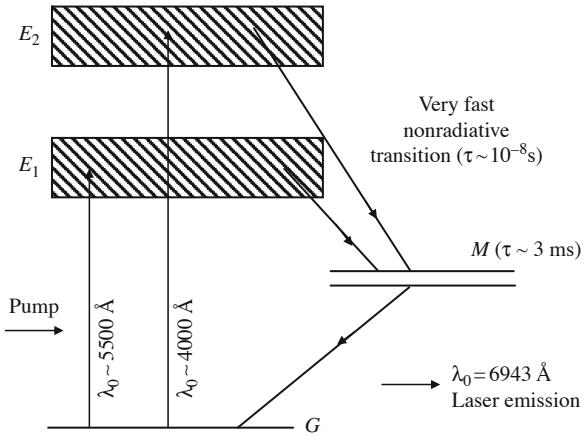
In Chapters 12–14 we shall discuss in detail fiber lasers, semiconductor lasers and coherent sources based on non-linear optical effect, namely parametric oscillators.

11.2 Ruby Lasers

The first laser to be operated successfully was the ruby laser which was fabricated by Maiman in 1960. Ruby, which is the lasing medium, consists of a matrix of aluminum oxide in which some of the aluminum ions are replaced by chromium ions. It is the energy levels of the chromium ions which take part in the lasing action. Typical concentrations of chromium ions are ~0.05% by weight. The energy level diagram of the chromium ion is shown in Fig. 11.1. As is evident from figure this a three-level laser.¹ The pumping of the chromium ions is performed with the help of flash lamp (e.g., a xenon or krypton flashlamp) and the chromium ions in the ground state absorb radiation around wavelengths of 5500 Å and 4000 Å and are excited to the levels marked E_1 and E_2 . The chromium ions excited to these levels relax rapidly through a non-radiative transition (in a time $\sim 10^{-8}$ – 10^{-9} s) to the level

¹The level M actually consists of a pair of levels corresponding to wavelengths of 6943 and 6929 Å. However, laser action takes place only on the 6943 Å line because of higher inversion.

Fig. 11.1 The energy levels of the chromium ions in the ruby laser



marked M which is the upper laser level. The level M is a metastable level with a lifetime of ~ 3 ms. Laser emission occurs between level M and the ground state G at an output wavelength of $\lambda_0 = 6943 \text{ \AA}$.

The flashlamp operation of the laser leads to a pulsed output of the laser. As soon as the flashlamp stops operating the population of the upper level is depleted very rapidly and lasing action stops till the arrival of the next flash. Even during the short period of a few tens of microseconds in which the laser is oscillating, the output is a highly irregular function of time with the intensity having random amplitude fluctuations of varying duration as shown in Fig. 11.2. This is called laser spiking, the formation of which can be understood as follows: when the pump is turned on, the intensity of light at the laser transition is small and hence the pump builds up the inversion rapidly. Although under steady-state conditions the inversion cannot exceed the threshold inversion, on a transient basis it can go beyond the threshold value due to the absence of sufficient laser radiation in the cavity which causes stimulated emission. Thus the inversion goes beyond threshold when the radiation density in the cavity builds up rapidly. Since the inversion is greater than threshold, the radiation density goes beyond the steady-state value which in turn depletes the upper level population and reduces the inversion below threshold. This leads to an interruption of laser oscillation till the pump can again create an inversion beyond threshold. This cycle repeats itself to produce the characteristic spiking in lasers.

Figure 11.3 shows a typical setup of a flashlamp pumped pulsed ruby laser. The helical flashlamp is surrounded by a cylindrical reflector to direct the pump light onto the ruby rod efficiently. The ruby rod length is typically 2–20 cm with diameters of 0.1–2 cm. As we have seen in Section 5.3, typical input electrical energies required are in the range of 10–20 kJ. In addition to the helical flashlamp pumping scheme shown in Fig. 11.3, one may use other pumping schemes such as that shown in Fig. 11.4 in which the pump lamp and the laser rod are placed along the foci of

Fig. 11.2 Temporal output power variations of a ruby laser beam leading to what is referred to as laser spiking. The three figures show regular spiking, partially regular spiking and irregular spiking. (Adapted with permission from Sacchi and Svelto (1965) © 1965 IEEE)

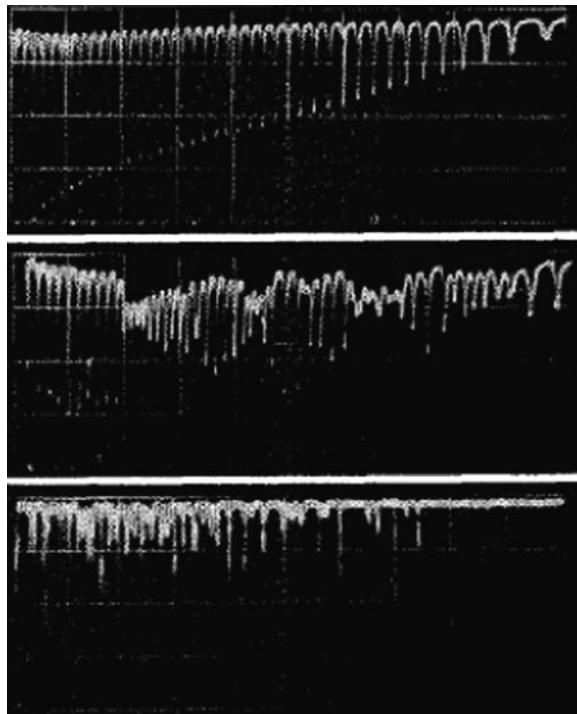


Fig. 11.3 A typical setup of a flashlamp pumped-pulsed ruby laser. The flashlamp is covered by a cylindrical reflector for efficient coupling of the pump light to the ruby rod

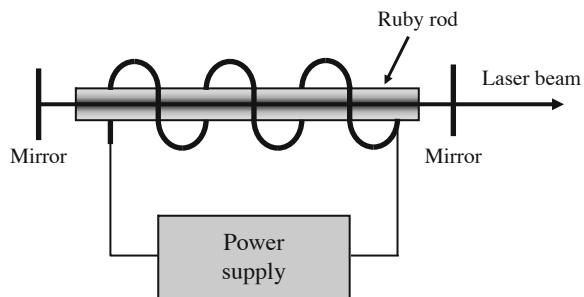
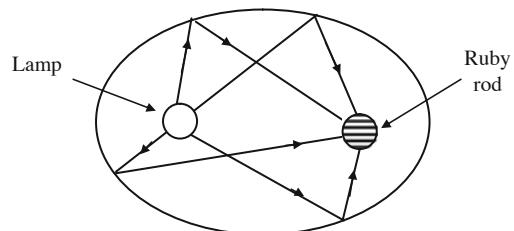


Fig. 11.4 Elliptical pump cavity in which the lamp and the ruby rod are placed along the foci of the elliptical cylindrical reflector



an elliptical cylindrical reflector. It is well known that the elliptical reflector focuses the light emerging from one focus into the other focus of the ellipse, thus leading to an efficient focusing of pump light on the laser rod.

In spite of the fact that the ruby laser is a three-level laser, it still is one of the important practical lasers. The absorption bands of ruby are very well matched with the emission spectra of practically available flashlamps so that an efficient use of the pump can be made. It also has a favorable combination of a long lifetime and a narrow linewidth. The ruby laser is also attractive from an application point of view since its output lies in the visible region where photographic emulsions and photodetectors are much more sensitive than they are in the infrared region. Ruby lasers find applications in pulsed holography, in laser ranging, etc.

11.3 Neodymium-Based Lasers

The Nd:YAG laser (YAG stands for yttrium aluminum garnet which is $\text{Y}_3\text{Al}_5\text{O}_{12}$) and the Nd:glass laser are two very important solid-state laser systems in which the energy levels of the neodymium ion take part in laser emission. They both correspond to a four-level laser. Using neodymium ions in a YAG or glass host has specific advantages and applications.

- (a) Since glass has an amorphous structure the fluorescent linewidth of emission is very large leading to a high value of the laser threshold. On the other hand YAG is a crystalline material and the corresponding linewidth is much smaller which implies much lower thresholds for laser oscillation.
- (b) The fact that the linewidth in the case of the glass host is much larger than in the case of the YAG host can be made use of in the production of ultrashort pulses using mode locking since as discussed in Section 7.7.3, the pulsewidth obtainable by mode locking is the inverse of the oscillating linewidth.
- (c) The larger linewidth in glass leads to a smaller amplification coefficient and thus the capability of storing a larger amount of energy before the occurrence of saturation. This is especially important in obtaining very high-energy pulses using *Q*-switching.
- (d) Other advantages of the glass host are the excellent optical quality and excellent uniformity of doping that can be obtained and also the range of glasses with different properties that can be used for solving specific design problems.
- (e) As compared to YAG, glass has a much lower thermal conductivity which may lead to induced birefringence and optical distortion.

From the above discussion we can see that for continuous or very high pulse repetition rate operation the Nd:YAG laser will be preferred over Nd:glass. On the other hand for high energy-pulsed operation, Nd:glass lasers maybe preferred. In the following we discuss some specific characteristics of Nd:YAG and Nd:glass laser systems.

11.3.1 Nd:YAG Laser

The Nd:YAG laser is a four-level laser and the energy level diagram of the neodymium ion is shown in Fig. 11.5. The laser emission occurs at $\lambda_0 \approx 1.06 \mu\text{m}$. Since the energy difference between the lower laser level and the ground level is $\sim 0.26 \text{ eV}$, the ratio of its population to that of the ground state at room temperature ($T = 300 \text{ K}$) is $e^{-\Delta E / k_B T} \approx e^{-9} \ll 1$. Thus the lower laser level is almost unpopulated and hence inversion is easy to achieve. The main pump bands for excitation of the neodymium ions are in the 0.81 and 0.75 μm wavelength regions and pumping is done using arc lamps (e.g., the Krypton arc lamp). Typical neodymium ion concentrations used are $\sim 1.38 \times 10^{20} \text{ cm}^{-3}$. The spontaneous lifetime corresponding to the laser transition is 550 μs and the emission line corresponds to homogeneous broadening and has a width $\Delta v \sim 1.2 \times 10^{11} \text{ Hz}$ which corresponds to $\Delta\lambda \sim 4.5 \text{ \AA}$. We have shown in Section 5.4 that the Nd:YAG laser has a much lower threshold of oscillation than a ruby laser.

With the availability of high-power compact and efficient semiconductor lasers, efficient pumping of Nd ions to upper laser level can be accomplished using laser diodes. This leads to very compact diode pumped Nd-based lasers. Diode laser pumping is simpler than lamp pumping and also produces much less heat in the laser medium leading to increased overall efficiency. Since the laser diode output is narrow band unlike a normal lamp, the output at 808 nm can be efficiently used for pumping. Typical output powers of 150 W are commercially available. In fact an intracavity second-harmonic generator can efficiently convert the laser wavelength to 532 nm (the second harmonic of 1064 nm of Nd:YAG) leading to very efficient green lasers.

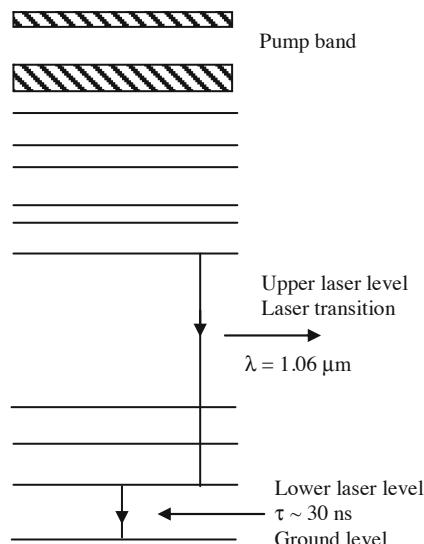


Fig. 11.5 The energy levels of neodymium ion in the Nd:YAG laser

Nd:YAG lasers find many applications in range finders, illuminators with *Q*-switched operation giving about 10–50 pulses per second with output energies in the range of 100 mJ per pulse, and pulse width ~ 10 ns. They also find applications in resistor trimming, scribing, micromachining operations as well as welding, hole drilling, etc.

11.3.2 Nd:Glass Laser

The Nd:glass laser is again a four-level laser system with a laser emission around 1.06 μm . Typical neodymium ion concentrations are $\sim 2.8 \times 10^{20} \text{ cm}^{-3}$ and various silicate and phosphate glasses are used as the host material. Since glass has an amorphous structure different neodymium ions situated at different sites have slightly different surroundings. This leads to an inhomogeneous broadening and the resultant linewidth is $\Delta\nu \sim 7.5 \times 10^{12} \text{ Hz}$ which corresponds to $\Delta\lambda \sim 260 \text{ \AA}$. This width is much larger than in Nd:YAG lasers and consequently the threshold pump powers are also much higher. The spontaneous lifetime of the laser transition is $\sim 300 \mu\text{s}$.

Nd-doped fiber lasers are also efficient. Chapter 12 discusses fiber lasers and primarily erbium-doped fiber lasers. But similar analysis can also be carried out for Nd-doped fiber lasers.

Nd:glass lasers are more suitable for high energy-pulsed operation such as in laser fusion where the requirement is of subnanosecond pulses with an energy content of several kilojoules (i.e., peak powers of several tens of terawatts). Other applications are in welding or drilling operations requiring high pulse energies.

Table 11.1 gives a comparison of some important characteristics of ruby, Nd:YAG, and Nd:glass laser systems.

Table 11.1 Comparison of ruby, Nd:YAG, and Nd:glas laser systems

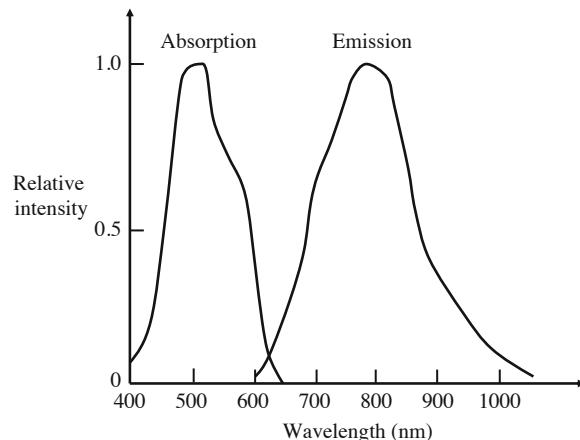
Laser	Ruby	Nd:YAG	Nd:glass
Wavelength (\AA)	6943	10,641	10,623
Spontaneous lifetime (μs)	3000	240	300
Active ion concentration (cm^{-3})	1.58×10^{19}	1.38×10^{20}	2.83×10^{20}
Linewidth (GHz) (\AA)	330 5.5	120 4.0	7500 260
Population inversion density for 1% gain/ $\text{cm}(\text{cm}^{-3})$	4×10^7 $+ 7.6 \times 10^{18}$	1.1×10^{16}	3.3×10^{17}
Index of refraction (n) (at laser λ)	$n_o = 1.763$ $n_e = 1.755$	1.82	1.55
Major pump bands (\AA)	4040 5540	5800 7500	5800 7500
		8100	8100

Table adapted from Koechner (1976).

11.4 Titanium Sapphire Laser

Titanium sapphire (Ti:sapphire) laser is one of the most important solid-state lasers since it is a continuously tunable laser from about 650 to 1100 nm and due to its large gain bandwidth can produce mode-locked pulses in the tens of femtosecond regime. Titanium sapphire laser consists of titanium-doped sapphire (Al_2O_3), in which the energy levels of titanium take part in laser action. The titanium sapphire laser is pumped by another laser, usually argon ion laser emitting in the wavelength region of 514 nm; other lasers such as frequency-doubled Nd:YAG laser emitting 532 nm wavelength is also used for pumping the laser. Figure 11.6 shows the absorption and emission spectra of titanium sapphire. The broad absorption and emission spectra are evident. The broad emission spectrum leads to a broad gain spectrum which in turn allows for ultrashort pulse generation using mode locking techniques.

Fig. 11.6 Absorption and emission spectrum of titanium sapphire laser medium [Adapted from <http://www.olympusmicro.com/primer/techniques/fluorescence/multiphoton/images/tisapspectra.jpg>]



11.5 The He–Ne Laser

The first gas laser to be operated successfully was the He–Ne laser. As we discussed earlier in solid-state lasers, the pumping is usually done using a flashlamp or a continuous high-power lamp. Such a technique is efficient if the lasing system has broad absorption bands. In gas lasers since the atoms are characterized by sharp energy levels as compared to those in solids, one generally uses an electrical discharge to pump the atoms.

The He–Ne laser consists of a long and narrow discharge tube (diameter ~ 2 –8 mm and length 10–100 cm) which is filled with helium and neon with typical pressures of 1 torr² and 0.1 torr. The actual lasing atoms are the neon atoms and as we shall discuss helium is used for a selective pumping of the upper laser level

²Torr is a unit of pressure and 1 torr = 1 mm Hg.

Fig. 11.7 A typical He–Ne laser with external mirrors. The ends of the discharge tube are fitted with Brewster windows

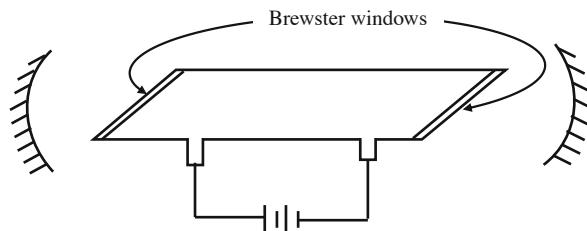
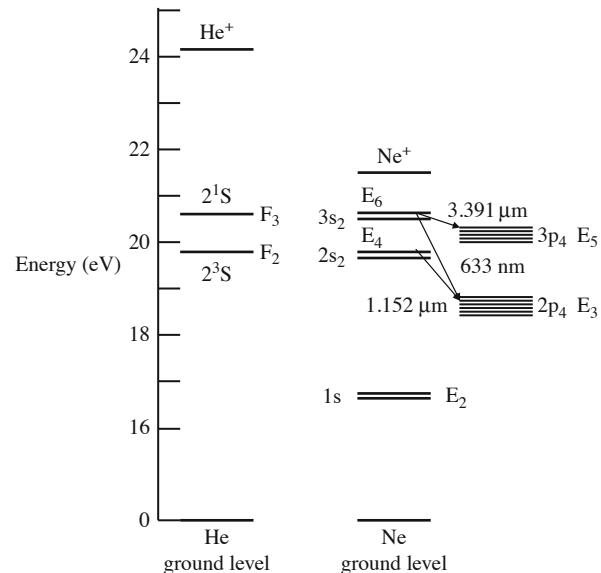


Fig. 11.8 Low lying energy levels of helium and neon taking part in the He–Ne laser



of neon. The laser resonator may consist of either internal or external mirrors (see Fig. 11.7). Figure 11.8 shows the energy levels of helium and neon. When an electrical discharge is passed through the gas, the electrons which are accelerated down the tube collide with helium and neon atoms and excite them to higher energy levels. The helium atoms tend to accumulate at levels F_2 and F_3 due to their long lifetimes of $\sim 10^{-4}$ and 5×10^{-6} s, respectively. Since the levels E_4 and E_6 of neon atoms have almost the same energy as F_2 and F_3 , excited helium atoms colliding with neon atoms in the ground state can excite the neon atoms to E_4 and E_6 . Since the pressure of helium is ten times that of neon, the levels E_4 and E_6 of neon are selectively populated as compared to other levels of neon.

Transition between E_6 and E_3 produces the very popular 6328 Å line of the He–Ne laser. Neon atoms de-excite through spontaneous emission from E_3 to E_2 (lifetime $\sim 10^{-8}$ s). Since this time is shorter than the lifetime of level E_6 (which is $\sim 10^{-7}$ s) one can achieve steady-state population inversion between E_6 and E_3 .

Level E_2 is metastable and thus tends to collect atoms. The atoms from this level relax back to the ground level mainly through collisions with the walls of the tube. Since E_2 is metastable it is possible for the atoms in this level to absorb the spontaneously emitted radiation in the $E_3 \rightarrow E_2$ transition to be re-excited to E_3 . This will have the effect of reducing the inversion. It is for this reason that the gain in this laser transition is found to increase with decreasing tube diameter.

The other two important wavelengths from the He-Ne laser are 1.15 and 3.39 μm , which correspond to the $E_4 \rightarrow E_3$ and $E_6 \rightarrow E_5$ transitions. It is interesting to observe that both 3.39 μm and 6328 Å transitions share the same upper laser level. Now since the 3.39 μm transition corresponds to a much lower frequency than the 6328 Å line, the Doppler broadening is much smaller at 3.39 μm and also since gain depends inversely on v^2 (see Eq. (4.26)), the gain at 3.39 μm is much higher than at 6328 Å. Thus due to the very large gain, oscillations will normally tend to occur at 3.39 μm rather than at 6328 Å. Once the laser starts to oscillate at 3.39 μm , further build up of population in E_6 is not possible. The laser can be made to oscillate at 6328 Å by either using optical elements in the path which strongly absorb the 3.39 μm wavelength or increasing the linewidth through the Zeeman effect by applying an inhomogeneous magnetic field across the tube.

If the resonator mirrors are placed outside the discharge tube then reflections from the ends of the discharge tube can be avoided by placing the windows at the Brewster angle (see Fig. 11.7). In such a case the beam polarized in the plane of incidence suffers no reflection at the windows while the perpendicular polarization suffers reflection losses. This leads to a polarized output of the laser.

11.6 The Argon Ion Laser

In an argon ion laser, one uses the energy levels of the ionized argon atom and the laser emits various discrete lines in the 3500–5200 Å wavelength region. Figure 11.9

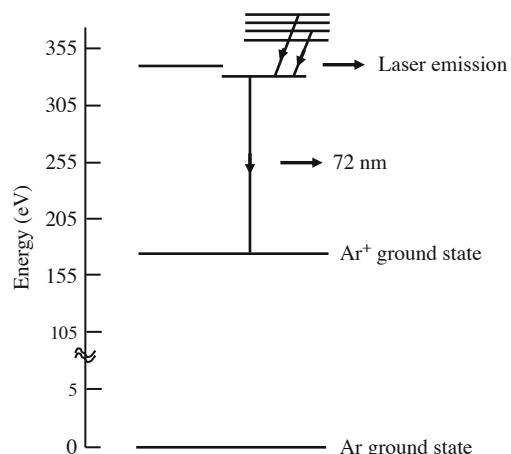


Fig. 11.9 Some of the levels taking part in the laser transition corresponding to the argon ion laser

shows some of the energy levels taking part in the laser transition. The argon atoms have to be first ionized and then excited to the higher energy levels of the ion. Because of the large energies involved in this, the argon ion laser discharge is very intense; typical values being 40 A at 165 V. A particular wavelength out of the many possible lines is chosen by placing a dispersive prism inside the cavity close to one of the mirrors. Rotation of the prism-mirror system provides feedback only at the wavelength which is incident normally on the mirror. Typical output power in a continuous wave argon ion laser is 3–5 W. Some of the important emission wavelengths include 5145 Å, 4965 Å, 4880 Å, 4765 Å, and 4579 Å.

11.7 The CO₂ Laser

The lasers discussed above use transitions among the various excited electronic states of an atom or an ion. In a CO₂ laser one uses the transitions occurring between different vibrational states of the carbon dioxide molecule. Figure 11.10 shows the carbon dioxide molecule consisting of a central carbon atom with two oxygen atoms attached one on either side. Such a molecule can vibrate in the three independent modes of vibration shown in Fig. 11.10.

These correspond to the symmetric stretch, the bending, and the asymmetric stretch modes. Each of these modes is characterized by a definite frequency of vibration. According to basic quantum mechanics these vibrational degrees of freedom are quantized, i.e., when a molecule vibrates in any of the modes it can have only a discrete set of energies. Thus if we call ν_1 the frequency corresponding to the symmetric stretch mode then the molecule can have energies of only

$$E_1 = \left(m + \frac{1}{2} \right) h\nu_1, \quad m = 0, 1, 2, \dots \quad (11.1)$$

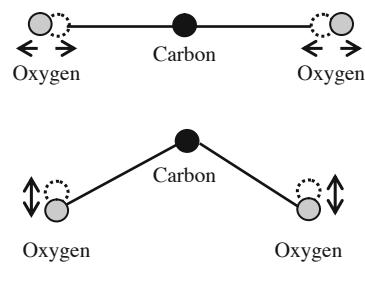
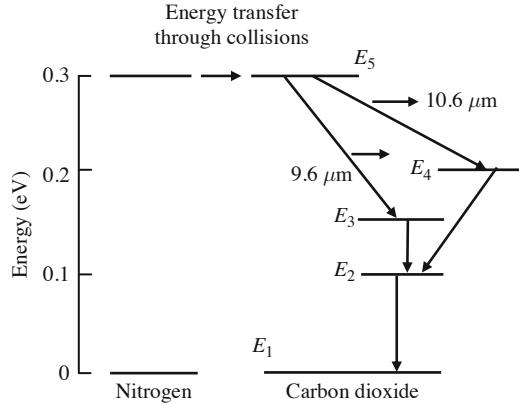


Fig. 11.10 The three independent modes of vibration of the carbon dioxide molecule

Fig. 11.11 The low lying vibrational levels of nitrogen and carbon dioxide molecules. Energy transfer from excited nitrogen molecules to carbon dioxide molecules results in the excitation of carbon dioxide molecules. Important lasing transitions occur at 9.6 and 10.6 μm



when it vibrates in the symmetric stretch mode. Thus the degree of excitation is characterized by the integer m when the carbon dioxide molecule vibrates in the symmetric stretch mode. In general, since the carbon dioxide molecule can vibrate in a combination of the three modes the state of vibration can be described by three integers ($m n q$); the three integers correspond, respectively, to the degree of excitation in the symmetric stretch, bending, and asymmetric stretch modes, respectively. Figure 11.11 shows the various vibrational energy levels taking part in the laser transition.

The laser transition at 10.6 μm occurs between the (001) and (100) levels of carbon dioxide. The excitation of the carbon dioxide molecules to the long-lived level (001) occurs both through collisional transfer from nearly resonant excited nitrogen molecules and also from the cascading down of carbon dioxide molecules from higher energy levels.

The CO₂ laser possesses an extremely high efficiency of ~30%. This is because of efficient pumping to the (001) level and also because all the energy levels involved are close to the ground level. Thus the atomic quantum efficiency which is the ratio of the energy difference corresponding to the laser transition to the energy difference of the pump transition, i.e.,

$$\eta = \frac{E_5 - E_4}{E_5 - E_1}$$

is quite high (~45%). Thus a large portion of the input power can be converted into useful laser power.

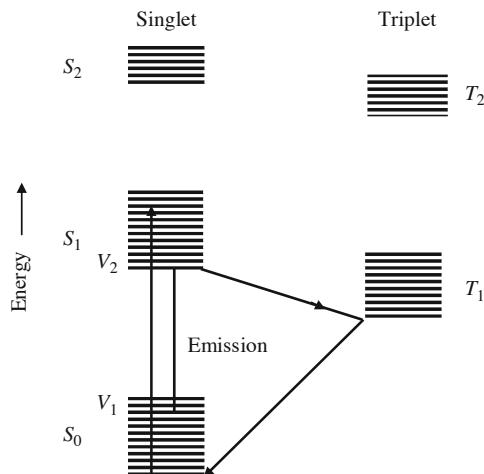
Output powers of several watts to several kilowatts can be obtained from CO₂ lasers. High-power CO₂ lasers find applications in materials processing, welding, hole drilling, cutting, etc., because of their very high output power. In addition, the atmospheric attenuation is low at 10.6 μm which leads to some applications of CO₂ lasers in open air communications.

11.8 Dye Lasers

One of the most widely used tunable lasers in the visible region is the organic dye laser. The dyes used in the lasers are organic substances which are dissolved in solvents such as water, ethyl alcohol, methanol, and ethylene glycol. These dyes exhibit strong and broad absorption and fluorescent spectra and because of this they can be made tunable. By choosing different dyes one can obtain tunability from 3000 Å to 1.2 μm.

The levels taking part in the absorption and lasing correspond to the various vibrational sublevels of different electronic states of the dye molecule. Figure 11.12 shows a typical energy level diagram of a dye in which S_0 is the ground state, S_1 is the first excited singlet state, and T_1, T_2 are the excited triplet states of the dye molecule. Each state consists of a large number of closely spaced vibrational and rotational sublevels. Because of strong interaction with the solvent, the closely spaced sublevels are collision broadened to such an extent that they almost form a continuum.

Fig. 11.12 Typical energy level diagram of a dye molecule



When dye molecules in the solvent are irradiated by visible or ultraviolet radiation then the molecules are excited to the various sublevels of the state S_1 . Due to collisions with the solvent molecules, the molecules excited to higher vibrational and rotational states of S_1 relax very quickly (in times $\sim 10^{-11}$ – 10^{-12} s) to the lowest level V_2 of the state S_1 . Molecules from this level emit spontaneously and de-excite to the different sublevels of S_0 . Thus the fluorescent spectrum is found to be red shifted against the absorption spectrum.

Problems

Problem 11.1 Consider mode locking an Nd:YAG laser and an Nd:glass laser. Compare the minimum pulse widths obtainable using these two lasers. If the length of the resonators in both cases is 20 cm, what typical pulse widths are possible by mode locking these two lasers?

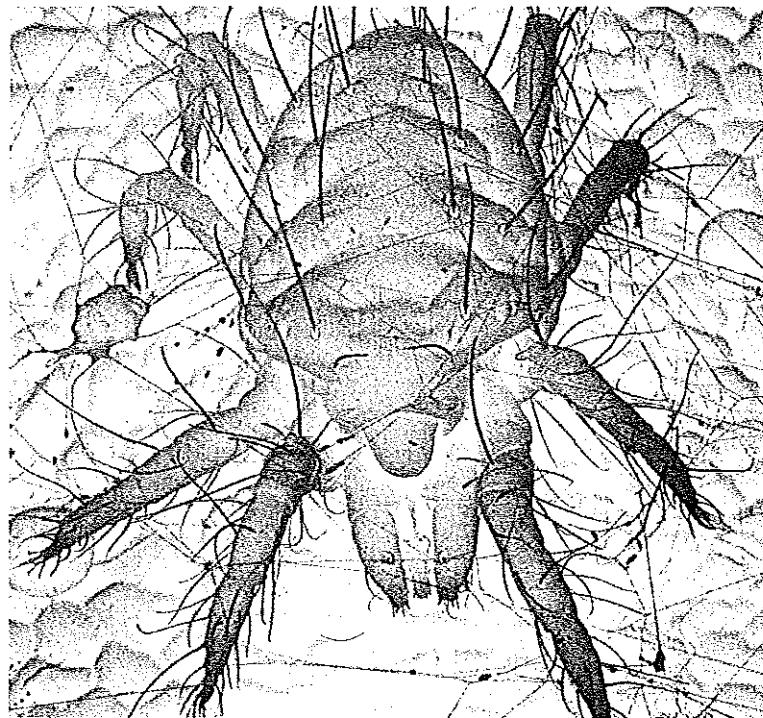
Problem 11.2 Estimate the Doppler-broadened linewidth of a carbon dioxide laser system assuming a temperature of 300 K.

Problem 11.3 Consider a He–Ne laser having a cavity length of 30 cm and oscillating over a bandwidth of 1500 MHz. What will be the coherence length of the laser?

Problem 11.4 Compare the Doppler-broadened linewidth of the 632.8 and 3.39 mm emission lines of He–Ne laser. Assume a temperature of 300 K.

CHAPTER 3

Wave Properties of Particles



In a scanning electron microscope, an electron beam that scans a specimen causes secondary electrons to be ejected in numbers that vary with the angle of the surface. A suitable data display suggests the three-dimensional form of the specimen. The high resolution of this image of a red spider mite on a leaf is a consequence of the wave nature of moving electrons.

3.1 DE BROGLIE WAVES

A moving body behaves in certain ways as though it has a wave nature

3.2 WAVES OF WHAT?

Waves of probability

3.3 DESCRIBING A WAVE

A general formula for waves

3.4 PHASE AND GROUP VELOCITIES

A group of waves need not have the same velocity as the waves themselves

3.5 PARTICLE DIFFRACTION

An experiment that confirms the existence of de Broglie waves

3.6 PARTICLE IN A BOX

Why the energy of a trapped particle is quantized

3.7 UNCERTAINTY PRINCIPLE I

We cannot know the future because we cannot know the present

3.8 UNCERTAINTY PRINCIPLE II

A particle approach gives the same result

3.9 APPLYING THE UNCERTAINTY PRINCIPLE

A useful tool, not just a negative statement

Looking back, it may seem odd that two decades passed between the 1905 discovery of the particle properties of waves and the 1924 speculation that particles might show wave behavior. It is one thing, however, to suggest a revolutionary concept to explain otherwise mysterious data and quite another to suggest an equally revolutionary concept without a strong experimental mandate. The latter is just what Louis de Broglie did in 1924 when he proposed that moving objects have wave as well as particle characteristics. So different was the scientific climate at the time from that around the turn of the century that de Broglie's ideas soon received respectful attention, whereas the earlier quantum theory of light of Planck and Einstein had been largely ignored despite its striking empirical support. The existence of de Broglie waves was experimentally demonstrated by 1927, and the duality principle they represent provided the starting point for Schrödinger's successful development of quantum mechanics in the previous year.

3.1 DE BROGLIE WAVES

A moving body behaves in certain ways as though it has a wave nature

A photon of light of frequency ν has the momentum

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

since $\lambda\nu = c$. The wavelength of a photon is therefore specified by its momentum according to the relation

Photon wavelength
$$\lambda = \frac{h}{p} \quad (3.1)$$

De Broglie suggested that Eq. (3.1) is a completely general one that applies to material particles as well as to photons. The momentum of a particle of mass m and velocity v is $p = \gamma mv$, and its de Broglie wavelength is accordingly

De Broglie wavelength
$$\lambda = \frac{h}{\gamma mv} \quad (3.2)$$



Louis de Broglie (1892–1987), although coming from a French family long identified with diplomacy and the military and initially a student of history, eventually followed his older brother Maurice in a career in physics. His doctoral thesis in 1924 contained the proposal that moving bodies have wave properties that complement their particle properties: these "seemingly incompatible conceptions can each represent an

aspect of the truth. . . . They may serve in turn to represent the facts without ever entering into direct conflict." Part of de Broglie's inspiration came from Bohr's theory of the hydrogen atom, in which the electron is supposed to follow only certain orbits around the nucleus. "This fact suggested to me the idea that electrons . . . could not be considered simply as particles but that periodicity must be assigned to them also." Two years later Erwin Schrödinger used the concept of de Broglie waves to develop a general theory that he and others applied to explain a wide variety of atomic phenomena. The existence of de Broglie waves was confirmed in diffraction experiments with electron beams in 1927, and in 1929 de Broglie received the Nobel Prize.

The greater the particle's momentum, the shorter its wavelength. In Eq. (3.2) γ is the relativistic factor

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

As in the case of em waves, the wave and particle aspects of moving bodies can never be observed at the same time. We therefore cannot ask which is the "correct" description. All that can be said is that in certain situations a moving body resembles a wave and in others it resembles a particle. Which set of properties is most conspicuous depends on how its de Broglie wavelength compares with its dimensions and the dimensions of whatever it interacts with.

Example 3.1

Find the de Broglie wavelengths of (a) a 46-g golf ball with a velocity of 30 m/s, and (b) an electron with a velocity of 10^7 m/s.

Solution

(a) Since $v \ll c$, we can let $\gamma = 1$. Hence

$$\lambda = \frac{h}{mv} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{(0.046 \text{ kg})(30 \text{ m/s})} = 4.8 \times 10^{-34} \text{ m}$$

The wavelength of the golf ball is so small compared with its dimensions that we would not expect to find any wave aspects in its behavior.

(b) Again $v \ll c$, so with $m = 9.1 \times 10^{-31}$ kg, we have

$$\lambda = \frac{h}{mv} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{(9.1 \times 10^{-31} \text{ kg})(10^7 \text{ m/s})} = 7.3 \times 10^{-11} \text{ m}$$

The dimensions of atoms are comparable with this figure—the radius of the hydrogen atom, for instance, is 5.3×10^{-11} m. It is therefore not surprising that the wave character of moving electrons is the key to understanding atomic structure and behavior.

Example 3.2

Find the kinetic energy of a proton whose de Broglie wavelength is 1.000 fm = 1.000×10^{-15} m, which is roughly the proton diameter.

Solution

A relativistic calculation is needed unless pc for the proton is much smaller than the proton rest energy of $E_0 = 0.938$ GeV. To find out, we use Eq. (3.2) to determine pc :

$$pc = (\gamma mv)c = \frac{hc}{\lambda} = \frac{(4.136 \times 10^{-15} \text{ eV} \cdot \text{s})(2.998 \times 10^8 \text{ m/s})}{1.000 \times 10^{-15} \text{ m}} = 1.240 \times 10^9 \text{ eV} \\ = 1.2410 \text{ GeV}$$

Since $pc > E_0$ a relativistic calculation is required. From Eq. (1.24) the total energy of the proton is

$$E = \sqrt{E_0^2 + p^2c^2} = \sqrt{(0.938 \text{ GeV})^2 + (1.2340 \text{ GeV})^2} = 1.555 \text{ GeV}$$

The corresponding kinetic energy is

$$KE = E - E_0 = (1.555 - 0.938) \text{ GeV} = 0.617 \text{ GeV} = 617 \text{ MeV}$$

De Broglie had no direct experimental evidence to support his conjecture. However, he was able to show that it accounted in a natural way for the energy quantization—the restriction to certain specific energy values—that Bohr had had to postulate in his 1913 model of the hydrogen atom. (This model is discussed in Chap. 4.) Within a few years Eq. (3.2) was verified by experiments involving the diffraction of electrons by crystals. Before we consider one of these experiments, let us look into the question of what kind of wave phenomenon is involved in the matter waves of de Broglie.

3.2 WAVES OF WHAT?

Waves of probability

In water waves, the quantity that varies periodically is the height of the water surface. In sound waves, it is pressure. In light waves, electric and magnetic fields vary. What is it that varies in the case of matter waves?

The quantity whose variations make up matter waves is called the **wave function**, symbol Ψ (the Greek letter psi). The value of the wave function associated with a moving body at the particular point x, y, z in space at the time t is related to the likelihood of finding the body there at the time.



Max Born (1882–1970) grew up in Breslau, then a German city but today part of Poland, and received a doctorate in applied mathematics at Göttingen in 1907. Soon afterward he decided to concentrate on physics, and was back in Göttingen in 1909 as a lecturer. There he worked on various aspects of the theory of crystal lattices, his “central interest” to which he often returned in later years. In 1915, at

Planck’s recommendation, Born became professor of physics in Berlin where, among his other activities, he played piano to Einstein’s violin. After army service in World War I and a period at Frankfurt University, Born was again in Göttingen, now as professor of physics. There a remarkable center of theoretical physics developed under his leadership: Heisenberg and Pauli were among his assistants and Fermi, Dirac, Wigner, and Goeppert were among those who worked with him, just to name future Nobel Prize winners. In those days, Born wrote, “There was complete freedom of teaching and learning in German universities, with no class examinations, and no control of students. The University just offered lectures and the student had to decide for himself which he wished to attend.”

Born was a pioneer in going from “the bright realm of classical physics into the still dark and unexplored underworld of the new quantum mechanics;” he was the first to use the latter term. From Born came the basic concept that the wave function Ψ of a particle is related to the probability of finding it. He began with an idea of Einstein, who “sought to make the duality of particles (light quanta or photons) and waves comprehensible by interpreting the square of the optical wave amplitude as probability density for the occurrence of photons. This idea could at once be extended to the Ψ -function: $|\Psi|^2$ must represent the probability density for electrons (or other particles). To assert this was easy; but how was it to be proved? For this purpose atomic scattering processes suggested themselves.” Born’s development of the quantum theory of atomic scattering (collisions of atoms with various particles) not only verified his “new way of thinking about the phenomena of nature” but also founded an important branch of theoretical physics.

Born left Germany in 1933 at the start of the Nazi period, like so many other scientists. He became a British subject and was associated with Cambridge and then Edinburgh universities until he retired in 1953. Finding the Scottish climate harsh and wishing to contribute to the democratization of postwar Germany, Born spent the rest of his life in Bad Pyrmont, a town near Göttingen. His textbooks on modern physics and on optics were standard works on these subjects for many years.

The wave function Ψ itself, however, has no direct physical significance. There is a simple reason why Ψ cannot be interpreted in terms of an experiment. The probability that something be in a certain place at a given time must lie between 0 (the object is definitely not there) and 1 (the object is definitely there). An intermediate probability, say 0.2, means that there is a 20% chance of finding the object. But the amplitude of a wave can be negative as well as positive, and a negative probability, say -0.2 , is meaningless. Hence Ψ by itself cannot be an observable quantity.

This objection does not apply to $|\Psi|^2$, the square of the absolute value of the wave function, which is known as **probability density**:

The probability of experimentally finding the body described by the wave function Ψ at the point x, y, z , at the time t is proportional to the value of $|\Psi|^2$ there at t .

A large value of $|\Psi|^2$ means the strong possibility of the body's presence, while a small value of $|\Psi|^2$ means the slight possibility of its presence. As long as $|\Psi|^2$ is not actually 0 somewhere, however, there is a definite chance, however small, of detecting it there. This interpretation was first made by Max Born in 1926.

There is a big difference between the probability of an event and the event itself. Although we can speak of the wave function Ψ that describes a particle as being spread out in space, this does not mean that the particle itself is thus spread out. When an experiment is performed to detect electrons, for instance, a whole electron is either found at a certain time and place or it is not; there is no such thing as a 20 percent of an electron. However, it is entirely possible for there to be a 20 percent chance that the electron be found at that time and place, and it is this likelihood that is specified by $|\Psi|^2$.

W. L. Bragg, the pioneer in x-ray diffraction, gave this loose but vivid interpretation: "The dividing line between the wave and particle nature of matter and radiation is the moment 'now.' As this moment steadily advances through time it coagulates a wavy future into a particle past. . . . Everything in the future is a wave, everything in the past is a particle." If "the moment 'now'" is understood to be the time a measurement is performed, this is a reasonable way to think about the situation. (The philosopher Søren Kierkegaard may have been anticipating this aspect of modern physics when he wrote, "Life can only be understood backwards, but it must be lived forwards.")

Alternatively, if an experiment involves a great many identical objects all described by the same wave function Ψ , the *actual density* (number per unit volume) of objects at x, y, z at the time t is proportional to the corresponding value of $|\Psi|^2$. It is instructive to compare the connection between Ψ and the density of particles it describes with the connection discussed in Sec. 2.4 between the electric field E of an electromagnetic wave and the density N of photons associated with the wave.

While the wavelength of the de Broglie waves associated with a moving body is given by the simple formula $\lambda = h/\gamma mv$, to find their amplitude Ψ as a function of position and time is often difficult. How to calculate Ψ is discussed in Chap. 5 and the ideas developed there are applied to the structure of the atom in Chap. 6. Until then we can assume that we know as much about Ψ as each situation requires.

3.3 DESCRIBING A WAVE

A general formula for waves

How fast do de Broglie waves travel? Since we associate a de Broglie wave with a moving body, we expect that this wave has the same velocity as that of the body. Let us see if this is true.

If we call the de Broglie wave velocity v_p , we can apply the usual formula

$$v_p = \nu\lambda$$

to find v_p . The wavelength λ is simply the de Broglie wavelength $\lambda = h/\gamma mv$. To find the frequency, we equate the quantum expression $E = \hbar\nu$ with the relativistic formula for total energy $E = \gamma mc^2$ to obtain

$$\begin{aligned} \hbar\nu &= \gamma mc^2 \\ \nu &= \frac{\gamma mc^2}{h} \end{aligned}$$

The de Broglie wave velocity is therefore

De Broglie phase
velocity $v_p = \nu\lambda = \left(\frac{\gamma mc^2}{h}\right)\left(\frac{h}{\gamma mv}\right) = \frac{c^2}{v}$ (3.3)

Because the particle velocity v must be less than the velocity of light c , the de Broglie waves always travel faster than light! In order to understand this unexpected result, we must look into the distinction between **phase velocity** and **group velocity**. (Phase velocity is what we have been calling wave velocity.)

Let us begin by reviewing how waves are described mathematically. For simplicity we consider a string stretched along the x axis whose vibrations are in the y direction, as in Fig. 3.1, and are simple harmonic in character. If we choose $t = 0$ when the displacement y of the string at $x = 0$ is a maximum, its displacement at any future time t at the same place is given by the formula

$$y = A \cos 2\pi\nu t \quad (3.4)$$

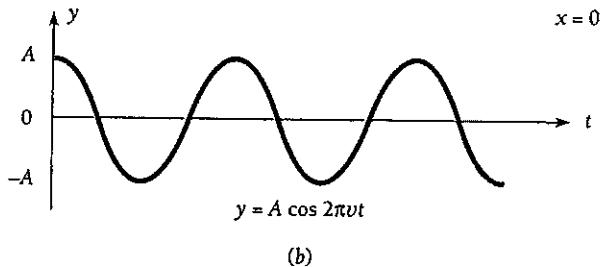
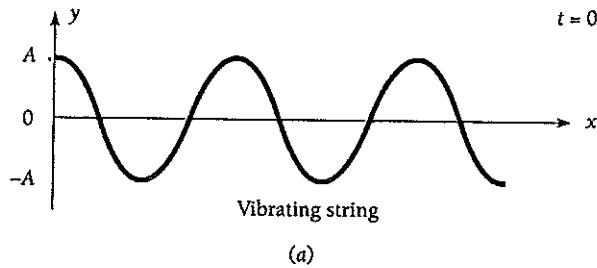


Figure 3.1 (a) The appearance of a wave in a stretched string at a certain time. (b) How the displacement of a point on the string varies with time.

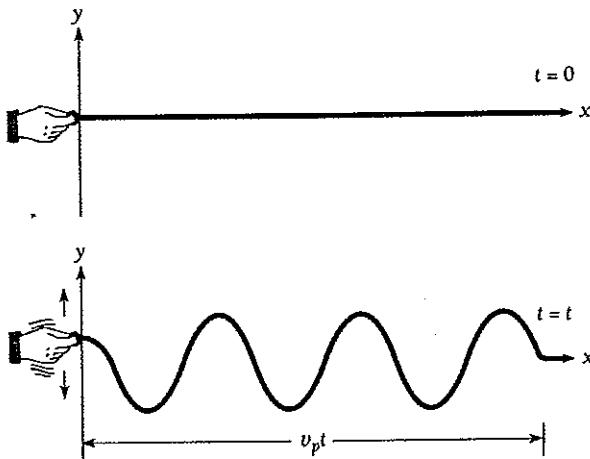


Figure 3.2 Wave propagation.

where A is the amplitude of the vibrations (that is, their maximum displacement on either side of the x axis) and ν their frequency.

Equation (3.4) tells us what the displacement of a single point on the string is as a function of time t . A complete description of wave motion in a stretched string, however, should tell us what y is at any point on the string at any time. What we want is a formula giving y as a function of both x and t .

To obtain such a formula, let us imagine that we shake the string at $x = 0$ when $t = 0$, so that a wave starts to travel down the string in the $+x$ direction (Fig. 3.2). This wave has some speed v_p that depends on the properties of the string. The wave travels the distance $x = v_p t$ in the time t , so the time interval between the formation of the wave at $x = 0$ and its arrival at the point x is x/v_p . Hence the displacement y of the string at x at any time t is exactly the same as the value of y at $x = 0$ at the earlier time $t - x/v_p$. By simply replacing t in Eq. (3.4) with $t - x/v_p$, then, we have the desired formula giving y in terms of both x and t :

$$\text{Wave formula} \quad y = A \cos 2\pi\nu \left(t - \frac{x}{v_p} \right) \quad (3.5)$$

As a check, we note that Eq. (3.5) reduces to Eq. (3.4) at $x = 0$.

Equation (3.5) may be rewritten

$$y = A \cos 2\pi \left(\nu t - \frac{\nu x}{v_p} \right)$$

Since the wave speed v_p is given by $v_p = \nu\lambda$ we have

$$\text{Wave formula} \quad y = A \cos 2\pi \left(\nu t - \frac{x}{\lambda} \right) \quad (3.6)$$

Equation (3.6) is often more convenient to use than Eq. (3.5).

Perhaps the most widely used description of a wave, however, is still another form of Eq. (3.5). The quantities angular frequency ω and wave number k are defined by the formulas

Angular frequency $\omega = 2\pi\nu$ (3.7)

Wave number $k = \frac{2\pi}{\lambda} \quad \omega$ (3.8)

The unit of ω is the radian per second and that of k is the radian per meter. Angular frequency gets its name from uniform circular motion, where a particle that moves around a circle ν times per second sweeps out $2\pi\nu$ rad/s. The wave number is equal to the number of radians corresponding to a wave train 1 m long, since there are 2π rad in one complete wave.

In terms of ω and k , Eq. (3.5) becomes

Wave formula $y = A \cos(\omega t - kx)$ (3.9)

In three dimensions k becomes a vector \mathbf{k} normal to the wave fronts and x is replaced by the radius vector \mathbf{r} . The scalar product $\mathbf{k} \cdot \mathbf{r}$ is then used instead of kx in Eq. (3.9).



3.7 UNCERTAINTY PRINCIPLE 1

We cannot know the future because we cannot know the present

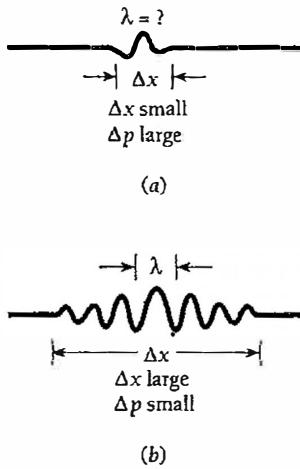


Figure 3.12 (a) A narrow de Broglie wave group. The position of the particle can be precisely determined, but the wavelength (and hence the particle's momentum) cannot be established because there are not enough waves to measure accurately. (b) A wide wave group. Now the wavelength can be precisely determined but not the position of the particle.

To regard a moving particle as a wave group implies that there are fundamental limits to the accuracy with which we can measure such "particle" properties as position and momentum.

To make clear what is involved, let us look at the wave group of Fig. 3.3. The particle that corresponds to this wave group may be located anywhere within the group at a given time. Of course, the probability density $|\Psi|^2$ is a maximum in the middle of the group, so it is most likely to be found there. Nevertheless, we may still find the particle anywhere that $|\Psi|^2$ is not actually 0.

The narrower its wave group, the more precisely a particle's position can be specified (Fig. 3.12a). However, the wavelength of the waves in a narrow packet is not well defined; there are not enough waves to measure λ accurately. This means that since $\lambda = h/\gamma mv$, the particle's momentum γmv is not a precise quantity. If we make a series of momentum measurements, we will find a broad range of values.

On the other hand, a wide wave group, such as that in Fig. 3.12b, has a clearly defined wavelength. The momentum that corresponds to this wavelength is therefore a precise quantity, and a series of measurements will give a narrow range of values. But where is the particle located? The width of the group is now too great for us to be able to say exactly where the particle is at a given time.

Thus we have the **uncertainty principle**:

It is impossible to know both the exact position and exact momentum of an object at the same time.

This principle, which was discovered by Werner Heisenberg in 1927, is one of the most significant of physical laws.

A formal analysis supports the above conclusion and enables us to put it on a quantitative basis. The simplest example of the formation of wave groups is that given in Sec. 3.4, where two wave trains slightly different in angular frequency ω and wave number k were superposed to yield the series of groups shown in Fig. 3.4. A moving body corresponds to a single wave group, not a series of them, but a single wave group can also be thought of in terms of the superposition of trains of harmonic waves. However, an infinite number of wave trains with different frequencies, wave numbers, and amplitudes is required for an isolated group of arbitrary shape, as in Fig. 3.13.

At a certain time t , the wave group $\Psi(x)$ can be represented by the **Fourier integral**

$$\Psi(x) = \int_0^{\infty} g(k) \cos kx \, dk \quad (3.19)$$

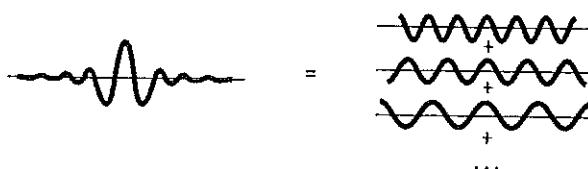


Figure 3.13 An isolated wave group is the result of superposing an infinite number of waves with different wavelengths. The narrower the wave group, the greater the range of wavelengths involved. A narrow de Broglie wave group thus means a well-defined position (Δx smaller) but a poorly defined wavelength and a large uncertainty Δp in the momentum of the particle the group represents. A wide wave group means a more precise momentum but a less precise position.

where the function $g(k)$ describes how the amplitudes of the waves that contribute to $\Psi(x)$ vary with wave number k . This function is called the **Fourier transform** of $\Psi(x)$, and it specifies the wave group just as completely as $\Psi(x)$ does. Figure 3.14 contains graphs of the Fourier transforms of a pulse and of a wave group. For comparison, the Fourier transform of an infinite train of harmonic waves is also included. There is only a single wave number in this case, of course.

Strictly speaking, the wave numbers needed to represent a wave group extend from $k = 0$ to $k = \infty$, but for a group whose length Δx is finite, the waves whose amplitudes $g(k)$ are appreciable have wave numbers that lie within a finite interval Δk . As Fig. 3.14 indicates, the narrower the group, the broader the range of wave numbers needed to describe it, and vice versa.

The relationship between the distance Δx and the wave-number spread Δk depends upon the shape of the wave group and upon how Δx and Δk are defined. The minimum value of the product $\Delta x \Delta k$ occurs when the envelope of the group has the familiar bell shape of a Gaussian function. In this case the Fourier transform happens to be a Gaussian function also. If Δx and Δk are taken as the standard deviations of the respective functions $\Psi(x)$ and $g(k)$, then this minimum value is $\Delta x \Delta k = \frac{1}{2}$. Because wave groups in general do not have Gaussian forms, it is more realistic to express the relationship between Δx and Δk as

$$\Delta x \Delta k \geq \frac{1}{2} \quad (3.20)$$

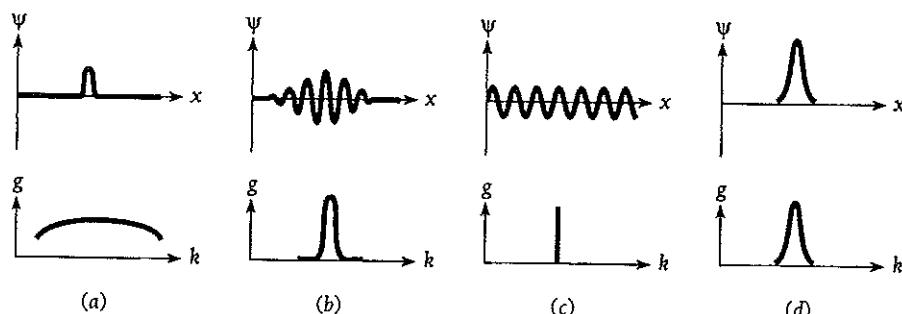


Figure 3.14 The wave functions and Fourier transforms for (a) a pulse, (b) a wave group, (c) a wave train, and (d) a Gaussian distribution. A brief disturbance needs a broader range of frequencies to describe it than a disturbance of greater duration. The Fourier transform of a Gaussian function is also a Gaussian function.

Gaussian Function

When a set of measurements is made of some quantity x in which the experimental errors are random, the result is often a Gaussian distribution whose form is the bell-shaped curve shown in Fig. 3.15. The standard deviation σ of the measurements is a measure of the spread of x values about the mean of x_0 , where σ equals the square root of the average of the squared deviations from x_0 . If N measurements were made,

$$\text{Standard deviation} \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_0)^2}$$

The width of a Gaussian curve at half its maximum value is 2.35σ .

The Gaussian function $f(x)$ that describes the above curve is given by

$$\text{Gaussian function} \quad f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-x_0)^2/2\sigma^2}$$

where $f(x)$ is the probability that the value x be found in a particular measurement. Gaussian functions occur elsewhere in physics and mathematics as well. (Gabriel Lippmann had this to say about the Gaussian function: "Experimentalists think that it is a mathematical theorem while mathematicians believe it to be an experimental fact.")

The probability that a measurement lie inside a certain range of x values, say between x_1 and x_2 , is given by the area of the $f(x)$ curve between these limits. This area is the integral

$$P_{x_1 x_2} = \int_{x_1}^{x_2} f(x) dx$$

An interesting question is what fraction of a series of measurements has values within a standard deviation of the mean value x_0 . In this case $x_1 = x_0 - \sigma$ and $x_2 = x_0 + \sigma$, and

$$P_{x_0 \pm \sigma} = \int_{x_0 - \sigma}^{x_0 + \sigma} f(x) dx = 0.683$$

Hence 68.3 percent of the measurements fall in this interval, which is shaded in Fig. 3.15. A similar calculation shows that 95.4 percent of the measurements fall within two standard deviations of the mean value.

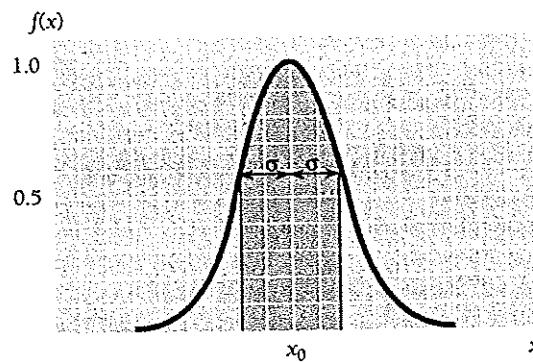


Figure 3.15 A Gaussian distribution. The probability of finding a value of x is given by the Gaussian function $f(x)$. The mean value of x is x_0 , and the total width of the curve at half its maximum value is 2.35σ , where σ is the standard deviation of the distribution. The total probability of finding a value of x within a standard deviation of x_0 is equal to the shaded area and is 68.3 percent.

The de Broglie wavelength of a particle of momentum p is $\lambda = h/p$ and the corresponding wave number is

$$k = \frac{2\pi}{\lambda} = \frac{2\pi p}{h}$$

In terms of wave number the particle's momentum is therefore

$$p = \frac{hk}{2\pi}$$

Hence an uncertainty Δk in the wave number of the de Broglie waves associated with the particle results in an uncertainty Δp in the particle's momentum according to the formula

$$\Delta p = \frac{h \Delta k}{2\pi}$$

Since $\Delta x \Delta k \geq \frac{1}{2}$, $\Delta k \geq 1/(2\Delta x)$ and

Uncertainty principle
$$\Delta x \Delta p \geq \frac{h}{4\pi} \quad (3.21)$$

This equation states that the product of the uncertainty Δx in the position of an object at some instant and the uncertainty Δp in its momentum component in the x direction at the same instant is equal to or greater than $h/4\pi$.

If we arrange matters so that Δx is small, corresponding to a narrow wave group, then Δp will be large. If we reduce Δp in some way, a broad wave group is inevitable and Δx will be large.



Werner Heisenberg (1901–1976) was born in Duisberg, Germany, and studied theoretical physics at Munich, where he also became an enthusiastic skier and mountaineer. At Göttingen in 1924 as an assistant to Max Born, Heisenberg became uneasy about mechanical models of the atom: "Any picture of the atom that our imagination is able to invent is for that very reason defective," he later remarked. Instead he conceived an abstract approach using matrix algebra. In 1925, together with Born and Pascual Jordan, Heisenberg developed this approach into a consistent theory of quantum mechanics, but it was so difficult to understand and apply that it had very little impact on physics at the time. Schrödinger's wave formulation of quantum mechanics the following year was much more successful; Schrödinger and others soon showed that the wave and matrix versions of quantum mechanics were mathematically equivalent.

In 1927, working at Bohr's institute in Copenhagen, Heisenberg developed a suggestion by Wolfgang Pauli into the uncertainty principle. Heisenberg initially felt that this principle was a consequence of the disturbances inevitably produced by any

measuring process. Bohr, on the other hand, thought that the basic cause of the uncertainties was the wave-particle duality, so that they were built into the natural world rather than solely the result of measurement. After much argument Heisenberg came around to Bohr's view. (Einstein, always skeptical about quantum mechanics, said after a lecture by Heisenberg on the uncertainty principle: "Marvelous, what ideas the young people have these days. But I don't believe a word of it.") Heisenberg received the Nobel Prize in 1932.

Heisenberg was one of the very few distinguished scientists to remain in Germany during the Nazi period. In World War II he led research there on atomic weapons, but little progress had been made by the war's end. Exactly why remains unclear, although there is no evidence that Heisenberg, as he later claimed, had moral qualms about creating such weapons and more or less deliberately dragged his feet. Heisenberg recognized early that "an explosive of unimaginable consequences" could be developed, and he and his group should have been able to have gotten farther than they did. In fact, alarmed by the news that Heisenberg was working on an atomic bomb, the U.S. government sent the former Boston Red Sox catcher Moe Berg to shoot Heisenberg during a lecture in neutral Switzerland in 1944. Berg, sitting in the second row, found himself uncertain from Heisenberg's remarks about how advanced the German program was, and kept his gun in his pocket.

These uncertainties are due not to inadequate apparatus but to the imprecise character in nature of the quantities involved. Any instrumental or statistical uncertainties that arise during a measurement only increase the product $\Delta x \Delta p$. Since we cannot know exactly both where a particle is right now and what its momentum is, we cannot say anything definite about where it will be in the future or how fast it will be moving then. We cannot know the future for sure because we cannot know the present for sure. But our ignorance is not total: we can still say that the particle is more likely to be in one place than another and that its momentum is more likely to have a certain value than another.

H-Bar

The quantity $h/2\pi$ appears often in modern physics because it turns out to be the basic unit of angular momentum. It is therefore customary to abbreviate $h/2\pi$ by the symbol \hbar ("h-bar"):

$$\hbar = \frac{h}{2\pi} = 1.054 \times 10^{-34} \text{ J} \cdot \text{s}$$

In the remainder of this book \hbar is used in place of $h/2\pi$. In terms of \hbar , the uncertainty principle becomes

Uncertainty principle	$\Delta x \Delta p \geq \frac{\hbar}{2}$	(3.22)
-----------------------	--	--------

Example 3.6

A measurement establishes the position of a proton with an accuracy of $\pm 1.00 \times 10^{-11} \text{ m}$. Find the uncertainty in the proton's position 1.00 s later. Assume $v \ll c$.

Solution

Let us call the uncertainty in the proton's position Δx_0 at the time $t = 0$. The uncertainty in its momentum at this time is therefore, from Eq. (3.22),

$$\Delta p \geq \frac{\hbar}{2\Delta x_0}$$

Since $v \ll c$, the momentum uncertainty is $\Delta p = \Delta(mv) = m \Delta v$ and the uncertainty in the proton's velocity is

$$\Delta v = \frac{\Delta p}{m} \geq \frac{\hbar}{2m \Delta x_0}$$

The distance x the proton covers in the time t cannot be known more accurately than

$$\Delta x = t \Delta v \geq \frac{\hbar t}{2m \Delta x_0}$$

Hence Δx is inversely proportional to Δx_0 : the more we know about the proton's position at $t = 0$, the less we know about its later position at $t > 0$. The value of Δx at $t = 1.00 \text{ s}$ is

$$\begin{aligned} \Delta x &\geq \frac{(1.054 \times 10^{-34} \text{ J} \cdot \text{s})(1.00 \text{ s})}{(2)(1.672 \times 10^{-27} \text{ kg})(1.00 \times 10^{-11} \text{ m})} \\ &\geq 3.15 \times 10^3 \text{ m} \end{aligned}$$

This is 3.15 km—nearly 2 mi! What has happened is that the original wave group has spread out to a much wider one (Fig. 3.16). This occurred because the phase velocities of the component waves vary with wave number and a large range of wave numbers must have been present to produce the narrow original wave group. See Fig. 3.14.

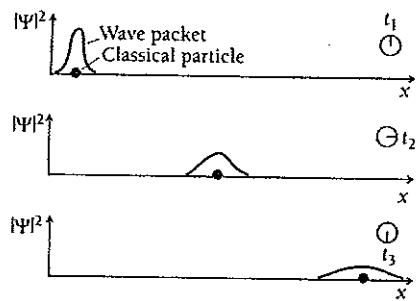


Figure 3.16 The wave packet that corresponds to a moving packet is a composite of many individual waves, as in Fig. 3.13. The phase velocities of the individual waves vary with their wave lengths. As a result, as the particle moves, the wave packet spreads out in space. The narrower the original wavepacket—that is, the more precisely we know its position at that time—the more it spreads out because it is made up of a greater span of waves with different phase velocities.

3.8 UNCERTAINTY PRINCIPLE II

A particle approach gives the same result

The uncertainty principle can be arrived at from the point of view of the particle properties of waves as well as from the point of view of the wave properties of particles.

We might want to measure the position and momentum of an object at a certain moment. To do so, we must touch it with something that will carry the required information back to us. That is, we must poke it with a stick, shine light on it, or perform some similar act. The measurement process itself thus requires that the object be interfered with in some way. If we consider such interferences in detail, we are led to the same uncertainty principle as before even without taking into account the wave nature of moving bodies.

Suppose we look at an electron using light of wavelength λ , as in Fig. 3.17. Each photon of this light has the momentum h/λ . When one of these photons bounces off the electron (which must happen if we are to "see" the electron), the electron's

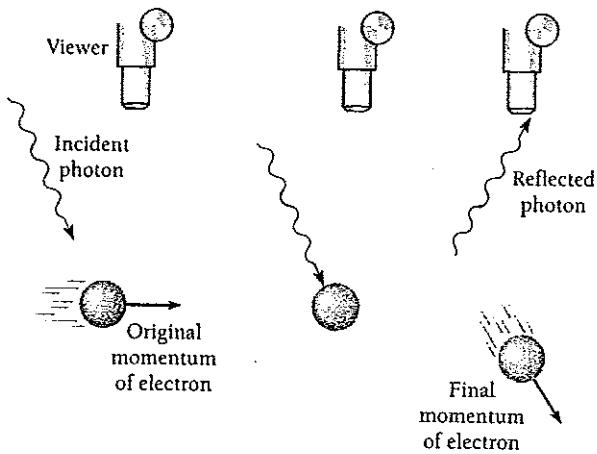


Figure 3.17 An electron cannot be observed without changing its momentum.

original momentum will be changed. The exact amount of the change Δp cannot be predicted, but it will be of the same order of magnitude as the photon momentum h/λ . Hence

$$\Delta p \approx \frac{h}{\lambda} \quad (3.23)$$

The longer the wavelength of the observing photon, the smaller the uncertainty in the electron's momentum.

Because light is a wave phenomenon as well as a particle phenomenon, we cannot expect to determine the electron's location with perfect accuracy regardless of the instrument used. A reasonable estimate of the minimum uncertainty in the measurement might be one photon wavelength, so that

$$\Delta x \geq \lambda \quad (3.24)$$

The shorter the wavelength, the smaller the uncertainty in location. However, if we use light of short wavelength to increase the accuracy of the position measurement, there will be a corresponding decrease in the accuracy of the momentum measurement because the higher photon momentum will disturb the electron's motion to a greater extent. Light of long wavelength will give a more accurate momentum but a less accurate position.

Combining Eqs. (3.23) and (3.24) gives

$$\Delta x \Delta p \geq h \quad (3.25)$$

This result is consistent with Eq. (3.22), $\Delta x \Delta p \geq \hbar/2$.

Arguments like the preceding one, although superficially attractive, must be approached with caution. The argument above implies that the electron can possess a definite position and momentum at any instant and that it is the measurement process that introduces the indeterminacy in $\Delta x \Delta p$. On the contrary, this indeterminacy is inherent in the nature of a moving body. The justification for the many "derivations" of this kind is first, they show it is impossible to imagine a way around the uncertainty principle; and second, they present a view of the principle that can be appreciated in a more familiar context than that of wave groups.

3.9 APPLYING THE UNCERTAINTY PRINCIPLE

A useful tool, not just a negative statement

Planck's constant \hbar is so small that the limitations imposed by the uncertainty principle are significant only in the realm of the atom. On such a scale, however, this principle is of great help in understanding many phenomena. It is worth keeping in mind that the lower limit of $\hbar/2$ for $\Delta x \Delta p$ is rarely attained. More usually $\Delta x \Delta p \geq \hbar$, or even (as we just saw) $\Delta x \Delta p \geq h$.

Example 3.7

A typical atomic nucleus is about 5.0×10^{-15} m in radius. Use the uncertainty principle to place a lower limit on the energy an electron must have if it is to be part of a nucleus.

Solution

Letting $\Delta x = 5.0 \times 10^{-5}$ m we have

$$\Delta p \geq \frac{\hbar}{2\Delta x} \geq \frac{1.054 \times 10^{-34} \text{ J} \cdot \text{s}}{(2)(5.0 \times 10^{-5} \text{ m})} \geq 1.1 \times 10^{-20} \text{ kg} \cdot \text{m/s}$$

If this is the uncertainty in a nuclear electron's momentum, the momentum p itself must be at least comparable in magnitude. An electron with such a momentum has a kinetic energy KE many times greater than its rest energy mc^2 . From Eq. (1.24) we see that we can let $KE = pc$ here to a sufficient degree of accuracy. Therefore

$$KE = pc \geq (1.1 \times 10^{-20} \text{ kg} \cdot \text{m/s})(3.0 \times 10^8 \text{ m/s}) \geq 3.3 \times 10^{-12} \text{ J}$$

Since $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$, the kinetic energy of an electron must exceed 20 MeV if it is to be inside a nucleus. Experiments show that the electrons emitted by certain unstable nuclei never have more than a small fraction of this energy, from which we conclude that nuclei cannot contain electrons. The electron an unstable nucleus may emit comes into being at the moment the nucleus decays (see Secs. 11.3 and 12.5).

Example 3.8

A hydrogen atom is 5.3×10^{-11} m in radius. Use the uncertainty principle to estimate the minimum energy an electron can have in this atom.

Solution

Here we find that with $\Delta x = 5.3 \times 10^{-11}$ m,

$$\Delta p \geq \frac{\hbar}{2\Delta x} \geq 9.9 \times 10^{-25} \text{ kg} \cdot \text{m/s}$$

An electron whose momentum is of this order of magnitude behaves like a classical particle, and its kinetic energy is

$$KE = \frac{p^2}{2m} \geq \frac{(9.9 \times 10^{-25} \text{ kg} \cdot \text{m/s})^2}{(2)(9.1 \times 10^{-31} \text{ kg})} \geq 5.4 \times 10^{-19} \text{ J}$$

which is 3.4 eV. The kinetic energy of an electron in the lowest energy level of a hydrogen atom is actually 13.6 eV.

Energy and Time

Another form of the uncertainty principle concerns energy and time. We might wish to measure the energy E emitted during the time interval Δt in an atomic process. If the energy is in the form of em waves, the limited time available restricts the accuracy with which we can determine the frequency ν of the waves. Let us assume that the minimum uncertainty in the number of waves we count in a wave group is one wave. Since the frequency of the waves under study is equal to the number of them we count divided by the time interval, the uncertainty $\Delta\nu$ in our frequency measurement is

$$\Delta\nu \geq \frac{1}{\Delta t}$$

The corresponding energy uncertainty is

$$\Delta E = h \Delta\nu$$

and so

$$\Delta E \geq \frac{h}{\Delta t} \quad \text{or} \quad \Delta E \Delta t \geq h$$

A more precise calculation based on the nature of wave groups changes this result to

Uncertainties in energy and time	$\Delta E \Delta t \geq \frac{\hbar}{2}$	(3.26)
-------------------------------------	--	--------

Equation (3.26) states that the product of the uncertainty ΔE in an energy measurement and the uncertainty Δt in the time at which the measurement is made is equal to or greater than $\hbar/2$. This result can be derived in other ways as well and is a general one not limited to em waves.

Example 3.9

An "excited" atom gives up its excess energy by emitting a photon of characteristic frequency, as described in Chap. 4. The average period that elapses between the excitation of an atom and the time it radiates is 1.0×10^{-8} s. Find the inherent uncertainty in the frequency of the photon.

Solution

The photon energy is uncertain by the amount

$$\Delta E \geq \frac{\hbar}{2\Delta t} \geq \frac{1.054 \times 10^{-34} \text{ J} \cdot \text{s}}{2(1.0 \times 10^{-8} \text{ s})} \geq 5.3 \times 10^{-27} \text{ J}$$

The corresponding uncertainty in the frequency of light is

$$\Delta\nu = \frac{\Delta E}{h} \geq 8 \times 10^6 \text{ Hz}$$

This is the irreducible limit to the accuracy with which we can determine the frequency of the radiation emitted by an atom. As a result, the radiation from a group of excited atoms does not appear with the precise frequency ν . For a photon whose frequency is, say, 5.0×10^{14} Hz, $\Delta\nu/\nu = 1.6 \times 10^{-8}$. In practice, other phenomena such as the doppler effect contribute more than this to the broadening of spectral lines.

EXERCISES

It is only the first step that takes the effort. —Marquise du Deffand

3.1 De Broglie Waves

1. A photon and a particle have the same wavelength. Can anything be said about how their linear momenta compare? About how the photon's energy compares with the particle's total energy? About how the photon's energy compares with the particle's kinetic energy?
2. Find the de Broglie wavelength of (a) an electron whose speed is 1.0×10^8 m/s, and (b) an electron whose speed is 2.0×10^8 m/s.
3. Find the de Broglie wavelength of a 1.0-mg grain of sand blown by the wind at a speed of 20 m/s.
4. Find the de Broglie wavelength of the 40-keV electrons used in a certain electron microscope.
5. By what percentage will a nonrelativistic calculation of the de Broglie wavelength of a 100-keV electron be in error?
6. Find the de Broglie wavelength of a 1.00-MeV proton. Is a relativistic calculation needed?
7. The atomic spacing in rock salt, NaCl, is 0.282 nm. Find the kinetic energy (in eV) of a neutron with a de Broglie wavelength of 0.282 nm. Is a relativistic calculation needed? Such neutrons can be used to study crystal structure.
8. Find the kinetic energy of an electron whose de Broglie wavelength is the same as that of a 100-keV x-ray.
9. Green light has a wavelength of about 550 nm. Through what potential difference must an electron be accelerated to have this wavelength?
10. Show that the de Broglie wavelength of a particle of mass m and kinetic energy KE is given by

$$\lambda = \frac{hc}{\sqrt{KE(KE + 2mc^2)}}$$

11. Show that if the total energy of a moving particle greatly exceeds its rest energy, its de Broglie wavelength is nearly the same as the wavelength of a photon with the same total energy.
12. (a) Derive a relativistically correct formula that gives the de Broglie wavelength of a charged particle in terms of the potential difference V through which it has been accelerated.
(b) What is the nonrelativistic approximation of this formula, valid for $eV \ll mc^2$?

3.4 Phase and Group Velocities

13. An electron and a proton have the same velocity. Compare the wavelengths and the phase and group velocities of their de Broglie waves.
14. An electron and a proton have the same kinetic energy. Compare the wavelengths and the phase and group velocities of their de Broglie waves.

15. Verify the statement in the text that, if the phase velocity is the same for all wavelengths of a certain wave phenomenon (that is, there is no dispersion), the group and phase velocities are the same.
16. The phase velocity of ripples on a liquid surface is $\sqrt{2\pi S/\lambda\rho}$, where S is the surface tension and ρ the density of the liquid. Find the group velocity of the ripples.
17. The phase velocity of ocean waves is $\sqrt{g\lambda/2\pi}$, where g is the acceleration of gravity. Find the group velocity of ocean waves.
18. Find the phase and group velocities of the de Broglie waves of an electron whose speed is $0.900c$.
19. Find the phase and group velocities of the de Broglie waves of an electron whose kinetic energy is 500 keV.
20. Show that the group velocity of a wave is given by $v_g = d\nu/d(1/\lambda)$.
21. (a) Show that the phase velocity of the de Broglie waves of a particle of mass m and de Broglie wavelength λ is given by

$$v_p = c\sqrt{1 + \left(\frac{mc\lambda}{\hbar}\right)^2}$$

(b) Compare the phase and group velocities of an electron whose de Broglie wavelength is exactly 1×10^{-13} m.

22. In his original paper, de Broglie suggested that $E = h\nu$ and $p = h/\lambda$, which hold for electromagnetic waves, are also valid for moving particles. Use these relationships to show that the group velocity v_g of a de Broglie wave group is given by dE/dp , and with the help of Eq. (1.24), verify that $v_g = v$ for a particle of velocity v .

3.5 Particle Diffraction

23. What effect on the scattering angle in the Davisson-Germer experiment does increasing the electron energy have?
24. A beam of neutrons that emerges from a nuclear reactor contains neutrons with a variety of energies. To obtain neutrons with an energy of 0.050 eV, the beam is passed through a crystal whose atomic planes are 0.20 nm apart. At what angles relative to the original beam will the desired neutrons be diffracted?
25. In Sec. 3.5 it was mentioned that the energy of an electron entering a crystal increases, which reduces its de Broglie wavelength. Consider a beam of 54-eV electrons directed at a nickel target. The potential energy of an electron that enters the target changes by 26 eV. (a) Compare the electron speeds outside and inside the target. (b) Compare the respective de Broglie wavelengths.
26. A beam of 50-keV electrons is directed at a crystal and diffracted electrons are found at an angle of 50° relative to the original beam. What is the spacing of the atomic planes of the crystal? A relativistic calculation is needed for λ .

3.6 Particle in a Box

27. Obtain an expression for the energy levels (in MeV) of a neutron confined to a one-dimensional box 1.00×10^{-14} m wide. What is the neutron's minimum energy? (The diameter of an atomic nucleus is of this order of magnitude.)
28. The lowest energy possible for a certain particle trapped in a certain box is 1.00 eV. (a) What are the next two higher energies the particle can have? (b) If the particle is an electron, how wide is the box?
29. A proton in a one-dimensional box has an energy of 400 keV in its first excited state. How wide is the box?

3.7 Uncertainty Principle I

3.8 Uncertainty Principle II

3.9 Applying the Uncertainty Principle

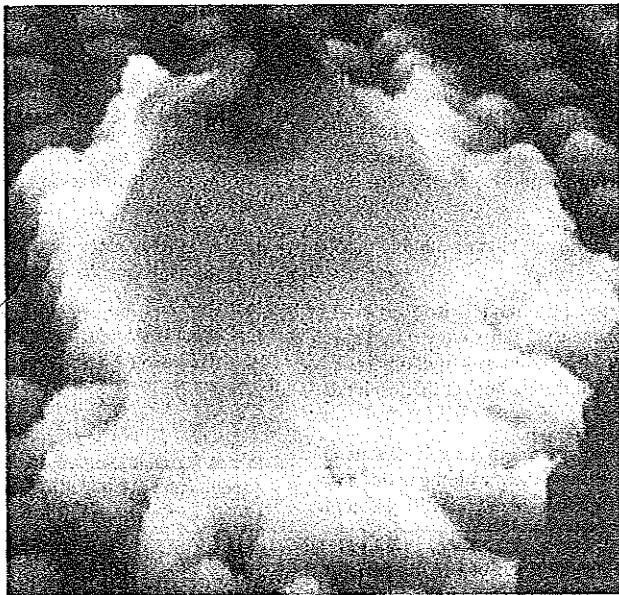
30. Discuss the prohibition of $E = 0$ for a particle trapped in a box L wide in terms of the uncertainty principle. How does the minimum momentum of such a particle compare with the momentum uncertainty required by the uncertainty principle if we take $\Delta x = L$?
31. The atoms in a solid possess a certain minimum zero-point energy even at 0 K, while no such restriction holds for the molecules in an ideal gas. Use the uncertainty principle to explain these statements.
32. Compare the uncertainties in the velocities of an electron and a proton confined in a 1.00-nm box.
33. The position and momentum of a 1.00-keV electron are simultaneously determined. If its position is located to within 0.100 nm, what is the percentage of uncertainty in its momentum?
34. (a) How much time is needed to measure the kinetic energy of an electron whose speed is 10.0 m/s with an uncertainty of no more than 0.100 percent? How far will the electron have traveled in this period of time? (b) Make the same calculations

for a 1.00-g insect whose speed is the same. What do these sets of figures indicate?

35. How accurately can the position of a proton with $v \ll c$ be determined without giving it more than 1.00 keV of kinetic energy?
36. (a) Find the magnitude of the momentum of a particle in a box in its n th state. (b) The minimum change in the particle's momentum that a measurement can cause corresponds to a change of ± 1 in the quantum number n . If $\Delta x = L$, show that $\Delta p \Delta x \geq \hbar/2$.
37. A marine radar operating at a frequency of 9400 MHz emits groups of electromagnetic waves 0.0800 μs in duration. The time needed for the reflections of these groups to return indicates the distance to a target. (a) Find the length of each group and the number of waves it contains. (b) What is the approximate minimum bandwidth (that is, spread of frequencies) the radar receiver must be able to process?
38. An unstable elementary particle called the eta meson has a rest mass of 549 MeV/c² and a mean lifetime of 7.00×10^{-19} s. What is the uncertainty in its rest mass?
39. The frequency of oscillation of a harmonic oscillator of mass m and spring constant C is $\nu = \sqrt{C/m}/2\pi$. The energy of the oscillator is $E = p^2/2m + Cx^2/2$, where p is its momentum when its displacement from the equilibrium position is x . In classical physics the minimum energy of the oscillator is $E_{\min} = 0$. Use the uncertainty principle to find an expression for E in terms of x only and show that the minimum energy is actually $E_{\min} = \hbar\nu/2$ by setting $dE/dx = 0$ and solving for E_{\min} .
40. (a) Verify that the uncertainty principle can be expressed in the form $\Delta L \Delta \theta \geq \hbar/2$, where ΔL is the uncertainty in the angular momentum of a particle and $\Delta \theta$ is the uncertainty in its angular position. (Hint: Consider a particle of mass m moving in a circle of radius r at the speed v , for which $L = mv r$.) (b) At what uncertainty in L will the angular position of a particle become completely indeterminate?

CHAPTER 5

Quantum Mechanics



Scanning tunneling micrograph of gold atoms on a carbon (graphite) substrate.
The cluster of gold atoms is about 1.5 nm across and three atoms high.

5.1 QUANTUM MECHANICS

Classical mechanics is an approximation of quantum mechanics

5.2 THE WAVE EQUATION

It can have a variety of solutions, including complex ones

5.3 SCHRÖDINGER'S EQUATION: TIME-DEPENDENT FORM

A basic physical principle that cannot be derived from anything else

5.4 LINEARITY AND SUPERPOSITION

Wave functions add, not probabilities

5.5 EXPECTATION VALUES

How to extract information from a wave function

5.6 OPERATORS

Another way to find expectation values

5.7 SCHRÖDINGER'S EQUATION: STEADY-STATE FORM

Eigenvalues and eigenfunctions

5.8 PARTICLE IN A BOX

How boundary conditions and normalization determine wave functions

5.9 FINITE POTENTIAL WELL

The wave function penetrates the walls, which lowers the energy levels

5.10 TUNNEL EFFECT

A particle without the energy to pass over a potential barrier may still tunnel through it

5.11 HARMONIC OSCILLATOR

Its energy levels are evenly spaced

APPENDIX: THE TUNNEL EFFECT

Although the Bohr theory of the atom, which can be extended further than was done in Chap. 4, is able to account for many aspects of atomic phenomena, it has a number of severe limitations as well. First of all, it applies only to hydrogen and one-electron ions such as He^+ and Li^{2+} —it does not even work for ordinary helium. The Bohr theory cannot explain why certain spectral lines are more intense than others (that is, why certain transitions between energy levels have greater probabilities of occurrence than others). It cannot account for the observation that many spectral lines actually consist of several separate lines whose wavelengths differ slightly. And perhaps most important, it does not permit us to obtain what a really successful theory of the atom should make possible: an understanding of how individual atoms interact with one another to endow macroscopic aggregates of matter with the physical and chemical properties we observe.

The preceding objections to the Bohr theory are not put forward in an unfriendly way, for the theory was one of those seminal achievements that transform scientific thought, but rather to emphasize that a more general approach to atomic phenomena is required. Such an approach was developed in 1925 and 1926 by Erwin Schrödinger, Werner Heisenberg, Max Born, Paul Dirac, and others under the apt name of quantum mechanics. “The discovery of quantum mechanics was nearly a total surprise. It described the physical world in a way that was fundamentally new. It seemed to many of us a miracle,” noted Eugene Wigner, one of the early workers in the field. By the early 1930s the application of quantum mechanics to problems involving nuclei, atoms, molecules, and matter in the solid state made it possible to understand a vast body of data (“a large part of physics and the whole of chemistry,” according to Dirac) and—vital for any theory—led to predictions of remarkable accuracy. Quantum mechanics has survived every experimental test thus far of even its most unexpected conclusions.

5.1 QUANTUM MECHANICS

Classical mechanics is an approximation of quantum mechanics

The fundamental difference between classical (or Newtonian) mechanics and quantum mechanics lies in what they describe. In classical mechanics, the future history of a particle is completely determined by its initial position and momentum together with the forces that act upon it. In the everyday world these quantities can all be determined well enough for the predictions of Newtonian mechanics to agree with what we find.

Quantum mechanics also arrives at relationships between observable quantities, but the uncertainty principle suggests that the nature of an observable quantity is different in the atomic realm. Cause and effect are still related in quantum mechanics, but what they concern needs careful interpretation. In quantum mechanics the kind of certainty about the future characteristic of classical mechanics is impossible because the initial state of a particle cannot be established with sufficient accuracy. As we saw in Sec. 3.7, the more we know about the position of a particle now, the less we know about its momentum and hence about its position later.

The quantities whose relationships quantum mechanics explores are *probabilities*. Instead of asserting, for example, that the radius of the electron’s orbit in a ground-state hydrogen atom is always exactly 5.3×10^{-11} m, as the Bohr theory does, quantum mechanics states that this is the *most probable* radius. In a suitable experiment most trials will yield a different value, either larger or smaller, but the value most likely to be found will be 5.3×10^{-11} m.

Quantum mechanics might seem a poor substitute for classical mechanics. However, classical mechanics turns out to be just an approximate version of quantum mechanics. The certainties of classical mechanics are illusory, and their apparent agreement with experiment occurs because ordinary objects consist of so many individual atoms that departures from average behavior are unnoticeable. Instead of two sets of physical principles, one for the macroworld and one for the microworld, there is only the single set included in quantum mechanics.

Wave Function

As mentioned in Chap. 3, the quantity with which quantum mechanics is concerned is the wave function Ψ of a body. While Ψ itself has no physical interpretation, the square of its absolute magnitude $|\Psi|^2$ evaluated at a particular place at a particular time is proportional to the probability of finding the body there at that time. The linear momentum, angular momentum, and energy of the body are other quantities that can be established from Ψ . The problem of quantum mechanics is to determine Ψ for a body when its freedom of motion is limited by the action of external forces.

Wave functions are usually complex with both real and imaginary parts. A probability, however, must be a positive real quantity. The probability density $|\Psi|^2$ for a complex Ψ is therefore taken as the product $\Psi^*\Psi$ of Ψ and its complex conjugate Ψ^* . The complex conjugate of any function is obtained by replacing $i (= \sqrt{-1})$ by $-i$ wherever it appears in the function. Every complex function Ψ can be written in the form

$$\text{Wave function} \quad \Psi = A + iB$$

where A and B are real functions. The complex conjugate Ψ^* of Ψ is

$$\text{Complex conjugate} \quad \Psi^* = A - iB$$

$$\text{and so} \quad |\Psi|^2 = \Psi^*\Psi = A^2 - i^2B^2 = A^2 + B^2$$

since $i^2 = -1$. Hence $|\Psi|^2 = \Psi^*\Psi$ is always a positive real quantity, as required.

Normalization

Even before we consider the actual calculation of Ψ , we can establish certain requirements it must always fulfill. For one thing, since $|\Psi|^2$ is proportional to the probability density P of finding the body described by Ψ , the integral of $|\Psi|^2$ over all space must be finite—the body is somewhere, after all. If

$$\int_{-\infty}^{\infty} |\Psi|^2 dV = 0$$

the particle does not exist, and the integral obviously cannot be ∞ and still mean anything. Furthermore, $|\Psi|^2$ cannot be negative or complex because of the way it is defined. The only possibility left is that the integral be a finite quantity if Ψ is to describe properly a real body.

It is usually convenient to have $|\Psi|^2$ be equal to the probability density P of finding the particle described by Ψ , rather than merely be proportional to P . If $|\Psi|^2$ is to

equal P , then it must be true that

$$\text{Normalization} \quad \int_{-\infty}^{\infty} |\Psi|^2 dV = 1 \quad (5.1)$$

since if the particle exists somewhere at all times,

$$\int_{-\infty}^{\infty} P dV = 1$$

A wave function that obeys Eq. (5.1) is said to be **normalized**. Every acceptable wave function can be normalized by multiplying it by an appropriate constant; we shall shortly see how this is done.

Well-Behaved Wave Functions

Besides being normalizable, Ψ must be single-valued, since P can have only one value at a particular place and time, and continuous. Momentum considerations (see Sec. 5.6) require that the partial derivatives $\partial\Psi/\partial x$, $\partial\Psi/\partial y$, $\partial\Psi/\partial z$ be finite, continuous, and single-valued. Only wave functions with all these properties can yield physically meaningful results when used in calculations, so only such "well-behaved" wave functions are admissible as mathematical representations of real bodies. To summarize:

- 1 Ψ must be continuous and single-valued everywhere.
- 2 $\partial\Psi/\partial x$, $\partial\Psi/\partial y$, $\partial\Psi/\partial z$ must be continuous and single-valued everywhere.
- 3 Ψ must be normalizable, which means that Ψ must go to 0 as $x \rightarrow \pm\infty$, $y \rightarrow \pm\infty$, $z \rightarrow \pm\infty$ in order that $\int |\Psi|^2 dV$ over all space be a finite constant.

These rules are not always obeyed by the wave functions of particles in model situations that only approximate actual ones. For instance, the wave functions of a particle in a box with infinitely hard walls do not have continuous derivatives at the walls, since $\Psi = 0$ outside the box (see Fig. 5.4). But in the real world, where walls are never infinitely hard, there is no sharp change in Ψ at the walls (see Fig. 5.7) and the derivatives are continuous. Exercise 7 gives another example of a wave function that is not well-behaved.

Given a normalized and otherwise acceptable wave function Ψ , the probability that the particle it describes will be found in a certain region is simply the integral of the probability density $|\Psi|^2$ over that region. Thus for a particle restricted to motion in the x direction, the probability of finding it between x_1 and x_2 is given by

$$\text{Probability} \quad P_{x_1, x_2} = \int_{x_1}^{x_2} |\Psi|^2 dx \quad (5.2)$$

We will see examples of such calculations later in this chapter and in Chap. 6.

5.2 THE WAVE EQUATION

It can have a variety of solutions, including complex ones

Schrödinger's equation, which is the fundamental equation of quantum mechanics in the same sense that the second law of motion is the fundamental equation of Newtonian mechanics, is a wave equation in the variable Ψ .

Before we tackle Schrödinger's equation, let us review the wave equation

$$\text{Wave equation} \quad \frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} \quad (5.3)$$

which governs a wave whose variable quantity is y that propagates in the x direction with the speed v . In the case of a wave in a stretched string, y is the displacement of the string from the x axis; in the case of a sound wave, y is the pressure difference; in the case of a light wave, y is either the electric or the magnetic field magnitude. Equation (5.3) can be derived from the second law of motion for mechanical waves and from Maxwell's equations for electromagnetic waves.

Partial Derivatives

Suppose we have a function $f(x, y)$ of two variables, x and y , and we want to know how f varies with only one of them, say x . To find out, we differentiate f with respect to x while treating the other variable y as a constant. The result is the partial derivative of f with respect to x , which is written $\partial f / \partial x$

$$\frac{\partial f}{\partial x} = \left(\frac{df}{dx} \right)_{y=\text{constant}}$$

The rules for ordinary differentiation hold for partial differentiation as well. For instance, if $f = cx^2$,

$$\frac{df}{dx} = 2cx$$

and so, if $f = yx^2$,

$$\frac{\partial f}{\partial x} = \left(\frac{df}{dx} \right)_{y=\text{constant}} = 2yx$$

The partial derivative of $f = yx^2$ with respect to the other variable, y , is

$$\frac{\partial f}{\partial y} = \left(\frac{df}{dy} \right)_{x=\text{constant}} = x^2$$

Second order partial derivatives occur often in physics, as in the wave equation. To find $\partial^2 f / \partial x^2$, we first calculate $\partial f / \partial x$ and then differentiate again, still keeping y constant:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right)$$

For $f = yx^2$,

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} (2yx) = 2y$$

Similarly

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} (x^2) = 0$$

Solutions of the wave equation may be of many kinds, reflecting the variety of waves that can occur—a single traveling pulse, a train of waves of constant amplitude and wavelength, a train of superposed waves of the same amplitudes and wavelengths, a train of superposed waves of different amplitudes and wavelengths,

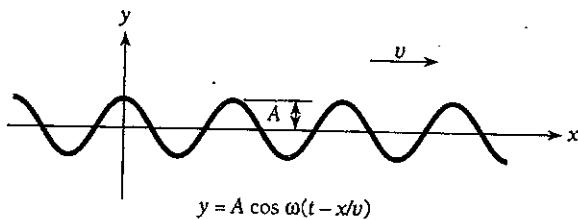


Figure 5.1 Waves in the xy plane traveling in the $+x$ direction along a stretched string lying on the x axis.

a standing wave in a string fastened at both ends, and so on. All solutions must be of the form

$$y = F\left(t \pm \frac{x}{v}\right) \quad (5.4)$$

where F is any function that can be differentiated. The solutions $F(t - x/v)$ represent waves traveling in the $+x$ direction, and the solutions $F(t + x/v)$ represent waves traveling in the $-x$ direction.

Let us consider the wave equivalent of a "free particle," which is a particle that is not under the influence of any forces and therefore pursues a straight path at constant speed. This wave is described by the general solution of Eq. (5.3) for undamped (that is, constant amplitude A), monochromatic (constant angular frequency ω) harmonic waves in the $+x$ direction, namely

$$y = Ae^{-i\omega(t-x/v)} \quad (5.5)$$

In this formula y is a complex quantity, with both real and imaginary parts.

Because

$$e^{-i\theta} = \cos \theta - i \sin \theta$$

Eq. (5.5) can be written in the form

$$y = A \cos \omega \left(t - \frac{x}{v}\right) - iA \sin \omega \left(t - \frac{x}{v}\right) \quad (5.6)$$

Only the real part of Eq. (5.6) [which is the same as Eq. (3.5)] has significance in the case of waves in a stretched string. There y represents the displacement of the string from its normal position (Fig. 5.1), and the imaginary part of Eq. (5.6) is discarded as irrelevant.

Example 5.1

Verify that Eq. (5.5) is a solution of the wave equation.

Solution

The derivative of an exponential function e^u is

$$\frac{d}{dx}(e^u) = e^u \frac{du}{dx}$$

The partial derivative of y with respect to x (which means t is treated as a constant) from Eq. (5.5) is therefore

$$\frac{\partial y}{\partial x} = \frac{i\omega}{v} y$$

and the second partial derivative is

$$\frac{\partial^2 y}{\partial x^2} = \frac{i^2 \omega^2}{v^2} y = -\frac{\omega^2}{v^2} y$$

since $i^2 = -1$. The partial derivative of y with respect to t (now holding x constant) is

$$\frac{\partial y}{\partial t} = -i\omega y$$

and the second partial derivative is

$$\frac{\partial^2 y}{\partial t^2} = i^2 \omega^2 y = -\omega^2 y$$

Combining these results gives

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}$$

which is Eq. (5.3). Hence Eq. (5.5) is a solution of the wave equation.

5.3 SCHRÖDINGER'S EQUATION: TIME-DEPENDENT FORM

A basic physical principle that cannot be derived from anything else

In quantum mechanics the wave function Ψ corresponds to the wave variable y of wave motion in general. However, Ψ , unlike y , is not itself a measurable quantity and may therefore be complex. For this reason we assume that Ψ for a particle moving freely in the $+x$ direction is specified by

$$\Psi = A e^{-i\omega(t-x/v)} \quad (5.7)$$

Replacing ω in the above formula by $2\pi\nu$ and v by $\lambda\nu$ gives

$$\Psi = A e^{-2\pi i(\nu t - x/\lambda)} \quad (5.8)$$

This is convenient since we already know what ν and λ are in terms of the total energy E and momentum p of the particle being described by Ψ . Because

$$E = \hbar\nu = 2\pi\hbar\nu \quad \text{and} \quad \lambda = \frac{\hbar}{p} = \frac{2\pi\hbar}{p}$$

we have

$$\text{Free particle} \quad \Psi = A e^{-(i/\hbar)(Et - px)} \quad (5.9)$$

Equation (5.9) describes the wave equivalent of an unrestricted particle of total energy E and momentum p moving in the $+x$ direction, just as Eq. (5.5) describes, for example, a harmonic displacement wave moving freely along a stretched string.

The expression for the wave function Ψ given by Eq. (5.9) is correct only for freely moving particles. However, we are most interested in situations where the motion of a particle is subject to various restrictions. An important concern, for example, is an electron bound to an atom by the electric field of its nucleus. What we must now do is obtain the fundamental differential equation for Ψ , which we can then solve for Ψ in a specific situation. This equation, which is Schrödinger's equation, can be arrived at in various ways, but it *cannot* be rigorously derived from existing physical principles:

the equation represents something new. What will be done here is to show one route to the wave equation for Ψ and then to discuss the significance of the result.

We begin by differentiating Eq. (5.9) for Ψ twice with respect to x , which gives

$$\frac{\partial^2 \Psi}{\partial x^2} = -\frac{p^2}{\hbar^2} \Psi$$

$$p^2 \Psi = -\hbar^2 \frac{\partial^2 \Psi}{\partial x^2} \quad (5.10)$$

Differentiating Eq. (5.9) once with respect to t gives

$$\frac{\partial \Psi}{\partial t} = -\frac{iE}{\hbar} \Psi$$

$$E\Psi = -\frac{\hbar}{i} \frac{\partial \Psi}{\partial t} \quad (5.11)$$

At speeds small compared with that of light, the total energy E of a particle is the sum of its kinetic energy $p^2/2m$ and its potential energy U , where U is in general a function of position x and time t :

$$E = \frac{p^2}{2m} + U(x, t) \quad (5.12)$$

The function U represents the influence of the rest of the universe on the particle. Of course, only a small part of the universe interacts with the particle to any extent; for



Erwin Schrödinger (1887–1961) was born in Vienna to an Austrian father and a half-English mother and received his doctorate at the university there. After World War I, during which he served as an artillery officer, Schrödinger had appointments at several German universities before becoming professor of physics in Zurich, Switzerland. Late in November, 1925, Schrödinger gave a talk on de Broglie's notion that a moving particle has a wave character. A colleague remarked to him afterward that to deal properly with a wave, one needs a wave equation. Schrödinger took this to heart, and a few weeks later he was "struggling with a new atomic theory. If only I knew more mathematics! I am very optimistic about this thing and expect that if I can only . . . solve it, it will be *very* beautiful." (Schrödinger was not the only physicist to find the mathematics he needed difficult; the eminent mathematician David Hilbert said at about this time, "Physics is much too hard for physicists.")

The struggle was successful, and in January 1926 the first of four papers on "Quantization as an Eigenvalue Problem" was completed. In this epochal paper Schrödinger introduced the equation that bears his name and solved it for the hydrogen atom,

thereby opening wide the door to the modern view of the atom which others had only pushed ajar. By June Schrödinger had applied wave mechanics to the harmonic oscillator, the diatomic molecule, the hydrogen atom in an electric field, the absorption and emission of radiation, and the scattering of radiation by atoms and molecules. He had also shown that his wave mechanics was mathematically equivalent to the more abstract Heisenberg-Born-Jordan matrix mechanics.

The significance of Schrödinger's work was at once realized. In 1927 he succeeded Planck at the University of Berlin but left Germany in 1933, the year he received the Nobel Prize, when the Nazis came to power. He was at Dublin's Institute for Advanced Study from 1939 until his return to Austria in 1956. In Dublin, Schrödinger became interested in biology, in particular the mechanism of heredity. He seems to have been the first to make definite the idea of a genetic code and to identify genes as long molecules that carry the code in the form of variations in how their atoms are arranged. Schrödinger's 1944 book *What Is Life?* was enormously influential, not only by what it said but also by introducing biologists to a new way of thinking—that of the physicist—about their subject. *What Is Life?* started James Watson on his search for "the secret of the gene," which he and Francis Crick (a physicist) discovered in 1953 to be the structure of the DNA molecule.

instance, in the case of the electron in a hydrogen atom, only the electric field of the nucleus must be taken into account.

Multiplying both sides of Eq. (5.12) by the wave function Ψ gives

$$E\Psi = \frac{p^2\Psi}{2m} + U\Psi \quad (5.13)$$

Now we substitute for $E\Psi$ and $p^2\Psi$ from Eqs. (5.10) and (5.11) to obtain the time-dependent form of Schrödinger's equation:

Time-dependent
Schrödinger
equation in one
dimension

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + U\Psi \quad (5.14)$$

In three dimensions the time-dependent form of Schrödinger's equation is

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} \right) + U\Psi \quad (5.15)$$

where the particle's potential energy U is some function of x, y, z , and t .

Any restrictions that may be present on the particle's motion will affect the potential-energy function U . Once U is known, Schrödinger's equation may be solved for the wave function Ψ of the particle, from which its probability density $|\Psi|^2$ may be determined for a specified x, y, z, t .

Validity of Schrödinger's Equation

Schrödinger's equation was obtained here using the wave function of a freely moving particle (potential energy $U = \text{constant}$). How can we be sure it applies to the general case of a particle subject to arbitrary forces that vary in space and time [$U = U(x, y, z, t)$]? Substituting Eqs. (5.10) and (5.11) into Eq. (5.13) is really a wild leap with no formal justification; this is true for all other ways in which Schrödinger's equation can be arrived at, including Schrödinger's own approach.

What we must do is postulate Schrödinger's equation, solve it for a variety of physical situations, and compare the results of the calculations with the results of experiments. If both sets of results agree, the postulate embodied in Schrödinger's equation is valid. If they disagree, the postulate must be discarded and some other approach would then have to be explored. In other words,

Schrödinger's equation cannot be derived from other basic principles of physics; it is a basic principle in itself.

What has happened is that Schrödinger's equation has turned out to be remarkably accurate in predicting the results of experiments. To be sure, Eq. (5.15) can be used only for nonrelativistic problems, and a more elaborate formulation is needed when particle speeds near that of light are involved. But because it is in accord with experience within its range of applicability, we must consider Schrödinger's equation as a valid statement concerning certain aspects of the physical world.

It is worth noting that Schrödinger's equation does not increase the number of principles needed to describe the workings of the physical world. Newton's second law

of motion $F = ma$, the basic principle of classical mechanics, can be derived from Schrödinger's equation provided the quantities it relates are understood to be averages rather than precise values. (Newton's laws of motion were also not derived from any other principles. Like Schrödinger's equation, these laws are considered valid in their range of applicability because of their agreement with experiment.)

5.4 LINEARITY AND SUPERPOSITION

Wave functions add, not probabilities

An important property of Schrödinger's equation is that it is linear in the wave function Ψ . By this is meant that the equation has terms that contain Ψ and its derivatives but no terms independent of Ψ or that involve higher powers of Ψ or its derivatives. As a result, a linear combination of solutions of Schrödinger's equation for a given system is also itself a solution. If Ψ_1 and Ψ_2 are two solutions (that is, two wave functions that satisfy the equation), then

$$\Psi = a_1\Psi_1 + a_2\Psi_2$$

is also a solution, where a_1 and a_2 are constants (see Exercise 8). Thus the wave functions Ψ_1 and Ψ_2 obey the superposition principle that other waves do (see Sec. 2.1) and we conclude that interference effects can occur for wave functions just as they can for light, sound, water, and electromagnetic waves. In fact, the discussions of Secs. 3.4 and 3.7 assumed that de Broglie waves are subject to the superposition principle.

Let us apply the superposition principle to the diffraction of an electron beam. Figure 5.2a shows a pair of slits through which a parallel beam of monoenergetic electrons pass on their way to a viewing screen. If slit 1 only is open, the result is the intensity variation shown in Fig. 5.2b that corresponds to the probability density

$$P_1 = |\Psi_1|^2 = \Psi_1^* \Psi_1$$

If slit 2 only is open, as in Fig. 5.2c, the corresponding probability density is

$$P_2 = |\Psi_2|^2 = \Psi_2^* \Psi_2$$

We might suppose that opening both slits would give an electron intensity variation described by $P_1 + P_2$, as in Fig. 5.2d. However, this is not the case because in quantum

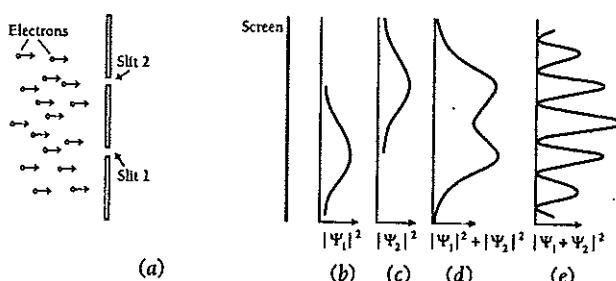


Figure 5.2 (a) Arrangement of double-slit experiment. (b) The electron intensity at the screen with only slit 1 open. (c) The electron intensity at the screen with only slit 2 open. (d) The sum of the intensities of (b) and (c). (e) The actual intensity at the screen with slits 1 and 2 both open. The wave functions Ψ_1 and Ψ_2 add to produce the intensity at the screen, not the probability densities $|\Psi_1|^2$ and $|\Psi_2|^2$.

mechanics wave functions add, *not* probabilities. Instead the result with both slits open is as shown in Fig. 5.2e, the same pattern of alternating maxima and minima that occurs when a beam of monochromatic light passes through the double slit of Fig. 2.4.

The diffraction pattern of Fig. 5.2e arises from the superposition Ψ of the wave functions Ψ_1 and Ψ_2 of the electrons that have passed through slits 1 and 2:

$$\Psi = \Psi_1 + \Psi_2$$

The probability density at the screen is therefore

$$\begin{aligned} P &= |\Psi|^2 = |\Psi_1 + \Psi_2|^2 = (\Psi_1^* + \Psi_2^*)(\Psi_1 + \Psi_2) \\ &= \Psi_1^*\Psi_1 + \Psi_2^*\Psi_2 + \Psi_1^*\Psi_2 + \Psi_2^*\Psi_1 \\ &= P_1 + P_2 + \Psi_1^*\Psi_2 + \Psi_2^*\Psi_1 \end{aligned}$$

The two terms at the right of this equation represent the difference between Fig. 5.2d and e and are responsible for the oscillations of the electron intensity at the screen. In Sec. 6.8 a similar calculation will be used to investigate why a hydrogen atom emits radiation when it undergoes a transition from one quantum state to another of lower energy.

5.5 EXPECTATION VALUES

How to extract information from a wave function

Once Schrödinger's equation has been solved for a particle in a given physical situation, the resulting wave function $\Psi(x, y, z, t)$ contains all the information about the particle that is permitted by the uncertainty principle. Except for those variables that are quantized this information is in the form of probabilities and not specific numbers.

As an example, let us calculate the **expectation value** $\langle x \rangle$ of the position of a particle confined to the x axis that is described by the wave function $\Psi(x, t)$. This is the value of x we would obtain if we measured the positions of a great many particles described by the same wave function at some instant t and then averaged the results.

To make the procedure clear, we first answer a slightly different question: What is the average position \bar{x} of a number of identical particles distributed along the x axis in such a way that there are N_1 particles at x_1 , N_2 particles at x_2 , and so on? The average position in this case is the same as the center of mass of the distribution, and so

$$\bar{x} = \frac{N_1x_1 + N_2x_2 + N_3x_3 + \dots}{N_1 + N_2 + N_3 + \dots} = \frac{\sum N_i x_i}{\sum N_i} \quad (5.16)$$

When we are dealing with a single particle, we must replace the number N_i of particles at x_i by the probability P_i that the particle be found in an interval dx at x_i . This probability is

$$P_i = |\Psi_i|^2 dx \quad (5.17)$$

where Ψ_i is the particle wave function evaluated at $x = x_i$. Making this substitution and changing the summations to integrals, we see that the expectation value of the

position of the single particle is

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} x |\Psi|^2 dx}{\int_{-\infty}^{\infty} |\Psi|^2 dx} \quad (5.18)$$

If Ψ is a normalized wave function, the denominator of Eq. (5.18) equals the probability that the particle exists somewhere between $x = -\infty$ and $x = \infty$ and therefore has the value 1. In this case

Expectation value
for position

$$\langle x \rangle = \int_{-\infty}^{\infty} x |\Psi|^2 dx \quad (5.19)$$

Example 5.2

A particle limited to the x axis has the wave function $\Psi = ax$ between $x = 0$ and $x = 1$; $\Psi = 0$ elsewhere. (a) Find the probability that the particle can be found between $x = 0.45$ and $x = 0.55$. (b) Find the expectation value $\langle x \rangle$ of the particle's position.

Solution

(a) The probability is

$$\int_{x_1}^{x_2} |\Psi|^2 dx = a^2 \int_{0.45}^{0.55} x^2 dx = a^2 \left[\frac{x^3}{3} \right]_{0.45}^{0.55} = 0.0251a^2$$

(b) The expectation value is

$$\langle x \rangle = \int_0^1 x |\Psi|^2 dx = a^2 \int_0^1 x^3 dx = a^2 \left[\frac{x^4}{4} \right]_0^1 = \frac{a^2}{4}$$

The same procedure as that followed above can be used to obtain the expectation value $\langle G(x) \rangle$ of any quantity—for instance, potential energy $U(x)$ —that is a function of the position x of a particle described by a wave function Ψ . The result is

$$\text{Expectation value} \quad \langle G(x) \rangle = \int_{-\infty}^{\infty} G(x) |\Psi|^2 dx \quad (5.20)$$

The expectation value $\langle p \rangle$ for momentum cannot be calculated this way because, according to the uncertainty principles, no such function as $p(x)$ can exist. If we specify x , so that $\Delta x = 0$, we cannot specify a corresponding p since $\Delta x \Delta p \geq \hbar/2$. The same problem occurs for the expectation value $\langle E \rangle$ for energy because $\Delta E \Delta t \geq \hbar/2$ means that, if we specify t , the function $E(t)$ is impossible. In Sec. 5.6 we will see how $\langle p \rangle$ and $\langle E \rangle$ can be determined.

In classical physics no such limitation occurs, because the uncertainty principle can be neglected in the macroworld. When we apply the second law of motion to the motion of a body subject to various forces, we expect to get $p(x, t)$ and $E(x, t)$ from the solution as well as $x(t)$. Solving a problem in classical mechanics gives us the entire future course of the body's motion. In quantum physics, on the other hand, all we get directly by applying Schrödinger's equation to the motion of a particle is the wave function Ψ , and the future course of the particle's motion—like its initial state—is a matter of probabilities instead of certainties.

5.6 OPERATORS

Another way to find expectation values

A hint as to the proper way to evaluate $\langle p \rangle$ and $\langle E \rangle$ comes from differentiating the free-particle wave function $\Psi = Ae^{-(i/\hbar)(Et - px)}$ with respect to x and to t . We find that

$$\frac{\partial \Psi}{\partial x} = \frac{i}{\hbar} p \Psi$$

$$\frac{\partial \Psi}{\partial t} = -\frac{i}{\hbar} E \Psi$$

which can be written in the suggestive forms

$$p\Psi = \frac{\hbar}{i} \frac{\partial}{\partial x} \Psi \quad (5.21)$$

$$E\Psi = i\hbar \frac{\partial}{\partial t} \Psi \quad (5.22)$$

Evidently the dynamical quantity p in some sense corresponds to the differential operator $(\hbar/i) \partial/\partial x$ and the dynamical quantity E similarly corresponds to the differential operator $i\hbar \partial/\partial t$.

An operator tells us what operation to carry out on the quantity that follows it. Thus the operator $i\hbar \partial/\partial t$ instructs us to take the partial derivative of what comes after it with respect to t and multiply the result by $i\hbar$. Equation (5.22) was on the postmark used to cancel the Austrian postage stamp issued to commemorate the 100th anniversary of Schrödinger's birth.

It is customary to denote operators by using a caret, so that \hat{p} is the operator that corresponds to momentum p and \hat{E} is the operator that corresponds to total energy E . From Eqs. (5.21) and (5.22) these operators are

$$\text{Momentum operator} \quad \hat{p} = \frac{\hbar}{i} \frac{\partial}{\partial x} \quad (5.23)$$

$$\text{Total-energy operator} \quad \hat{E} = i\hbar \frac{\partial}{\partial t} \quad (5.24)$$

Though we have only shown that the correspondences expressed in Eqs. (5.23) and (5.24) hold for free particles, they are entirely general results whose validity is the same as that of Schrödinger's equation. To support this statement, we can replace the equation $E = KE + U$ for the total energy of a particle with the operator equation

$$\hat{E} = \hat{K}\hat{E} + \hat{U} \quad (5.25)$$

The operator \hat{U} is just $U(\Psi)$. The kinetic energy KE is given in terms of momentum p by

$$KE = \frac{p^2}{2m}$$

and so we have

$$\text{Kinetic-energy operator } \hat{K}E = \frac{\hat{p}^2}{2m} = \frac{1}{2m} \left(i \frac{\partial}{\partial x} \right)^2 = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \quad (5.26)$$

Equation (5.25) therefore reads

$$i\hbar \frac{\partial}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U \quad (5.27)$$

Now we multiply the identity $\Psi = \Psi$ by Eq. (5.27) and obtain

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + U\Psi$$

which is Schrödinger's equation. Postulating Eqs. (5.23) and (5.24) is equivalent to postulating Schrödinger's equation.

Operators and Expectation Values

Because p and E can be replaced by their corresponding operators in an equation, we can use these operators to obtain expectation values for p and E . Thus the expectation value for p is

$$\langle p \rangle = \int_{-\infty}^{\infty} \Psi^* \hat{p} \Psi \, dx = \int_{-\infty}^{\infty} \Psi^* \left(\frac{\hbar}{i} \frac{\partial}{\partial x} \right) \Psi \, dx = \frac{\hbar}{i} \int_{-\infty}^{\infty} \Psi^* \frac{\partial \Psi}{\partial x} \, dx \quad (5.28)$$

and the expectation value for E is

$$\langle E \rangle = \int_{-\infty}^{\infty} \Psi^* \hat{E} \Psi \, dx = \int_{-\infty}^{\infty} \Psi^* \left(i\hbar \frac{\partial}{\partial t} \right) \Psi \, dx = i\hbar \int_{-\infty}^{\infty} \Psi^* \frac{\partial \Psi}{\partial t} \, dx \quad (5.29)$$

Both Eqs. (5.28) and (5.29) can be evaluated for any acceptable wave function $\Psi(x, t)$.

Let us see why expectation values involving operators have to be expressed in the form

$$\langle p \rangle = \int_{-\infty}^{\infty} \Psi^* \hat{p} \Psi \, dx$$

The other alternatives are

$$\int_{-\infty}^{\infty} \hat{p} \Psi^* \Psi \, dx = \frac{\hbar}{i} \int_{-\infty}^{\infty} \frac{\partial}{\partial x} (\Psi^* \Psi) \, dx = \frac{\hbar}{i} \left[\Psi^* \Psi \right]_{-\infty}^{\infty} = 0$$

since Ψ^* and Ψ must be 0 at $x = \pm\infty$, and

$$\int_{-\infty}^{\infty} \Psi^* \Psi \hat{p} \, dx = \frac{\hbar}{i} \int_{-\infty}^{\infty} \Psi^* \Psi \frac{\partial}{\partial x} \, dx$$

which makes no sense. In the case of algebraic quantities such as x and $V(x)$, the order of factors in the integrand is unimportant, but when differential operators are involved, the correct order of factors must be observed.

Every observable quantity G characteristic of a physical system may be represented by a suitable quantum-mechanical operator \hat{G} . To obtain this operator, we express G in terms of x and p and then replace p by $(\hbar/i) \partial/\partial x$. If the wave function Ψ of the system is known, the expectation value of $G(x, p)$ is

$$\begin{array}{ll} \text{Expectation value} & \langle G(x, p) \rangle = \int_{-\infty}^{\infty} \Psi^* \hat{G} \Psi \, dx \\ \text{of an operator} & \end{array} \quad (5.30)$$

In this way all the information about a system that is permitted by the uncertainty principle can be obtained from its wave function Ψ .

5.7 SCHRÖDINGER'S EQUATION: STEADY-STATE FORM

Eigenvalues and eigenfunctions

In a great many situations the potential energy of a particle does not depend on time explicitly; the forces that act on it, and hence U , vary with the position of the particle only. When this is true, Schrödinger's equation may be simplified by removing all reference to t .

We begin by noting that the one-dimensional wave function Ψ of an unrestricted particle may be written

$$\Psi = A e^{-(i/\hbar)(Et - px)} = A e^{-(iE/\hbar)t} e^{(ip/\hbar)x} = \psi e^{-(iE/\hbar)t} \quad (5.31)$$

Evidently Ψ is the product of a time-dependent function $e^{-(iE/\hbar)t}$ and a position-dependent function ψ . As it happens, the time variations of *all* wave functions of particles acted on by forces independent of time have the same form as that of an unrestricted particle. Substituting the Ψ of Eq. (5.31) into the time-dependent form of Schrödinger's equation, we find that

$$E\psi e^{-(iE/\hbar)t} = -\frac{\hbar^2}{2m} e^{-(iE/\hbar)t} \frac{\partial^2 \psi}{\partial x^2} + U\psi e^{-(iE/\hbar)t}$$

Dividing through by the common exponential factor gives

$$\begin{array}{ll} \text{Steady-state} & \\ \text{Schrödinger equation} & \frac{\partial^2 \psi}{\partial x^2} + \frac{2m}{\hbar^2} (E - U)\psi = 0 \\ \text{in one dimension} & \end{array} \quad (5.32)$$

Equation (5.32) is the **steady-state form of Schrödinger's equation**. In three dimensions it is

$$\begin{array}{ll} \text{Steady-state} & \\ \text{Schrödinger} & \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{2m}{\hbar^2} (E - U)\psi = 0 \\ \text{equation in three} & \\ \text{dimensions} & \end{array} \quad (5.33)$$

An important property of Schrödinger's steady-state equation is that, if it has one or more solutions for a given system, each of these wave functions corresponds to a specific value of the energy E . Thus energy quantization appears in wave mechanics as a natural element of the theory, and energy quantization in the physical world is revealed as a universal phenomenon characteristic of *all* stable systems.

A familiar and quite close analogy to the manner in which energy quantization occurs in solutions of Schrödinger's equation is with standing waves in a stretched string of length L that is fixed at both ends. Here, instead of a single wave propagating indefinitely in one direction, waves are traveling in both the $+x$ and $-x$ directions simultaneously. These waves are subject to the condition (called a boundary condition) that the displacement y always be zero at both ends of the string. An acceptable function $y(x, t)$ for the displacement must, with its derivatives (except at the ends), be as well-behaved as ψ and its derivatives—that is, be continuous, finite, and single-valued. In this case y must be real, not complex, as it represents a directly measurable quantity. The only solutions of the wave equation, Eq. (5.3), that are in accord with these various limitations are those in which the wavelengths are given by

$$\lambda_n = \frac{2L}{n+1} \quad n = 0, 1, 2, 3, \dots$$

as shown in Fig. 5.3. It is the combination of the wave equation and the restrictions placed on the nature of its solution that leads us to conclude that $y(x, t)$ can exist only for certain wavelengths λ_n .

Eigenvalues and Eigenfunctions

The values of energy E_n for which Schrödinger's steady-state equation can be solved are called **eigenvalues** and the corresponding wave functions ψ_n are called **eigenfunctions**. (These terms come from the German *Eigenwert*, meaning "proper or characteristic value," and *Eigenfunktion*, "proper or characteristic function.") The discrete energy levels of the hydrogen atom

$$E_n = -\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2}\left(\frac{1}{n^2}\right) \quad n = 1, 2, 3, \dots$$

are an example of a set of eigenvalues. We shall see in Chap. 6 why these particular values of E are the only ones that yield acceptable wave functions for the electron in the hydrogen atom.

An important example of a dynamical variable other than total energy that is found to be quantized in stable systems is angular momentum L . In the case of the hydrogen atom, we shall find that the eigenvalues of the magnitude of the total angular momentum are specified by

$$L = \sqrt{l(l+1)}\hbar \quad l = 0, 1, 2, \dots, (n-1)$$

Of course, a dynamical variable G may not be quantized. In this case measurements of G made on a number of identical systems will not yield a unique result but instead a spread of values whose average is the expectation value

$$\langle G \rangle = \int_{-\infty}^{\infty} G|\psi|^2 dx$$

In the hydrogen atom, the electron's position is not quantized, for instance, so that we must think of the electron as being present in the vicinity of the nucleus with a certain probability $|\psi|^2$ per unit volume but with no predictable position or even orbit in the classical sense. This probabilistic statement does not conflict with the fact that

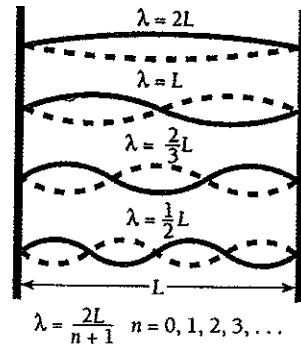


Figure 5.3 Standing waves in a stretched string fastened at both ends.

experiments performed on hydrogen atoms always show that each one contains a whole electron, not 27 percent of an electron in a certain region and 73 percent elsewhere. The probability is one of finding the electron, and although this probability is smeared out in space, the electron itself is not.

Operators and Eigenvalues

The condition that a certain dynamical variable G be restricted to the discrete values G_n —in other words, that G be quantized—is that the wave functions ψ_n of the system be such that

$$\text{Eigenvalue equation} \quad \hat{G}\psi_n = G_n\psi_n \quad (5.34)$$

where \hat{G} is the operator that corresponds to G and each G_n is a real number. When Eq. (5.34) holds for the wave functions of a system, it is a fundamental postulate of quantum mechanics that any measurement of G can only yield one of the values G_n . If measurements of G are made on a number of identical systems all in states described by the particular eigenfunction ψ_k , each measurement will yield the single value G_k .

Example 5.3

An eigenfunction of the operator d^2/dx^2 is $\psi = e^{2x}$. Find the corresponding eigenvalue.

Solution

Here $\hat{G} = d^2/dx^2$, so

$$\hat{G}\psi = \frac{d^2}{dx^2}(e^{2x}) = \frac{d}{dx} \left[\frac{d}{dx}(e^{2x}) \right] = \frac{d}{dx}(2e^{2x}) = 4e^{2x}$$

But $e^{2x} = \psi$, so

$$\hat{G}\psi = 4\psi$$

From Eq. (5.34) we see that the eigenvalue G here is just $G = 4$.

In view of Eqs. (5.25) and (5.26) the total-energy operator \hat{E} of Eq. (5.24) can also be written as

Hamiltonian operator	$\hat{H} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U \quad (5.35)$
-------------------------	--

and is called the **Hamiltonian operator** because it is reminiscent of the Hamiltonian function in advanced classical mechanics, which is an expression for the total energy of a system in terms of coordinates and momenta only. Evidently the steady-state Schrödinger equation can be written simply as

Schrödinger's equation	$\hat{H}\psi_n = E_n\psi_n \quad (5.36)$
---------------------------	--

Table 5.1 Operators Associated with Various Observable Quantities

Quantity	Operator
Position, x	x
Linear momentum, p	$\frac{\hbar}{i} \frac{\partial}{\partial x}$
Potential energy, $U(x)$	$U(x)$
Kinetic energy, $KE = \frac{p^2}{2m}$	$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$
Total energy, E	$i\hbar \frac{\partial}{\partial t}$
Total energy (Hamiltonian form), H	$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x)$

so we can say that the various E_n are the eigenvalues of the Hamiltonian operator \hat{H} . This kind of association between eigenvalues and quantum-mechanical operators is quite general. Table 5.1 lists the operators that correspond to various observable quantities.

5.8 PARTICLE IN A BOX

How boundary conditions and normalization determine wave functions

To solve Schrödinger's equation, even in its simpler steady-state form, usually requires elaborate mathematical techniques. For this reason the study of quantum mechanics has traditionally been reserved for advanced students who have the required proficiency in mathematics. However, since quantum mechanics is the theoretical structure whose results are closest to experimental reality, we must explore its methods and applications to understand modern physics. As we shall see, even a modest mathematical background is enough for us to follow the trains of thought that have led quantum mechanics to its greatest achievements.

The simplest quantum-mechanical problem is that of a particle trapped in a box with infinitely hard walls. In Sec. 3.6 we saw how a quite simple argument yields the energy levels of the system. Let us now tackle the same problem in a more formal way, which will give us the wave function ψ_n that corresponds to each energy level.

We may specify the particle's motion by saying that it is restricted to traveling along the x axis between $x = 0$ and $x = L$ by infinitely hard walls. A particle does not lose energy when it collides with such walls, so that its total energy stays constant. From a formal point of view the potential energy U of the particle is infinite on both sides of the box, while U is a constant—say 0 for convenience—on the inside (Fig. 5.4). Because the particle cannot have an infinite amount of energy, it cannot exist outside the box, and so its wave function ψ is 0 for $x \leq 0$ and $x \geq L$. Our task is to find what ψ is within the box, namely, between $x = 0$ and $x = L$.

Within the box Schrödinger's equation becomes

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} E \psi = 0 \quad (5.37)$$

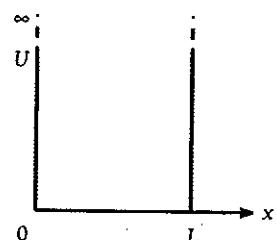


Figure 5.4 A square potential well with infinitely high barriers at each end corresponds to a box with infinitely hard walls.

since $U = 0$ there. (The total derivative $d^2\psi/dx^2$ is the same as the partial derivative $\partial^2\psi/\partial x^2$ because ψ is a function only of x in this problem.) Equation (5.37) has the solution

$$\psi = A \sin \frac{\sqrt{2mE}}{\hbar} x + B \cos \frac{\sqrt{2mE}}{\hbar} x \quad (5.38)$$

which we can verify by substitution back into Eq. (5.37). A and B are constants to be evaluated.

This solution is subject to the boundary conditions that $\psi = 0$ for $x = 0$ and for $x = L$. Since $\cos 0 = 1$, the second term cannot describe the particle because it does not vanish at $x = 0$. Hence we conclude that $B = 0$. Since $\sin 0 = 0$, the sine term always yields $\psi = 0$ at $x = 0$, as required, but ψ will be 0 at $x = L$ only when

$$\frac{\sqrt{2mE}}{\hbar} L = n\pi \quad n = 1, 2, 3, \dots \quad (5.39)$$

This result comes about because the sines of the angles $\pi, 2\pi, 3\pi, \dots$ are all 0.

From Eq. (5.39) it is clear that the energy of the particle can have only certain values, which are the eigenvalues mentioned in the previous section. These eigenvalues, constituting the **energy levels** of the system, are found by solving Eq. (5.39) for E_n , which gives

$$\text{Particle in a box} \quad E_n = \frac{n^2\pi^2\hbar^2}{2mL^2} \quad n = 1, 2, 3, \dots \quad (5.40)$$

Equation (5.40) is the same as Eq. (3.18) and has the same interpretation [see the discussion that follows Eq. (3.18) in Sec. 3.6].

Wave Functions

The wave functions of a particle in a box whose energies are E_n are, from Eq. (5.38) with $B = 0$,

$$\psi_n = A \sin \frac{\sqrt{2mE_n}}{\hbar} x \quad (5.41)$$

Substituting Eq. (5.40) for E_n gives

$$\psi_n = A \sin \frac{n\pi x}{L} \quad (5.42)$$

for the eigenfunctions corresponding to the energy eigenvalues E_n .

It is easy to verify that these eigenfunctions meet all the requirements discussed in Sec. 5.1: for each quantum number n , ψ_n is a finite, single-valued function of x , and ψ_n and $\partial\psi_n/\partial x$ are continuous (except at the ends of the box). Furthermore, the integral

of $|\psi_n|^2$ over all space is finite, as we can see by integrating $|\psi_n|^2 dx$ from $x = 0$ to $x = L$ (since the particle is confined within these limits). With the help of the trigonometric identity $\sin^2 \theta = \frac{1}{2}(1 - \cos 2\theta)$ we find that

$$\begin{aligned} \int_{-\infty}^{\infty} |\psi_n|^2 dx &= \int_0^L |\psi_n|^2 dx = A^2 \int_0^L \sin^2 \left(\frac{n\pi x}{L} \right) dx \\ &= \frac{A^2}{2} \left[\int_0^L dx - \int_0^L \cos \left(\frac{2n\pi x}{L} \right) dx \right] \\ &= \frac{A^2}{2} \left[x - \left(\frac{L}{2n\pi} \right) \sin \frac{2n\pi x}{L} \right]_0^L = A^2 \left(\frac{L}{2} \right) \end{aligned} \quad (5.43)$$

To normalize ψ we must assign a value to A such that $|\psi_n|^2 dx$ is equal to the probability $P dx$ of finding the particle between x and $x + dx$, rather than merely proportional to $P dx$. If $|\psi_n|^2 dx$ is to equal $P dx$, then it must be true that

$$\int_{-\infty}^{\infty} |\psi_n|^2 dx = 1 \quad (5.44)$$

Comparing Eqs. (5.43) and (5.44), we see that the wave functions of a particle in a box are normalized if

$$A = \sqrt{\frac{2}{L}} \quad (5.45)$$

The normalized wave functions of the particle are therefore

$$\text{Particle in a box} \quad \psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \quad n = 1, 2, 3, \dots \quad (5.46)$$

The normalized wave functions ψ_1 , ψ_2 , and ψ_3 together with the probability densities $|\psi_1|^2$, $|\psi_2|^2$, and $|\psi_3|^2$ are plotted in Fig. 5.5. Although ψ_n may be negative as well as positive, $|\psi_n|^2$ is never negative and, since ψ_n is normalized, its value at a given x is equal to the probability density of finding the particle there. In every case $|\psi_n|^2 = 0$ at $x = 0$ and $x = L$, the boundaries of the box.

At a particular place in the box the probability of the particle being present may be very different for different quantum numbers. For instance, $|\psi_1|^2$ has its maximum value of $2/L$ in the middle of the box, while $|\psi_2|^2 = 0$ there. A particle in the lowest energy level of $n = 1$ is most likely to be in the middle of the box, while a particle in the next higher state of $n = 2$ is never there! Classical physics, of course, suggests the same probability for the particle being anywhere in the box.

The wave functions shown in Fig. 5.5 resemble the possible vibrations of a string fixed at both ends, such as those of the stretched string of Fig. 5.2. This follows from the fact that waves in a stretched string and the wave representing a moving particle are described by equations of the same form, so that when identical restrictions are placed upon each kind of wave, the formal results are identical.

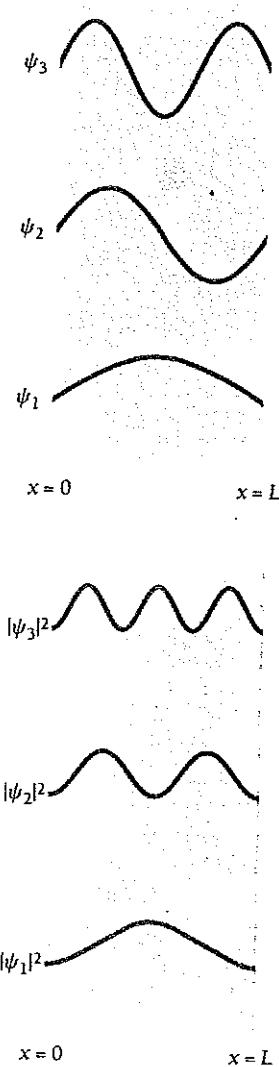


Figure 5.5 Wave functions and probability densities of a particle confined to a box with rigid walls.

Example 5.4

Find the probability that a particle trapped in a box L wide can be found between $0.45L$ and $0.55L$ for the ground and first excited states.

Solution

This part of the box is one-tenth of the box's width and is centered on the middle of the box (Fig. 5.6). Classically we would expect the particle to be in this region 10 percent of the time. Quantum mechanics gives quite different predictions that depend on the quantum number of the particle's state. From Eqs. (5.2) and (5.46) the probability of finding the particle between x_1 and x_2 when it is in the n th state is

$$\begin{aligned} P_{x_1, x_2} &= \int_{x_1}^{x_2} |\psi_n|^2 dx = \frac{2}{L} \int_{x_1}^{x_2} \sin^2 \frac{n\pi x}{L} dx \\ &= \left[\frac{x}{L} - \frac{1}{2n\pi} \sin \frac{2n\pi x}{L} \right]_{x_1}^{x_2} \end{aligned}$$

Here $x_1 = 0.45L$ and $x_2 = 0.55L$. For the ground state, which corresponds to $n = 1$, we have

$$P_{x_1, x_2} = 0.198 = 19.8 \text{ percent}$$

This is about twice the classical probability. For the first excited state, which corresponds to $n = 2$, we have

$$P_{x_1, x_2} = 0.0065 = 0.65 \text{ percent}$$

This low figure is consistent with the probability density of $|\psi_n|^2 = 0$ at $x = 0.5L$.

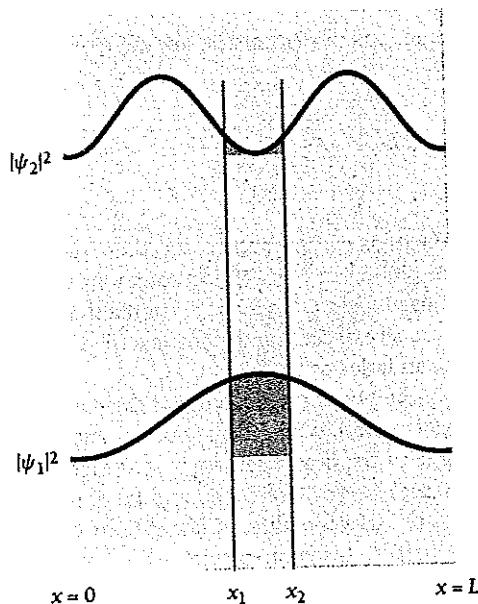


Figure 5.6 The probability P_{x_1, x_2} of finding a particle in the box of Fig. 5.5 between $x_1 = 0.45L$ and $x_2 = 0.55L$ is equal to the area under the $|\psi|^2$ curves between these limits.

Example 5.5

Find the expectation value $\langle x \rangle$ of the position of a particle trapped in a box L wide.

Solution

From Eqs. (5.19) and (5.46) we have

$$\begin{aligned}\langle x \rangle &= \int_{-\infty}^{\infty} x |\psi|^2 dx = \frac{2}{L} \int_0^L x \sin^2 \frac{n\pi x}{L} dx \\ &= \frac{2}{L} \left[\frac{x^2}{4} - \frac{x \sin(2n\pi x/L)}{4n\pi/L} - \frac{\cos(2n\pi x/L)}{8(n\pi/L)^2} \right]_0^L\end{aligned}$$

Since $\sin n\pi = 0$, $\cos 2n\pi = 1$, and $\cos 0 = 1$, for all the values of n the expectation value of x is

$$\langle x \rangle = \frac{2}{L} \left(\frac{L^2}{4} \right) = \frac{L}{2}$$

This result means that the average position of the particle is the middle of the box in all quantum states. There is no conflict with the fact that $|\psi|^2 = 0$ at $L/2$ in the $n = 2, 4, 6, \dots$ states because $\langle x \rangle$ is an *average*, not a probability, and it reflects the symmetry of $|\psi|^2$ about the middle of the box.

Momentum

Finding the momentum of a particle trapped in a one-dimensional box is not as straightforward as finding $\langle x \rangle$. Here

$$\begin{aligned}\psi^* &= \psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \\ \frac{d\psi}{dx} &= \sqrt{\frac{2}{L}} \frac{n\pi}{L} \cos \frac{n\pi x}{L}\end{aligned}$$

and so, from Eq. (5.30),

$$\begin{aligned}\langle p \rangle &= \int_{-\infty}^{\infty} \psi^* \hat{p} \psi dx = \int_{-\infty}^{\infty} \psi^* \left(\frac{\hbar}{i} \frac{d}{dx} \right) \psi dx \\ &= \frac{\hbar}{i} \frac{2}{L} \frac{n\pi}{L} \int_0^L \sin \frac{n\pi x}{L} \cos \frac{n\pi x}{L} dx\end{aligned}$$

We note that

$$\int \sin ax \cos ax dx = \frac{1}{2a} \sin^2 ax$$

With $a = n\pi/L$ we have

$$\langle p \rangle = \frac{\hbar}{iL} \left[\sin^2 \frac{n\pi x}{L} \right]_0^L = 0$$

since

$$\sin^2 0 = \sin^2 n\pi = 0 \quad n = 1, 2, 3, \dots$$

The expectation value $\langle p \rangle$ of the particle's momentum is 0.

At first glance this conclusion seems strange. After all, $E = p^2/2m$, and so we would anticipate that

Momentum eigenvalues for trapped particle

$$p_n = \pm \sqrt{2mE_n} = \pm \frac{n\pi\hbar}{L} \quad (5.47)$$

The \pm sign provides the explanation: The particle is moving back and forth, and so its *average* momentum for any value of n is

$$p_{av} = \frac{(+n\pi\hbar/L) + (-n\pi\hbar/L)}{2} = 0$$

which is the expectation value.

According to Eq. (5.47) there should be two momentum eigenfunctions for every energy eigenfunction, corresponding to the two possible directions of motion. The general procedure for finding the eigenvalues of a quantum-mechanical operator, here \hat{p} , is to start from the eigenvalue equation

$$\hat{p}\psi_n = p_n\psi_n \quad (5.48)$$

where each p_n is a real number. This equation holds only when the wave functions ψ_n are eigenfunctions of the momentum operator \hat{p} , which here is

$$\hat{p} = \frac{\hbar}{i} \frac{d}{dx}$$

We can see at once that the energy eigenfunctions

$$\psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}$$

are not also momentum eigenfunctions, because

$$\frac{\hbar}{i} \frac{d}{dx} \left(\sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \right) = \frac{\hbar}{i} \frac{n\pi}{L} \sqrt{\frac{2}{L}} \cos \frac{n\pi x}{L} \neq p_n \psi_n$$

To find the correct momentum eigenfunctions, we note that

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{1}{2i} e^{i\theta} - \frac{1}{2i} e^{-i\theta}$$

Hence each energy eigenfunction can be expressed as a linear combination of the two wave functions

Momentum eigenfunctions for trapped particle

$$\psi_n^+ = \frac{1}{2i} \sqrt{\frac{2}{L}} e^{in\pi x/L} \quad (5.49)$$

$$\psi_n^- = \frac{1}{2i} \sqrt{\frac{2}{L}} e^{-in\pi x/L} \quad (5.50)$$

Inserting the first of these wave functions in the eigenvalue equation, Eq. (5.48), we have

$$\hat{p}\psi_n^+ = p_n^+ \psi_n^+$$

$$\frac{\hbar}{i} \frac{d}{dx} \psi_n^+ = \frac{\hbar}{i} \frac{1}{2i} \sqrt{\frac{2}{L}} \frac{in\pi}{L} e^{in\pi x/L} = \frac{n\pi\hbar}{L} \psi_n^+ = p_n^+ \psi_n^+$$

so that

$$p_n^+ = +\frac{n\pi\hbar}{L} \quad (5.51)$$

Similarly the wave function ψ_n^- leads to the momentum eigenvalues

$$p_n^- = -\frac{n\pi\hbar}{L} \quad (5.52)$$

We conclude that ψ_n^+ and ψ_n^- are indeed the momentum eigenfunctions for a particle in a box, and that Eq. (5.47) correctly states the corresponding momentum eigenvalues.

5.9 FINITE POTENTIAL WELL

The wave function penetrates the walls, which lowers the energy levels

Potential energies are never infinite in the real world, and the box with infinitely hard walls of the previous section has no physical counterpart. However, potential wells with barriers of finite height certainly do exist. Let us see what the wave functions and energy levels of a particle in such a well are.

Figure 5.7 shows a potential well with square corners that is U high and L wide and contains a particle whose energy E is less than U . According to classical mechanics, when the particle strikes the sides of the well, it bounces off without entering regions I and III. In quantum mechanics, the particle also bounces back and forth, but now it has a certain probability of penetrating into regions I and III even though $E < U$.

In regions I and III Schrödinger's steady-state equation is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} (E - U)\psi = 0$$

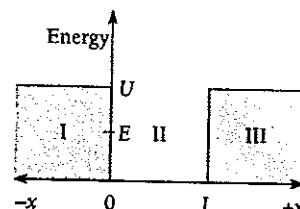


Figure 5.7 A square potential well with finite barriers. The energy E of the trapped particle is less than the height U of the barriers.

which we can rewrite in the more convenient form

$$\frac{d^2\psi}{dx^2} - a^2\psi = 0 \quad \begin{cases} x < 0 \\ x > L \end{cases} \quad (5.53)$$

where

$$a = \frac{\sqrt{2m(U - E)}}{\hbar} \quad (5.54)$$

The solutions to Eq. (5.53) are real exponentials:

$$\psi_I = Ce^{ax} + De^{-ax} \quad (5.55)$$

$$\psi_{III} = Fe^{ax} + Ge^{-ax} \quad (5.56)$$

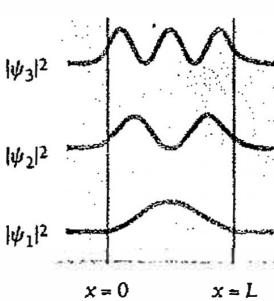
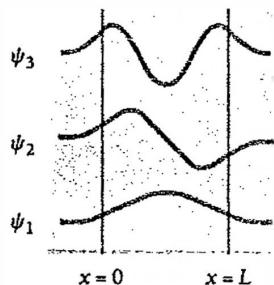


Figure 5.8 Wave functions and probability densities of a particle in a finite potential well. The particle has a certain probability of being found outside the wall.

Both ψ_I and ψ_{III} must be finite everywhere. Since $e^{-ax} \rightarrow \infty$ as $x \rightarrow -\infty$ and $e^{ax} \rightarrow \infty$ as $x \rightarrow \infty$, the coefficients D and F must therefore be 0. Hence we have

$$\psi_I = Ce^{ax} \quad (5.57)$$

$$\psi_{III} = Ge^{-ax} \quad (5.58)$$

These wave functions decrease exponentially inside the barriers at the sides of the well.

Within the well Schrödinger's equation is the same as Eq. (5.37) and its solution is again

$$\psi_{II} = A \sin \frac{\sqrt{2mE}}{\hbar} x + B \cos \frac{\sqrt{2mE}}{\hbar} x \quad (5.59)$$

In the case of a well with infinitely high barriers, we found that $B = 0$ in order that $\psi = 0$ at $x = 0$ and $x = L$. Here, however, $\psi_{II} = C$ at $x = 0$ and $\psi_{II} = G$ at $x = L$, so both the sine and cosine solutions of Eq. (5.59) are possible.

For either solution, both ψ and $d\psi/dx$ must be continuous at $x = 0$ and $x = L$: the wave functions inside and outside each side of the well must not only have the same value where they join but also the same slopes, so they match up perfectly. When these boundary conditions are taken into account, the result is that exact matching only occurs for certain specific values E_n of the particle energy. The complete wave functions and their probability densities are shown in Fig. 5.8.

Because the wavelengths that fit into the well are longer than for an infinite well of the same width (see Fig. 5.5), the corresponding particle momenta are lower (we recall that $\lambda = h/p$). Hence the energy levels E_n are lower for each n than they are for a particle in an infinite well.

Appendix to Chapter 5

The Tunnel Effect

We consider the situation that was shown in Fig. 5.9 of a particle of energy $E < U$ that approaches a potential barrier U high and L wide. Outside the barrier in regions I and III Schrödinger's equation for the particle takes the forms

$$\frac{d^2\psi_1}{dx^2} + \frac{2m}{\hbar^2} E\psi_1 = 0 \quad (5.73)$$

$$\frac{d^2\psi_{III}}{dx^2} + \frac{2m}{\hbar^2} E\psi_{III} = 0 \quad (5.74)$$

The solutions to these equations that are appropriate here are

$$\psi_1 = Ae^{ik_1 x} + Be^{-ik_1 x} \quad (5.75)$$

$$\psi_{III} = Fe^{ik_1 x} + Ge^{-ik_1 x} \quad (5.76)$$

where

$$\text{Wave number outside barrier} \quad k_1 = \frac{\sqrt{2mE}}{\hbar} = \frac{p}{\hbar} = \frac{2\pi}{\lambda} \quad (5.77)$$

is the wave number of the de Broglie waves that represent the particles outside the barrier.

Because

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

$$e^{-i\theta} = \cos \theta - i \sin \theta$$

these solutions are equivalent to Eq. (5.38)—the values of the coefficients are different in each case, of course—but are in a more suitable form to describe particles that are not trapped.

The various terms in Eqs. (5.75) and (5.76) are not hard to interpret. As was shown schematically in Fig. 5.9, $Ae^{ik_1 x}$ is a wave of amplitude A incident from the left on the barrier. Hence we can write

$$\text{Incoming wave} \quad \psi_{I+} = Ae^{ik_1 x} \quad (5.78)$$

This wave corresponds to the incident beam of particles in the sense that $|\psi_{I+}|^2$ is their probability density. If v_{I+} is the group velocity of the incoming wave, which equals the velocity of the particles, then

$$S = |\psi_{I+}|^2 v_{I+}$$

is the flux of particles that arrive at the barrier. That is, S is the number of particles per second that arrive there.

At $x = 0$ the incident wave strikes the barrier and is partially reflected, with

$$\text{Reflected wave} \quad \psi_{I-} = Be^{-ik_I x} \quad (5.79)$$

representing the reflected wave. Hence

$$\psi_I = \psi_{I+} + \psi_{I-} \quad (5.80)$$

On the far side of the barrier ($x > L$) there can only be a wave

$$\text{Transmitted wave} \quad \psi_{III+} = Fe^{ik_3 x} \quad (5.81)$$

traveling in the $+x$ direction at the velocity v_{III+} since region III contains nothing that could reflect the wave. Hence $G = 0$ and

$$\psi_{III} = \psi_{III+} = Fe^{ik_3 x} \quad (5.82)$$

The transmission probability T for a particle to pass through the barrier is the ratio

$$\text{Transmission probability} \quad T = \frac{|\psi_{III+}|^2 v_{III+}}{|\psi_{I+}|^2 v_{I+}} = \frac{FF^* v_{III+}}{AA^* v_{I+}} \quad (5.83)$$

between the flux of particles that emerges from the barrier and the flux that arrives at it. In other words, T is the fraction of incident particles that succeed in tunneling through the barrier. Classically $T = 0$ because a particle with $E < U$ cannot exist inside the barrier; let us see what the quantum-mechanical result is.

In region II Schrödinger's equation for the particles is

$$\frac{d^2\psi_{II}}{dx^2} + \frac{2m}{\hbar^2} (E - U) \psi_{II} = \frac{d^2\psi_{II}}{dx^2} - \frac{2m}{\hbar^2} (U - E) \psi_{II} = 0 \quad (5.84)$$

Since $U > E$ the solution is

$$\text{Wave function inside barrier} \quad \psi_{II} = Ce^{-k_2 x} + De^{k_2 x} \quad (5.85)$$

where the wave number inside the barrier is

$$\text{Wave number inside barrier} \quad k_2 = \frac{\sqrt{2m(U - E)}}{\hbar} \quad (5.86)$$

Since the exponents are real quantities, ψ_{II} does not oscillate and therefore does not represent a moving particle. However, the probability density $|\psi_{II}|^2$ is not zero, so there is a finite probability of finding a particle within the barrier. Such a particle may emerge into region III or it may return to region I.

Applying the Boundary Conditions

In order to calculate the transmission probability T we have to apply the appropriate boundary conditions to ψ_I , ψ_{II} , and ψ_{III} . Fig. 5.14 shows the wave functions in regions I, II, and III. As discussed earlier, both ψ and its derivative $\partial\psi/\partial x$ must be continuous everywhere. With reference to Fig. 5.14, these conditions mean that for a perfect fit at each side of the barrier, the wave functions inside and outside must have the same value and the same slope. Hence at the left-hand side of the barrier

$$\text{Boundary conditions} \quad \left. \begin{array}{l} \psi_I = \psi_{II} \\ \frac{d\psi_I}{dx} = \frac{d\psi_{II}}{dx} \end{array} \right\} x = 0 \quad (5.87)$$

$$\text{at } x = 0 \quad \left. \begin{array}{l} \psi_{II} = \psi_{III} \\ \frac{d\psi_{II}}{dx} = \frac{d\psi_{III}}{dx} \end{array} \right\} x = L \quad (5.88)$$

and at the right-hand side

$$\text{Boundary conditions} \quad \left. \begin{array}{l} \psi_{II} = \psi_{III} \\ \frac{d\psi_{II}}{dx} = \frac{d\psi_{III}}{dx} \end{array} \right\} x = L \quad (5.89)$$

$$\text{at } x = L \quad \left. \begin{array}{l} \psi_{II} = \psi_{III} \\ \frac{d\psi_{II}}{dx} = \frac{d\psi_{III}}{dx} \end{array} \right\} x = L \quad (5.90)$$

Now we substitute ψ_I , ψ_{II} , and ψ_{III} from Eqs. (5.75), (5.81), and (5.85) into the above equations. This yields in the same order

$$A + B = C + D \quad (5.91)$$

$$ik_1 A - ik_1 B = -k_2 C + k_2 D \quad (5.92)$$

$$Ce^{-k_2 L} + De^{k_2 L} = Fe^{ik_1 L} \quad (5.93)$$

$$-k_2 Ce^{-k_2 L} + k_2 De^{k_2 L} = ik_1 F e^{ik_1 L} \quad (5.94)$$

Equations (5.91) to (5.94) may be solved for (A/F) to give

$$\left(\frac{A}{F} \right) = \left[\frac{1}{2} + \frac{i}{4} \left(\frac{k_2}{k_1} - \frac{k_1}{k_2} \right) \right] e^{(ik_1 + k_2)L} + \left[\frac{1}{2} - \frac{i}{4} \left(\frac{k_2}{k_1} - \frac{k_1}{k_2} \right) \right] e^{(ik_1 - k_2)L} \quad (5.95)$$

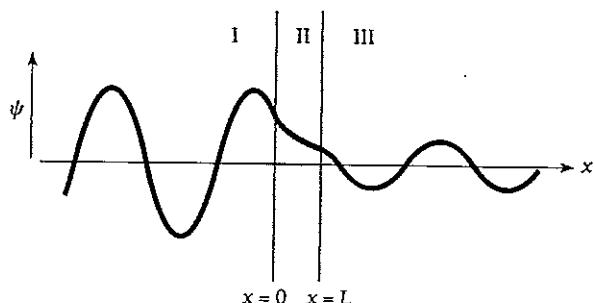


Figure 5.14 At each wall of the barrier, the wave functions inside and outside it must match up perfectly, which means that they must have the same values and slopes there.

Let us assume that the potential barrier U is high relative to the energy E of the incident particles. If this is the case, then $k_2/k_1 > k_1/k_2$ and

$$\frac{k_2}{k_1} - \frac{k_1}{k_2} \approx \frac{k_2}{k_1} \quad (5.96)$$

Let us also assume that the barrier is wide enough for ψ_{II} to be severely weakened between $x = 0$ and $x = L$. This means that $k_2 L \gg 1$ and

$$e^{k_2 L} \gg e^{-k_2 L}$$

Hence Eq. (5.95) can be approximated by

$$\left(\frac{A}{F}\right) = \left(\frac{1}{2} + \frac{i k_2}{4 k_1}\right) e^{(ik_1+k_2)L} \quad (5.97)$$

The complex conjugate of (A/F) , which we need to compute the transmission probability T , is found by replacing i by $-i$ wherever it occurs in (A/F) :

$$\left(\frac{A}{F}\right)^* = \left(\frac{1}{2} - \frac{i k_2}{4 k_1}\right) e^{(-ik_1+k_2)L} \quad (5.98)$$

Now we multiply (A/F) and $(A/F)^*$ to give

$$\frac{AA^*}{FF^*} = \left(\frac{1}{4} + \frac{k_2^2}{16 k_1^2}\right) e^{2k_2 L}$$

Here $v_{III+} = v_{I+}$ so $v_{III+}/v_{I+} = 1$ in Eq. (5.83), which means that the transmission probability is

$$\text{Transmission probability } T = \frac{FF^* v_{I+}}{AA^* v_{I+}} = \left(\frac{AA^*}{FF^*}\right)^{-1} = \left[\frac{16}{4 + (k_2/k_1)^2}\right] e^{-2k_2 L} \quad (5.99)$$

From the definitions of k_1 , Eq. (5.77), and of k_2 , Eq. (5.86), we see that

$$\left(\frac{k_2}{k_1}\right)^2 = \frac{2m(U-E)/\hbar^2}{2mE/\hbar^2} = \frac{U}{E} - 1 \quad (5.100)$$

This formula means that the quantity in brackets in Eq. (5.99) varies much less with E and U than does the exponential. The bracketed quantity, furthermore, always is of the order of magnitude of 1 in value. A reasonable approximation of the transmission probability is therefore

Approximate transmission probability	$T = e^{-2k_2 L}$	(5.101)
--	-------------------	---------

as stated in Sec. 5.10.

EXERCISES

Press on, and faith will catch up with you. —Jean D'Alembert

5.1 Quantum Mechanics

1. Which of the wave functions in Fig. 5.15 cannot have physical significance in the interval shown? Why not?
2. Which of the wave functions in Fig. 5.16 cannot have physical significance in the interval shown? Why not?

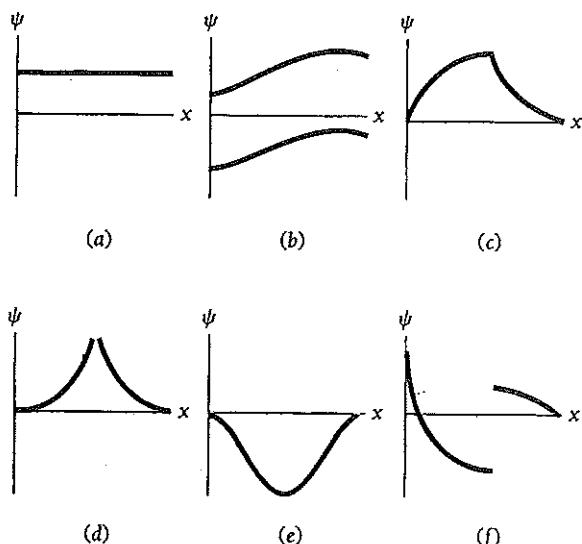


Figure 5.15

3. Which of the following wave functions cannot be solutions of Schrödinger's equation for all values of x ? Why not? (a) $\psi = A \sec x$; (b) $\psi = A \tan x$; (c) $\psi = Ae^{x^2}$; (d) $\psi = Ae^{-x^2}$.
4. Find the value of the normalization constant A for the wave function $\psi = Axe^{-x^2/2}$.
5. The wave function of a certain particle is $\psi = A \cos^2 x$ for $-\pi/2 < x < \pi/2$. (a) Find the value of A . (b) Find the probability that the particle be found between $x = 0$ and $x = \pi/4$.

5.2 The Wave Equation

6. The formula $y = A \cos \omega(t - x/v)$, as we saw in Sec. 3.3, describes a wave that moves in the $+x$ direction along a stretched string. Show that this formula is a solution of the wave equation, Eq. (5.3).
7. As mentioned in Sec. 5.1, in order to give physically meaningful results in calculations a wave function and its partial derivatives must be finite, continuous, and single-valued, and in addition must be normalizable. Equation (5.9) gives the wave function of a particle moving freely (that is, with no forces acting on it) in the $+x$ direction as

$$\Psi = Ae^{-(i/\hbar)(Et - px)}$$

where E is the particle's total energy and p is its momentum. Does this wave function meet all the above requirements? If not, could a linear superposition of such wave functions meet these requirements? What is the significance of such a superposition of wave functions?

5.4 Linearity and Superposition

8. Prove that Schrödinger's equation is linear by showing that

$$\Psi = a_1\Psi_1(x, t) + a_2\Psi_2(x, t)$$

is also a solution of Eq. (5.14) if Ψ_1 and Ψ_2 are themselves solutions.

5.6 Operators

9. Show that the expectation values $\langle px \rangle$ and $\langle xp \rangle$ are related by

$$\langle px \rangle - \langle xp \rangle = \frac{\hbar}{i}$$

This result is described by saying that p and x do not commute and it is intimately related to the uncertainty principle.

10. An eigenfunction of the operator d^2/dx^2 is $\sin nx$, where $n = 1, 2, 3, \dots$. Find the corresponding eigenvalues.

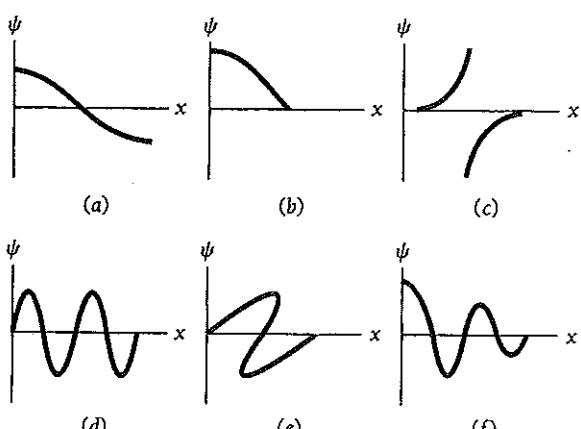


Figure 5.16

5.7 Schrödinger's Equation: Steady-State Form

11. Obtain Schrödinger's steady-state equation from Eq. (3.5) with the help of de Broglie's relationship $\lambda = h/mv$ by letting $y = \psi$ and finding $\partial^2\psi/\partial x^2$.

5.8 Particle in a Box

12. According to the correspondence principle, quantum theory should give the same results as classical physics in the limit of large quantum numbers. Show that as $n \rightarrow \infty$, the probability of finding the trapped particle of Sec. 5.8 between x and $x + \Delta x$ is $\Delta x/L$ and so is independent of x , which is the classical expectation.
13. One of the possible wave functions of a particle in the potential well of Fig. 5.17 is sketched there. Explain why the wavelength and amplitude of ψ vary as they do.

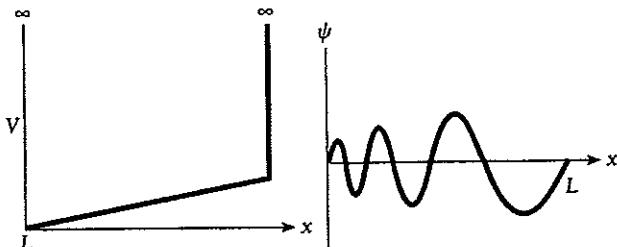


Figure 5.17

14. In Sec. 5.8 a box was considered that extends from $x = 0$ to $x = L$. Suppose the box instead extends from $x = x_0$ to $x = x_0 + L$, where $x_0 \neq 0$. Would the expression for the wave functions of a particle in this box be any different from those in the box that extends from $x = 0$ to $x = L$? Would the energy levels be different?

15. An important property of the eigenfunctions of a system is that they are orthogonal to one another, which means that

$$\int_{-\infty}^{\infty} \psi_n \psi_m dV = 0 \quad n \neq m$$

Verify this relationship for the eigenfunctions of a particle in a one-dimensional box given by Eq. (5.46).

16. A rigid-walled box that extends from $-L$ to L is divided into three sections by rigid interior walls at $-x$ and x , where $x < L$. Each section contains one particle in its ground state. (a) What is the total energy of the system as a function of x ? (b) Sketch $E(x)$ versus x . (c) At what value of x is $E(x)$ a minimum?
17. As shown in the text, the expectation value $\langle x \rangle$ of a particle trapped in a box L wide is $L/2$, which means that its average position is the middle of the box. Find the expectation value $\langle x^2 \rangle$.
18. As noted in Exercise 8, a linear combination of two wave functions for the same system is also a valid wave function. Find the normalization constant B for the combination

$$\psi = B \left(\sin \frac{\pi x}{L} + \sin \frac{2\pi x}{L} \right)$$

of the wave functions for the $n = 1$ and $n = 2$ states of a particle in a box L wide.

19. Find the probability that a particle in a box L wide can be found between $x = 0$ and $x = L/n$ when it is in the n th state.
20. In Sec. 3.7 the standard deviation σ of a set of N measurements of some quantity x was defined as

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

- (a) Show that, in terms of expectation values, this formula can be written as

$$\sigma = \sqrt{\langle (x - \langle x \rangle)^2 \rangle} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

- (b) If the uncertainty in position of a particle in a box is taken as the standard deviation, find the uncertainty in the expectation value $\langle x \rangle = L/2$ for $n = 1$. (c) What is the limit of Δx as n increases?

21. A particle is in a cubic box with infinitely hard walls whose edges are L long (Fig. 5.18). The wave functions of the particle are given by

$$\psi = A \sin \frac{n_x \pi x}{L} \sin \frac{n_y \pi y}{L} \sin \frac{n_z \pi z}{L} \quad \begin{matrix} n_x = 1, 2, 3, \dots \\ n_y = 1, 2, 3, \dots \\ n_z = 1, 2, 3, \dots \end{matrix}$$

Find the value of the normalization constant A .

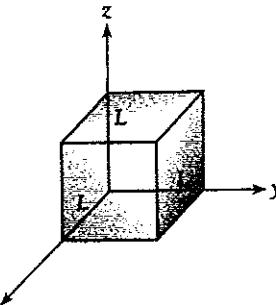


Figure 5.18 A cubic box.

22. The particle in the box of Exercise 21 is in its ground state of $n_x = n_y = n_z = 1$. (a) Find the probability that the particle will be found in the volume defined by $0 \leq x \leq L/4$, $0 \leq y \leq L/4$, $0 \leq z \leq L/4$. (b) Do the same for $L/2$ instead of $L/4$.
23. (a) Find the possible energies of the particle in the box of Exercise 21 by substituting its wave function ψ in Schrödinger's equation and solving for E . (Hint: Inside the box $U = 0$.) (b) Compare the ground-state energy of a particle in a one-dimensional box of length L with that of a particle in the three-dimensional box.

5.10 Tunnel Effect

24. Electrons with energies of 0.400 eV are incident on a barrier 3.00 eV high and 0.100 nm wide. Find the approximate probability for these electrons to penetrate the barrier.

25. A beam of electrons is incident on a barrier 6.00 eV high and 0.200 nm wide. Use Eq. (5.60) to find the energy they should have if 1.00 percent of them are to get through the barrier.

5.11 Harmonic Oscillator

26. Show that the energy-level spacing of a harmonic oscillator is in accord with the correspondence principle by finding the ratio $\Delta E_n/E_n$ between adjacent energy levels and seeing what happens to this ratio as $n \rightarrow \infty$.
27. What bearing would you think the uncertainty principle has on the existence of the zero-point energy of a harmonic oscillator?
28. In a harmonic oscillator, the particle varies in position from $-A$ to $+A$ and in momentum from $-p_0$ to $+p_0$. In such an oscillator, the standard deviations of x and p are $\Delta x = A/\sqrt{2}$ and $\Delta p = p_0/\sqrt{2}$. Use this observation to show that the minimum energy of a harmonic oscillator is $\frac{1}{2}h\nu$.
29. Show that for the $n = 0$ state of a harmonic oscillator whose classical amplitude of motion is A , $y = 1$ at $x = A$, where y is the quantity defined by Eq. (5.67).
30. Find the probability density $|\psi_0|^2 dx$ at $x = 0$ and at $x = \pm A$ of a harmonic oscillator in its $n = 0$ state (see Fig. 5.13).
31. Find the expectation values $\langle x \rangle$ and $\langle x^2 \rangle$ for the first two states of a harmonic oscillator.
32. The potential energy of a harmonic oscillator is $U = \frac{1}{2}kx^2$. Show that the expectation value $\langle U \rangle$ of U is $E_0/2$ when the oscillator is in the $n = 0$ state. (This is true of all states of the harmonic oscillator, in fact.) What is the expectation value of the oscillator's kinetic energy? How do these results compare with the classical values of \bar{U} and \bar{K} ?
33. A pendulum with a 1.00-g bob has a massless string 250 mm long. The period of the pendulum is 1.00 s. (a) What is its zero-point energy? Would you expect the zero-point oscillations to be detectable? (b) The pendulum swings with a very small

amplitude such that its bob rises a maximum of 1.00 mm above its equilibrium position. What is the corresponding quantum number?

34. Show that the harmonic-oscillator wave function ψ_1 is a solution of Schrödinger's equation.
35. Repeat Exercise 34 for ψ_2 .
36. Repeat Exercise 34 for ψ_3 .
- Appendix: The Tunnel Effect**
37. Consider a beam of particles of kinetic energy E incident on a potential step at $x = 0$ that is U high, where $E > U$ (Fig. 5.19). (a) Explain why the solution $D e^{-ikx}$ (in the notation of appendix) has no physical meaning in this situation, so that $D = 0$. (b) Show that the transmission probability here is $T = CC^*v'/AA^*v_1 = 4k_1^2/(k_1 + k')^2$. (c) A 1.00-mA beam of electrons moving at 2.00×10^6 m/s enters a region with a sharply defined boundary in which the electron speeds are reduced to 1.00×10^6 m/s by a difference in potential. Find the transmitted and reflected currents.
38. An electron and a proton with the same energy E approach a potential barrier whose height U is greater than E . Do they have the same probability of getting through? If not, which has the greater probability?

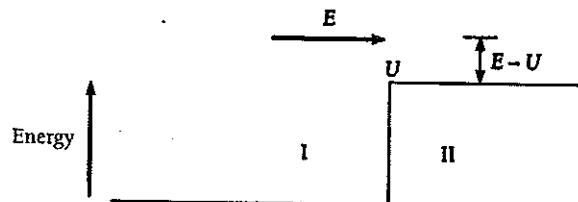
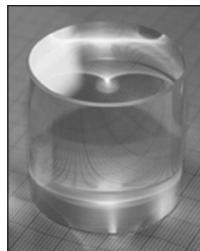
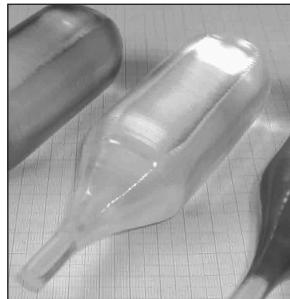
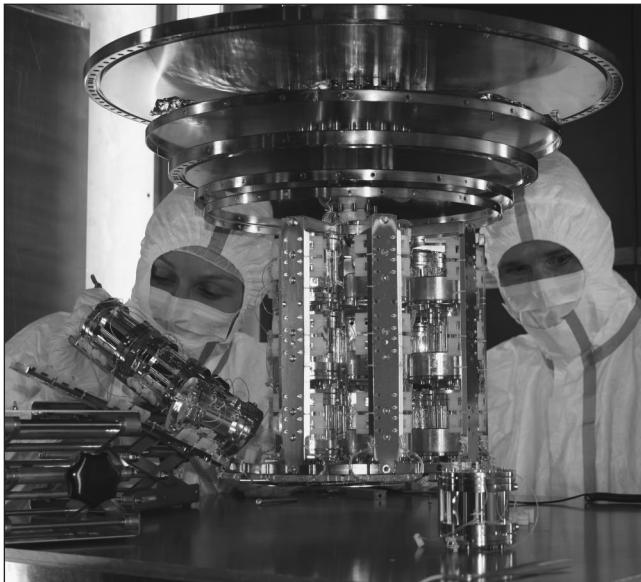


Figure 5.19



Left: GaAs ingots and wafers. GaAs is a III–V compound semiconductor because Ga and As are from Groups III and V, respectively.
Right: An $\text{In}_x\text{Ga}_{1-x}\text{As}$ (a III–V compound semiconductor)-based photodetector.

Left: Courtesy of Sumitomo Electric Industries. Right: Courtesy of Thorlabs.



Left: A detector structure that will be used to detect dark matter particles. Each individual cylindrical detector has a CaWO_4 single crystal, similar to that shown on the bottom right. These crystals are called scintillators, and convert high-energy radiation to light. The Czochralski technique is used to grow the crystal shown on top right, which is a CaWO_4 ingot. The detector crystal is cut from this ingot.

Left: Courtesy of Max Planck Institute for Physics. Right: Reproduced from Andreas Erb and Jean-Come Lanfranchi, *CrystEngCom*, 15, 2301, 2015, by permission of the Royal Society of Chemistry. All rights reserved.

CHAPTER**1**

Elementary Materials Science Concepts¹

Understanding the basic building blocks of matter has been one of the most intriguing endeavors of humankind. Our understanding of interatomic interactions has now reached a point where we can quite comfortably explain the macroscopic properties of matter, based on quantum mechanics and electrostatic interactions between electrons and ionic nuclei in the material. There are many properties of materials that can be explained by a classical treatment of the subject. In this chapter, as well as in Chapter 2, we treat the interactions in a material from a classical perspective and introduce a number of elementary concepts. These concepts do not invoke any quantum mechanics, which is a subject of modern physics and is introduced in Chapter 3. Although many useful engineering properties of materials can be treated with hardly any quantum mechanics, it is impossible to develop the science of electronic materials and devices without modern physics.

1.9 THE CRYSTALLINE STATE

1.9.1 TYPES OF CRYSTALS

A **crystalline solid** is a solid in which the atoms bond with each other in a regular pattern to form a periodic collection (or array) of atoms, as shown for the copper crystal in Figure 1.32. The most important property of a crystal is **periodicity**, which leads to what is termed **long-range order**. In a crystal, the local bonding geometry is repeated many times at regular intervals, to produce a periodic array of atoms that constitutes the crystal structure. The location of each atom is well known by virtue of periodicity. There is therefore a long-range order, since we can always predict the atomic arrangement anywhere in the crystal. Nearly all metals, many ceramics and semiconductors, and various polymers are crystalline solids in the sense that the atoms or molecules are positioned on a **periodic array of points in space**.

All crystals can be described in terms of a lattice and a basis.¹⁵ A **lattice** is an infinite *periodic* array of geometric points in space, without any atoms. When we place an identical group of atoms (or molecules), called a **basis**, at each lattice point, we obtain the actual **crystal structure**. The crystal is thus a lattice plus a basis at each lattice point. In the copper crystal in Figure 1.32a, each lattice point has one Cu atom and the basis is a single Cu atom. As apparent from Figure 1.32a, the lattice of the copper crystal has **cubic symmetry** and is one of many possible lattices.

Since the crystal is essentially a periodic repetition of a small volume (or cell) of atoms in three dimensions, it is useful to identify the repeating unit so that the

¹⁵ Lattice is a purely imaginary geometric concept whose only requirement is that the infinite array of points has periodicity. In many informal discussions, the term lattice or crystal lattice is used to mean the crystal structure itself. These concepts are further developed in Section 1.14 under Additional Topics.

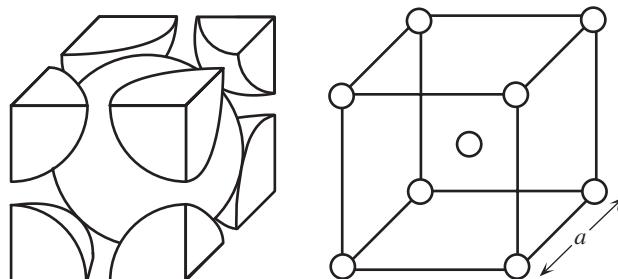
crystal properties can be described through this unit. The **unit cell** is the most convenient small cell in the crystal structure that carries the properties of the crystal. The repetition of the unit cell in three dimensions generates the whole crystal structure, as is apparent in Figure 1.32a for the copper crystal.

The unit cell of the copper crystal is cubic with Cu atoms at its corners and one Cu atom at the center of each face, as indicated in Figure 1.32b. The unit cell of Cu is thus said to have a **face-centered cubic (FCC)** structure. The Cu atoms are shared with neighboring unit cells. Effectively, then, only one-eighth of a corner atom is in the unit cell and one-half of the face-centered atom belongs to the unit cell, as shown in Figure 1.32b. This means there are effectively four atoms in the unit cell. The length of the cubic unit cell is termed the **lattice parameter** a of the crystal structure. For Cu, for example, a is 0.362 nm, whereas the radius R of the Cu atom in the crystal is 0.128 nm. Assuming the Cu atoms are spheres that touch each other, we can geometrically relate a and R . For clarity, it is often more convenient to draw the unit cell with the spheres reduced, as in Figure 1.32c.

The FCC crystal structure of Cu is known as a **close-packed crystal structure** because the Cu atoms are packed as closely as possible, as is apparent in Figure 1.32a and b. The volume of the FCC unit cell is 74 percent full of atoms, which is the maximum packing possible with identical spheres. By comparison, iron has a **body-centered cubic (BCC)** crystal structure, and its unit cell is shown in Figure 1.33. The BCC unit cell has Fe atoms at its corners and one Fe atom at the center of the cell. The volume of the BCC unit cell is 68 percent full of atoms, which is lower than the maximum possible packing.

The FCC crystal structure is only one way to pack the atoms as closely as possible. For example, in zinc, the atoms are arranged as closely as possible in a hexagonal symmetry, to form the **hexagonal close-packed (HCP) structure** shown in Figure 1.34a. This structure corresponds to packing spheres as closely as possible first as one layer A, as shown in Figure 1.34b. You can visualize this by arranging six pennies as closely as possible on a table top. On top of layer A we can place an identical layer B, with the spheres taking up the voids on layer A, as depicted in Figure 1.34b. The third layer can be placed on top of B and lined up with layer A. The stacking sequence is therefore ABAB. . . . A unit cell for the HCP structure is shown in Figure 1.34c, which shows that this is not a cubic structure. The unit cell

Figure 1.33 Body-centered cubic (BCC) crystal structure.
(a) A BCC unit cell with close-packed hard spheres representing the Fe atoms. (b) A reduced-sphere unit cell.



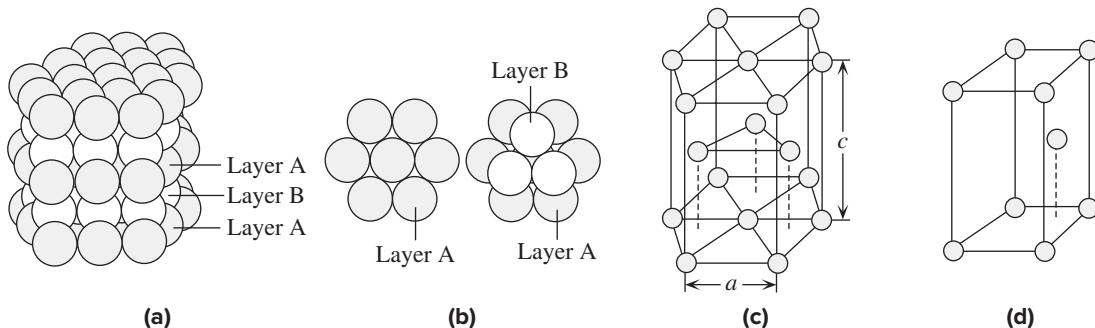


Figure 1.34 The hexagonal close-packed (HCP) crystal structure. (a) The hexagonal close-packed (HCP) structure. A collection of many Zn atoms. Color difference distinguishes layers (stacks). (b) The stacking sequence of closely packed layers is ABAB. (c) A unit cell with reduced spheres. (d) The smallest unit cell with reduced spheres.

shown, although convenient, is not the smallest unit cell. The smallest unit cell for the HCP structure is shown in Figure 1.34d and is called the **hexagonal unit cell**. The repetition of this unit cell will generate the whole HCP structure. The atomic packing density in the HCP crystal structure is 74 percent, which is the same as that in the FCC structure.

Covalently bonded solids, such as silicon and germanium, have a diamond crystal structure brought about by the directional nature of the covalent bond, as shown in Figure 1.35 (see also Figure 1.6). The rigid local bonding geometry of four Si–Si bonds in the tetrahedral configuration forces the atoms to form what is called the **diamond cubic crystal structure**. The unit cell in this case can be identified with the cubic structure. Although there are atoms at each corner and at the center of each face, indicating an FCC-like structure, there are four atoms within the cell as well. Thus, there are eight atoms in the unit cell. The diamond unit cell can actually be described in terms of an FCC lattice (a geometric arrangement of points) with each lattice point having a basis of two Si atoms. If we place the two Si atoms at each site appropriately, for example, one right at the lattice point, and the other displaced from it by a quarter lattice distance $a/4$ along the cube edges, we can easily generate the diamond unit cell. In the copper crystal, each FCC lattice point has one Cu atom, whereas in the Si crystal each lattice point has two Si atoms; thus there are $4 \times 2 = 8$ atoms in the diamond unit cell.

In the GaAs crystal, as in the silicon crystal, each atom forms four directional bonds with its neighbors. The unit cell looks like a diamond cubic, as indicated in Figure 1.36 but with the Ga and As atoms alternating positions. This unit cell is termed the **zinc blende** structure after ZnS, which has this type of unit cell. Many important compound semiconductors have this crystal structure, GaAs being the most commonly known. The zinc blende unit cell can also be described in terms of a fundamental FCC lattice and a basis that has two atoms, Zn and S (or Ga and As). For example, we can place one S at each lattice point and one Zn atom displaced from the Zn by $a/4$ along the cube edges.

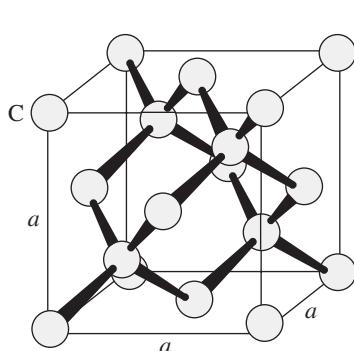


Figure 1.35 The diamond unit cell which is cubic. The cell has eight atoms.

Gray Sn (α -Sn) and the elemental semiconductors Ge and Si have this crystal structure.

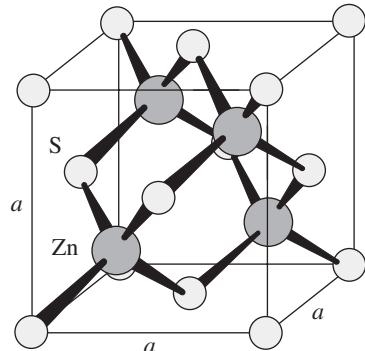


Figure 1.36 The zinc blende (ZnS) cubic crystal structure.

Many important compound crystals have the zinc blende structure. Examples: AlAs, GaAs, GaP, GaSb, InAs, InP, InSb, ZnS, ZnTe.

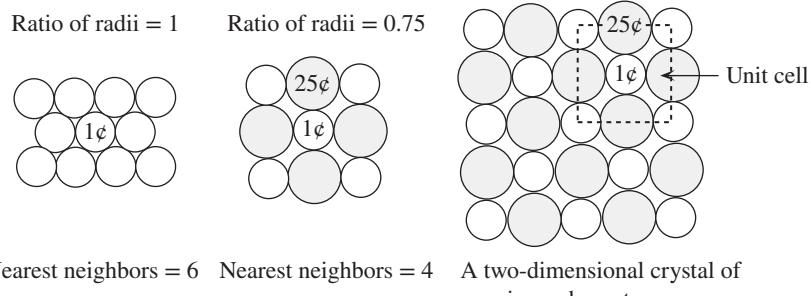


Figure 1.37 Packing of coins on a table top to build a two-dimensional crystal.

In ionic solids, the cations (*e.g.*, Na^+) and the anions (Cl^-) attract each other nondirectionally. The crystal structure depends on how closely the opposite ions can be brought together and how the same ions can best avoid each other while maintaining long-range order, or maintaining symmetry. These depend on the relative charge and relative size per ion.

To demonstrate the importance of the size effect in two dimensions, consider identical coins, say pennies (1-cent coins). At most, we can make six pennies touch one penny, as shown in Figure 1.37. On the other hand, if we use quarters¹⁶

¹⁶ Although many are familiar with the United States coinage, any two coins with a size ratio of about 0.75 would work out the same.

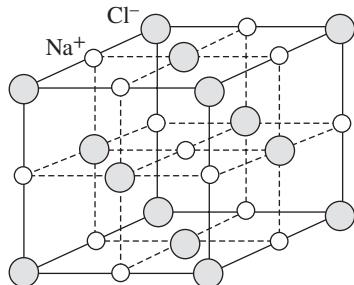


Figure 1.38 A possible reduced-sphere unit cell for the NaCl (rock salt) crystal.

An alternative unit cell may have Na^+ and Cl^- interchanged. Examples: AgCl , CaO , CsF , LiF , LiCl , NaF , NaCl , KF , KCl , and MgO .

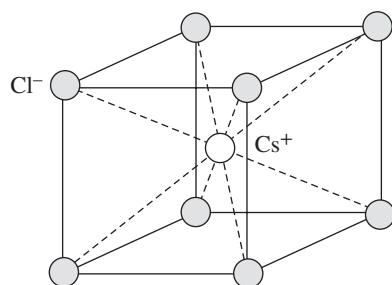


Figure 1.39 A possible reduced-sphere unit cell for the CsCl crystal.

An alternative unit cell may have Cs^+ and Cl^- interchanged. Examples: CsCl , CsBr , CsI , TlCl , TlBr , and TlI .

(25-cent coins) to touch one penny, at most only five quarters can do so. However, this arrangement cannot be extended to the construction of a two-dimensional crystal with periodicity. To fulfill the long-range symmetry requirement for crystals, we can only use four quarters to touch the penny and thereby build a two-dimensional “penny–quarter” crystal, which is shown in the figure. In the two-dimensional crystal, a penny has four quarters as nearest neighbors; similarly, a quarter has four pennies as nearest neighbors. A convenient unit cell is a square cell with one-quarter of a penny at each corner and a full penny at the center (as shown in the figure).

The three-dimensional equivalent of the unit cell of the penny–quarter crystal is the **NaCl unit cell** shown in Figure 1.38. The Na^+ ion is about half the size of the Cl^- ion, which permits six nearest neighbors while maintaining long-range order. The repetition of this unit cell in three dimensions generates the whole NaCl crystal, which was depicted in Figure 1.9b.

A similar unit cell with Na^+ and Cl^- interchanged is also possible and equally convenient. We can therefore describe the whole crystal with two interpenetrating FCC unit cells, each having oppositely charged ions at the corners and face centers. Many ionic solids have the rock salt (NaCl) crystal structure.

When the cation and anions have equal charges and are about the same size, as in the CsCl crystal, the unit cell is called the **CsCl unit cell**, which is shown in Figure 1.39. Each cation is surrounded by eight anions (and vice versa), which are at the corners of a cube. This is not a true BCC unit cell because the atoms at various BCC lattice points are different. (As discussed in Section 1.14, CsCl has a simple cubic lattice with a basis that has one Cl^- ion and one Cs^+ ion.)

Table 1.3 summarizes some of the important properties of the main crystal structures considered in this section.

Table 1.3 Properties of some important crystal structures

Crystal Structure	a and R (R is the Radius of the Atom)	Coordination Number (CN)	Number of Atoms per Unit Cell	Atomic Packing Factor	Examples
Simple cubic	$a = 2R$	6	1	0.52	No metals (Except Po)
BCC	$a = \frac{4R}{\sqrt{3}}$	8	2	0.68	Many metals: α -Fe, Cr, Mo, W
FCC	$a = \frac{4R}{\sqrt{2}}$	12	4	0.74	Many metals: Ag, Au, Cu, Pt
HCP	$a = 2R$ $c = 1.633a$	12	2	0.74	Many metals: Co, Mg, Ti, Zn
Diamond	$a = \frac{8R}{\sqrt{3}}$	4	8	0.34	Covalent solids: Diamond, Ge, Si, α -Sn
Zinc blende		4	8	0.34	Many covalent and ionic solids. Many compound semiconductors. ZnS, GaAs, GaSb, InAs, InSb
NaCl		6	4 cations 4 anions	0.67 (NaCl)	Ionic solids such as NaCl, AgCl, LiF, MgO, CaO Ionic packing factor depends on relative sizes of ions.
CsCl		8	1 cation 1 anion		Ionic solids such as CsCl, CsBr, CsI

EXAMPLE 1.16

THE COPPER (FCC) CRYSTAL Consider the FCC unit cell of the copper crystal shown in Figure 1.40.

- How many atoms are there per unit cell?
- If R is the radius of the Cu atom, show that the lattice parameter a is given by $a = R2\sqrt{2}$.
- Calculate the **atomic packing factor** (APF) defined by

$$\text{APF} = \frac{\text{Volume of atoms in unit cell}}{\text{Volume of unit cell}}$$

- Calculate the **atomic concentration** (number of atoms per unit volume) in Cu and the density of the crystal given that the atomic mass of Cu is 63.55 g mol^{-1} and the radius of the Cu atom is 0.128 nm .

SOLUTION

- There are four atoms per unit cell. The Cu atom at each corner is shared with eight other adjoining unit cells. Each Cu atom at the face center is shared with the neighboring unit cell. Thus, the number of atoms in the unit cell = 8 corners ($\frac{1}{8}$ atom) + 6 faces ($\frac{1}{2}$ atom) = 4 atoms.
- Consider the unit cell shown in Figure 1.40 and one of the cubic faces. The face is a square of side a and the diagonal is $\sqrt{a^2 + a^2}$ or $a\sqrt{2}$. The diagonal has one atom at the center of diameter $2R$, which touches two atoms centered at the corners. The diagonal, going from corner to corner, is therefore $R + 2R + R$. Thus, $4R = a\sqrt{2}$ and $a = 4R/\sqrt{2} = R2\sqrt{2}$. Therefore, $a = 0.3620 \text{ nm}$.

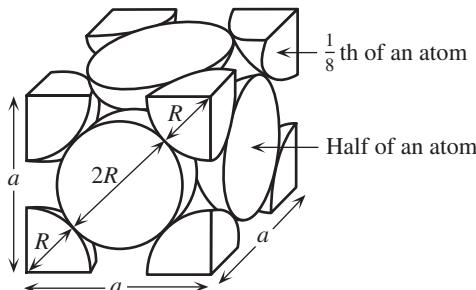


Figure 1.40 The FCC unit cell.
The atomic radius is R and the lattice parameter is a .

$$\begin{aligned}
 c. \quad \text{APF} &= \frac{(\text{Number of atoms in unit cell}) \times (\text{Volume of atom})}{\text{Volume of unit cell}} \\
 &= \frac{4 \times \frac{4}{3}\pi R^3}{a^3} = \frac{\frac{4^2}{3}\pi R^3}{(R2\sqrt{2})^3} = \frac{4^2\pi}{3(2\sqrt{2})^3} = 0.74
 \end{aligned}$$

d. In general, if there are x atoms in the unit cell, the atomic concentration is

$$n_{\text{at}} = \frac{\text{Number of atoms in unit cell}}{\text{Volume of unit cell}} = \frac{x}{a^3}$$

Thus, for Cu

$$n_{\text{at}} = \frac{4}{(0.3620 \times 10^{-7} \text{ cm})^3} = 8.43 \times 10^{22} \text{ cm}^{-3}$$

There are x atoms in the unit cell, and each atom has a mass of M_{at}/N_A grams. The density ρ is

$$\rho = \frac{\text{Mass of all atoms in unit cell}}{\text{Volume of unit cell}} = \frac{x \left(\frac{M_{\text{at}}}{N_A} \right)}{a^3}$$

that is,

$$\rho = \frac{n_{\text{at}} M_{\text{at}}}{N_A} = \frac{(8.43 \times 10^{22} \text{ cm}^{-3})(63.55 \text{ g mol}^{-1})}{6.022 \times 10^{23} \text{ mol}^{-1}} = 8.9 \text{ g cm}^{-3}$$

Note that the expression $\rho = (n_{\text{at}} M_{\text{at}})/N_A$ is particularly useful in finding the atomic concentration n_{at} from the density since the latter can be easily measured or available in various data resources.

1.9.2 CRYSTAL DIRECTIONS AND PLANES

There can be a number of possibilities for choosing a unit cell for a given crystal structure, as is apparent in Figure 1.34c and d for the HCP crystal. As a convention, we generally represent the **geometry of the unit cell** as a parallelepiped with sides a , b , and c and angles α , β , and γ , as depicted in Figure 1.41a. The sides a , b , and c and angles α , β , and γ are referred to as the **lattice parameters**. To establish a reference frame and to apply three-dimensional geometry, we insert an xyz coordinate system. The x , y , and z axes follow the edges of the parallelepiped and the origin is

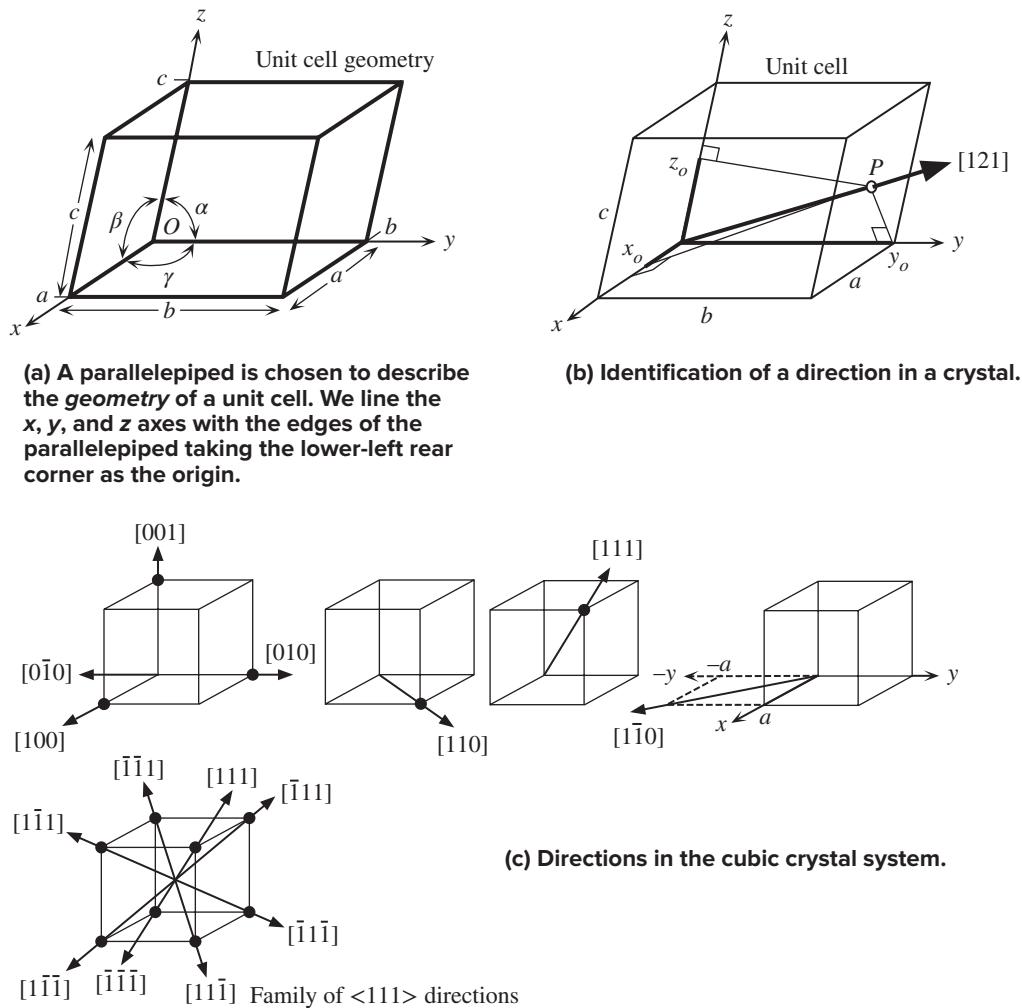


Figure 1.41

at the lower-left rear corner of the cell. The unit cell extends along the x axis from 0 to a , along y from 0 to b , and along z from 0 to c .

For Cu and Fe, the unit-cell geometry has $a = b = c$, $\alpha = \beta = \gamma = 90^\circ$, and cubic symmetry. For Zn, the unit cell has hexagonal geometry, with $a = b \neq c$, $\alpha = \beta = 90^\circ$, and $\gamma = 120^\circ$, as shown in Figure 1.34d.

In explaining crystal properties, we must frequently specify a direction in a crystal, or a particular plane of atoms. Many properties, for example, the elastic modulus, electrical resistivity, magnetic susceptibility, etc., are directional within the crystal. We use the convention described here for labeling crystal directions based on three-dimensional geometry.

All parallel vectors have the same indices. Therefore, the direction to be labeled can be moved to pass through the origin of the unit cell. As an example, Figure 1.41b

shows a direction whose indices are to be determined. A point P on the vector can be expressed by the coordinates x_o , y_o , z_o where x_o , y_o , and z_o are projections from point P onto the x , y , and z axes, respectively, as shown in Figure 1.41b. It is generally convenient to place P where the line cuts a surface (though this is not necessary). We can express these coordinates in terms of the lattice parameters a , b , and c , respectively. We then have three coordinates, say x_1 , y_1 , and z_1 , for point P in terms of a , b , and c . For example, if

$$x_o, y_o, z_o \quad \text{are} \quad \frac{1}{2}a, b, \frac{1}{2}c$$

then P is at

$$x_1, y_1, z_1 \quad i.e., \quad \frac{1}{2}, 1, \frac{1}{2}$$

We then multiply or divide these numbers until we have the smallest integers (which may include 0). If we call these integers u , v , and w , then the direction is written in square brackets without commas as $[uvw]$. If any integer is a negative number, we use a bar on top of that integer. For the particular direction in Figure 1.41b, we therefore have $[121]$.

Some of the important directions in a cubic lattice are shown in Figure 1.41c. For example, the x , y , and z directions in the cube are $[100]$, $[010]$, and $[001]$, as shown. Reversing a direction simply changes the sign of each index. The negative x , y , and z directions are $[\bar{1}00]$, $[0\bar{1}0]$, and $[00\bar{1}]$, respectively.

Certain directions in the crystal are equivalent because the differences between them are based only on our arbitrary decision for labeling x , y , and z directions. For example, $[100]$ and $[010]$ are different simply because of the way in which we labeled the x and y axes. Indeed, directional properties of a material (*e.g.*, elastic modulus, and dielectric susceptibility) along the edge of the cube $[100]$ are invariably the same as along the other edges, for example, $[010]$ and $[001]$. All of these directions along the edges of the cube constitute a **family of directions**, which is any set of directions considered to be equivalent. We label a family of directions, for example, $[100]$, $[010]$, $[001]$, . . . , by using a common notation, triangular brackets. Thus, $\langle 100 \rangle$ represents the family of six directions, $[100]$, $[010]$, $[001]$, $[\bar{1}00]$, $[0\bar{1}0]$, and $[00\bar{1}]$ in a cubic crystal. Similarly, the family of diagonal directions in the cube, shown in Figure 1.41c, is denoted $\langle 111 \rangle$.

We also frequently need to describe a particular plane in a crystal. Figure 1.42 shows a general unit cell with a plane to be labeled. We use the following convention, called the **Miller indices of a plane**, for this purpose.

We take the intercepts x_o , y_o , and z_o of the plane on the x , y , and z axes, respectively. If the plane passes through the origin, we can use another convenient parallel plane, or simply shift the origin to another point. All planes that have been shifted by a lattice parameter have identical Miller indices.

We express the intercepts x_o , y_o , and z_o in terms of the lattice parameters a , b , and c , respectively, to obtain x_1 , y_1 , and z_1 . We then invert these numbers. Taking the reciprocals, we obtain

$$\frac{1}{x_1}, \frac{1}{y_1}, \frac{1}{z_1}$$

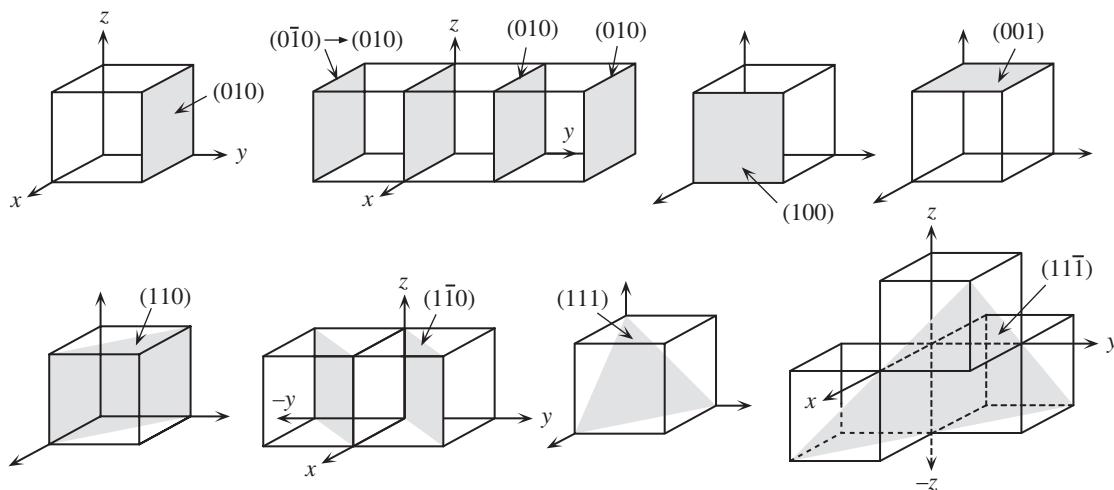
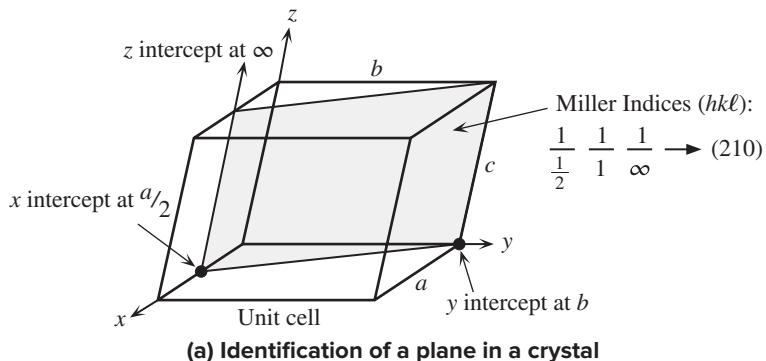


Figure 1.42 Labeling of crystal planes and typical examples in the cubic lattice.

We then clear all fractions, without reducing to lowest integers, to obtain a set of integers, say h , k , and ℓ . We then put these integers into parentheses, without commas, that is, (hkl) . For the plane in Figure 1.42a, we have

Intercepts x_o , y_o , and z_o are $\frac{1}{2}a$, $1b$, and ∞c .

Intercepts x_1 , y_1 , and z_1 , in terms of a , b , and c , are $\frac{1}{2}$, 1 , and ∞ .

Reciprocals $1/x_1$, $1/y_1$, and $1/z_1$ are $1/\frac{1}{2}$, $1/1$, $1/\infty = 2$, 1 , 0 .

This set of numbers does not have fractions, so it is not necessary to clear fractions. Hence, the Miller indices (hkl) are (210) .

If there is a negative integer due to a negative intercept, a bar is placed across the top of the integer. Also, if parallel planes differ only by a shift that involves a multiple number of lattice parameters, then these planes may be assigned the same Miller indices. For example, the plane $(\bar{0}\bar{1}0)$ is the xz plane that cuts the y axis at $-b$.

If we shift the plane along y by two lattice parameters ($2b$), it will cut the y axis at b and the Miller indices will become (010). In terms of the unit cell, the (010) plane is the same as the (010) plane, as shown in Figure 1.42b. Note that not all parallel planes are identical. Planes can have the same Miller indices *only* if they are separated by a multiple of the lattice parameter. For example, the (010) plane is not identical to the (020) plane, even though they are geometrically parallel. In terms of the unit cell, plane (010) is a face of the unit cell cutting the y axis at b , whereas (020) is a plane that is halfway inside the unit cell, cutting the y axis at $\frac{1}{2}b$. The planes contain different numbers of atoms. The (020) plane cannot be shifted by the lattice parameter b to coincide with plane (010).

It is apparent from Figure 1.42b that in the case of the cubic crystal, the $[hkl]$ direction is always perpendicular to the (hkl) plane.

Certain planes in the crystal belong to a **family of planes** because their indices differ only as a consequence of the arbitrary choice of axis labels. For example, the indices of the (100) plane become (010) if we switch the x and y axes. All the (100), (010), and (001) planes, and hence the parallel $(\bar{1}00)$, $(0\bar{1}0)$, $(00\bar{1})$ planes, form a family of planes, conveniently denoted by curly brackets as $\{100\}$.

Frequently we need to know the number of atoms per unit area on a given plane (hkl) . For example, if the surface concentration of atoms is high on one plane, then that plane may encourage oxide growth more rapidly than another plane where there are less atoms per unit area. **Planar concentration of atoms** is the number of atoms per unit area, that is, the surface concentration of atoms, on a given plane in the crystal. Among the $\{100\}$, $\{110\}$, and $\{111\}$ planes in FCC crystals, the most densely packed planes, those with the highest planar concentration, are $\{111\}$ planes and the least densely packed are $\{110\}$.

MILLER INDICES AND PLANAR CONCENTRATION Consider the plane shown in Figure 1.43a, which passes through one side of a face and the center of an opposite face in the FCC lattice. The plane passes through the origin at the lower-left rear corner. We therefore shift the origin to say point O' at the lower-right rear corner of the unit cell. In terms of a , the plane cuts the x , y , and z axes at ∞ , -1 , $\frac{1}{2}$, respectively. We take the reciprocals to obtain, 0 , -1 , 2 . Therefore, the Miller indices are (012).

EXAMPLE 1.17

To calculate the planar concentration $n_{(hkl)}$ on a given (hkl) plane, we consider a bound area A of the (hkl) plane within the unit cell as in Figure 1.43b. Only atoms whose centers lie on A are involved in $n_{(hkl)}$. For each atom, we then evaluate what portion of the atomic cross section (a circle in two dimensions) cut by the plane (hkl) is contained within A . Consider the Cu FCC crystal with $a = 0.3620 \text{ nm}$.

The (100) plane corresponds to a cube face and has an area $A = a^2$. There is one full atom at the center; that is, the (100) plane cuts through one full atom, one full circle in two dimensions, at the face center as in Figure 1.43b. However, not all corner atoms are within A . Only a quarter of a circle is within the bound area A in Figure 1.43b.

$$\text{Number of atoms in } A = (4 \text{ corners}) \times (\frac{1}{4} \text{ atom}) + 1 \text{ atom at face center} = 2$$

Planar concentration $n_{(100)}$ of (100) is

$$n_{(100)} = \frac{4(\frac{1}{4}) + 1}{a^2} = \frac{2}{a^2} = \frac{2}{(0.3620 \times 10^{-9} \text{ m})^2} = 15.3 \text{ atoms nm}^{-2}$$

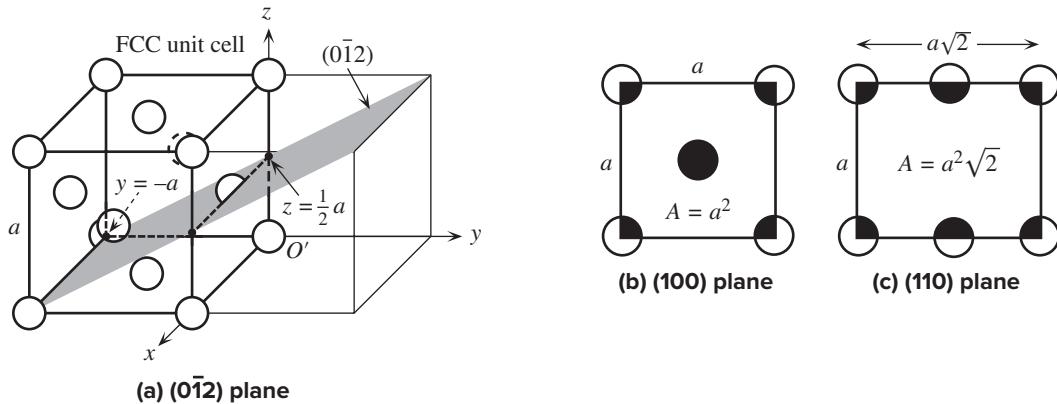


Figure 1.43 The $(0\bar{1}2)$ plane and planar concentrations in an FCC crystal.

Consider the (110) plane as in Figure 1.43c. The number of atoms in the area $A = (a)(a\sqrt{2})$ defined by two face diagonals and two cube sides is

$$(4 \text{ corners}) \times (\frac{1}{4} \text{ atom}) + (2 \text{ face diagonals}) \times (\frac{1}{2} \text{ atom at diagonal center}) = 2$$

Planar concentration on (110) is

$$n_{(110)} = \frac{4(\frac{1}{4}) + 2(\frac{1}{2})}{(a)(a\sqrt{2})} = \frac{2}{a^2\sqrt{2}} = 10.8 \text{ atoms nm}^{-2}$$

Similarly, for the (111) plane, $n_{(111)}$ is $17.0 \text{ atoms nm}^{-2}$. Clearly the (111) planes are the most and (110) planes are the least densely packed among the (100) , (110) , and (111) planes.

We can estimate the surface concentration n_{surf} of atoms from the bulk concentration n_{bulk} . The quantity $1/n_{\text{bulk}}^{1/3}$ represents the separation of the atoms d . Taking each atom to be a cube then d is the side of this cube. An atom on the surface occupies an area d^2 and therefore the surface concentration is $1/d^2$, or $n_{\text{surf}} \approx n_{\text{bulk}}^{2/3}$. Using $n_{\text{bulk}} = 8.43 \times 10^{28} \text{ m}^{-3}$ for copper from Example 1.16, $n_{\text{surf}} \approx n_{\text{bulk}}^{2/3} = (8.43 \times 10^{28} \text{ m}^{-3})^{2/3} = 1.92 \times 10^{19} \text{ m}^{-2}$ which is 19 atoms nm^{-2} . This is roughly the same order of magnitude as planar concentrations above and not too far out from $n_{(111)}$. (Question 1.4 explores this further.)

ADDITIONAL TOPICS

1.14 BRAVAIS LATTICES

An infinite periodic array of geometric points in space defines a **space lattice** or simply a **lattice**. Strictly, a lattice does not contain any atoms or molecules because it is simply an imaginary array of geometric points. A two-dimensional *simple square* lattice is shown in Figure 1.73a. In three dimensions, Figure 1.73a would correspond to the simple cubic (SC) lattice. The actual crystal is obtained from the lattice by placing an identical group of atoms (or molecules) at each lattice point. The identical group of atoms is called the **basis** of the crystal structure. Thus, conceptually,

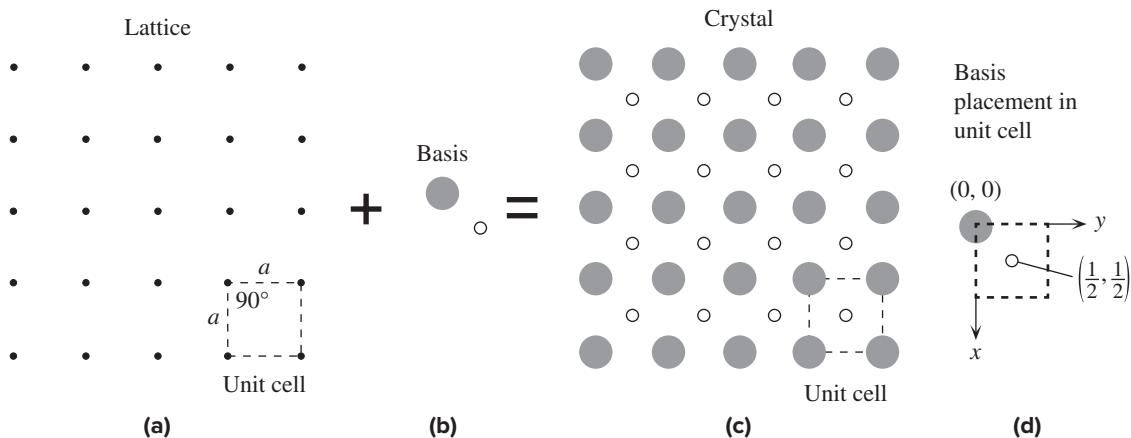


Figure 1.73 (a) A simple square lattice. The unit cell is a square with a side a . (b) Basis has two atoms. (c) Crystal = Lattice + Basis. The unit cell is a simple square with two atoms. (d) Placement of basis atoms in the crystal unit cell.

as illustrated in Figure 1.73a to c,

$$\text{Crystal} = \text{Lattice} + \text{Basis}$$

The unit cell of the two-dimensional lattice in Figure 1.73a is a square which is characterized by the length a of one of the sides; a is called a **lattice parameter**. A given lattice can generate different *patterns* of atoms depending on the basis. The lattice in Figure 1.73a with the two-atom basis in Figure 1.73b produces the crystal in Figure 1.73c. Although the latter crystal appears as a body-centered square (similar to BCC in three dimensions), it is nonetheless a *simple square lattice* with two atoms comprising the basis. Suppose that the basis had only one atom; then the crystal would appear as the simple square lattice in Figure 1.73a (with each point now being an atom). The *patterns* in Figure 1.73a and c are different but the underlying lattice is the same. Because they have the same lattice, the two crystals would have certain identical symmetries. For example, for both crystals, a rotation by 90° about a lattice point would produce the same crystal structure.

To fully characterize the crystal, we also have to specify the locations of the basis atoms in the unit cell as in Figure 1.73d. By convention, we place a Cartesian coordinate system at the rear-left corner of the unit cell with the x and y axes along the square edges. We indicate the coordinates (x_i, y_i) of each i th atom in terms of the lattice parameters along x and y . Thus, the atoms in the unit cell in Figure 1.73d are at $(0, 0)$ and at $(\frac{1}{2}, \frac{1}{2})$. The CsCl unit cell in Figure 1.39 appears as BCC, but it can be described by a SC lattice and a basis that has one Cl^- ion and one Cs^+ ion. The ions in the SC unit cell are located at $(0, 0, 0)$ and at the cell center at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Similarly, the NaCl crystal in Figure 1.38 is an FCC lattice with a basis of Na^+ and Cl^- ions.

The diamond unit cell of silicon is an FCC lattice with two Si atoms constituting the basis. The two Si atoms are placed at $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Most of the important III-V compound semiconductors such as GaAs, AlAs, InAs, InP, etc., which are widely used in numerous optoelectronic devices, have the zinc blende

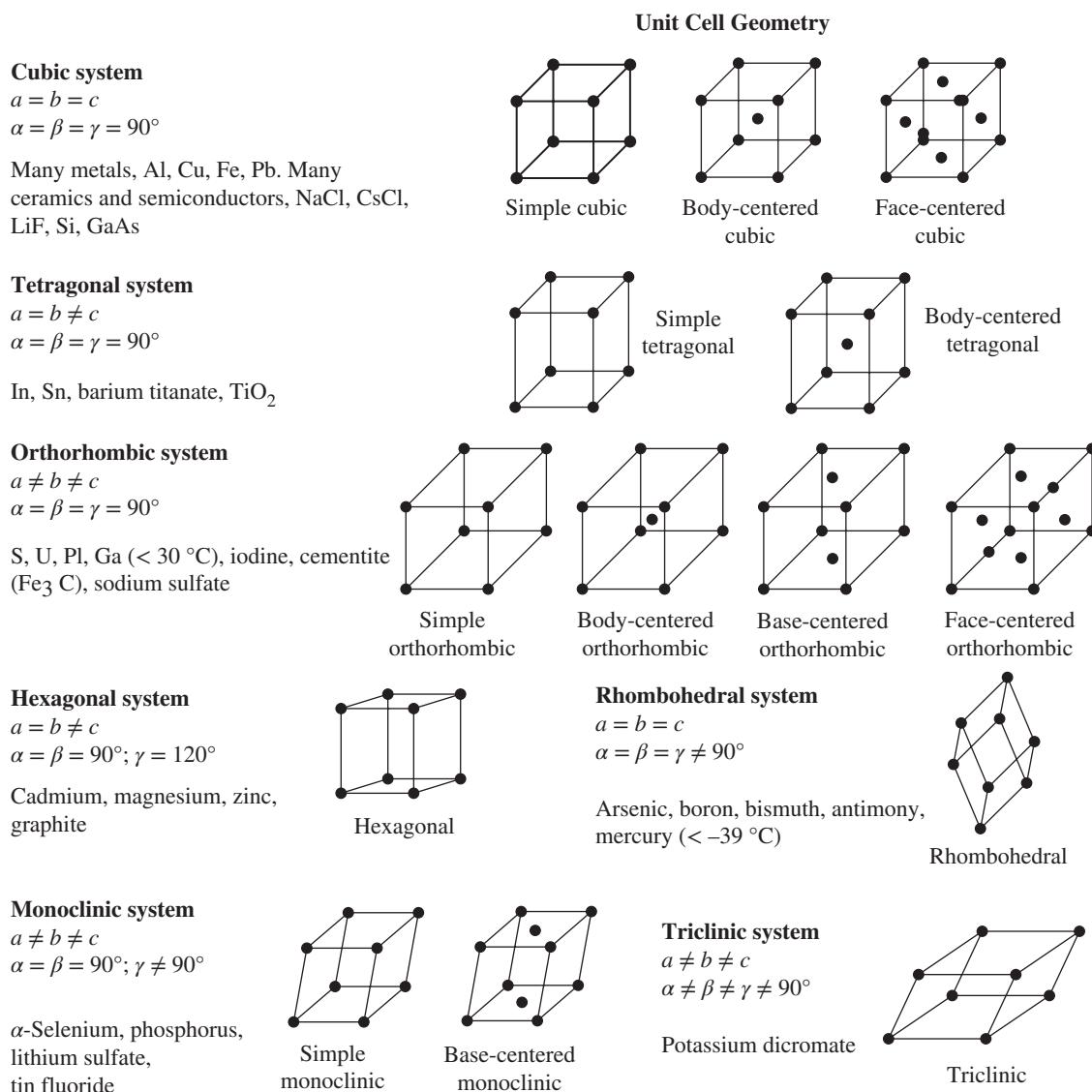


Figure 1.74 The seven crystal systems (unit-cell geometries) and fourteen Bravais lattices.

(ZnS) unit cell. The zinc blende unit cell consists of an FCC lattice and a basis that has the Zn and S atoms placed at $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, respectively.

We generally represent the *geometry* of the unit cell of a lattice as a parallelepiped with sides a, b, c and angles α, β, γ as depicted in Figure 1.41a. In the case of copper and iron, the geometry of the unit cell has $a = b = c$, $\alpha = \beta = \gamma = 90^\circ$, and cubic symmetry. For Zn, the unit cell has hexagonal geometry with $a = b \neq c$, $\alpha = \beta = 90^\circ$, and $\gamma = 120^\circ$ as shown in Figure 1.34d. Based on different lattice parameters, there are *seven* possible distinct unit-cell geometries, which we call

crystal systems each with a particular distinct symmetry. The seven crystal systems are depicted in Figure 1.74 with typical examples. We are already familiar with the cubic and hexagonal systems. The seven crystal systems only categorize the unit cells based on the geometry of the unit cell and not in terms of the symmetry and periodicity of the lattice points. (One should not confuse the unit-cell geometry with the lattice, which is a periodic array of points.) In the cubic system, for example, there are three possible distinct lattices corresponding to SC, BCC, and FCC which are shown in Figure 1.74. All three have the same cubic geometry: $a = b = c$ and $\alpha = \beta = \gamma = 90^\circ$.

Many distinctly different lattices, or distinct patterns of points, exist in three dimensions. There are 14 distinct lattices whose unit cells have one of the seven geometries as indicated in Figure 1.74. Each of these is called a **Bravais lattice**. The copper crystal, for example, has the FCC Bravais lattice, but arsenic, antimony, and bismuth crystals have the rhombohedral Bravais lattice. Tin's unit cell belongs to the **tetragonal** crystal system, and its crystal lattice is a **body-centered tetragonal (BCT)**.

DEFINING TERMS

Activated state is the state that occurs temporarily during a transformation or reaction when the reactant atoms or molecules come together to form a particular arrangement (intermediate between reactants and products) that has a higher potential energy than the reactants. The potential energy barrier between the activated state and the reactants is the activation energy.

Activation energy is the potential energy barrier against the formation of a product. In other words, it is the minimum energy that the reactant atom or molecule must have to be able to reach the activated state and hence form a product.

Amorphous solid is a solid that exhibits no crystalline structure or long-range order. It only possesses a short-range order in the sense that the nearest neighbors of an atom are well defined by virtue of chemical bonding requirements.

Anion is an atom that has gained negative charge by virtue of accepting one or more electrons. Usually, atoms of nonmetallic elements can gain electrons easily to become anions. Anions become attracted to the anode (positive terminal) in ionic conduction. Typical anions are the halogen ions F^- , Cl^- , Br^- , and I^- .

Atomic mass (or relative atomic mass or atomic weight) M_{at} of an element is the average atomic

mass, in atomic mass units (amu), of all the naturally occurring isotopes of the element. Atomic masses are listed in the Periodic Table. The amount of an element that has 6.022×10^{23} atoms (the Avogadro number of atoms) has a mass in grams equal to the atomic mass.

Atomic mass unit (amu) is a convenient mass measurement equal to one-twelfth of the mass of a neutral carbon atom that has a mass number of $A = 12$ (6 protons and 6 neutrons). It has been found that $amu = 1.66054 \times 10^{-27}$ kg, which is equivalent to $10^{-3}/N_A$, where N_A is Avogadro's number.

Atomic packing factor (APF) is the fraction of volume actually occupied by atoms in a crystal.

Avogadro's number (N_A) is the number of atoms in exactly 12 g of carbon-12. It is 6.022×10^{23} . Since atomic mass is defined as one-twelfth of the mass of the carbon-12 atom, the N_A number of atoms of any substance has a mass equal to the atomic mass M_{at} , in grams.

Basis represents an atom, a molecule, or a collection of atoms, that is placed at each lattice point to generate the true crystal structure of a substance. All crystals are thought of as a lattice with each point occupied by a basis.

Bond energy or **binding energy** is the work (or energy) needed to separate two atoms infinitely from their equilibrium separation in the molecule or solid.

Bulk modulus K is volume stress (pressure) needed per unit elastic volume strain and is defined by $p = -K\Delta$, where p is the applied volume stress (pressure) and Δ is the volume strain. K indicates the extent to which a body can be reversibly (and hence elastically) deformed in volume by an applied pressure.

Cation is an atom that has gained positive charge by virtue of losing one or more electrons. Usually, metal atoms can lose electrons easily to become cations. Cations become attracted to the cathode (negative terminal) in ionic conduction, as in gaseous discharge. The alkali metals, Li, Na, K, . . . , easily lose their valence electron to become cations, Li^+ , Na^+ , K^+ , . . .

Coordination number is the number of nearest neighbors around a given atom in the crystal.

Covalent bond is the sharing of a pair of valence electrons between two atoms. For example, in H_2 , the two hydrogen atoms share their electrons, so that each has a closed shell.

Crystal is a three-dimensional periodic arrangement of atoms, molecules, or ions. A characteristic property of the crystal structure is its periodicity and a degree of symmetry. For each atom, the number of neighbors and their exact orientations are well defined; otherwise the periodicity will be lost. Therefore, a long-range order results from strict adherence to a well-defined bond length and relative bond angle (that is, exact orientation of neighbors).

Crystallization is a process by which crystals of a substance are formed from another phase of that substance. Examples are solidification just below the fusion temperature from the melt, or condensation of the molecules from the vapor phase onto a substrate. The crystallization process initially requires the formation of small crystal nuclei, which contain a limited number (perhaps 10^3 – 10^4) of atoms or molecules of the substance. Following nucleation, the nuclei grow by atomic diffusion from the melt or vapor.

Diffusion is the migration of atoms by virtue of their random thermal motions.

Diffusion coefficient is a measure of the rate at which atoms diffuse. The rate depends on the nature of the diffusion process and is typically temperature dependent. The diffusion coefficient is defined as the magnitude of diffusion flux density per unit concentration gradient.

Dislocation is a line imperfection within a crystal that extends over many atomic distances.

Edge dislocation is a line imperfection within a crystal that occurs when an additional, short plane of atoms does not extend as far as its neighbors. The edge of this short plane constitutes a line of atoms where the bonding is irregular, that is, a line of imperfection called an edge dislocation.

Elastic modulus or **Young's modulus** (Y) is a measure of the ease with which a solid can be elastically deformed. The greater Y is, the more difficult it is to deform the solid elastically. When a solid of length ℓ is subjected to a tensile stress σ (force per unit area), the solid will extend elastically by an amount $\delta\ell$ where $\delta\ell/\ell$ is the strain ϵ . Stress and strain are related by $\sigma = Ye$, so Y is the stress needed per unit elastic strain.

Electric dipole moment is formed when a positive charge $+Q$ is separated from a negative charge $-Q$ of equal magnitude. Even though the net charge is zero, there is nonetheless an electric dipole moment formed by the two charges $-Q$ and $+Q$ being separated by a finite distance. Just as two charges exert a Coulombic force on each other, two dipoles also exert an electrostatic force on each other that depends on the separation of dipoles and their relative orientation.

Electron affinity represents the energy that is needed to add an electron to a neutral atom to create a negative ion (*anion*). When an electron is added to Cl to form Cl^- , energy is actually released.

Electronegativity is a relative measure of the ability of an atom to attract the electrons in a bond it forms with another atom. The *Pauling scale of electronegativity* assigns an electronegativity value (a pure number) X to various elements, the highest being 4 for F, and the lowest values being for the alkali metal atoms, for which X are less than 1. The difference $X_A - X_B$ in the electronegativities of two atoms A and B is a measure of the polar or ionic character of the bond $A-B$ between A and B . A molecule $A-B$ would be polar, that is, possess a dipole moment, if X_A and X_B are different.

Equilibrium between two systems requires mechanical, thermal, and chemical equilibrium. Mechanical equilibrium means that the pressure should be the same in the two systems, so that one does not expand at the expense of the other. Thermal equilibrium implies that both have the same temperature. Equilibrium within a single-phase substance (*e.g.*, steam only or hydrogen gas only) implies uniform pressure and temperature within the system.

Equilibrium state of a system is the state in which the pressure and temperature in the system are uniform throughout. We say that the system possesses mechanical and thermal equilibrium.

Eutectic composition is an alloy composition of two elements that results in the lowest melting temperature compared to any other composition. A eutectic solid has a structure that is a mixture of two phases. The eutectic structure is usually special, such as alternating lamellae.

Face-centered cubic (FCC) lattice is a cubic lattice that has one lattice point at each corner of a cube and one at the center of each face. If there is a chemical species (atom or a molecule) at each lattice point, then the structure is an FCC crystal structure.

Frenkel defect is an ionic crystal imperfection that occurs when an ion moves into an interstitial site, thereby creating a vacancy in its original site. The imperfection is therefore a pair of point defects.

Flux density is the rate of flow of particles in a particular direction per unit area.

Grain is an individual crystal within a polycrystalline material. Within a grain, the crystal structure and orientation are the same everywhere and the crystal is oriented in one direction only.

Grain boundary is a surface region between differently oriented, adjacent grain crystals. The grain boundary contains a lattice mismatch between adjacent grains.

Heat is the amount of energy transferred from one system to another (or between the system and its surroundings) as a result of a temperature difference. Heat is not a new form of energy, but rather the transfer of energy from one body to another by virtue of the random motions of their molecules. When a hot body is in

contact with a cold body, energy is transferred from the hot body to the cold one. The energy that is transferred is the excess mean kinetic energy of the molecules in the hot body. Molecules in the hot body have a higher mean kinetic energy and vibrate more violently. As a result of the collisions between the molecules, there is a net transfer of energy (heat) from the hot body to the cold one, until the molecules in both bodies have the same mean kinetic energy, that is, until their temperatures become equal.

Heat capacity at constant volume is the increase in the total energy E of the system per degree increase in the temperature of the system with the volume remaining constant: $C = (\partial E / \partial T)_V$. Thus, the heat added to the system does no mechanical work due to a volume change but increases the internal energy. **Molar heat capacity** is the heat capacity for 1 mole of a substance. **Specific heat capacity** is the heat capacity per unit mass.

Interstitial site (interstice) is an unoccupied space between the atoms (or ions, or molecules) in a crystal.

Ionization energy is the energy required to remove an electron from a neutral atom; normally the most outer electron that has the least binding energy to the nucleus is removed to ionize an atom.

Isomorphous describes a structure that is the same everywhere (*from iso*, uniform, and *morphology* structure).

Isotropic substance is a material that has the same property in all directions.

Kinetic molecular theory assumes that the atoms and molecules of all substances (gases, liquids, and solids) above absolute zero of temperature are in constant motion. Monatomic molecules (*e.g.*, He, Ne) in a gas exhibit constant and random translational motion, whereas the atoms in a solid exhibit constant vibrational motion.

Lattice is a regular array of points in space with a discernible periodicity. There are 14 distinct lattices possible in three-dimensional space. When an atom or molecule is placed at each lattice point, the resulting regular structure is a crystal structure.

Lattice parameters are (a) the lengths of the sides of the unit cell, and (b) the angles between the sides.

Mean free path is the mean distance a molecule in a gas travels freely before it collides with another molecule. The mean free path depends on the concentration of molecules, which depends on the pressure and temperature.

Mechanical work is qualitatively defined as the energy expended in displacing a constant force through a distance. When a force \mathbf{F} is moved a distance $d\mathbf{x}$, work done $dW = \mathbf{F} \cdot d\mathbf{x}$. When we lift a body such as an apple of mass m (100 g) by a distance h (1 m), we do work by an amount $F \Delta x = mgh$ (1 J), which is then stored as the gravitational potential energy of the body. We have transferred energy from ourselves to the potential energy of the body by exchanging energy with it in the form of work. Further, in lifting the apple, the molecules have been displaced in orderly fashion, all upwards. Work therefore involves an orderly displacement of atoms and molecules of a substance in complete contrast to heat. When the volume V of a substance changes by dV when the pressure is P , the mechanical work involved is $P dV$ and is called the **PV work**.

Metallic bonding is the binding of metal atoms in a crystal through the attraction between the positive metal ions and the mobile valence electrons in the crystal. The valence electrons permeate the space between the ions.

Miller indices (hkl) are indices that conveniently identify parallel planes in a crystal. Consider a plane with the intercepts, x_1 , y_1 , and z_1 , in terms of lattice parameters a , b , and c . (For a plane passing through the origin, we shift the origin or use a parallel plane.) Then, (hkl) are obtained by taking the reciprocals of x_1 , y_1 , and z_1 , and clearing all fractions.

Miscibility of two substances is a measure of the mutual solubility of those two substances when they are in the same phase, such as liquid.

Mole of a substance is that amount of the substance that contains N_A number of atoms (or molecules), where N_A is Avogadro's number (6.023×10^{23}). One mole of a substance has a mass equal to its atomic (molecular) mass, in grams. For example, 1 mole of copper contains 6.023×10^{23} atoms and has a mass of 63.55 g.

Phase of a system is a homogeneous portion of the chemical system that has the same composition,

structure, and properties everywhere. In a given chemical system, one phase may be in contact with another phase of the system. For example, iced water at 0 °C will have solid and liquid phases in contact. Each phase, solid ice and liquid water, has a distinct structure.

Phase diagram is a temperature versus composition diagram in which the existence and coexistence of various phases are identified by regions and lines. Between the liquidus and solidus lines, for example, the material is a heterogeneous mixture of the liquid and solid phases.

Physical vapor deposition (PVD) involves the heating and evaporation of a source material in a vacuum chamber so that the vapor can be condensed onto a substrate of choice, placed facing the source. The result is a thin film of the source material on the substrate.

Planar concentration of atoms is the number of atoms per unit area on a given (hkl) plane in a crystal.

Polarization is the separation of positive and negative charges in a system, which results in a net electric dipole moment.

Polymorphism or **allotropy** is a material attribute that allows the material to possess more than one crystal structure. Each possible crystal structure is called a polymorph. Generally, the structure of the polymorph depends on the temperature and pressure, as well as on the method of preparation of the solid. (For example, diamond can be prepared from graphite by the application of very high pressures.)

Primary bond is a strong interatomic bond, typically greater than 1 eV/atom, that involves ionic, covalent, or metallic bonding.

Property is a system characteristic or an attribute that we can measure. Pressure, volume, temperature, mass, energy, electrical resistivity, magnetization, polarization, and color are all properties of matter. Properties such as pressure, volume, and temperature can only be attributed to a system of many particles (which we treat as a continuum). Note that heat and work are not properties of a substance; instead, they represent energy transfers involved in producing changes in the properties.

Saturated solution is a solution that has the maximum possible amount of solute dissolved in a given amount of solvent at a specified temperature and pressure.

Schottky defect is an ionic crystal imperfection that occurs when a pair of ions is missing, that is, when there is a cation and anion pair vacancy.

Screw dislocation is a crystal defect that occurs when one portion of a perfect crystal is twisted or skewed with respect to another portion on only one side of a line.

Secondary bond is a weak bond, typically less than 0.1 eV/atom, which is due to dipole–dipole interactions between the atoms or molecules.

Solid solution is a homogeneous crystalline phase that contains two or more chemical components.

Solute is the minor chemical component of a solution; the component that is usually added in small amounts to a solvent to form a solution.

Solvent is the major chemical component of a solution.

Stoichiometric compounds are compounds with an integer ratio of atoms, as in CaF₂, in which two fluorine atoms bond with one calcium atom.

Strain is a relative measure of the deformation a material exhibits under an applied stress. Under an applied tensile (or compressive) stress, strain ϵ is the change in the length per unit original length. When a shear stress is applied, the deformation involves a shear angle. **Shear strain** is the tangent of the shear angle

that is developed by the application of the shearing stress. **Volume strain** Δ is the change in the volume per unit original volume.

Stress is force per unit area. When the applied force F is perpendicular to the area A , stress $\sigma = F/A$ is either tensile or compressive. If the applied force is tangential to the area, then stress is **shear stress**, $\tau = F/A$.

Thermal expansion is the change in the length or volume of a substance due to a change in the temperature.

Linear coefficient of thermal expansion λ is the fractional change in the length per unit temperature change or $\Delta L/L_0 = \lambda \Delta T$. Volume coefficient of expansion α_V is the fractional change in the volume per unit temperature change; $\alpha_V \approx 3\lambda$.

Unit cell is the most convenient small cell in a crystal structure that carries the characteristics of the crystal. The repetition of the unit cell in three dimensions generates the whole crystal structure.

Vacancy is a point defect in a crystal, where a normally occupied lattice site is missing an atom.

Valence electrons are the electrons in the outer shell of an atom. Since they are the farthest away from the nucleus, they are the first electrons involved in atom-to-atom interactions.

Young's modulus see **elastic modulus**.

QUESTIONS AND PROBLEMS

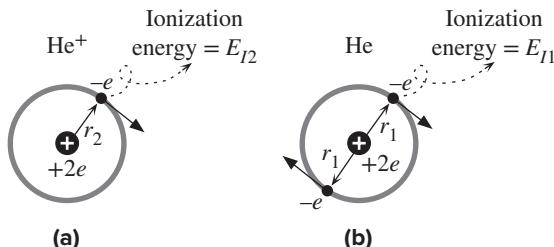
1.1 Virial theorem The Li atom has a nucleus with a $+3e$ positive charge, which is surrounded by a full 1s shell with two electrons, and a single valence electron in the outer 2s subshell. The atomic radius of the Li atom is about 0.17 nm. Using the Virial theorem, and assuming that the valence electron sees the nuclear $+3e$ shielded by the two 1s electrons, that is, a net charge of $+e$, estimate the ionization energy of Li (the energy required to free the 2s electron). Compare this value with the experimental value of 5.39 eV. Suppose that the actual nuclear charge seen by the valence electron is not $+e$ but a little higher, say $+1.25e$, due to the imperfect shielding provided by the closed 1s shell. What would be the new ionization energy? What is your conclusion?

1.2 Virial theorem and the He atom In Example 1.1, we calculated the radius of the H-atom using the Virial theorem. First consider the He⁺ atom, which as shown in Figure 1.75a has one electron in the K-shell orbiting the nucleus. Take the PE and the KE as zero when the electrons and the nucleus are infinitely separated. The nucleus has a charge of $+2e$ and there is one electron orbiting the nucleus at a radius r_2 . Using the Virial theorem show that the energy of the He⁺ ion is

$$E(\text{He}^+) = -(1/2) \frac{2e^2}{4\pi\epsilon_0 r_2} \quad [1.48]$$

Energy of
He⁺ ion

Figure 1.75 (a) A classical view of a He^+ ion. There is one electron in the K-shell orbiting the nucleus that has a charge $+2e$. (b) The He atom. There are two electrons in the K-shell. Due to their mutual repulsion, they orbit to void each other.



Now consider the He-atom shown in Figure 1.75b. There are two electrons. Each electron interacts with the nucleus (at a distance r_1) and the other electron (at a distance $2r_1$). Using the Virial theorem show that the energy of the He atom is

Energy of
He atom

$$E(\text{He}) = -(1/2) \left[\frac{7e^2}{8\pi\epsilon_0 r_1} \right] \quad [1.49]$$

The first ionization energy E_{I1} is defined as the energy required to remove one electron from the He atom. The second ionization energy E_{I2} is the energy required to remove the second (last) electron from He^+ . Both are shown in Figure 1.75. These have been measured and given as $E_{I1} = 2372 \text{ kJ mol}^{-1}$ and $E_{I2} = 5250 \text{ kJ mol}^{-1}$. Find the radii r_1 and r_2 for He and He^+ . Note that the first ionization energy provides sufficient energy to take He to He^+ , that is, $\text{He} \rightarrow \text{He}^+ + e^-$ absorbs 2372 kJ mol^{-1} . How does your r_1 value compare with the often quoted He radius of 31 pm?

1.3 Atomic mass and molar fractions

- a. Consider a multicomponent alloy containing N elements. If w_1, w_2, \dots, w_N are the weight fractions of components 1, 2, ..., N in the alloy and M_1, M_2, \dots, M_N are the respective atomic masses of the elements, show that the atomic fraction of the i th component is given by

Weight to atomic
percentage

$$n_i = \frac{w_i/M_i}{w_1/M_1 + w_2/M_2 + \dots + w_N/M_N}$$

- b. Suppose that a substance (compound or an alloy) is composed of N elements, A, B, C, \dots and that we know their atomic (or molar) fractions n_A, n_B, n_C, \dots . Show that the weight fractions w_A, w_B, w_C, \dots are given by

Atomic to weight
percentage

$$w_A = \frac{n_A M_A}{n_A M_A + n_B M_B + n_C M_C + \dots}$$

$$w_B = \frac{n_B M_B}{n_A M_A + n_B M_B + n_C M_C + \dots}$$

- c. Consider the semiconducting II–VI compound cadmium selenide, CdSe . Given the atomic masses of Cd and Se, find the weight fractions of Cd and Se in the compound and grams of Cd and Se needed to make 100 grams of CdSe .
- d. A Se–Te–P glass alloy has the composition 77 wt.% Se, 20 wt.% Te, and 3 wt.% P. Given their atomic masses, what are the atomic fractions of these constituents?

1.4 Mean atomic separation, surface concentration, and density

There are many instances where we only wish to use reasonable estimates for the mean separation between the host atoms in a crystal and the mean separation between impurities in the crystal. These can be related in a simple way to the atomic concentration of the host atoms and atomic concentration of the impurity atoms, respectively. The final result does not depend on the sample geometry or volume. Sometimes, we need to know the number of atoms per unit area n_s on the surface of a solid given the number of atoms per unit volume in the bulk, n_b . Consider a crystal of the material of interest which is a cube of side L as shown in Figure 1.76. To each atom, we can attribute a portion of the whole volume, which is a

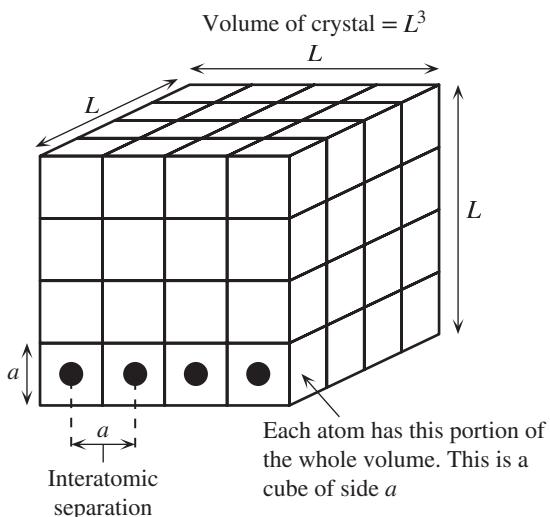


Figure 1.76 Consider a crystal that has volume L^3 . This volume is proportioned to each atom, which is a cube of side a^3 .

cube of side a . Thus, each atom is considered to occupy a volume of a^3 . Suppose that there are N atoms in the volume L^3 . Thus, $L^3 = Na^3$.

- If n_b is the bulk concentration of atoms, show that the mean separation a between the atoms is given by $a = 1/n_b^{1/3}$.
 - Show that the surface concentration n_s of atoms is given by $n_s = n_b^{2/3}$.
 - Show that the density of the solid is given by $\rho = n_b M_{\text{at}}/N_A$ where M_{at} is the atomic mass. Calculate the atomic concentration in Si from its density (2.33 g cm^{-3}).
 - A silicon crystal has been doped with phosphorus. The P concentration in the crystal is 10^{16} cm^{-3} . P atoms substitute for Si atoms and are randomly distributed in the crystal. What is the mean separation between the P atoms?
- 1.5 The covalent bond** Consider the H_2 molecule in a simple way as two touching H atoms, as depicted in Figure 1.77. Does this arrangement have a lower energy than two separated H atoms? Suppose that electrons totally correlate their motions so that they move to avoid each other as in the snapshot in Figure 1.77. The radius r_o of the hydrogen atom is 0.0529 nm . The electrostatic potential energy of two charges Q_1 and Q_2 separated by a distance r is given by $Q_1 Q_2 / (4\pi\epsilon_0 r)$. Using the virial theorem as in Example 1.1 consider the following:
- Calculate the total electrostatic potential energy PE of all the charges when they are arranged as shown in Figure 1.77. In evaluating the PE of the whole collection of charges you must consider all pairs of charges and, at the same time, avoid double counting of interactions between the same pair of charges. The total PE is the sum of the following: electron 1 interacting with the proton at a distance r_o on the left, proton at r_o on the right, and electron 2 at a distance $2r_o$ + electron 2 interacting with a proton at r_o and another proton at $3r_o$ + two protons, separated by $2r_o$, interacting with each other. Is this configuration energetically favorable?

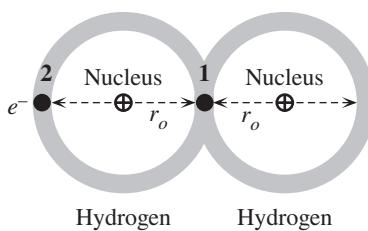


Figure 1.77 A simplified view of the covalent bond in H_2 . A snapshot at one instant.

- b. Given that in the isolated H atom the *PE* is $2 \times (-13.6 \text{ eV})$, calculate the change in *PE* in going from two isolated H atoms to the H_2 molecule. Using the virial theorem, find the change in the total energy and hence the covalent bond energy. How does this compare with the experimental value of 4.51 eV?

- 1.6 Ionic bonding and CsCl** The potential energy *E* per $\text{Cs}^+ - \text{Cl}^-$ pair within the CsCl crystal depends on the interionic separation *r* in the same fashion as in the NaCl crystal,

Energy per ion pair in ionic crystals

$$E(r) = -\frac{e^2 M}{4\pi\epsilon_0 r} + \frac{B}{r^m} \quad [1.50]$$

where for CsCl, $M = 1.763$, $B = 1.192 \times 10^{-104} \text{ J m}^9$ or $7.442 \times 10^{-5} \text{ eV nm}^9$, and $m = 9$. Find the equilibrium separation (r_o) of the ions in the crystal and the ionic bonding energy, that is, the ionic cohesive energy, and compare the latter value to the experimental value of 657 kJ mol^{-1} . Given that the *ionization energy* of Cs is 3.89 eV and the *electron affinity* of Cl (energy released when an electron is added) is 3.61 eV, calculate the atomic cohesive energy of the CsCl crystal as joules per mole.

- 1.7 Ionic bonding and LiCl** Equation 1.50 can be used to represent the *PE* of the ion pair inside the LiCl crystal. LiCl has the NaCl structure with $M = 1.748$, $m = 7.30$, $B = 2.34 \times 10^{-89} \text{ J m}^{7.30}$. Further, the ionization energy of Li ($\text{Li} \rightarrow \text{Li}^+ + e^-$) is $520.2 \text{ kJ mol}^{-1}$. The electron affinity of Cl (energy released in $\text{Cl} + e^- \rightarrow \text{Cl}^-$) is $348.7 \text{ kJ mol}^{-1}$ (a) Calculate the equilibrium separation of ions in the LiCl crystal. (b) Calculate the bonding energy per ion pair in the LiCl crystal. (c) Calculate the atomic cohesive energy of the LiCl crystal. (c) Calculate the density of LiCl.

- 1.8 Madelung constant** If we were to examine the NaCl crystal in three dimensions, we would find that each Na^+ ion has

- 6 Cl^- ions as *nearest* neighbors at a distance *r*,
- 12 Na^+ ions as *second* nearest neighbors at a distance $r\sqrt{2}$,
- 8 Cl^- ions as *third* nearest neighbors at a distance $r\sqrt{3}$,

and so on. Show that the electrostatic potential energy of the Na^+ atom can be written as

$$E(r) = -\frac{e^2}{4\pi\epsilon_0 r} \left[6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \dots \right] = -\frac{e^2 M}{4\pi\epsilon_0 r}$$

where *M*, called the **Madelung constant**, is given by the summation in the square brackets for this particular ionic crystal structure (NaCl). Calculate *M* for the first three terms and compare it with $M = 1.7476$, its value had we included the higher terms. What is your conclusion?

- *1.9 Bonding and bulk modulus** In general, the potential energy *E* per atom, or per ion pair, in a crystal as a function of interatomic (interionic) separation *r* can be written as the sum of an attractive *PE* and a repulsive *PE*,

General PE curve for bonding

$$E(r) = -\frac{A}{r^n} + \frac{B}{r^m} \quad [1.51]$$

where *A* and *n* are constants characterizing the attractive *PE* and *B* and *m* are constants characterizing the repulsive *PE*. This energy is minimum when the crystal is in equilibrium. The magnitude of the minimum energy and its location r_o define the bonding energy and the equilibrium interatomic (or interionic) separation, respectively.

When a pressure *P* is applied to a solid, its original volume V_o shrinks to *V* by an amount $\Delta V = V - V_o$. The bulk modulus *K* relates the volume strain $\Delta V/V$ to the applied pressure *P* by

Bulk modulus definition

$$P = -K \frac{\Delta V}{V_o} \quad [1.52]$$

The bulk modulus *K* is related to the energy curve. In its simplest form (assuming a simple cubic unit cell), *K* can be estimated from Equation 1.51 by

Bulk modulus

$$K = \frac{1}{9cr_o} \left[\frac{d^2 E}{dr^2} \right]_{r=r_o} \quad [1.53]$$

where c is a numerical factor, of the order of unity, given by b/p where p is the number of atoms or ion pairs in the unit cell and b is a numerical factor that relates the cubic unit cell lattice parameter a_o to the equilibrium interatomic (interionic) separation r_o by $b = a_o^3/r_o^3$.

- a. Show that the bond energy and equilibrium separation are given by

$$E_{\text{bond}} = \frac{A}{r_o^n} \left(1 - \frac{n}{m} \right) \quad \text{and} \quad r_o = \left(\frac{Bm}{An} \right)^{1/(m-n)}$$

- b. Show that the bulk modulus is given by

$$K = \frac{An}{9cr_o^{n+3}}(m-n) \quad \text{or} \quad K = \frac{mnE_{\text{bond}}}{9cr_o^3}$$

- c. For a NaCl-type crystal, Na^+ and Cl^- ions touch along the cube edge so that $r_o = (a_o/2)$. Thus, $a^3 = 2^3 r_o^3$ and $b = 2^3 = 8$. There are four ion pairs in the unit cell, $p = 4$. Thus, $c = b/p = 8/4 = 2$. Using the values from Example 1.3, calculate the bulk modulus of NaCl.

- *1.10 **Van der Waals bonding** Below 24.5 K, Ne is a crystalline solid with an FCC structure. The interatomic interaction energy per atom can be written as

$$E(r) = -2\varepsilon \left[14.45 \left(\frac{\sigma}{r} \right)^6 - 12.13 \left(\frac{\sigma}{r} \right)^{12} \right] \quad (\text{eV/atom})$$

where ε and σ are constants that depend on the polarizability, the mean dipole moment, and the extent of overlap of core electrons. For crystalline Ne, $\varepsilon = 3.121 \times 10^{-3}$ eV and $\sigma = 0.274$ nm.

- a. Show that the equilibrium separation between the atoms in an inert gas crystal is given by $r_o = (1.090)\sigma$. What is the equilibrium interatomic separation in the Ne crystal?
b. Find the bonding energy per atom in solid Ne.
c. Calculate the density of solid Ne (atomic mass = 20.18).

1.11 **Kinetic molecular theory**

- a. In a particular Ar-ion laser tube the gas pressure due to Ar atoms is about 0.1 torr at 25 °C when the laser is off. What is the concentration of Ar atoms per cm^3 at 25 °C in this laser? (760 torr = 1 atm = 1.013×10^5 Pa.)
b. In the He–Ne laser tube He and Ne gases are mixed and sealed. The total pressure P in the gas is given by contributions arising from He and Ne atoms:

$$P = P_{\text{He}} + P_{\text{Ne}}$$

where P_{He} and P_{Ne} are the *partial pressures* of He and Ne in the gas mixture, that is, pressures due to He and Ne gases alone,

$$P_{\text{He}} = \frac{N_{\text{He}}}{N_A} \left(\frac{RT}{V} \right) \quad \text{and} \quad P_{\text{Ne}} = \frac{N_{\text{Ne}}}{N_A} \left(\frac{RT}{V} \right)$$

In a particular He–Ne laser tube the ratio of He and Ne atoms is 7:1, and the total pressure is about 1 torr at 22 °C. Calculate the concentrations of He and Ne atoms in the gas at 22 °C. What is the pressure at an operating temperature of 130 °C?

- 1.12 **Kinetic molecular theory** Calculate the effective (rms) speeds of the He and Ne atoms in the He–Ne gas laser tube at room temperature (300 K).

- *1.13 **Kinetic molecular theory and the Ar-ion laser** An argon-ion laser has a laser tube that contains Ar atoms that produce the laser emission when properly excited by an electrical discharge. Suppose that the gas temperature inside the tube is 1300 °C (very hot).

- a. Calculate the mean speed (v_{av}), rms velocity ($v_{\text{rms}} = \sqrt{v^2}$), and the rms speed ($v_{\text{rms},x} = \sqrt{v_x^2}$) in one particular direction of the Ar atoms in the laser tube, assuming 1300 °C. (See Example 1.11.)
b. Consider a light source that is emitting waves and is moving toward an observer, somewhat like a whistling train moving toward a passenger. If f_o is the frequency of the light waves emitted at the source, then, due to the *Doppler effect*, the observer measures a higher frequency f

that depends on the velocity v_{Ar} of the source moving toward the observer and the speed c of light,

$$f = f_0 \left(1 + \frac{v_{\text{Ar}}}{c} \right)$$

It is the Ar ions that emit the laser output light in the Ar-ion laser. The emission wavelength $\lambda_o = c/f_0$ is 514.5 nm. Calculate the wavelength λ registered by an observer for those atoms that are moving with a mean speed v_{av} toward the observer. Those atoms that are moving away from the observer will result in a lower observed frequency because v_{Ar} will be negative. Estimate the range of all possible wavelengths (the difference between the longest and the shortest wavelengths) that can be emitted by the Ar-ion laser around 514.5 nm.

- 1.14 Heat capacity of gases** Table 1.9 shows the experimental values of the molar heat capacity for a few gases at 25 °C. Assume that we can neglect the vibrations of the atoms in the molecules. For each gas calculate the expected heat capacity from translational and rotational degrees of freedom only. Use the difference between the calculated values above and experimental values in Table 1.9 to find the contribution from molecular vibrations. What is your conclusion?

Table 1.9 Heat capacities for some gases at room temperature at constant volume, C_V in $\text{J mol}^{-1} \text{K}^{-1}$

Gas	Ar	Ne	Cl_2	O_2	N_2	CO_2	CH_4	SF_6
C_V	12.5	12.7	25.6	21.0	20.8	28.9	27.4	89.0

- *1.15 Degrees of freedom in a gas molecule** A monatomic molecule such as Ar has only three degrees of freedom (DOF) for motion along the three independent directions x , y , and z . In a system in which there are two independent atoms such Cl and Cl, the total number of DOF f is 6 because each atom has 3 degrees of freedom. Once we form a Cl_2 molecule, the original 6 DOF in KE are partitioned as shown in Figure 1.78. The Cl_2 molecule has 3 translational degrees of freedom, 2 rotational and 1 vibrational, summing to the original 6. The vibrational degree of freedom itself has KE and PE terms with each having an average of $\frac{1}{2}kT$ so that a vibrational degree of freedom actually has kT of energy rather than $\frac{1}{2}kT$. The PE term arises from the stretching and compression of the bond (which acts like a spring) during the vibrations. Put differently, each vibrational DOF has two “subdegrees” of freedom associated with KE and PE terms, each of which has an average of $(1/2)kT$ of energy. Let n_a be the number of atoms in a molecule. Then $3n_a$ is the total number of kinetic energy based DOF. There will always be 3 translational DOF for the molecule and at most 3 rotational degrees of freedom. There may be one or more vibrational DOF because there may be many ways in which the atoms in

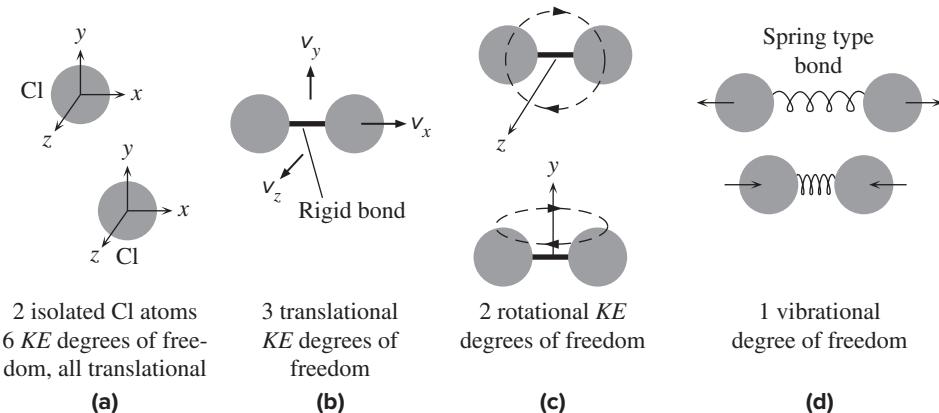


Figure 1.78 The partitioning of degrees of freedom in a diatomic molecule.

the molecule can vibrate, but there is a maximum. If f_{rot} and f_{vib} are the rotational and vibrational DOF, then

$$3n_a = 3 + f_{\text{rot}} + f_{\text{vib}}.$$

- What is the vibrational DOF for Cl_2 ? What is the maximum molar heat capacity at constant volume C_V for Cl_2 ? Given Table 1.9, what is the vibrational contribution?
- What is the vibrational DOF for SF_6 ? The molar heat capacity at constant volume for the SF_6 gas at 300 K is $89 \text{ J mol}^{-1} \text{ K}^{-1}$ but at 700 K, it is $141 \text{ J mol}^{-1} \text{ K}^{-1}$. How many vibrational DOF do you need to explain the observations at these two temperatures?

- 1.16 Dulong–Petit rule for metals** Consider the room temperature experimental specific heats of those 22 metals listed in Table 1.10. They are listed in increasing atomic mass from Li to Bi. Plot c_s versus $1/M_{\text{at}}$ and find the best line that goes through the origin. What is the slope of this best line? Now consider Be, which is a very light metal. It has $c_s = 1.825 \text{ J g}^{-1} \text{ K}^{-1}$, $M_{\text{at}} = 9.012 \text{ g mol}^{-1}$. What is its molar heat capacity? What is your conclusion? (To avoid points cluttering in one region of the plot, you can also try a log–log plot.)

Table 1.10 Specific heat capacity in $\text{J g}^{-1} \text{ K}^{-1}$ and atomic mass for selected metals at 25 °C

Metal	Li	Na	Mg	Al	K	Ca	Ti	V	Cr	Fe	Co
M_{at}	6.94	22.99	24.3	26.98	39.1	40.08	47.87	50.94	51.99	55.85	58.93
c_s	3.58	1.228	1.023	0.897	0.757	0.647	0.523	0.489	0.449	0.444	0.421
Metal	Cu	Zn	Zr	Mo	Ag	Sb	Ta	W	Au	Pb	Bi
M_{at}	58.93	65.39	91.22	95.94	107.86	121.76	180.95	183.84	196.97	207.2	208.99
c_s	0.385	0.388	0.278	0.251	0.235	0.207	0.14	0.132	0.129	0.13	0.122

1.17 Heat capacity

- Calculate the heat capacity per mole and per gram of N_2 gas, neglecting the vibrations of the molecule. How does this compare with the experimental value of $0.743 \text{ J g}^{-1} \text{ K}^{-1}$?
- Calculate the heat capacity per mole and per gram of CO_2 gas, neglecting the vibrations of the molecule. How does this compare with the experimental value of $0.648 \text{ J K}^{-1} \text{ g}^{-1}$? Assume that the CO_2 molecule is linear (O—C—O) so that it has two rotational degrees of freedom.
- Based on the Dulong–Petit rule, calculate the heat capacity per mole and per gram of solid silver. How does this compare with the experimental value of $0.235 \text{ J K}^{-1} \text{ g}^{-1}$?
- Based on the Dulong–Petit rule, calculate the heat capacity per mole and per gram of the silicon crystal. How does this compare with the experimental value of $0.71 \text{ J K}^{-1} \text{ g}^{-1}$?

- 1.18 Dulong–Petit atomic heat capacity** Express the Dulong–Petit rule for the molar heat capacity as heat capacity per atom and in the units of eV K^{-1} per atom, called the **atomic heat capacity**. CsI is an ionic crystal used in optical applications that require excellent infrared transmission at very long wavelengths (up to 55 μm). It has the CsCl crystal structure with one Cs^+ and one I^- ion in the unit cell. Calculate the specific heat capacity of CsI and compare it with the experimental value of $0.20 \text{ J K}^{-1} \text{ g}^{-1}$. What is your conclusion?

1.19 Dulong–Petit specific heat capacity of alloys and compounds

- Calculate the specific heat capacity of Pb–Sn solder assuming that its composition is 38 wt.% Pb and 62 wt.% Sn.
- Calculate the specific heat capacities of Pb and Sn individually as c_{sA} and c_{sB} , respectively, and then calculate the c_s for the alloy using

$$c_s = c_{sA}w_A + c_{sB}w_B$$

where w_A and w_B are the weight fractions of A (Pb) and B (Sn) in the alloy (solder). Compare your result with part (a). What is your conclusion?

Alloy specific
heat capacity

- c. ZnSe is an important optical material (used in infrared windows and lenses and high-power CO₂ laser optics) and also an important II–VI semiconductor that can be used to fabricate blue-green laser diodes. Calculate the specific heat capacity of ZnSe, and compare the calculation to the experimental value of 0.345 J K⁻¹ g⁻¹.
- 1.20 Molecular collisions** Consider the atmosphere as made up from 80% N₂ and 20% O₂ gases. At a pressure P , the N₂ and O₂ gases will have partial pressure of P_N and P_O respectively so that $P = P_N + P_O$. If n_N and n_O are the concentration of N₂ and O₂ molecules respectively then $P_N = n_N kT$, and $P_O = n_O kT$. Consider a vacuum chamber in which the total pressure is 10⁻⁵ torr. Assume 27 °C.
- Calculate the concentrations of N₂ and O₂ gases in the chamber.
 - Suppose that we simply consider the collisions of N₂ with N₂ and O₂ with O₂ and neglect N₂ and O₂ collisions. Calculate the mean free path for N₂ and O₂ molecules. See Table 1.11.
 - What are the mean free paths for each gas if the gas were in the container alone at 10⁻⁵ torr?
 - Obviously the calculation in *b* is not correct because we neglected collisions between N₂ and O₂. Suppose that we try to improve our calculations by using some average value for the collisional radius r by averaging that involves the relative concentrations of molecular species in the tank, that is,

$$r = \frac{r_1 n_1 + r_2 n_2}{n_1 + n_2}$$

where the subscript 1 refers to molecular species 1 (N₂) and 2 to species 2 (O₂) and we take $n = n_1 + n_2$ in the mean free path equation since all molecules are involved in the collisions. Calculate the mean free path using these parameters. What is your conclusion? (See also Question 1.11)

Table 1.11 Radii for molecular or atomic collisions in gases

Molecule or Atom	He	Ne	Ar	N ₂	O ₂	CO ₂
$r(\text{nm})$	0.100	0.117	0.143	0.158	0.148	0.230

| SOURCE: Moore, Walter J., *Physical Chemistry*, 5th Ed. London: Longman, 1971.

- 1.21 SF₆ insulating gas in HV switchgear** SF₆ (sulfur hexafluoride) is a gas that has excellent insulating properties and is widely used in high voltage electric power applications, such as gas insulated switchgear and circuit breakers up to megavolts. Six F atoms surround the S atom so that there are 6 bonds in total along $\pm x$ and $\pm y$ and $\pm z$ directions. The SF₆ gas in a particular sealed switchgear container has a pressure of 500 kPa (roughly 5 atm). Assume the temperature is 27 °C (300 K). (a) What is the SF₆ concentration in the tank? (b) What is the heat capacity C_V at constant volume per mole, assuming that we can neglect all vibrations of the molecule (but not rotations)? How does this compare with the reported experimental value in Table 1.9? (c) The diameter of the SF₆ molecule is roughly 0.48 nm. What is the mean free path of SF₆ molecules in the container?
- *1.22 Mean free path and gas discharge in Ar-ion laser** Consider the collisions of a free electron with the molecules of a gas inside a laser tube. The much lighter electron is much faster than the heavier gas molecules. From an electron's perspective, the molecules look stationary. Suppose that the electron has just collided with a gas molecule. It moves off in a particular direction and travels a distance ℓ , the mean free path of the electron, and collides again with another or a second molecule, as shown in Figure 1.79. As long as the electron is within the cross-sectional area S of the second molecule, it will collide with it. Clearly, within the volume $S\ell$, there must be at least one molecule inasmuch as the electron collides once after traveling the distance ℓ . If n is the concentration of molecules, then $nS\ell = 1$, so that

$$\ell = \frac{1}{n\pi r^2}$$

Consider the argon gas inside an Ar-ion laser tube. The pressure of the gas in the tube is roughly 0.1 torr. The gas temperature during operation is approximately 1300 °C. A large applied electric field E accelerates a free electron somewhere in the gas. As the electron accelerates, it gains energy from the

Mean free path
of electrons
colliding with
atoms or
molecules

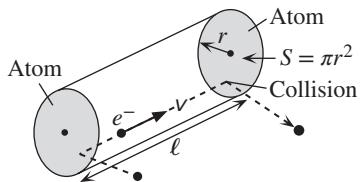


Figure 1.79 The mean free path of an electron in a gas. The electron has a negligible size compared with the scattering gas atom and the electron is much faster than the gas atom. Assume the gas atoms are stationary in determining the mean free path ℓ .

field and when it impacts an Ar atom, it ionizes it to Ar^+ and releases a free electron that can also be accelerated, and so on. The ionization energy of the Ar atom is 15.8 eV. The radius of an Ar atom is approximately 0.143 nm. (See Table 1.11) (a) What is the concentration of Ar atoms in the tube? (b) What is the mean free path of collisions between Ar atoms? (c) What is the mean free path of an electron colliding with Ar atoms? (d) Suppose that the electron is traveling along the force of the field, $F = eE$, so that it gains an energy $Fd\ell$ in moving a distance $d\ell$. What should be the electric field that would impart sufficient energy to the electron over a distance ℓ so that upon collision it may be able to ionize the Ar atom.²⁴

1.23 Thermal expansion

- If λ is the thermal expansion coefficient, show that the thermal expansion coefficient for an area is 2λ . Consider an aluminum square sheet of area 1 cm^2 . If the thermal expansion coefficient of Al at room temperature (25°C) is about $24 \times 10^{-6} \text{ K}^{-1}$, at what temperature is the percentage change in the area $+1\%$?
- A particular incandescent light bulb (100 W, 120 V) has a tungsten (W) filament of length 57.9 cm and a diameter of 63.5 μm . Calculate the length of the filament at 2300°C , the approximate operating temperature of the filament inside the bulb. The linear expansion coefficient λ of W is approximately $4.50 \times 10^{-6} \text{ K}^{-1}$ at 300 K. How would you improve your calculation?

1.24 Thermal expansion of Si

The expansion coefficient of silicon over the temperature range 120–1500 K is given by Okada and Tokumaru (1984) as

$$\lambda = 3.725 \times 10^{-6} [1 - e^{-5.88 \times 10^{-3}(T-124)}] + 5.548 \times 10^{-10} T$$

where λ is in K^{-1} (or $^\circ\text{C}^{-1}$) and T is in kelvins.

- By expanding the above function around 20°C (293 K) show that,
- $$\lambda = 2.5086 \times 10^{-6} + (8.663 \times 10^{-9})(T - 293) - (2.3839 \times 10^{-11})(T - 293)^2$$
- The change $\delta\rho$ in the density due to a change δT in the temperature, from Example 1.9, is given by

$$\delta\rho = -\rho_0\alpha_V \delta T = -3\rho_0\lambda \delta T$$

Given the density of Si as 2.329 g cm^{-3} at 20°C , calculate the density at 1000°C by using the full expression and by using the polynomials expansion of λ . What is your conclusion?

Silicon linear
expansion
coefficient

Silicon linear
expansion
coefficient

1.25 Thermal expansion of GaP and GaAs

- GaP has the zinc blende structure. The linear expansion coefficient in GaP has been measured as follows: $\lambda = 4.65 \times 10^{-6} \text{ K}^{-1}$ at 300 K; $5.27 \times 10^{-6} \text{ K}^{-1}$ at 500 K; $5.97 \times 10^{-6} \text{ K}^{-1}$ at 800 K. Calculate the coefficients, A, B, and C in

$$\frac{dL}{L_o dT} = \lambda(T) = A + B(T - T_o) + C(T - T_o)^2 + \dots$$

where $T_o = 300 \text{ K}$. The lattice constant of GaP, a , at 27°C is 0.5451 nm. Calculate the lattice constant at 300°C .

²⁴ The actual description is quite involved. The electrons in the gas would be moving around randomly and at the same time accelerating due to the presence of an applied field. We will examine this in Chapter 2. Further, the approach in this question is highly simplified to highlight the concept and find very rough estimates rather than carry out accurate calculations. In fact, the cross section that is involved in the ionization of an Ar atom is smaller than the actual cross section of the Ar atom, because the projectile electron may not necessarily ionize the Ar atom during its interactions with it. (The cross section also depends on the energy of the electron.)

GaAs linear expansion coefficient

- b. The linear expansion coefficient of GaAs over 200–1000 K is given by

$$\lambda = 4.25 \times 10^{-6} + (5.82 \times 10^{-9})T - (2.82 \times 10^{-12})T^2$$

where T is in kelvins. The lattice constant a at 300 K is 0.56533 nm. Calculate the lattice constant and the density at -40°C .

- 1.26 Bimetal cantilever devices** Consider two thin plate strips of equal length that are welded to each other as shown in Figure 1.80a. Suppose that metal B has a large thermal expansion coefficient λ_B than λ_A . A and B are of equal length L_o at T_o . When the temperature increases by ΔT , B extends more than A so that the extension in lengths can only be accommodated if the two-metal system bends to form an arc of a circle centered at O as in Figure 1.80b. Center-to-center separation of the strips is $d/2$ so that the strip thickness is d and the two metals are assumed to have the same thickness. Suppose that L_A and L_B are the new lengths (along the center of the strip), then

$$L_A = L_o(1 + \lambda_A \Delta T) \quad \text{and} \quad L_B = L_o(1 + \lambda_B \Delta T)$$

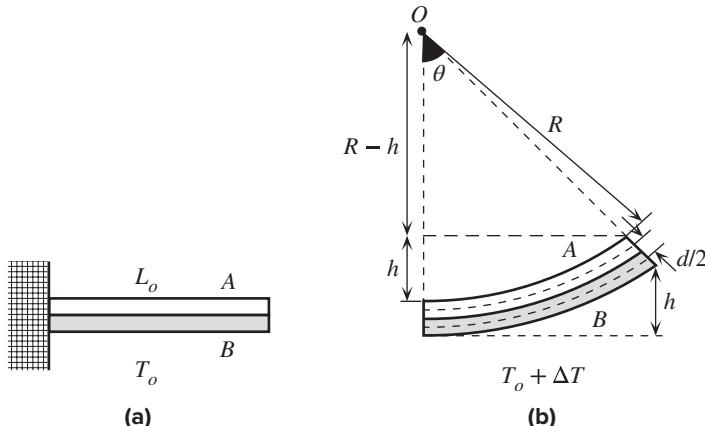


Figure 1.80 (a) Two different thin metals of identical length at T_o . (b) The lengths are different at a higher temperature. B expands more than A . The two metals bend to form an arc of a circle centered at O with a radius R . The arc subtends an angle θ at O .

Bending of a bimetallic strip

Both lengths have the same angle θ at O as shown in Figure 1.80b. Show that

$$\theta = \frac{2L_o(\lambda_B - \lambda_A)\Delta T}{d}$$

Show that the deflection h (very small) can be calculated from the geometry once we know θ , that is, for small θ

$$h \approx \frac{1}{2}L_o\theta$$

(Hint: $L_o/R \approx \sin \theta$ and $(R - h)/R \approx \cos \theta$ and then expand in terms of small θ)

Consider a steel-brass bimetallic strip cantilever as in Figure 1.80a, that is 1 mm thick and 100 mm long. The thermal expansion coefficient for steel is $10 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$, and for brass, it is $20 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$. If the bimetallic strip is flat at 20°C , what is the deflection at 100°C ?

- 1.27 Electrical noise** Consider an amplifier with a bandwidth B of 5 kHz, corresponding to a typical speech bandwidth. Assume the input resistance of the amplifier is $1 \text{ M}\Omega$. What is the rms noise voltage at the input? What will happen if the bandwidth is doubled to 10 kHz? What is your conclusion?

- 1.28 Thermal activation** A certain chemical oxidation process (e.g., SiO_2) has an activation energy of 2 eV atom $^{-1}$.

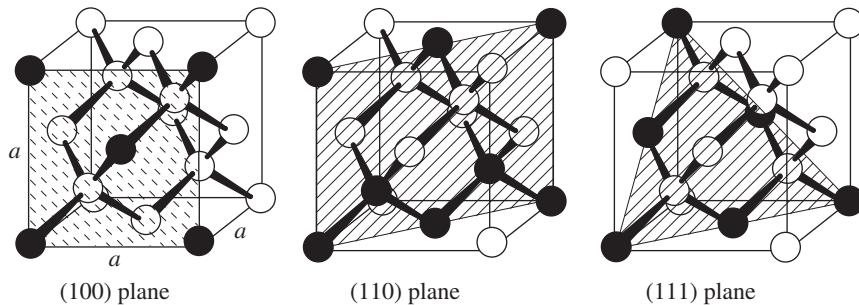
- Consider the material exposed to pure oxygen gas at a pressure of 1 atm at 27°C . Estimate how many oxygen molecules per unit volume will have energies in excess of 2 eV? (Consider the numerical integration of Equation 1.26.)
- If the temperature is 900°C , estimate the number of oxygen molecules with energies more than 2 eV. What happens to this concentration if the pressure is doubled?

- 1.29 Diffusion in Si** The diffusion coefficient of boron (B) atoms in a single crystal of Si has been measured to be $1.5 \times 10^{-18} \text{ m}^2 \text{ s}^{-1}$ at 1000°C and $1.1 \times 10^{-16} \text{ m}^2 \text{ s}^{-1}$ at 1200°C .
- What is the activation energy for the diffusion of B, in eV/atom?
 - What is the preexponential constant D_o ?
 - What is the rms distance (in micrometers) diffused in 1 hour by the B atom in the Si crystal at 1200°C and 1000°C ?
 - The diffusion coefficient of B in polycrystalline Si has an activation energy of 2.4–2.5 eV/atom and $D_o = (1.5 - 6) \times 10^{-7} \text{ m}^2 \text{ s}^{-1}$. What constitutes the diffusion difference between the single crystal sample and the polycrystalline sample?
- 1.30 Diffusion in SiO_2** The diffusion coefficient of P atoms in SiO_2 has an activation energy of 2.30 eV/atom and $D_o = 5.73 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$. What is the rms distance diffused in 1 hour by P atoms in SiO_2 at 1200°C ?
- 1.31 BCC and FCC crystals**
- Molybdenum has the BCC crystal structure, a density of 10.22 g cm^{-3} , and an atomic mass of 95.94 g mol^{-1} . What is the atomic concentration, lattice parameter a , and atomic radius of molybdenum?
 - Gold has the FCC crystal structure, a density of 19.3 g cm^{-3} , and an atomic mass of $196.97 \text{ g mol}^{-1}$. What is the atomic concentration, lattice parameter a , and atomic radius of gold?
- 1.32 BCC and FCC crystals**
- Tungsten (W) has the BCC crystal structure. The radius of the W atom is 0.1371 nm. The atomic mass of W is 183.8 amu (g mol^{-1}). Calculate the number of W atoms per unit volume and density of W.
 - Platinum (Pt) has the FCC crystal structure. The radius of the Pt atom is 0.1386 nm. The atomic mass of Pt is 195.09 amu (g mol^{-1}). Calculate the number of Pt atoms per unit volume and density of Pt.
- 1.33 Planar and surface concentrations** Niobium (Nb) has the BCC crystal with a lattice parameter $a = 0.3294 \text{ nm}$. Find the planar concentrations as the number of atoms per nm^2 of the (100), (110), and (111) planes. Which plane has the most concentration of atoms per unit area? Sometimes the number of atoms per unit area n_{surface} on the surface of a crystal is estimated by using the relation $n_{\text{surface}} = n_{\text{bulk}}^{2/3}$, where n_{bulk} is the concentration of atoms in the bulk. Compare n_{surface} values with the planar concentrations that you calculated and comment on the difference. [Note: The BCC (111) plane does not cut through the center atom and the (111) has one-sixth of an atom at each corner.]
- 1.34 Diamond and zinc blende** Si has the diamond and GaAs has the zinc blende crystal structure. Given the lattice parameters of Si and GaAs, $a = 0.543 \text{ nm}$ and $a = 0.565 \text{ nm}$, respectively, and the atomic masses of Si, Ga, and As as 28.08, 69.73, and 74.92, respectively, calculate the density of Si and GaAs. What is the atomic concentration (atoms per unit volume) in each crystal?
- 1.35 Zinc blende, NaCl , and CsCl**
- InAs is a III–V semiconductor that has the zinc blende structure with a lattice parameter of 0.606 nm . Given the atomic masses of In ($114.82 \text{ g mol}^{-1}$) and As (74.92 g mol^{-1}), find the density.
 - CdO has the NaCl crystal structure with a lattice parameter of 0.4695 nm . Given the atomic masses of Cd ($112.41 \text{ g mol}^{-1}$) and O (16.00 g mol^{-1}), find the density.
 - KCl has the same crystal structure as NaCl . The lattice parameter a of KCl is 0.629 nm . The atomic masses of K and Cl are 39.10 g mol^{-1} and 35.45 g mol^{-1} , respectively. Calculate the density of KCl.
- 1.36 Crystallographic directions and planes** Consider the cubic crystal system.
- Show that the line $[hkl]$ is perpendicular to the (hkl) plane.
 - Show that the spacing between adjacent (hkl) planes is given by

$$d = \frac{a}{\sqrt{h^2 + k^2 + \ell^2}}$$

1.37 Si and SiO_2

- Given the Si lattice parameter $a = 0.543 \text{ nm}$, calculate the number of Si atoms per unit volume, in nm^{-3} .
- Calculate the number of atoms per m^2 and per nm^2 on the (100), (110), and (111) planes in the Si crystal as shown in Figure 1.81. Which plane has the most number of atoms per unit area?
- The density of SiO_2 is 2.27 g cm^{-3} . Given that its structure is amorphous, calculate the number of molecules per unit volume, in nm^{-3} . Compare your result with (a) and comment on what happens when the surface of an Si crystal oxidizes. The atomic masses of Si and O are 28.09 and 16, respectively.

**Figure 1.81** Diamond cubic crystal structure and planes.

Determine what portion of a black-colored atom belongs to the plane that is hatched.

1.38 Vacancies in metals

- The energy of formation of a vacancy in the copper crystal is about 1 eV. Calculate the concentration of vacancies at room temperature (300 K) and just below the melting temperature, 1084 °C. Neglect the change in the density which is small.
- Table 1.12 shows the energies of vacancy formation in various metals with *close-packed* crystal structures and the melting temperature T_m . Plot E_v in eV versus T_m in kelvins, and explore if there is a correlation between E_v and T_m . Some materials engineers take E_v to be very roughly $10kT_m$. Do you think that they are correct? (Justify.)

Table 1.12 Energy of formation of vacancies for selected metals

	Metal								
	Al	Ag	Au	Cu	Mg	Pt	Pb	Ni	Pd
Crystal	FCC	FCC	FCC	FCC	HCP	FCC	FCC	FCC	FCC
E_v (eV)	0.70–0.76	1.0–1.1	0.90–0.98	1–1.28	0.79–0.89	1.3–1.5	0.55	1.63–1.79	1.54–1.85
T_m (°C)	660	962	1064	1085	650	1768	328	1455	1555

- 1.39 Vacancies in silicon** In device fabrication, Si is frequently doped by the diffusion of impurities (dopants) at high temperatures, typically 950–1200 °C. The energy of vacancy formation in the Si crystal is about 3.6 eV. What is the equilibrium concentration of vacancies in a Si crystal at 1000 °C? Neglect the change in the density with temperature which is less than 1 percent in this case.

- 1.40 Pb–Sn solder** Consider the soldering of two copper components. When the solder melts, it wets both metal surfaces. If the surfaces are not clean or have an oxide layer, the molten solder cannot wet the surfaces and the soldering fails. Assume that soldering takes place at 250 °C, and consider the diffusion of Sn atoms into the copper (the Sn atom is smaller than the Pb atom and hence diffuses more easily).

- The diffusion coefficient of Sn in Cu at two temperatures is $D = 1.69 \times 10^{-9} \text{ cm}^2 \text{ hr}^{-1}$ at 400°C and $D = 2.48 \times 10^{-7} \text{ cm}^2 \text{ hr}^{-1}$ at 650°C . Calculate the rms distance diffused by an Sn atom into the copper, assuming the cooling process takes 10 seconds.
- What should be the composition of the solder if it is to begin freezing at 250°C ?
- What are the components (phases) in this alloy at 200°C ? What are the compositions of the phases and their relative weights in the alloy?
- What is the microstructure of this alloy at 25°C ? What are weight fractions of the α and β phases assuming near equilibrium cooling?

1.41 Pb–Sn solder Consider 50% Pb–50% Sn solder.

- Sketch the temperature-time profile and the microstructure of the alloy at various stages as it is cooled from the melt.
- At what temperature does the solid melt?
- What is the temperature range over which the alloy is a mixture of melt and solid? What is the structure of the solid?
- Consider the solder at room temperature following cooling from 182°C . Assume that the rate of cooling from 182°C to room temperature is faster than the atomic diffusion rates needed to change the compositions of the α and β phases in the solid. Assuming the alloy is 1 kg, calculate the masses of the following components in the solid:
 - The primary α .
 - α in the whole alloy.
 - α in the eutectic solid.
 - β in the alloy. (Where is the β -phase?)
- Calculate the specific heat of the solder given the atomic masses of Pb (207.2) and Sn (118.71).

1.42 Gruneisen's rule and metals Al and Cu both have metallic bonding and the same crystal structure. Assuming that the Gruneisen's parameter γ for Al is the same as that for Cu, $\gamma = 2.1$ (see Table 1.8), estimate the linear expansion coefficient λ of Al, given that its bulk modulus $K = 75 \text{ GPa}$, $c_s = 900 \text{ J K}^{-1} \text{ kg}^{-1}$, and $\rho = 2.7 \text{ g cm}^{-3}$. Compare your estimate with the experimental value of $23.5 \times 10^{-6} \text{ K}^{-1}$.

1.43 Heat capacity and the thermal expansion coefficient of diamond Given that diamond has a bulk modulus of 443 GPa, specific heat capacity of $0.51 \text{ J g}^{-1} \text{ K}^{-1}$ and a density of 3.51 g cm^{-3} , estimate its linear expansion coefficient at room temperature taking the Grüneisen parameter as ~ 1 .



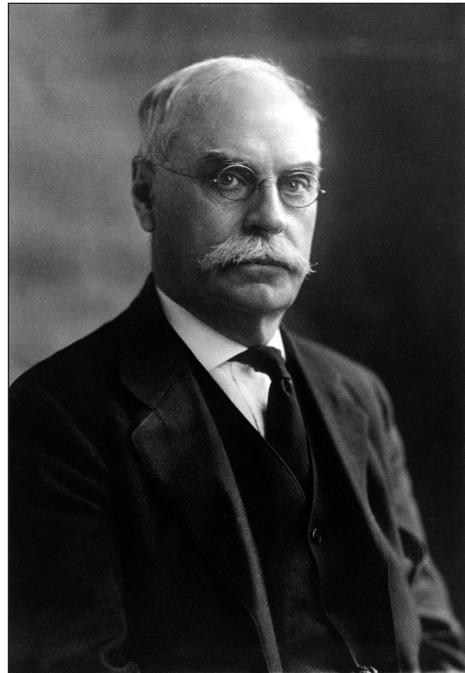
Right: Gordon Teal (Left) and Morgan Sparks fabricated the first grown-junction Ge transistor in 1950–1951 at Bell Labs. Gordon Teal started at Bell Labs but later moved to Texas Instruments where he led the development of the first commercial Si transistor; the first Si transistor was made at Bell Labs by Morris Tanenbaum. The Czochralski crystal growth of Ge and Si crystals was instrumental in the development of these transistors.

© Nokia Corporation.



Paul Drude (1863–1906) was a German physicist who is best known for his works on optics, and the electrical and optical properties of materials. He obtained his PhD from the University of Göttingen and held physics professorships at the University of Leipzig, University of Giessen and Humboldt University of Berlin. He proposed the electrical conduction model that bears his name around 1900.

| Courtesy of AIP Emilio Segrè Visual Archives, Physics Today Collection.



Edwin Hall (1855–1906) was an American physicist who obtained his PhD from Johns Hopkins University during which time he discovered the Hall effect (1879). Following his PhD (1880), he joined Harvard University as a Professor of Physics until his retirement in 1921.

| © Science & Society Picture Library/Getty Images.

CHAPTER**2**

Electrical and Thermal Conduction in Solids: Mainly Classical Concepts

Electrical conduction involves the motion of charges in a material under the influence of an applied electric field. A material can generally be classified as a conductor if it contains a large number of “free” or mobile charge carriers. In metals, due to the nature of metallic bonding, the valence electrons from the atoms form a sea of electrons that are free to move within the metal and are therefore called conduction electrons. In this chapter, we will treat the conduction electrons in metal as “free charges” that can be accelerated by an applied electric field. In the presence of an electric field, the conduction electrons attain an average velocity, called the drift velocity, that depends on the field. By applying Newton’s second law to electron motion and using such concepts as mean free time between electron collisions with lattice vibrations, crystal defects, impurities, etc., we will derive the fundamental equations that govern electrical conduction in solids. A key concept will be the drift mobility, which is a measure of the ease with which charge carriers in the solid drift under the influence of an external electric field.

Good electrical conductors, such as metals, are also known to be good thermal conductors. The conduction of thermal energy from higher to lower temperature regions in a metal involves the conduction electrons carrying the energy. Consequently, there is an innate relationship between the electrical and thermal conductivities, which is supported by theory and experiments.

2.1 CLASSICAL THEORY: THE DRUDE MODEL

The electric current density J is defined as the net amount of charge flowing across a unit area per unit time, that is,

Current density definition

$$J = \frac{\Delta q}{A \Delta t}$$

where Δq is the net quantity of charge flowing through an area A in time Δt . Figure 2.1 shows the net flow of electrons in a conductor section of cross-sectional area A in the presence of an applied field E_x . Notice that the direction of electron motion is opposite to that of the electric field E_x and of conventional current, because the electrons experience a Coulombic force eE_x in the x direction, due to their negative charge.

We know that the conduction electrons are actually moving around randomly¹ in the metal, but we will assume that as a result of the application of the electric field E_x , they all acquire a net velocity in the x direction. Otherwise, there would be no net flow of charge through area A .

The average velocity of the electrons in the x direction at time t is denoted $v_{dx}(t)$. This is called the **drift velocity**, which is the instantaneous velocity v_x in the x direction averaged over many electrons (perhaps, $\sim 10^{28} \text{ m}^{-3}$); that is

Definition of drift velocity

$$v_{dx} = \frac{1}{N}[v_{x1} + v_{x2} + v_{x3} + \dots + v_{xN}] \quad [2.1]$$

where v_{xi} is the x direction velocity of the i th electron, and N is the number of conduction electrons in the metal. Suppose that n is the number of electrons per unit volume in the conductor ($n = N/V$). In time Δt , electrons move a distance $\Delta x = v_{dx} \Delta t$, so the total charge Δq crossing the area A is $enA v_{dx} \Delta t$. This is valid because all the electrons within distance Δx pass through A ; thus, $n(A \Delta x)$ is the total number of electrons crossing A in time Δt .

The current density in the x direction is

$$J_x = \frac{\Delta q}{A \Delta t} = \frac{enA v_{dx} \Delta t}{A \Delta t} = env_{dx}$$

This general equation relates J_x to the average velocity v_{dx} of the electrons. It must be appreciated that the average velocity at one time may not be the same as at another time, because the applied field, for example, may be changing: $E_x = E_x(t)$. We therefore allow for a time-dependent current by writing

Current density and drift velocity

$$J_x(t) = env_{dx}(t) \quad [2.2]$$

To relate the current density J_x to the electric field E_x , we must examine the effect of the electric field on the motion of the electrons in the conductor. To do so, we will consider the copper crystal.

¹ All the conduction electrons are “free” within the metal and move around randomly, being scattered from vibrating metal ions, as we discuss in this chapter.

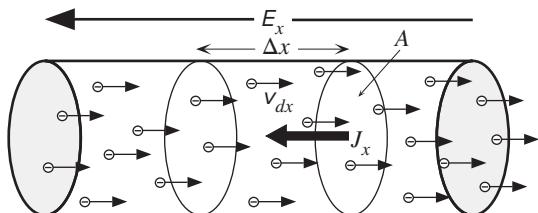


Figure 2.1 Drift of electrons in a conductor in the presence of an applied electric field.

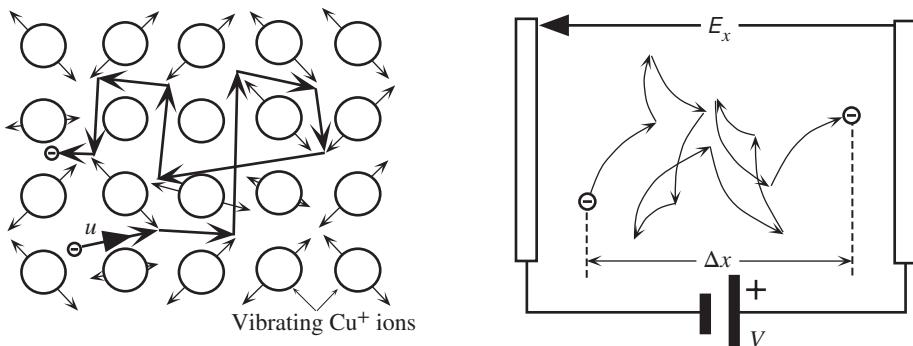
Electrons drift with an average velocity v_{dx} in the x direction.

The copper atom has a single valence electron in its $4s$ subshell, and this electron is loosely bound. The solid metal consists of positive ion cores, Cu^+ , at regular sites, in the face-centered cubic (FCC) crystal structure. The valence electrons detach themselves from their parents and wander around freely in the solid, forming a kind of electron cloud or gas. These mobile electrons are free to respond to an applied field, creating a current density J_x . The valence electrons in the electron gas are therefore **conduction electrons**.

The attractive forces between the negative electron cloud and the Cu^+ ions are responsible for metallic bonding and the existence of the solid metal. (This simplistic view of metal was depicted in Figure 1.7 for copper.) The electrostatic attraction between the conduction electrons and the positive metal ions, like the electrostatic attraction between the electron and the proton in the hydrogen atom, results in the conduction electron having both potential energy PE and kinetic energy KE . The conduction electrons move about the crystal lattice in the same way that gas atoms move randomly in a cylinder. Although the average KE for gas atoms is $\frac{3}{2}kT$, this is not the case for electrons in a metal, because these electrons strongly interact with the metal ions and with each other as a result of electrostatic interactions.

The mean KE of the conduction electrons in a metal is primarily determined by the electrostatic interaction of these electrons with the positive metal ions and also with each other. For most practical purposes, we will therefore neglect the temperature dependence of the mean KE compared with other factors that control the behavior of the conduction electrons in the metal crystal. We can speculate from Example 1.1, that the magnitude of mean KE must be comparable to the magnitude of the mean PE of electrostatic interaction² or, stated differently, to the metal bond energy which is several electron volts per atom. If u is the **mean speed** of the conduction electrons, then, from electrostatic interactions alone, we expect $\frac{1}{2}m_e u^2$ to be several electron volts which means that u is typically $\sim 10^6 \text{ m s}^{-1}$. This purely classical and intuitive reasoning is not sufficient, however, to show that the mean speed u is relatively temperature insensitive and much greater than that expected from kinetic molecular theory. The true reasons are quantum mechanical and are discussed in Chapter 4. (They arise from what is called the Pauli exclusion principle.)

² There is a theorem in classical mechanics called the **virial theorem**, which states that for a collection of particles, the mean KE has half the magnitude of the mean PE if the only forces acting on the particles are such that they follow an inverse square law dependence on the particle-particle separation (as in Coulombic and gravitational forces).



(a) A conduction electron in the electron gas moves about randomly in a metal (with a mean speed u) being frequently and randomly scattered by thermal vibrations of the atoms. In the absence of an applied field there is no net drift in any direction.

(b) In the presence of an applied field, E_x , there is a net drift along the x direction. This net drift along the force of the field is superimposed on the random motion of the electron. After many scattering events the electron has been displaced by a net distance, Δx , from its initial position toward the positive terminal.

Figure 2.2 Motion of a conduction electron in a metal.

In general, the copper crystal will not be perfect and the atoms will not be stationary. There will be crystal defects, vacancies, dislocations, impurities, etc., which will scatter the conduction electrons. More importantly, due to their thermal energy, the atoms will vibrate about their lattice sites (equilibrium positions), as depicted in Figure 2.2a. An electron will not be able to avoid collisions with vibrating atoms; consequently, it will be “scattered” from one atom to another. In the absence of an applied field, the path of an electron may be visualized as illustrated in Figure 2.2a, where scattering from lattice vibrations causes the electron to move randomly in the lattice. On those occasions when the electron reaches a crystal surface, it becomes “deflected” (or “bounced”) back into the crystal. Therefore, in the absence of a field, after some duration of time, the electron crosses its initial x plane position again. Over a long time, the electrons therefore show no net displacement in any one direction.

When the conductor is connected to a battery and an electric field is applied to the crystal, as shown in Figure 2.2b, the electron experiences an acceleration in the x direction in addition to its random motion, so after some time, it will drift a finite distance in the x direction. The electron accelerates along the x direction under the action of the force eE_x , and then it suddenly collides with a vibrating atom and loses the gained velocity. Therefore, there is an average velocity in the x direction, which, if calculated, determines the current via Equation 2.2. Note that since the electron experiences an acceleration in the x direction, its trajectory between collisions is a parabola, like the trajectory of a golf ball experiencing acceleration due to gravity.

To calculate the drift velocity v_{dx} of the electrons due to applied field E_x , we first consider the velocity v_{xi} of the i th electron in the x direction at time t . Suppose

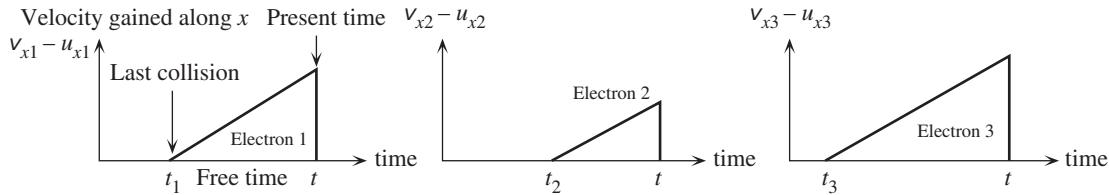


Figure 2.3 Velocity gained in the x direction at time t from the electric field (E_x) for three electrons.

There will be N electrons to consider in the metal.

its last collision was at time t_i ; therefore, for time $(t - t_i)$, it accelerated *free of collisions*, as indicated in Figure 2.3. Let u_{xi} be the velocity of electron i in the x direction just after the collision. We will call this the initial velocity. Since eE_x/m_e is the acceleration of the electron, the velocity v_{xi} in the x direction at time t will be

$$v_{xi} = u_{xi} + \frac{eE_x}{m_e}(t - t_i)$$

However, this is only for the i th electron. We need the average velocity v_{dx} for all such electrons along x . We average the expression for $i = 1$ to N electrons, as in Equation 2.1. We assume that immediately after a collision with a vibrating ion, the electron may move in any random direction; that is, it can just as likely move along the negative or positive x , so that u_{xi} averaged over many electrons is zero. Thus,

$$v_{dx} = \frac{1}{N}[v_{x1} + v_{x2} + \dots + v_{xN}] = \frac{eE_x}{m_e}\overline{(t - t_i)}$$
Drift velocity

where $\overline{(t - t_i)}$ is the **average free time** for N electrons between collisions.

Suppose that τ is the mean free time, or the **mean time between collisions** (also known as the **mean scattering time**). For some electrons, $(t - t_i)$ will be greater than τ , and for others, it will be shorter, as shown in Figure 2.3. Averaging $(t - t_i)$ for N electrons will be the same as τ . Thus, we can substitute τ for $(t - t_i)$ in the previous expression to obtain

$$v_{dx} = \frac{e\tau}{m_e}E_x$$
[2.3]
Drift velocity
and field

Equation 2.3 shows that the drift velocity increases linearly with the applied field. The constant of proportionality $e\tau/m_e$ has been given a special name and symbol. It is called the **drift mobility** μ_d , which is defined as

$$v_{dx} = \mu_d E_x$$
[2.4]
Definition of
drift mobility

where

$$\mu_d = \frac{e\tau}{m_e}$$
[2.5]
Drift mobility
and mean free
time

Equation 2.5 relates the drift mobility of the electrons to their mean scattering time τ . To reiterate, τ , which is also called the **relaxation time**, is directly related

to the microscopic processes that cause the scattering of the electrons in the metal; that is, lattice vibrations, crystal imperfections, and impurities, to name a few.

From the expression for the drift velocity v_{dx} , the current density J_x follows immediately by substituting Equation 2.4 into 2.2, that is,

Ohm's law

$$J_x = en\mu_d E_x \quad [2.6]$$

Therefore, the current density is proportional to the electric field and the conductivity σ is the term multiplying E_x , that is,

Unipolar conductivity

$$\sigma = en\mu_d \quad [2.7]$$

It is gratifying that by treating the electron as a particle and applying classical mechanics ($F = ma$), we are able to derive Ohm's law. We should note, however, that we assumed τ to be independent of the field.

Drift mobility is important because it is a widely used electronic parameter in semiconductor device physics. The drift mobility gauges how fast electrons will drift when driven by an applied field. If the electron is not highly scattered, then the mean free time between collisions will be long, τ will be large, and by Equation 2.5, the drift mobility will also be large; the electrons will therefore be highly mobile and be able to "respond" to the field. However, a large drift mobility does not necessarily imply high conductivity, because σ also depends on the concentration of conduction electrons n .

The mean time between collisions τ has further significance. Its reciprocal $1/\tau$ represents the **mean frequency of collisions or scattering events**; that is, $1/\tau$ is the mean probability per unit time that the electron will be scattered (see Example 2.1). Therefore, during a small time interval δt , the probability of scattering will be $\delta t/\tau$. The probability of scattering per unit time $1/\tau$ is time independent and depends only on the nature of the electron scattering mechanism.

There is one important assumption in the derivation of the drift velocity v_{dx} in Equation 2.3. We obtained v_{dx} by averaging the velocities v_{xi} of N electrons along x at one instant, as defined in Equation 2.1. The drift velocity therefore represents the average velocity of *all* the electrons along x at one instant; that is, v_{dx} is a number average at one instant. Figure 2.2b shows that after many collisions, after a time interval $\Delta t \gg \tau$, an electron would have been displaced by a net distance Δx along x . The term $\Delta x/\Delta t$ represents the effective velocity with which the electron drifts along x . It is an average velocity for one electron over many collisions, that is, over a long time (hence, $\Delta t \gg \tau$), so $\Delta x/\Delta t$ is a time average. Provided that Δt contains many collisions, it is reasonable to expect that the drift velocity $\Delta x/\Delta t$ from the time average for one electron is the same as the drift velocity v_{dx} per electron from averaging for all electrons at one instant, as in Equation 2.1, or

Drift velocity

$$\frac{\Delta x}{\Delta t} = v_{dx}$$

The two velocities are the same only under steady-state conditions ($\Delta t \gg \tau$). Example 2.4 derives v_{dx} for one electron and shows that it is the same as Equation 2.3.

PROBABILITY OF SCATTERING PER UNIT TIME AND THE MEAN FREE TIME If $1/\tau$ is defined as the mean probability per unit time that an electron is scattered, show that the mean time between collisions is τ .

EXAMPLE 2.1**SOLUTION**

Consider an infinitesimally small time interval dt at time t . Let N be the number of unscattered electrons at time t . The probability of scattering during dt is $(1/\tau) dt$, and the number of scattered electrons during dt is $N(1/\tau) dt$. The change dN in N is thus

$$dN = -N\left(\frac{1}{\tau}\right)dt$$

The negative sign indicates a reduction in N because, as electrons become scattered, N decreases. Integrating this equation, we can find N at any time t , given that at time $t = 0$, N_0 is the total number of unscattered electrons. Therefore,

$$N = N_0 \exp\left(-\frac{t}{\tau}\right)$$

Unscattered
electron
concentration

This equation represents the number of unscattered electrons at time t . It reflects an exponential decay law for the number of unscattered electrons. The above equation is called the probability distribution function for unscattered electrons in time. It is a probability distribution for free times.

The **mean free time** $\bar{\tau}$ can be calculated from the mathematical definition of $\bar{\tau}$,

$$\bar{\tau} = \frac{\int_0^\infty tN dt}{\int_0^\infty N dt} = \tau$$

Mean free
time

where we have used $N = N_0 \exp(-t/\tau)$. Clearly, $1/\tau$ is the **mean probability of scattering per unit time**.

It is left as an exercise to show that the exponential probability distribution of free times above can also be used to calculate the mean square time $\bar{\tau}^2$, which is $2\tau^2$.

ELECTRON DRIFT MOBILITY IN METALS Calculate the drift mobility and the mean scattering time of conduction electrons in copper at room temperature, given that the conductivity of copper is $5.9 \times 10^5 \Omega^{-1} \text{ cm}^{-1}$. The density of copper is 8.96 g cm^{-3} and its atomic mass is 63.5 g mol^{-1} . If the mean speed of the conduction electrons in Cu is roughly $1.6 \times 10^6 \text{ m s}^{-1}$, what is the mean free path between collisions?

EXAMPLE 2.2**SOLUTION**

We can calculate μ_d from $\sigma = en\mu_d$ because we already know the conductivity σ . The number of free electrons n per unit volume can be taken as equal to the number of Cu atoms per unit volume, if we assume that each Cu atom donates one electron to the conduction electron gas in the metal. One mole of copper has N_A (6.02×10^{23}) atoms and a mass of 63.5 g. Therefore, the number of copper atoms per unit volume is

$$n = \frac{dN_A}{M_{\text{at}}}$$

where d = density = 8.96 g cm^{-3} , and M_{at} = atomic mass = $63.5 \text{ (g mol}^{-1})$. Substituting for d , N_A , and M_{at} , we find $n = 8.5 \times 10^{22} \text{ electrons cm}^{-3}$.

The electron drift mobility is therefore

$$\mu_d = \frac{\sigma}{en} = \frac{5.9 \times 10^5 \Omega^{-1} \text{ cm}^{-1}}{[(1.6 \times 10^{-19} \text{ C})(8.5 \times 10^{22} \text{ cm}^{-3})]} = 43.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$$

From the drift mobility we can calculate the mean free time τ between collisions by using Equation 2.5,

$$\tau = \frac{\mu_d m_e}{e} = \frac{(43.4 \times 10^{-4} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1})(9.1 \times 10^{-31} \text{ kg})}{1.6 \times 10^{-19} \text{ C}} = 2.5 \times 10^{-14} \text{ s}$$

The mean speed u of the conduction electrons is about $1.6 \times 10^6 \text{ m s}^{-1}$, so that their mean free path $\ell = u\tau = 39 \text{ nm}$.

EXAMPLE 2.3

DRIFT VELOCITY AND MEAN SPEED What is the applied electric field that will impose a drift velocity equal to 0.1 percent of the mean speed u ($\sim 10^6 \text{ m s}^{-1}$) of conduction electrons in copper? What is the corresponding current density and current through a Cu wire of diameter 1 mm?

SOLUTION

The drift velocity of the conduction electrons is $v_{dx} = \mu_d E_x$, where μ_d is the drift mobility, which for copper is $43.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (see Example 2.2). With $v_{dx} = 0.001u = 10^3 \text{ m s}^{-1}$, we have

$$E_x = \frac{v_{dx}}{\mu_d} = \frac{10^3 \text{ m s}^{-1}}{43.4 \times 10^{-4} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}} = 2.3 \times 10^5 \text{ V m}^{-1} \quad \text{or} \quad 2.3 \text{ kV cm}^{-1}$$

This is an unattainably large electric field in a metal. Given the conductivity σ of copper, the equivalent current density is

$$J_x = \sigma E_x = (5.9 \times 10^7 \Omega^{-1} \text{ m}^{-1})(2.3 \times 10^5 \text{ V m}^{-1}) \\ = 1.4 \times 10^{13} \text{ A m}^{-2} \quad \text{or} \quad 1.4 \times 10^7 \text{ A mm}^{-2}$$

This means a current of $1.1 \times 10^7 \text{ A}$ through a 1 mm diameter wire! It is clear from this example that for all practical purposes, even under the highest working currents and voltages, the drift velocity is much smaller than the mean speed of the electrons. Consequently, when an electric field is applied to a conductor, for all practical purposes, the mean speed is unaffected.

EXAMPLE 2.4

DRIFT VELOCITY IN A FIELD: A CLOSER LOOK There is another way to explain the observed dependence of the drift velocity on the field, and Equation 2.3. Consider the path of a conduction electron in an applied field E_x as shown in Figure 2.4. Suppose that at time $t = 0$ the electron has just been scattered from a lattice vibration. Let u_{x1} be the initial velocity in the x direction just after this initial collision (to which we assign a collision number of zero). We will assume that immediately after a collision, the velocity of the electron is in a random *direction*. Suppose that the first collision occurs at time t_1 . Since eE_x/m_e is the acceleration, the distance s_1 covered in the x direction during the free time t_1 will be

$$s_1 = u_{x1} t_1 + \frac{1}{2} \left(\frac{eE_x}{m_e} \right) t_1^2$$

Distance
traversed
along x before
collision

At time t_1 , the electron collides with a lattice vibration (its first collision), and the velocity is randomized again to become u_{x2} . The whole process is then repeated during the next

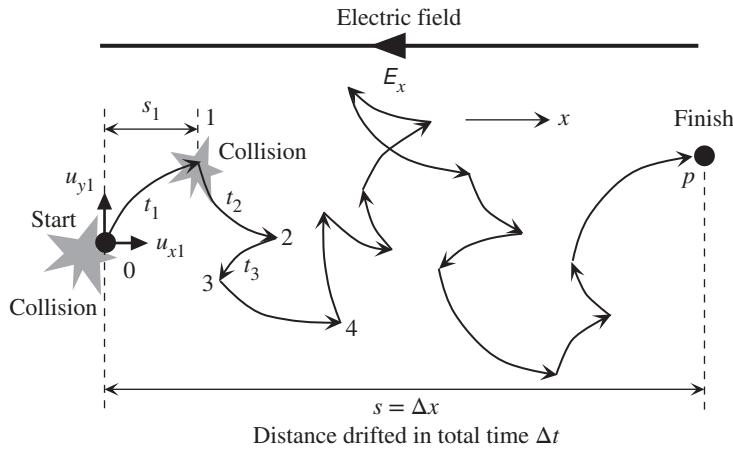


Figure 2.4 The motion of a single electron in the presence of an electric field E . During a time interval t_i , the electron traverses a distance s_i along x . After p collisions, it has drifted a distance $s = \Delta x$.

interval which lasts for a free time t_2 , and the electron traverses a distance s_2 along x , and so on. To find the overall distance traversed by the electron after p such scattering events, we sum all the above distances s_1, s_2, \dots for p free time intervals,

$$s = s_1 + s_2 + \dots + s_p = [u_{x1}t_1 + u_{x2}t_2 + \dots + u_{xp}t_p] + \frac{1}{2}\left(\frac{eE_x}{m_e}\right)[t_1^2 + t_2^2 + \dots + t_p^2] \quad [2.8]$$

Since after a collision the “initial” velocity u_x is always random, the first term has u_x values that are randomly negative and positive, so for many collisions (large p) the first term on the right-hand side of Equation 2.8 is nearly zero and can certainly be neglected compared with the second term. Thus, after many collisions, the net distance $s = \Delta x$ traversed in the x direction is given by the second term in Equation 2.8, which is the electric field induced displacement term. If \bar{t}^2 is the **mean square free time**, then

$$s = \frac{1}{2}\left(\frac{eE_x}{m_e}\right)p\bar{t}^2$$

where

$$\bar{t}^2 = \frac{1}{p}[t_1^2 + t_2^2 + \dots + t_p^2]$$

Suppose that τ is the **mean free time between collisions**, where $\tau = (t_1 + t_2 + \dots + t_p)/p$. We know from Example 2.1 that the probability that the electron will not be scattered, that is it is still free, decreases exponentially with time in which the mean free time \bar{t} is τ . Using the same probability distribution function, we easily can show that $\bar{t}^2 = 2(\bar{t})^2 = 2\tau^2$. So in terms of the mean free time τ between collisions, the overall distance $s = \Delta x$ drifted in the x direction after p collisions is

$$s = \frac{eE_x}{m_e}(p\tau^2)$$

Further, since the total time Δt taken for these p scattering events is simply $p\tau$, the drift velocity v_{dx} is given by $\Delta x/\Delta t$ or $s/(p\tau)$, that is,

$$v_{dx} = \frac{e\tau}{m_e}E_x \quad [2.9]$$

Distance
drifted after p
scattering
events

Mean square
free time
definition

Drift velocity
and mean free
time

¹ See Question 2.4 for the derivation.

Drift mobility and conductivity and mean free path

This is the same expression as Equation 2.3, except that τ is defined here as the average free time for a single electron over a long time, that is, over many collisions, whereas previously it was the mean free time averaged over many electrons. Further, in Equation 2.9 v_{dx} is an average drift for an electron over a long time, over many collisions. In Equation 2.1 v_{dx} is the average velocity averaged over all electrons at one instant. For all practical purposes, the two are equivalent. (The equivalence breaks down when we are interested in events over a time scale that is comparable to one scattering, $\sim 10^{-14}$ second.)

The drift mobility μ_d from Equation 2.9 is identical to that of Equation 2.5, $\mu_d = e\tau/m_e$. Suppose that the mean speed of the electrons (not the drift velocity) is u . Then an electron moves a distance $\ell = u\tau$ in mean free time τ , which is called the **mean free path**. The drift mobility and conductivity become,

$$\mu_d = \frac{e\ell}{m_e u} \quad \text{and} \quad \sigma = en\mu_d = \frac{e^2 n\ell}{m_e u} \quad [2.10]$$

Equations 2.3 and 2.10 both assume that after each collision the velocity is randomized. The scattering process, lattice scattering, is able to randomize the velocity in one single scattering. In general not all electron scattering processes can randomize the velocity in one scattering process. If it takes more than one collision to randomize the velocity, then the electron is able to carry with it some velocity gained from a previous collision and hence possesses a higher drift mobility. In such cases one needs to consider the effective mean free path a carrier has to move to eventually randomize the velocity gained; this is a point considered in Chapter 4 when we calculate the resistivity at low temperatures.

2.6 THERMAL CONDUCTION

2.6.1 THERMAL CONDUCTIVITY

Experience tells us that metals are both good electrical and good thermal conductors. We may therefore surmise that the free conduction electrons in a metal must also play a role in heat conduction. Our conjecture is correct for metals, but not for other materials. The transport of heat in a metal is accomplished by the electron gas (conduction electrons), whereas in nonmetals, the conduction is due to lattice vibrations.

When a metal piece is heated at one end, the amplitude of the atomic vibrations, and thus the average kinetic energy of the electrons, in this region increases, as depicted in Figure 2.19. Electrons gain energy from energetic atomic vibrations when the two collide. By virtue of their increased random motion, these energetic electrons then transfer the extra energy to the colder regions by colliding with the atomic vibrations there. Thus, electrons act as “energy carriers.”

The thermal conductivity of a material, as its name implies, measures the ease with which heat, that is, thermal energy, can be transported through the medium. Consider the metal rod shown in Figure 2.20, which is heated at one end. Heat will flow from the hot end to the cold end. Experiments show that the rate of heat flow, $Q' = dQ/dt$, through a thin section of thickness δx is proportional to the temperature gradient $\delta T/\delta x$ and the cross-sectional area A , so

*Fourier's law
of thermal
conduction*

$$Q' = -A\kappa \frac{\delta T}{\delta x} \quad [2.40]$$

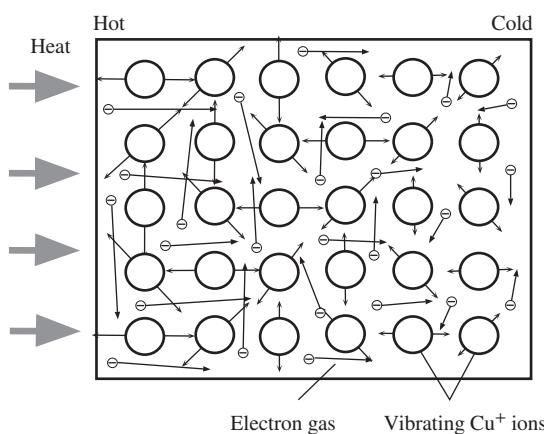


Figure 2.19 Thermal conduction in a metal involves transferring energy from the hot region to the cold region by conduction electrons.

More energetic electrons (shown with longer velocity vectors) from the hotter regions arrive at cooler regions, collide with lattice vibrations, and transfer their energy. Lengths of arrowed lines on atoms represent the magnitudes of atomic vibrations.

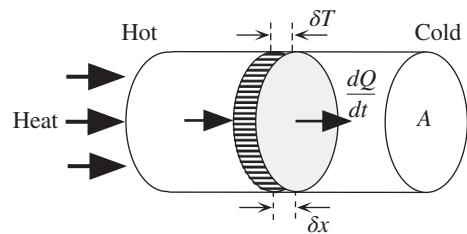


Figure 2.20 Heat flow in a metal rod heated at one end.

Consider the rate of heat flow, dQ/dt , across a thin section δx of the rod. The rate of heat flow is proportional to the temperature gradient $\delta T/\delta x$ and the cross-sectional area A .

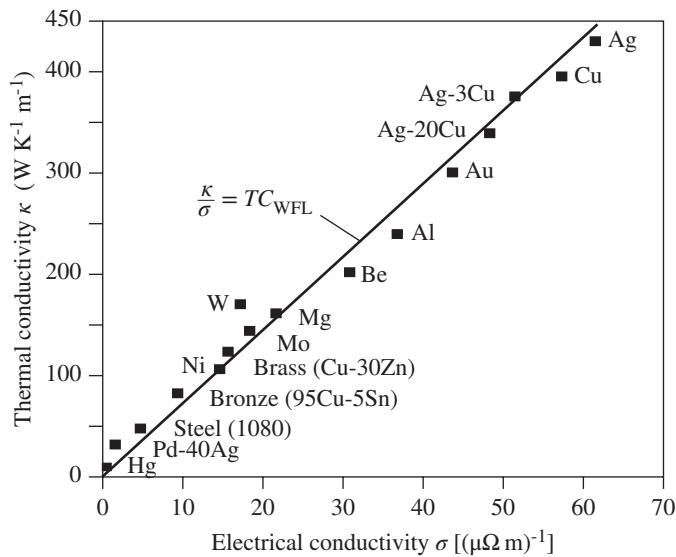


Figure 2.21 Thermal conductivity κ versus electrical conductivity σ for various metals (elements and alloys) at 20 °C.

The solid line represents the WFL law with $C_{\text{WFL}} \approx 2.44 \times 10^8 \text{ W } \Omega \text{ K}^{-2}$.

where κ is a material-dependent **constant of proportionality** that we call the **thermal conductivity**. The negative sign indicates that the heat flow direction is that of decreasing temperature. Equation 2.40 is often referred to as **Fourier's law** of heat conduction and is a defining equation for κ . The driving force for the heat flow is the temperature gradient $\delta T/\delta x$. If we compare Equation 2.40 with Ohm's law for the electric current I , we see that

$$I = -A\sigma \frac{\delta V}{\delta x} \quad [2.41]$$

Ohm's law of electrical conduction

which shows that in this case, the driving force is the potential gradient, that is, the electric field.¹² In metals, electrons participate in the processes of charge and heat transport, which are characterized by σ and κ , respectively. Therefore, it is not surprising to find that the two coefficients are related by the **Wiedemann–Franz–Lorenz law**,¹³ which is

$$\frac{\kappa}{\sigma T} = C_{\text{WFL}} \quad [2.42]$$

Wiedemann–Franz–Lorenz law

where $C_{\text{WFL}} = \pi^2 k^2 / 3e^2 = 2.44 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}$ is a constant called the **Lorenz number** (or the Wiedemann–Franz–Lorenz coefficient).

Experiments on a wide variety of metals, ranging from pure metals to various alloys, show that Equation 2.42 is reasonably well obeyed at close to room temperature and above, as illustrated in Figure 2.21. Since the electrical conductivity of

¹² Recall that $J = \sigma E$ which is equivalent to Equation 2.41.

¹³ Historically, Wiedemann and Franz noted in 1853 that κ/σ is the same for all metals at the same temperature. Lorenz in 1881 showed that κ/σ is proportional to the temperature with a proportionality constant that is nearly the same for many metals. The law stated in equation 2.42 reflects both observations. By the way, Lorenz, who was a Dane, should not be confused with Lorentz, who was Dutch.

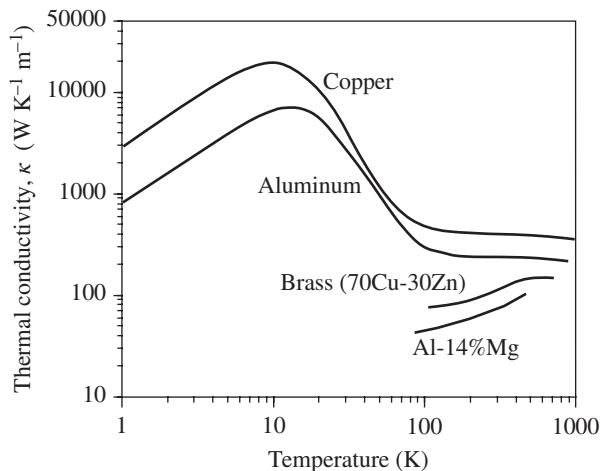


Figure 2.22 Thermal conductivity versus temperature for two pure metals (Cu and Al) and two alloys (brass and Al-14% Mg).

SOURCE: Data extracted from Touloukian, Y.S., et al., "Thermal Conductivity, Metallic Elements and Alloys," *Thermophysical Properties of Matter*, vol.1, 1970. New York, NY: Plenum, 1970.

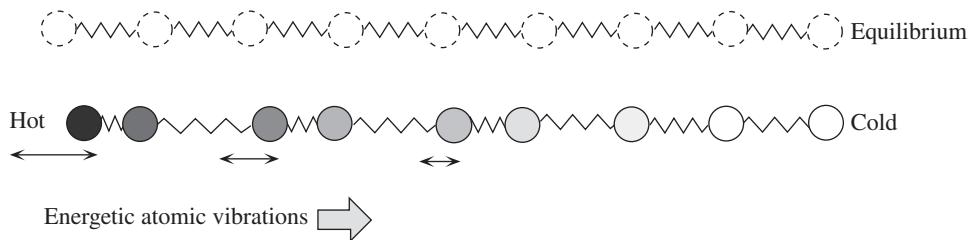


Figure 2.23 Conduction of heat in insulators involves the generation and propagation of atomic vibrations through the bonds that couple the atoms (an intuitive figure).

pure metals is inversely proportional to the temperature, we can immediately conclude that the thermal conductivity of these metals must be relatively temperature independent at room temperature and above.

Figure 2.22 shows the temperature dependence of κ for copper and aluminum down to the lowest temperatures. It can be seen that for these two metals, above ~ 100 K, the thermal conductivity becomes temperature independent, in agreement with Equation 2.42. Qualitatively, above ~ 100 K, κ is constant, because heat conduction depends essentially on the rate at which the electron transfers energy from one atomic vibration to another as it collides with them (Figure 2.19). This rate of energy transfer depends on the mean speed of the electron u , which increases only fractionally with the temperature. In fact, the fractionally small increase in u is more than sufficient to carry the energy from one collision to another and thereby excite more energetic lattice vibrations in the colder regions.

Nonmetals do not have any free conduction electrons inside the crystal to transfer thermal energy from hot to cold regions of the material. In nonmetals, the energy transfer involves lattice vibrations, that is, atomic vibrations of the crystal. We know that we can view the atoms and bonds in a crystal as balls connected together through springs as shown for one chain of atoms in Figure 2.23. As we know from the kinetic

Table 2.5 Typical thermal conductivities of various classes of materials at 25 °C

Material	κ (W m ⁻¹ K ⁻¹)	Material	κ (W m ⁻¹ K ⁻¹)
Pure metal		Ceramics and glasses	
Nb	52	Glass-borosilicate	0.75
Fe	80	Silica-fused (SiO ₂)	1.5
Zn	113	S ₃ N ₄	20
W	178	Alumina (Al ₂ O ₃)	30
Al	250	Sapphire (Al ₂ O ₃)	37
Cu	390	Beryllium (BeO)	260
Ag	420	Diamond	~1000
Metal alloys		Polymers	
Stainless steel	12–16	Polypropylene	0.12
55% Cu–45% Ni	19.5	PVC	0.17
70% Ni–30% Cu	25	Polycarbonate	0.22
1080 steel	50	Nylon 6,6	0.24
Bronze (95% Cu–5% Sn)	80	Teflon	0.25
Brass (63% Cu–37% Zn)	125	Polyethylene, low density	0.3
Dural (95% Al–4% Cu–1% Mg)	147	Polyethylene, high density	0.5

molecular theory, all the atoms would be vibrating and the average vibrational kinetic energy would be proportional to the temperature. Intuitively, as depicted in Figure 2.23, when we heat one end of a crystal, we set up large-amplitude atomic vibrations at this hot end. The springs *couple* the vibrations to neighboring atoms and thus allow the large-amplitude vibrations to propagate, as a **vibrational wave**, to the cooler regions of the crystal. If we were to grab the left-end atom in Figure 2.23 and vibrate it violently, we would be sending vibrational waves down the ball-spring-ball chain. The efficiency of heat transfer depends not only on the efficiency of coupling between the atoms, and hence on the nature of interatomic bonding, but also on how the vibrational waves propagate in the crystal and how they are scattered by crystal imperfections and by their interactions with other vibrational waves; this topic is discussed in Chapter 4. The stronger the coupling, the greater will be the thermal conductivity, a trend that is intuitive but also borne out by experiments. Diamond has an exceptionally strong covalent bond and also has a very high thermal conductivity; $\kappa \approx 1000$ W m⁻¹ K⁻¹. On the other hand, polymers have weak secondary bonding between the polymer chains and their thermal conductivities are very poor; $\kappa < 1$ W m⁻¹ K⁻¹.

The thermal conductivity, in general, depends on the temperature. Different classes of materials exhibit different κ values and also different κ versus T behavior. Table 2.5 summarizes κ at room temperature for various classes of materials. Notice how ceramics have a very large range of κ values.

THERMAL CONDUCTIVITY A 95/5 (95% Cu–5% Sn) bronze bearing made of powdered metal contains 15% (vol.%) porosity. Calculate its thermal conductivity at 300 K, given that the electrical conductivity of 95/5 bronze is 10^7 Ω⁻¹ m⁻¹.

EXAMPLE 2.20

SOLUTION

Recall that in Example 2.15, we found the electrical resistivity of the same bronze by using the mixture rule in Equation 2.32 in Section 2.4. We can use the same mixture rule again here, but we need the thermal conductivity of 95/5 bronze. From $\kappa/\sigma T = C_{WFL}$, we have

$$\kappa = \sigma T C_{WFL} = (1 \times 10^7)(300)(2.44 \times 10^{-8}) = 73.2 \text{ W m}^{-1} \text{ K}^{-1}$$

Thus, the effective thermal conductivity is

$$\frac{1}{\kappa_{\text{eff}}} = \frac{1}{\kappa_c} \left[\frac{1 + \frac{1}{2}\chi_d}{1 - \chi_d} \right] = \frac{1}{(73.2 \text{ W m}^{-1} \text{ K}^{-1})} \left[\frac{1 + \frac{1}{2}(0.15)}{1 - 0.15} \right]$$

so that

$$\kappa_{\text{eff}} = 57.9 \text{ W m}^{-1} \text{ K}^{-1}$$

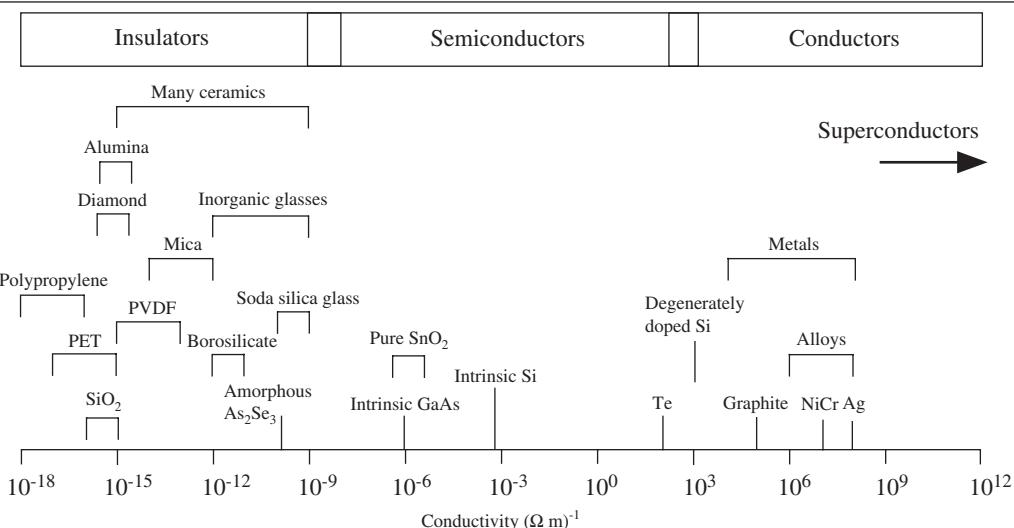


Figure 2.25 Range of conductivities exhibited by various materials.

DEFINING TERMS

Alloy is a metal that contains more than one element.

Brass is a copper-rich Cu–Zn alloy.

Bronze is a copper-rich Cu–Sn alloy.

Drift mobility is the drift velocity per unit applied field. If μ_d is the drift mobility, then the defining equation is $v_d = \mu_d E$, where v_d is the drift velocity and E is the field.

Drift velocity is the average electron velocity, over all the conduction electrons in the conductor, in the direction of an applied electrical force ($F = -eE$ for electrons). In the absence of an applied field, all the electrons move around randomly, and the average velocity over all the electrons in any direction is zero. With an applied field E_x , there is a net velocity per electron v_{dx} , in the direction opposite to the field,

where v_{dx} depends on E_x by virtue of $v_{dx} = \mu_d E_x$, where μ_d is the drift mobility.

Electrical conductivity (σ) is a property of a material that quantifies the ease with which charges flow inside the material along an applied electric field or a voltage gradient. The conductivity is the inverse of electrical resistivity ρ . Since charge flow is caused by a voltage gradient, σ is the rate of charge flow across a unit area per unit voltage gradient, $J = \sigma E$.

Electromigration is current density-induced diffusion of host metal atoms due to their repeated bombardment by conduction electrons at high current densities; the metal atoms migrate in the direction of electron flow.

Black's equation describes the mean time to failure of metal film interconnects due to electromigration failure.

Fourier's law states that the rate of heat flow Q' through a sample, due to thermal conduction, is proportional to the temperature gradient dT/dx and the cross-sectional area A , that is, $Q' = -\kappa A(dT/dx)$, where κ is the thermal conductivity.

Fuchs-Sondheimer equation describes the resistivity of a thin metal film in which scattering from the surfaces of the thin film becomes significant or dominant when the film thickness is comparable or smaller than the mean free path of electrons in the bulk crystal. The resistivity increases with decreasing film thickness.

Hall coefficient (R_H) is a parameter that gauges the magnitude of the Hall effect. If E_y is the electric field in the y direction, due to a current density J_x along x and a magnetic field B_z along z , then $R_H = E_y/J_x B_z$.

Hall effect is a phenomenon that occurs in a conductor carrying a current when the conductor is placed in a magnetic field perpendicular to the current. The charge carriers in the conductor are deflected by the magnetic field, giving rise to an electric field (Hall field) that is perpendicular to both the current and the magnetic field. If the current density J_x is along x and the magnetic field B_z is along z , then the Hall field is along either $+y$ or $-y$, depending on the polarity of the charge carriers in the material.

Heterogeneous mixture is a mixture in which the individual components remain physically separate and possess different chemical and physical properties; that is, a mixture of different phases.

Homogeneous mixture is a mixture of two or more chemical species in which the chemical properties (e.g., composition) and physical properties (e.g., density, heat capacity) are uniform throughout. A homogeneous mixture is a solution.

Interconnects are various thin metal conductors in a Si integrated circuit that connect various devices to implement the required wiring of the devices. In modern ICs, these interconnects are primarily electrode-deposited Cu films.

Ionic conduction is the migration of ions in the material as a result of field-directed diffusion. When a positive ion in an interstitial site jumps to a neighboring interstitial site in the direction of the field, it lowers its potential energy which is a favorable process. If it

jumps in the opposite direction, then it has to do work against the force of the field which is undesirable. Thus the diffusion of the positive ion is directed along the field.

Isomorphous phase diagram is a phase diagram for an alloy that has unlimited solid solubility.

Joule's law relates the power dissipated per unit volume P_{vol} by a current-carrying conductor to the applied field E and the current density J , such that $P_{\text{vol}} = JE = \sigma E^2$.

Lorentz force is the force experienced by a moving charge in a magnetic field. When a charge q is moving with a velocity \mathbf{v} in a magnetic field \mathbf{B} , the charge experiences a force \mathbf{F} that is proportional to the magnitude of its charge q , its velocity \mathbf{v} , and the field \mathbf{B} , such that $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$.

Magnetic field, magnetic flux density, or magnetic induction (\mathbf{B}) is a vector field quantity that describes the magnitude and direction of the *magnetic force* exerted on a moving charge or a current-carrying conductor. The magnetic force is essentially the Lorentz force and excludes the electrostatic force qE .

Magnetic permeability (μ) or simply permeability is a property of the medium that characterizes the effectiveness of a medium in generating as much magnetic field as possible for given external currents. It is the product of the permeability of free space (vacuum) or absolute permeability (μ_0) and relative permeability of the medium (μ_r), i.e., $\mu = \mu_0 \mu_r$.

Magnetometer is an instrument for measuring the magnitude of a magnetic field.

Matthiessen's rule gives the overall resistivity of a metal as the sum of individual resistivities due to scattering from thermal vibrations, impurities, and crystal defects. If the resistivity due to scattering from thermal vibrations is denoted ρ_T and the resistivities due to scattering from crystal defects and impurities can be lumped into a single resistivity term called the residual resistivity ρ_R , then $\rho = \rho_T + \rho_R$.

Mayadas-Shatzkes formula describes the resistivity of a thin metal film in which grain boundary scattering becomes significant or dominant; and the grain size is comparable or smaller than the mean free path of electrons in the bulk crystal. The resistivity increases with decreasing grain size.

Mean free path is the mean distance traversed by an electron between scattering events. If τ is the mean free time between scattering events and u is the mean speed of the electron, then the mean free path is $\ell = u\tau$.

Mean free time is the average time it takes to scatter a conduction electron. If t_i is the free time between collisions (between scattering events) for an electron labeled i , then $\tau = \bar{t}_i$ averaged over all the electrons. The drift mobility is related to the mean free time by $\mu_d = e\tau/m_e$. The reciprocal of the mean free time is the mean probability per unit time that a conduction electron will be scattered; in other words, the mean frequency of scattering events.

Nordheim's rule states that the resistivity of a solid solution (an isomorphous alloy) due to impurities ρ_I is proportional to the concentrations of the solute X and the solvent ($1 - X$).

Phase (in materials science) is a physically homogeneous portion of a materials system that has uniform physical and chemical characteristics.

Relaxation time is an equivalent term for the mean free time between scattering events.

Residual resistivity (ρ_R) is the contribution to the resistivity arising from scattering processes other than thermal vibrations of the lattice, for example, impurities, grain boundaries, dislocations, point defects.

Skin effect is an electromagnetic phenomenon that, at high frequencies, restricts ac current flow to near the surface of a conductor to reduce the energy stored in the magnetic field.

Solid solution is a crystalline material that is a homogeneous mixture of two or more chemical species. The mixing occurs at the atomic scale, as in mixing alcohol and water. Solid solutions can be substitutional (as in Cu–Ni) or interstitial (for example, C in Fe).

Stefan's law is a phenomenological description of the energy radiated (as electromagnetic waves) from a surface per second. When a surface is heated to a temperature T , it radiates net energy at a rate given by $P_{\text{radiated}} = \epsilon\sigma_s A(T^4 - T_0^4)$, where σ_s is Stefan's constant ($5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$), ϵ is the emissivity of the surface, A is the surface area, and T_0 is the ambient temperature.

Temperature coefficient of resistivity (TCR) (α_0) is defined as the fractional change in the electrical resistivity of a material per unit increase in the temperature with respect to some reference temperature T_0 .

Thermal conductivity (κ) is a property of a material that quantifies the ease with which heat flows along the material from higher to lower temperature regions. Since heat flow is due to a temperature gradient, κ is the rate of heat flow across a unit area per unit temperature gradient.

Thermal resistance (θ) is a measure of the difficulty with which heat conduction takes place along a material sample. The thermal resistance is defined as the temperature drop per unit heat flow, $\theta = \Delta T/Q'$. It depends on both the material and its geometry. If the heat losses from the surfaces are negligible, then $\theta = L/\kappa A$, where L is the length of the sample (along heat flow) and A is the cross-sectional area.

Thermally activated conductivity means that the conductivity increases in an exponential fashion with temperature as in $\sigma = \sigma_o \exp(-E_\sigma/kT)$ where E_σ is the activation energy.

Thin film is a conductor whose thickness is typically less than ~ 1 micron; the thickness is also much less than the width and length of the conductor. Typically thin films have a higher resistivity than the corresponding bulk material due to the grain boundary and surface scattering.

QUESTIONS AND PROBLEMS

- 2.1 **Electrical conduction** Na is a monovalent metal (BCC) with a density of 0.9712 g cm^{-3} . Its atomic mass is 22.99 g mol^{-1} . The drift mobility of electrons in Na is $53 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.
- Consider the collection of conduction electrons in the solid. If each Na atom donates one electron to the electron sea, estimate the mean separation between the electrons. (Note: If n is the concentration of particles, then the particles' mean separation $d = 1/n^{1/3}$.)

- b. Estimate the mean separation between an electron (e^-) and a metal ion (Na^+), assuming that most of the time the electron prefers to be between two neighboring Na^+ ions. What is the approximate Coulombic interaction energy (in eV) between an electron and an Na^+ ion?
- c. How does this electron/metal-ion interaction energy compare with the average thermal energy per particle, according to the kinetic molecular theory of matter? Do you expect the kinetic molecular theory to be applicable to the conduction electrons in Na? If the mean electron/metal-ion interaction energy is of the same order of magnitude as the mean KE of the electrons, what is the mean speed of electrons in Na? Why should the mean kinetic energy be comparable to the mean electron/metal-ion interaction energy?
- d. Calculate the electrical conductivity of Na and compare this with the experimental value of $2.1 \times 10^7 \Omega^{-1} \text{ m}^{-1}$ and comment on the difference.
- 2.2 Electrical conduction** The resistivity of aluminum at 25°C has been measured to be $2.72 \times 10^{-8} \Omega \text{ m}$. The thermal coefficient of resistivity of aluminum at 0°C is $4.29 \times 10^{-3} \text{ K}^{-1}$. Aluminum has a valency of 3, a density of 2.70 g cm^{-3} , and an atomic mass of 27.
- Calculate the resistivity of aluminum at -40°C .
 - What is the thermal coefficient of resistivity at -40°C ?
 - Estimate the mean free time between collisions for the conduction electrons in aluminum at 25°C , and hence estimate their drift mobility.
 - If the mean speed of the conduction electrons is about $2.0 \times 10^6 \text{ m s}^{-1}$, calculate the mean free path and compare this with the interatomic separation in Al (Al is FCC). What should be the thickness of an Al film that is deposited on an IC chip such that its resistivity is the same as that of bulk Al?
 - What is the percentage change in the power loss due to Joule heating of the aluminum wire when the temperature drops from 25°C to -40°C ?

- 2.3 Conduction in gold** Gold is in the same group as Cu and Ag. Assuming that each Au atom donates one conduction electron, calculate the drift mobility of the electrons in gold at 22°C . What is the mean free path of the conduction electrons if their mean speed is $1.4 \times 10^6 \text{ m s}^{-1}$? (Use ρ_o and α_o in Table 2.1.)

- 2.4 Mean free time between collisions** Let $1/\tau$ be the mean probability per unit time that a conduction electron in a metal collides with (or is scattered by) lattice vibrations, impurities, or defects, etc. Then the probability that an electron makes a collision in a small time interval δt is $\delta t/\tau$. Suppose that $n(t)$ is the concentration of electrons that have not yet collided. The change δn in the uncollided electron concentration is then $-n\delta t/\tau$. Thus, $\delta n = -n\delta t/\tau$, or $\delta n/n = -\delta t/\tau$. We can integrate this from $n = n_o$ at $t = 0$ to $n = n(t)$ at time t to find the concentration of uncollided electrons $n(t)$ at t

$$n(t) = n_o \exp(-t/\tau) \quad [2.84]$$

Show that the mean free time and mean square free time are given by

$$\bar{\tau} = \frac{\int_0^\infty t n(t) dt}{\int_0^\infty n(t) dt} = \tau \quad \text{and} \quad \bar{\tau}^2 = \frac{\int_0^\infty t^2 n(t) dt}{\int_0^\infty n(t) dt} = 2\tau^2 \quad [2.85]$$

What is your conclusion?

Concentration
of uncollided
electrons

Electron
scattering
statistics

- 2.5 Effective number of conduction electrons per atom**
- Electron drift mobility in tin (Sn) is $3.9 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. The room temperature (20°C) resistivity of Sn is about $110 \text{ n}\Omega \text{ m}$. Atomic mass M_{at} and density of Sn are $118.69 \text{ g mol}^{-1}$ and 7.30 g cm^{-3} , respectively. How many “free” electrons are donated by each Sn atom in the crystal? How does this compare with the position of Sn in Group IVB of the Periodic Table?
 - Consider the resistivity of few selected metals from Groups I to IV in the Periodic Table in Table 2.8. Calculate the number of conduction electrons contributed per atom and compare this with the location of the element in the Periodic Table. What is your conclusion?

Table 2.8 Selection of metals from Groups I to IV in the Periodic Table

Metal	Periodic Group	Valency	Density (g cm ⁻³)	Resistivity (nΩ m)	Mobility (cm ² V ⁻¹ s ⁻¹)
Na	IA	1	0.97	42.0	53
Mg	IIA	2	1.74	44.5	17
Ag	IB	1	10.5	15.9	56
Zn	IIB	2	7.14	59.2	8
Al	IIIB	3	2.7	26.5	12
Sn	IVB	4	7.30	110	3.9
Pb	IVB	4	11.4	206	2.3

| NOTE: Mobility from Hall-effect measurements.

- 2.6 Resistivity of Ta** Consider the resistivity of tantalum, which is summarized in Table 2.9. Plot ρ against T on a log–log plot and find n for the behavior $\rho \propto T^n$. Find the TCR at 0 and 25 °C. What is your conclusion?

Table 2.9 Resistivity of Ta

T (K)	200	273	293	298	300	400	500	600	700	800	900
ρ (nΩ m)	86.6	122	131	134	135	182	229	274	318	359	401

| SOURCE: Ed. Haynes, W.M., *CRC Handbook of Chemistry and Physics*, 96th Edition, 2015-2016. CRC Press.

- 2.7 TCR of isomorphous alloys** Determine the composition of the Cu–Ni alloy that will have a TCR of 4×10^{-4} K⁻¹, that is, a TCR that is an order of magnitude less than that of Cu. Over the composition range of interest, the resistivity of the Cu–Ni alloy can be calculated from $\rho_{\text{CuNi}} \approx \rho_{\text{Cu}} + C_{\text{eff}}X(1 - X)$, where C_{eff} , the effective Nordheim coefficient, is about 1310 nΩ m.
- 2.8 Resistivity of isomorphous alloys and Nordheim's rule** What are the maximum atomic and weight percentages of Cu that can be added to Au without exceeding a resistivity that is twice that of pure gold? What are the maximum atomic and weight percentages of Au that can be added to pure Cu without exceeding twice the resistivity of pure copper? (Alloys are normally prepared by mixing the elements in weight.)
- 2.9 Physical properties of alloys** Consider Cu–Sn alloys, called phosphor bronzes. Their properties are listed in Table 2.10 from the ASM Handbook. Plot these properties all in graph (using a log-scale for the properties axis) as a function of composition and deduce conclusions. How does κ/σ change? Compositions are wt %. Assume that Cu–Sn is a solid solution over this composition range.

Table 2.10 Selected properties of Cu with Sn at 20 °C

	ρ nΩ m	κ W m ⁻¹ K ⁻¹	c_s J kg ⁻¹ K ⁻¹	λ $\times 10^{-6}$ K ⁻¹	E GPa	d g cm ⁻³
Cu	17.1	391	385	17.0	115	8.94
98.7Cu-1.35Sn	36	208	380	17.8	117	8.89
92Cu-8Sn	133	62	380	18.2	110	8.80
90Cu-10Sn	157	50	380	18.4	110	8.78

| NOTE: ρ is resistivity, κ is thermal conductivity, c_s is specific heat capacity, λ is linear thermal expansion coefficient, E is Young's modulus and d is density.

- 2.10 Nordheim's rule and brass** Brass is a Cu–Zn alloy. Table 2.11 shows some typical resistivity values for various Cu–Zn compositions in which the alloy is a solid solution (up to 30% Zn).

- Plot ρ versus $X(1 - X)$. From the slope of the best-fit line find the mean (effective) Nordheim coefficient \bar{C} for Zn dissolved in Cu over this compositional range.
- Since X is the atomic fraction of Zn in brass, for each atom in the alloy, there are X Zn atoms and $(1 - X)$ Cu atoms. The conduction electrons consist of each Zn donating two electrons and each copper donating one electron.²⁷ Thus, there are $2(X) + 1(1 - X) = 1 + X$ conduction electrons per atom. Since the conductivity is proportional to the electron concentration, the combined Nordheim–Matthiessens rule must be scaled up by $(1 + X)$,

$$\rho_{\text{brass}} = \frac{\rho_o + CX(1 - X)}{(1 + X)}$$

Plot the data in Table 2.11 as $\rho(1 + X)$ versus $X(1 - X)$. From the best-fit line find C and ρ_o . What is your conclusion? (Compare the correlation coefficients of the best-fit lines in your two plots.)

Table 2.11 Cu–Zn brass alloys

Zn at.% in Cu–Zn	0	0.34	0.5	0.93	3.06	4.65	9.66	15.6	19.59	29.39
Resistivity nΩ m	17	18.1	18.84	20.7	26.8	29.9	39.1	49.0	54.8	63.5

| SOURCE: Fairbank, H.A., *Physical Review*, 66, 274, 1944.

- 2.11 Resistivity of solid solution metal alloys: testing Nordheim's rule** Nordheim's rule accounts for the increase in the resistivity resulting from the scattering of electrons from the random distribution of impurity (solute) atoms in the host (solvent) crystal. It can nonetheless be quite useful in approximately predicting the resistivity at one composition of a solid solution metal alloy, given the value at another composition. Table 2.12 lists some solid solution metal alloys and gives the resistivity ρ at one composition X and asks for a prediction ρ' based on Nordheim's rule at another composition X' . Fill in the table for ρ' and compare the predicted values with the experimental values, and comment.

Table 2.12 Resistivities of some solid solution metal alloys

	Alloy							
	Ag–Au	Au–Ag	Cu–Pd	Ag–Pd	Au–Pd	Pd–Pt	Pt–Pd	Cu–Ni
X (at.%)	8.8% Au	8.77% Ag	6.2% Pd	10.1% Pd	8.88% Pd	7.66% Pt	7.1% Pd	2.16% Ni
ρ_o (nΩ m)	16.2	22.7	17	16.2	22.7	108	105.8	17
ρ at X (nΩ m)	44.2	54.1	70.8	59.8	54.1	188.2	146.8	50
C_{eff}								
X'	15.4% Au	24.4% Ag	13% Pd	15.2% Pd	17.1% Pd	15.5% Pt	13.8% Pd	23.4% Ni
ρ' at X' (nΩ m)								
ρ' at X' (nΩ m)	66.3	107.2	121.6	83.8	82.2	244	181	300
Experimental								

| NOTE: First symbol (e.g., Ag in AgAu) is the matrix (solvent) and the second (Au) is the added solute. X is in at.%, converted from traditional weight percentages reported with alloys. C_{eff} is the effective Nordheim coefficient in $\rho = \rho_o + C_{\text{eff}} X(1 - X)$.

²⁷ The approach in Question 2.10 is an empirical and a classical way to try and account for the fact that as the Zn concentration increases, the resistivity does not increase at a rate demanded by the Nordheim equation. An intuitive correction is then done by increasing the conduction electron concentration with Zn, based on valency. There is, however, a modern physics explanation that involves not only scattering from the introduction of impurities (Zn), but also changes in something called the “Fermi surface and density of states at the Fermi energy”, which can be found in advanced solid state physics textbooks.

- *2.12 TCR and alloy resistivity** Table 2.13 shows the resistivity and TCR (α) of Cu–Ni alloys. Plot TCR versus $1/\rho$, and obtain the best-fit line. What is your conclusion? Consider the Matthiessen rule, and explain why the plot should be a straight line. What is the relationship between ρ_{Cu} , α_{Cu} , ρ_{CuNi} , and α_{CuNi} ? Can this be generalized?

Table 2.13 Cu–Ni alloys, resistivity, and TCR

	0	2	6	11	20
Resistivity (n Ω m)	17	50	100	150	300
TCR (ppm $^{\circ}\text{C}^{-1}$)	4270	1350	550	430	160

| NOTE: ppm-parts per million, i.e., 10^{-6} .

- 2.13 Hall effect measurements** The resistivity and the Hall coefficient of pure aluminum and Al with 1 at.% Si have been measured at 20 °C (293 K) as $\rho = 2.65 \mu\Omega \text{ cm}$, $R_H = -3.51 \times 10^{-11} \text{ m}^3 \text{ C}^{-1}$ for Al and $\rho = 3.33 \mu\Omega \text{ cm}$, $R_H = -3.16 \times 10^{-11} \text{ m}^3 \text{ C}^{-1}$ for 99 at.% Al-1 at% Si. The lattice parameters for the pure metal and the alloy are 0.4049 nm and 0.4074 nm. What does the simple Drude model predict for the drift mobility in these two metals? How many conduction electrons are there per atom? (Data from M Bradley and John Stringer, J. Phys. F: Metal Phys., 4, 839, 1974).

- 2.14 Hall effect and the Drude model** Table 2.14 shows the experimentally measured Hall coefficient and resistivities for various metals and their position in the periodic table. (a) Calculate the Hall mobility of each element. (b) Calculate the conduction electron concentration from the experimental value of R_H . (c) Find how many electrons per atom are contributed to the conduction electron gas in the metal per metal atom. What is your conclusion?

Table 2.14 Measured Hall coefficients for a few metals at 25 °C

	Li	Na	K	Cs	Cu	Ag	Au	Ca	Mg	Zn	Al	In
Group	I	I	I	I	IB	IB	IB	IIA	IIA	IIB	III	III
$R_H (\times 10^{-11} \text{ m}^3 \text{ C}^{-1})$	-15	-24.8	-42.8	-73.3	-5.4	-9.0	-7.2	-17.8	-8.3	+10.4	-3.4	-0.73
ρ (n Ω m)	92.8	48.8	73.9	208	17.1	16.7	22.6	33.6	44.8	60.1	27.1	83.7

| SOURCE: Hurd, C., *The Hall Coefficient of Metals and Alloys*, Plenum, New York, NY, 1972, along with other sources.

- 2.15 The Hall effect** Consider a rectangular sample, a metal or an n -type semiconductor, with a length L , width W , and thickness D . A current I is passed along L , perpendicular to the cross-sectional area WD . The face $W \times L$ is exposed to a magnetic field density B . A voltmeter is connected across the width, as shown in Figure 2.40, to read the Hall voltage V_H .

- a. Show that the Hall voltage recorded by the voltmeter is

Hall voltage

$$V_H = \frac{IB}{Den}$$

- b. Consider a 1-micron-thick strip of gold layer on an insulating substrate that is a candidate for a Hall probe sensor. If the current through the film is maintained at constant 100 mA, what is the magnetic field that can be recorded per μV of Hall voltage?

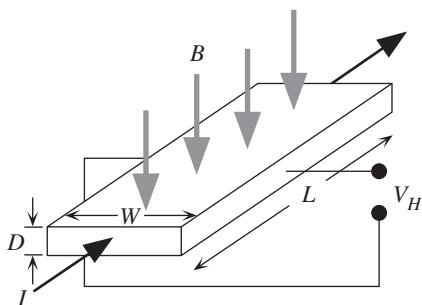


Figure 2.40 Hall effect in a rectangular material with length L , width W , and thickness D . The voltmeter is across the width W .

- 2.16 Electrical and thermal conductivity of In** Electron drift mobility in indium has been measured to be $6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. The room temperature (27°C) resistivity of In is $8.37 \times 10^{-8} \Omega\text{m}$, and its atomic mass and density are 114.82 amu or g mol^{-1} and 7.31 g cm^{-3} , respectively.
- Based on the resistivity value, determine how many free electrons are donated by each In atom in the crystal. How does this compare with the position of In in the Periodic Table (Group IIIB)?
 - If the mean speed of conduction electrons in In is $1.74 \times 10^8 \text{ cm s}^{-1}$, what is the mean free path?
 - Calculate the thermal conductivity of In. How does this compare with the experimental value of $81.6 \text{ W m}^{-1} \text{ K}^{-1}$?
- 2.17 Electrical and thermal conductivity of Ag** The electron drift mobility in silver has been measured to be $54 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at 27°C . The atomic mass and density of Ag are given as 107.87 amu or g mol^{-1} and 10.50 g cm^{-3} , respectively.
- Assuming that each Ag atom contributes one conduction electron, calculate the resistivity of Ag at 27°C . Compare this value with the measured value of $1.6 \times 10^{-8} \Omega\text{m}$ at the same temperature and suggest reasons for the difference.
 - Calculate the thermal conductivity of silver at 27°C and at 0°C .
- 2.18 Mixture rules** A 70% Cu–30% Zn brass electrical component has been made of powdered metal and contains 15 vol.% porosity. Assume that the pores are dispersed randomly. Given that the resistivity of 70% Cu–30% Zn brass is $62 \text{ n}\Omega \text{ m}$, calculate the effective resistivity of the brass component using the simple conductivity mixture rule, Equation 2.32, and the Reynolds and Hough rule.
- 2.19 Mixture rules**
- A certain carbon electrode used in electrical arcing applications is 47 percent porous. Given that the resistivity of graphite (in polycrystalline form) at room temperature is about $9.1 \mu\Omega \text{ m}$, estimate the effective resistivity of the carbon electrode using the appropriate Reynolds and Hough rule and the simple conductivity mixture rule. Compare your estimates with the measured value of $18 \mu\Omega \text{ m}$ and comment on the differences.
 - Silver particles are dispersed in a graphite paste to increase the effective conductivity of the paste. If the volume fraction of dispersed silver is 50 percent, what is the effective conductivity of this paste?
- 2.20 Ag–Ni alloys (contact materials) and the mixture rules** Silver alloys, particularly Ag alloys with the precious metals Pt, Pd, Ni, and Au, are extensively used as contact materials in various switches. Alloying Ag with other metals generally increases the hardness, wear resistance, and corrosion resistance at the expense of electrical and thermal conductivity. For example, Ag–Ni alloys are widely used as contact materials in switches in domestic appliances, control and selector switches, circuit breakers, and automotive switches up to several hundred amperes of current. Table 2.15 shows the resistivities of four Ag–Ni alloys used in make-and-break as well as disconnect contacts with current ratings up to $\sim 100 \text{ A}$.

Table 2.15 Resistivity of Ag–Ni contact alloys for switches

	Ni % in Ag–Ni alloy						
	0	10	15	20	30	40	100
d (g cm ⁻³)	10.49	10.25	10.15	10.05	9.8	9.7	8.91
ρ (nΩ m)	16.9	18.7	19.0	20.0	24.4	27.0	71.0

NOTE: Compositions are in wt.%. Ag–10% Ni means 90% Ag–10% Ni. d = density and ρ = resistivity. Use volume fraction of Ni = $w_{\text{Ni}}(d_{\text{alloy}}/d_{\text{Ni}})$, where w_{Ni} is the Ni weight fraction, to convert wt.% to volume %. Data combined from various sources.

- a. Ag–Ni is a two-phase alloy, a mixture of Ag-rich and Ni-rich phases. Using an appropriate mixture rule, predict the resistivity of the alloy and compare with the measured values in Table 2.15. Explain the difference between the predicted and experimental values.
- b. Compare the resistivity of Ag–10% Ni with that of Ag–10% Pd in Table 2.12. The resistivity of the Ag–Pd alloy is almost a factor of 3 greater. Ag–Pd is an isomorphous solid solution, whereas Ag–Ni is a two-phase mixture. Explain the difference in the resistivities of Ag–Ni and Ag–Pd.

2.21 Ag–W alloys (contact materials) and the mixture rule Silver–tungsten alloys are frequently used in heavy-duty switching applications (*e.g.*, current-carrying contacts and oil circuit breakers) and in arcing tips. Ag–W is a two-phase alloy, a mixture of Ag-rich and W-rich phases. The measured resistivity and density for various Ag–W compositions are summarized in Table 2.16.

- a. Plot the resistivity and density of the Ag–W alloy against the W content (wt.%)
- b. Show that the density of the mixture, d , is given by

$$d^{-1} = w_{\alpha}d_{\alpha}^{-1} + w_{\beta}d_{\beta}^{-1}$$

where w_{α} is the weight fraction of phase α , w_{β} is the weight fraction of phase β , d_{α} is the density of phase α , and d_{β} is the density of phase β .

- c. Show that the resistivity mixture rule is

$$\rho = \rho_{\alpha} \frac{dw_{\alpha}}{d_{\alpha}} + \rho_{\beta} \frac{dw_{\beta}}{d_{\beta}}$$

where ρ is the resistivity of the alloy (mixture), d is the density of the alloy (mixture), and subscripts α and β refer to phases α and β , respectively. Calculate d and plot it in *a* above.

- d. Calculate the density d and the resistivity ρ of the mixture for various values of W content (in wt.%) and plot the calculated values in the same graph as the experimental values. Use the Reynolds-Hough rule for mixtures in Equation 2.34. What is your conclusion?

Table 2.16 Dependence of resistivity in Ag–W alloy on composition as a function of wt.% W

	W (wt.%)												
	0	10	15	20	30	40	65	70	75	80	85	90	100
ρ (nΩ m)	16.2	18.6	19.7	20.9	22.7	27.6	35.5	38.3	40	46	47.9	53.9	55.6
d (g cm ⁻³)	10.5	10.75	10.95	11.3	12.0	12.35	14.485	15.02	15.325	16.18	16.6	17.25	19.1

I NOTE: ρ = resistivity and d = density.

- 2.22 Strain gauges** Consider a strain gauge that consists of a nichrome wire of resistivity $1100 \text{ n}\Omega \text{ m}$, TCR (α) = 0.0004 K^{-1} , a total length of 35 cm, and a diameter of $20 \mu\text{m}$. What is δR for a strain of 10^{-3} ? For nichrome, $\nu \approx 0.3$. What is δR if there is a temperature variation of $1 \text{ }^\circ\text{C}$, given that the linear thermal expansion coefficient is 15 ppm K^{-1} ?
- 2.23 Strain measurements** How would you use strain gauges in a Wheatstone bridge circuit to measure strains and reduce the effects of temperature variations? What would be the advantages and disadvantages of such a bridge circuit?
- 2.24 Strain gauges** Suppose you wish to construct a strain gauge from constantan, which is 55%Cu-45%Ni alloy. Constantan has a resistivity of $500 \text{ n}\Omega \text{ m}$, TCR (α) of $8 \times 10^{-6} \text{ K}^{-1}$, linear thermal expansion coefficient (λ) of $14.9 \times 10^{-6} \text{ K}^{-1}$, and a Poisson ratio ν of 0.3. Suppose that the strain gauge uses 50 cm of wire and the diameter is $5 \mu\text{m}$. What is δR for a strain of 10^{-3} ? What is δR if there is a temperature variation of $1 \text{ }^\circ\text{C}$?
- 2.25 Strain gauges** Consider the derivation of Equation 2.26 for metal strain gauges. Is the equation the same if the cross section that is a rectangle with dimensions $a \times b$ instead of a circular area of diameter D ? Does this equation depend on the shape of the cross section? What would be the advantage of using a gauge made from thin film strips on a carrier substrate that could be bonded to the structure under test? How important is the substrate in strain measurements?
- 2.26 Thermal coefficients of expansion and resistivity**

- a. Consider a thin metal wire of length L and diameter D . Its resistance is $R = \rho L/A$, where $A = \pi D^2/4$. By considering the temperature dependence of L , A , and ρ individually, show that

$$\frac{1}{R} \frac{dR}{dT} = \alpha_0 - \lambda_0$$

Change in R with temperature

where α_0 is the temperature coefficient of resistivity (TCR), and λ_0 is the temperature coefficient of linear expansion (thermal expansion coefficient or expansivity), that is,

$$\lambda_0 = L_0^{-1} \left(\frac{dL}{dT} \right)_{T=T_0} \quad \text{or} \quad \lambda_0 = D_0^{-1} \left(\frac{dD}{dT} \right)_{T=T_0}$$

Note: Consider differentiating $R = \rho L/[(\pi D^2)/4]$ with respect to T with each parameter, ρ , L , and D , having a temperature dependence.

Given that typically, for most pure metals, $\alpha_0 \approx 1/273 \text{ K}^{-1}$ and $\lambda_0 \approx 2 \times 10^{-5} \text{ K}^{-1}$, confirm that the temperature dependence of ρ controls R , rather than the temperature dependence of the geometry. Is it necessary to modify the given equation for a wire with a noncircular cross section?

- b. Is it possible to design a resistor from a suitable alloy such that its temperature dependence is almost nil? Consider the TCR of an alloy of two metals A and B , for which $\alpha_{AB} \approx \alpha_A \rho_A / \rho_{AB}$.
- 2.27 Thermal conduction** Consider brass alloys with an X atomic fraction of Zn. These alloys form a solid solution up to 30 at.%, and we can use the combined Matthiessen-Nordheim rule in Equation 2.21 to calculate the resistivity of the alloy. Take $C = 300 \text{ n}\Omega \text{ m}$ and $\rho_o = \rho_{\text{Cu}} = 17 \text{ n}\Omega \text{ m}$.
- a. An 80 at.% Cu–20 at.% Zn brass disk of 40 mm diameter and 5 mm thickness is used to conduct heat from a heat source to a heat sink.
- (1) Calculate the thermal resistance of the brass disk.
 - (2) If the disk is conducting heat at a rate of 100 W, calculate the temperature drop along the disk.
- b. What should be the composition of brass if the temperature drop across the disk is to be halved?
- 2.28 Thermal resistance** Consider a thin insulating disk made of mica to electrically insulate a semiconductor device from a conducting heat sink. Mica has $\kappa = 0.75 \text{ W m}^{-1} \text{ K}^{-1}$. The disk thickness is 0.1 mm, and the diameter is 10 mm. What is the thermal resistance of the disk? What is the temperature drop across the disk if the heat current through it is 5 W?

- *2.29 Thermal resistance** Consider a coaxial cable operating under steady-state conditions when the current flow through the inner conductor generates Joule heat at a rate $P = I^2R$. The heat generated per second by the core conductor flows through the dielectric; $Q' = I^2R$. The inner conductor reaches a temperature T_i , whereas the outer conductor is at T_o . Show that the thermal resistance θ of the hollow cylindrical insulation for heat flow in the radial direction is

Thermal
resistance of
hollow cylinder

$$\theta = \frac{(T_i - T_o)}{Q'} = \frac{\ln(b/a)}{2\pi\kappa L} \quad [2.86]$$

where a is the inside (core conductor) radius, b is the outside radius (outer conductor), κ is the thermal conductivity of the insulation, and L is the cable length. Consider a coaxial cable that has a copper core conductor and polyethylene (PE) dielectric with the following properties: Core conductor resistivity $\rho = 19 \text{ n}\Omega \text{ m}$, core radius $a = 4 \text{ mm}$, dielectric thickness $b - a = 3.5 \text{ mm}$, dielectric thermal conductivity $\kappa = 0.3 \text{ W m}^{-1} \text{ K}^{-1}$. The outside temperature T_o is 25°C . The cable is carrying a current of 500 A . What is the temperature of the inner conductor?

2.30 Temperature of a light bulb filament

- a. Consider a 100 W , 120 V incandescent bulb (lamp). The tungsten filament has a length of 0.579 m and a diameter of $63.5 \mu\text{m}$. Its resistivity at room temperature is $56 \text{ n}\Omega \text{ m}$. Given that the resistivity of the filament can be represented as

Resistivity of W

$$\rho = \rho_0 \left[\frac{T}{T_0} \right]^n \quad [2.87]$$

where T is the temperature in K, ρ_0 is the resistance of the filament at T_0 K, and $n = 1.24$ (Table 2.1), estimate the temperature of the bulb when it is operated at the rated voltage, that is, directly from the main outlet. Note that the bulb dissipates 100 W at 120 V .

- b. Suppose that the electrical power dissipated in the tungsten wire is totally radiated from the surface of the filament. The radiated power at the absolute temperature T can be described by Stefan's law

Radiated power

$$P_{\text{radiated}} = \epsilon\sigma_s A(T^4 - T_0^4) \quad [2.88]$$

where σ_s is Stefan's constant ($5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$), ϵ is the emissivity of the surface (0.35 for tungsten), A is the surface area of the tungsten filament, and T_0 is room temperature (293 K). Obviously, for $T > T_0$, $P_{\text{radiated}} = \epsilon\sigma_s AT^4$.

Assuming that all the electrical power is radiated from the surface, estimate the temperature of the filament and compare it with your answer in part (a).

- c. If the melting temperature of W is 3407°C , what is the voltage that guarantees that the light bulb will blow?

- 2.31 Superionic conduction in RbAg_4I_5** Figure 2.29 shows that the RbAg_4I_5 (rubidium silver iodide) crystal has a conductivity that is orders of magnitude higher than traditional ceramics and glasses in the same temperature range. Table 2.17 gives the conductivity of RbAg_4I_5 as a function of temperature. By carrying out a suitable plot, find the activation energy $E_a(\text{eV})$ and the pre-exponential constant A in the expression for ionic conduction, $\sigma = (A/T)\exp(-E_a/kT)$.

Table 2.17 Conductivity versus temperature data for a RbAg_4I_5 crystal

$T (\text{ }^\circ\text{C})$	25	27	34	51	56	65	75	77
$\sigma (\Omega^{-1} \text{ cm}^{-1})$	0.288	0.304	0.322	0.339	0.371	0.395	0.427	0.434
$T (\text{ }^\circ\text{C})$	87	89	92	107	121	132	134	147
$\sigma (\Omega^{-1} \text{ cm}^{-1})$	0.455	0.465	0.477	0.527	0.55	0.581	0.608	0.659

I SOURCE: Kim, K.S., and Piak, W., *Journal of Chemical & Engineering Data*, 20, 356, 1975.

- 2.32 Hall effect with ions in ionic crystals** By using various sensitive measurement techniques, it is possible to carry out Hall effect measurements on certain ionic crystals. Stuhrmann, Kreiterling and Funke in 2002 (Solid State Ionics, 154, 109) were able to measure the Hall voltage on superionic RbAg_4I_5 crystals in a magnetic field. The results at 100 °C indicate that the Hall coefficient is $5.7 \times 10^{-4} \text{ cm}^3 \text{ C}^{-1}$. The conductivity of the sample at the same temperature is $0.53 \Omega^{-1} \text{ cm}^{-1}$. The mobile charges are Ag^+ ions. What is the Hall mobility of Ag^+ ions? The Ag^+ concentration in the crystal can be estimated from the density of the crystal ($d = 5.35 \text{ g cm}^{-3}$) and is approximately $1.1 \times 10^{22} \text{ cm}^{-3}$. Assuming that all the ions are moving, what should be the drift mobility of Ag^+ ions at 100 °C? What is your conclusion?
- 2.33 Ionic conduction in soda-silicate glasses** Consider soda-silica glass of composition 25% Na_2O -75% SiO_2 that represents $(\text{Na}_2\text{O})_{0.25}(\text{SiO}_2)_{0.75}$. Its density is 2.39 g cm^{-3} . The diffusion coefficient D of Na^+ in this soda-silica at 350 °C is $3.38 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ and the Haven ratio f is 0.53. Calculate the conductivity of 25% Na_2O -75% SiO_2 glass at 350 °C and compare it with the value deduced from Figure 2.29.
- 2.34 Ionic conduction in borosilicate glasses** Table 2.18 shows the conductivities of four types of borosilicate glass identified as samples L, N, K, and C where L is 53.4 SiO_2 -25.8 B_2O_3 -20.8 Li_2O , N is 53.5 SiO_2 -26.1 B_2O_3 -20.4 Na_2O , K is 55.1 SiO_2 -25.8 B_2O_3 -19.1 K_2O , and C is 58.1 SiO_2 -24.7 B_2O_3 -17.2 Na_2O . The numbers represent molar percentages, *i.e.*, 55.1% SiO_2 , etc. The main difference between the samples is the alkaline ion species: L has Li^+ , N has Na^+ , K has K^+ , and C has Cs^+ mobile ions.
- Find the constant A , the activation energy E_σ for each sample. Plot E_σ versus the alkaline ion radius.
 - Calculate and compare the conductivities at the same temperature, say at 400 °C. Which are lower? Why? Plot semilogarithmically σ at 400 °C vs. ionic radius.
 - Find approximately the temperature for each glass so that all four glasses at this temperature have the same conductivity of $\sigma = 8.00 \times 10^{-6} \Omega^{-1} \text{ cm}^{-1}$. For example, T is 235 °C for glass L. What is your conclusion?

Table 2.18 Selected conductivities and properties of borosilicate glasses with different alkaline ions

Sample	Mobile Ion	Ionic Radius (nm)	σ_1 at T_1 $\Omega^{-1} \text{ cm}^{-1}$	σ_2 at T_2 $\Omega^{-1} \text{ cm}^{-1}$
L	Li^+	0.061	9.18×10^{-6} at 240 °C	8.86×10^{-4} at 490 °C
N	Na^+	0.086	1.54×10^{-7} at 190 °C	2.34×10^{-4} at 500 °C
K	K^+	0.139	2.22×10^{-8} at 220 °C	1.25×10^{-4} at 520 °C
C	Cs^+	0.160	5.43×10^{-9} at 230 °C	9.50×10^{-6} at 500 °C

NOTE: Conductivity and ionic radius values from Neyret, M., et al, *Journal of Non-Crystalline Solids*, 410, 74, 2015.

2.35 Skin effect

- What is the skin depth for a solid core copper wire carrying a current at 60 Hz? The resistivity of copper at 27 °C is $17 \text{ n}\Omega \text{ m}$. Its relative permeability $\mu_r \approx 1$. Is there any sense in using a conductor for power transmission which has a diameter more than 2 cm?
- What is the skin depth for a solid core iron wire carrying a current at 60 Hz? The resistivity of iron at 27 °C is $97 \text{ n}\Omega \text{ m}$. Assume that its relative permeability $\mu_r \approx 700$. How does this compare with the copper wire? Discuss why copper is preferred over iron for power transmission even though iron is nearly 100 times cheaper than copper.

- *2.36 Mayadas–Shatzkes thin film resistivity** Consider Equation 2.72 for the resistivity of a polycrystalline thin film in terms of β . Consider the expansion of Equation 2.72 around $\beta = 1$. If $\Delta\beta = \beta - 1$, then show that

$$(\rho_{\text{film}}/\rho_{\text{crystal}}) = 2.378 + 1.3475\Delta\beta + \dots$$

so that

Grain boundary scattering in thin films

$$\frac{\rho_{\text{film}}}{\rho_{\text{crystal}}} \approx 1.030 + 1.348\beta \quad [2.89]$$

Plot the actual expression for $(\rho_{\text{film}}/\rho_{\text{crystal}})$ versus β and then Equations 2.89 and 2.73a versus β and compare the two. What would be a range of values for which Equation 2.89 can be used with 3 percent error? What is your conclusion?

- 2.37 Polycrystalline copper films** Consider the data in Figure 2.38a, which are reproduced below in Table 2.19 in terms of the average grain size (d) and the resistivity of the film. Plot these on an excel graph. Plot the Mayadas–Shatzkes equation as a function of d on the same graph. You need to first calculate $\beta = (\lambda/d)R/(1-R)$ for each d value by assuming a particular R (e.g., $R = 0.4$) and then use Equation 2.72a. You can then modify R to bring the theoretical curve as close as possible to the experimental curve. What is your conclusion? Assume $\lambda = 40$ nm and $\rho_{\text{Cu}} = 17.3$ n Ω m.

Table 2.19 Dependence of the resistivity of polycrystalline films of copper on the grain size

d (nm)	189	168	139	140	128	107	99.3	59.8	44.3
ρ_{film} (n Ω m)	20.97	21.16	22.21	22.65	22.09	23.39	23.89	27.92	31.20

| SOURCE: Riedel, S., et al., *Microelectronic Engineering*, 33, 165, 1997.

2.38 Thin films

- Consider a polycrystalline copper film that has $R = 0.40$. What is the approximate mean grain size d in terms of the mean free path λ in the bulk that would lead to the polycrystalline Cu film having a resistivity that is $1.5\rho_{\text{bulk}}$. If the mean free path in the crystal is about 40 nm at room temperature, what is d ? (Assume $D \gg d$.)
- What is the thickness D of an epitaxial copper film in terms of λ in which surface scattering increases the film resistivity to $1.2\rho_{\text{bulk}}$ if the specular scattering fraction p is 0.1?

- 2.39 Thin films of Cu** Consider the resistivity of three types of Cu thin films as shown in Table 2.20. Thin films are one single crystal layer, and two polycrystalline layers with an average grain size shown in the table. All have the same thickness $D = 40$ nm. The resistivity measurements have an error bar (representing experimental scatter in data) that is roughly ± 3 percent. Suppose, we write Matthiessen's rule as

$$\rho_{\text{film}} = \rho_{\text{crystal}} + \Delta\rho_{\text{MS}} + \Delta\rho_{\text{SF}} \quad [2.90\text{a}]$$

or

$$\rho_{\text{film}}/\rho_{\text{crystal}} \approx 1 + (3/2)\beta + (3/8)(\lambda/D)(1 - p) \quad [2.90\text{b}]$$

where β is defined in Equation 2.72b, ρ_{crystal} is the bulk resistivity of the Cu crystal, and $\Delta\rho_{\text{MS}}$ and $\Delta\rho_{\text{SF}}$ are the contributions to resistivity arising from the scattering of electrons at the grain boundary and surfaces, respectively; that is, the Mayadas–Shatzkes and Fuchs–Sondheimer contributions, respectively.

Surface and grain boundary scattering in films

Surface and grain boundary scattering in films

Complete Table 2.20 by taking $\rho_{\text{crystal}} = 17.0 \text{ n}\Omega \text{ m}$ and assuming $p = 0$ and $R = 0.25$. What is your conclusion?

Table 2.20 The resistivity of three types of thin Cu films with the same thickness $D = 40 \text{ nm}$

$d \text{ (nm)}$	$\rho_{\text{film}} \text{ (n}\Omega \text{ m)}$	$\Delta\rho_{\text{MS}} \text{ (n}\Omega \text{ m)}$	$\Delta\rho_{\text{SF}} \text{ (n}\Omega \text{ m)}$	$\rho_{\text{crystal}} + \rho_{\text{MS}} + \rho_{\text{SF}}$
∞	24.8			
160	26.8			
40	29.1			

| SOURCE: Chawla, J.S., *Physical Review B*, 84, 235423, 2011. $d = \infty$ means a single crystal film.

- 2.40 Thin films of single crystal Cu on TiN** Thin single crystal films of Cu have been deposited onto a TiN (001) surface grown on a MgO crystal substrate by. Room temperature (25°C) resistivity measurements *in situ* (in vacuum) give the data in Table 2.21. How would you interpret the data? ($\lambda = 40 \text{ nm}$ for Cu)

Table 2.21 The resistivity of Cu single crystal thin films deposited on TiN (001) surface *in situ* in vacuum

$D \text{ (nm)}$	830	40.0	13.3	6.20
$\rho \text{ (n}\Omega \text{ m)} \text{ (vacuum)}$	17.1	21.0	29.7	44.4

| SOURCE: Chawla, J.S., et al., *Journal of Applied Physics*, 110, 043714, 2011.

- 2.41 Thin films of W** Thin single crystal films of W have been grown epitaxial on sapphire (Al_2O_3) substrates. The resistivity of a 187-nm-thick film is $64 \text{ n}\Omega \text{ m}$, which can be taken as the bulk resistivity. The W film with a thickness 19.9 nm has a resistivity of $86 \text{ n}\Omega \text{ m}$. If the mean free path λ in the bulk is 19.1 nm, what is the average p ?
- 2.42 Thin films of Cu on Si (100) surface** Different thickness polycrystalline Cu films have been deposited on the (100) surface of a Si crystal and their resistivities have been measured as summarized in Table 2.22. For these films, the average grain size d has been shown to be related to the film thickness D by $d \approx D/2.3$. Use Matthiessen's rule to combine Fuchs–Sondheimer and Mayadas–Shatzkes equations as in Equation 2.90b and plot ρ against $1/D$ and also ρ against D as a log-log plot on excel or a similar application. Plot the expected ρ in these graphs from Equation 2.90b by taking $p = 0$, $\lambda = 40 \text{ nm}$, and $R = 0.25$. Try a slightly greater and slightly lower R values (*e.g.*, 0.20 and 0.30) to see how the predicted curve changes with respect to the data. What is your conclusion?

Table 2.22 The resistivity of thin polycrystalline Cu films on the Si (100) surface

$D \text{ (nm)}$	407	222	170	120	101	85.4	68.5	51.2	34.1	17.2	8.59
$\rho \text{ (n}\Omega \text{ m)}$	19.8	20.8	20.0	22.1	23.5	27.9	30.7	32.2	50.4	70.5	126

| SOURCE: Lim, J.W., and Isshiki, M., *Journal of Applied Physics*, 99, 094909, 2006.

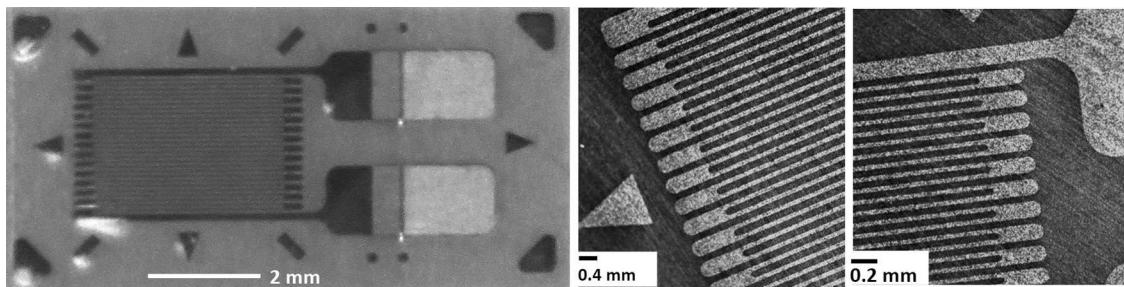
- 2.43 Interconnects** Consider a CMOS chip in which the interconnects are copper with a pitch P of 500 nm, interconnect thickness T of 400 nm, aspect ratio 1.4, and $H = X$. The dielectric is FSG with $\epsilon_r = 3.6$. Consider two cases, $L = 1$ mm and $L = 10$ mm, and calculate the overall effective interconnect capacitance C_{eff} and the RC delay time. Suppose that Al, which is normally Al with about 4 wt.% Cu in the microelectronics industry with a resistivity $31 \text{ n}\Omega \text{ m}$, is used as the interconnect. What is the corresponding RC delay time?
- *2.44 Thin 50 nm interconnects** Equation 2.76 is for conduction in a thin film of thickness D and assumes scattering from two surfaces, which shows that the increase in the resistivity $\Delta\rho_2 = \rho_{\text{bulk}} \frac{3}{8}(\lambda/D)(1-p)$. An interconnect line in an IC is not quite a thin film and has four surfaces (interfaces), because the thickness T of the conductor is comparable to the width W . If we assume $T = W$, we can very roughly take the resistivity increase with four surfaces as $\Delta\rho_4 \approx \Delta\rho_2 + \Delta\rho_2 \approx \rho_{\text{bulk}} \frac{3}{4}(\lambda/D)(1-p)$ in which $D = T$. (The exact expression is more complicated, but the latter will suffice for this problem.) In addition there will be a contribution from grain boundary scattering so that we need to use Equation 2.90a. For simplicity assume $T \approx W \approx X \approx H \approx 50$ nm, $\lambda = 40$ nm, $p = 0$ and $\epsilon_r = 3.6$. If the mean grain size d is roughly 30 nm and $R = 0.4$, estimate the resistivity of the interconnect and hence the RC delay for a 0.5 mm interconnect. (You can consider Equation 2.90b but the surface scattering now is from four surfaces as explained above.)
- 2.45 Electromigration** Although electromigration-induced failure in Cu metallization is less severe than in Al metallization, it can still lead to interconnect failure depending on current densities and the operating temperature. In a set of experiments carried out on electroplated Cu metallization lines, failure of the Cu interconnects have been examined under accelerated tests (at elevated temperatures). The mean lifetime t_{50} (time for 50 percent of the lines to break) have been measured as a function of current density J and temperature T at a given current density. The results are summarized in Table 2.23.
- Plot semilogarithmically t_{50} versus $1/T$ (T in Kelvins) for the first three interconnects. Al(Cu) and Cu ($1.3 \times 0.7 \mu\text{m}^2$) have single activation energies E_A . Calculate E_A for these interconnects. Cu ($1.3 \times 0.7 \mu\text{m}^2$) exhibits different activation energies for the high-and low-temperature regions. Estimate these E_A .
 - Plot on a log-log plot t_{50} versus J at 370 °C. Show that at low J , $n \approx 1.1$ and at high J , $n \approx 1.8$.

Table 2.23 Results of electromigration failure experiments on various Al and Cu interconnects

Al(Cu) [$J = 25 \text{ mA}/\mu\text{m}^2$, $A = 0.35 \times 0.2 (\mu\text{m})^2$]	Cu [$J = 25 \text{ mA}/\mu\text{m}^2$, $A = 0.24 \times 0.28 (\mu\text{m})^2$]	Cu [$J = 25 \text{ mA}/\mu\text{m}^2$, $A = 1.3 \times 0.7 (\mu\text{m})^2$]	Cu ($T = 370$ °C)				
T (°C)	t_{50} (hr)	T (°C)	t_{50} (hr)	T (°C)	t_{50} (hr)	$J \text{ mA } \mu\text{m}^{-2}$	t_{50} (hr)
365	0.11	397	2.87	395	40.3	3.54	131.5
300	0.98	354	12.8	360	196	11.7	25.2
259	5.73	315	70.53	314	825	24.8	14.9
233	15.7	269	180	285	2098	49.2	4.28
		232	899			74.1	2.29
						140	0.69

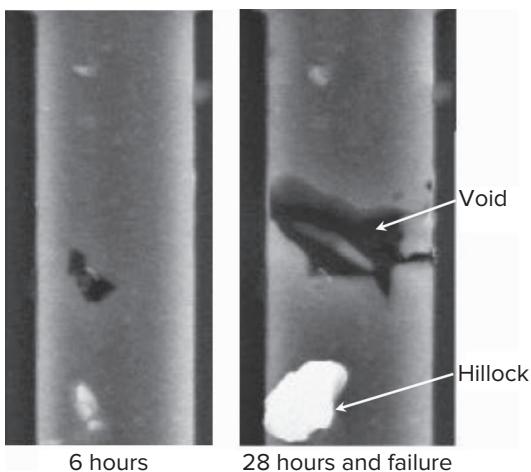
NOTE: $A = \text{width} \times \text{height}$ in micron².

SOURCE: Rosenberg, R., et al., *Annual Review of Materials Science*, 30, 229, 2000.



A commercial strain gauge by Micro-Measurements (Vishay Precision Group). This gauge has a maximum strain range of $\pm 5\%$. The overall resistance of the gauge is 350Ω . The gauge wire is a constantan alloy with a small thermal coefficient of resistance. The gauge wires are embedded in a polyimide polymer flexible substrate. The external solder pads are copper coated. Its useful temperature range is -75°C to $+175^\circ\text{C}$.

| Photo by S. Kasap.



Scanning electron microscope images of the growth of a hillock and a void in a polycrystalline aluminum interconnect line carrying a current of $2 \times 10^6 \text{ A cm}^{-2}$ at 230°C . The interconnect line was $8 \mu\text{m}$ wide and the mean grain size was $4 \mu\text{m}$. Left: After 6 hours. Right: After 28 hours and failure.

From K. Weyzig, H. Wendrock, A. Buerke and T. Kötter, "In-situ study of interconnect failures by electromigration inside a scanning electron microscope" *AIP Conference Proceedings*, 491, 89–99 (1999); with the permission of AIP Publishing.

Subject	Name of the Book	Author	Editor Chapters Needed	Sub sections to be omitted
1 Electricity and Magnetism	Fundamentals of Physics	Halliday And Resnik	10 21,22,23 and 24	22-5,22-9 and 24-8
2 Optics	Fundamentals of Physics	Halliday And Resnik	10 35, 36,33	33.1,33.2,33.3,33.5,33.6,33.7,
Optics				35-1, 35.3, 35.4, 35.5,36.3, 36.6,36.7
3 Lasers	Lasers - Fundamentals and Applications	K. Thyagarajan · Ajoy Ghatak	2 4, 5, 10, 11	
4 Quantum Mechanics	Modern Physics	Arther Biser	6 3 and 5	3.4,3.5,3.6,5.10, 5.11
5 Solid State Physics	Principles of Electronic Materials & Devices	S.O. Kasap	4 1 (sec. 1.9.1,1.9.2,1.14) ,2 (sec. 2.1,2.6.1,fig. 2.25)	1.1 to1.8,1.9.3 to 1.13, 1.15, 2.2 to 2.5,2.6.2 to 2.12
Additional portions on Thin film interference and Newton's rings are included in the compilation.				