Energy transmitted by ripples in the electric and magnetic fields

example 12

We’ll see in chapter 24 that phenomena like light, radio waves, and x-rays are all ripples in the electric and magnetic fields. The old method would give zero energy for a region of space containing a light wave but no charges. That would be wrong! We can only use the old method in static cases.

Now let’s give at least some justification for the other features of the expressions for energy density, \(-\frac{1}{8\pi G}|g|^2\) and \(\frac{1}{8\pi k}|E|^2\), besides the proportionality to the square of the field strength.

First, why the different plus and minus signs? The basic idea is that the signs have to be opposite in the gravitational and electric cases because there is an attraction between two positive masses (which are the only kind that exist), but two positive charges would repel. Since we’ve already seen examples where the positive sign in the electric energy makes sense, the gravitational energy equation must be the one with the minus sign.

It may also seem strange that the constants \(G\) and \(k\) are in the denominator. They tell us how strong the three different forces are, so shouldn’t they be on top? No. Consider, for instance, an alternative universe in which gravity is twice as strong as in ours. The numerical value of \(G\) is doubled. Because \(G\) is doubled, all the gravitational field strengths are doubled as well, which quadruples the quantity \(|g|^2\). In the expression \(-\frac{1}{8\pi G}|g|^2\), we have quadrupled something on top and doubled something on the bottom, which makes the energy twice as big. That makes perfect sense.

Discussion questions

A The figure shows a positive charge in the gap between two capacitor plates. First make a large drawing of the field pattern that would be formed by the capacitor itself, without the extra charge in the middle. Next, show how the field pattern changes when you add the particle at these two positions. Compare the energy of the electric fields in the two cases. Does this agree with what you would have expected based on your knowledge of electrical forces?

B Criticize the following statement: “A solenoid makes a charge in the space surrounding it, which dissipates when you release the energy.”

C In example 11, I argued that the fields surrounding a positive and negative charge contain less energy when the charges are closer together. Perhaps a simpler approach is to consider the two extreme possibilities: the case where the charges are infinitely far apart, and the one in which they are at zero distance from each other, i.e., right on top of each other. Carry out this reasoning for the case of (1) a positive charge and a negative charge of equal magnitude, (2) two positive charges of equal magnitude, (3) the gravitational energy of two equal masses.
22.5 Potential for nonuniform fields

The calculus-savvy reader will have no difficulty generalizing the field-potential relationship to the case of a varying field. The potential energy associated with a varying force is

$$\Delta PE = - \int F \, dx, \quad \text{[one dimension]}$$

so for electric fields we divide by $q$ to find

$$\Delta V = - \int E \, dx, \quad \text{[one dimension]}$$

Applying the fundamental theorem of calculus yields

$$E = - \frac{dV}{dx}. \quad \text{[one dimension]}$$

Potential associated with a point charge example 13

What is the potential associated with a point charge?

As derived previously in self-check A on page 641, the field is

$$|E| = \frac{kQ}{r^2}$$

The difference in potential between two points on the same radius line is

$$\Delta V = \int dV$$

$$= - \int E_x \, dx$$

In the general discussion above, $x$ was just a generic name for distance traveled along the line from one point to the other, so in this case $x$ really means $r$.

$$\Delta V = - \int_{r_1}^{r_2} E_r \, dr$$

$$= - \int_{r_1}^{r_2} \frac{kQ}{r^2} \, dr$$

$$= \left. \frac{kQ}{r} \right|_{r_1}^{r_2}$$

$$= \frac{kQ}{r_2} - \frac{kQ}{r_1}.$$  

The standard convention is to use $r_1 = \infty$ as a reference point, so that the potential at any distance $r$ from the charge is

$$V = \frac{kQ}{r}.$$
The interpretation is that if you bring a positive test charge closer to a positive charge, its electrical energy is increased; if it was released, it would spring away, releasing this as kinetic energy.

**self-check B**
Show that you can recover the expression for the field of a point charge by evaluating the derivative $E_x = -\frac{dV}{dx}$.

---

22.6 Two or three dimensions

The topographical map shown in figure u suggests a good way to visualize the relationship between field and potential in two dimensions. Each contour on the map is a line of constant height; some of these are labeled with their elevations in units of feet. Height is related to gravitational potential energy, so in a gravitational analogy, we can think of height as representing potential. Where the contour lines are far apart, as in the town, the slope is gentle. Lines close together indicate a steep slope.

If we walk along a straight line, say straight east from the town, then height (potential) is a function of the east-west coordinate $x$. Using the usual mathematical definition of the slope, and writing $V$ for the height in order to remind us of the electrical analogy, the slope along such a line is $\Delta V/\Delta x$. If the slope isn’t constant, we either need to use the slope of the $V - x$ graph, or use calculus and talk about the derivative $dV/dx$.

What if everything isn’t confined to a straight line? Water flows
downhill. Notice how the streams on the map cut perpendicularly through the lines of constant height.

It is possible to map potentials in the same way, as shown in figure v. The electric field is strongest where the constant-potential curves are closest together, and the electric field vectors always point perpendicular to the constant-potential curves.

Figure x shows some examples of ways to visualize field and potential patterns.

Mathematically, the calculus of section 22.5 generalizes to three dimensions as follows:

\[
E_x = -\frac{dV}{dx} \\
E_y = -\frac{dV}{dy} \\
E_z = -\frac{dV}{dz}
\]

**self-check C**

Imagine that the topographical map in figure w represents potential rather than height. (a) Consider the stream that starts near the center of the map. Determine the positive and negative signs of \( \frac{dV}{dx} \) and \( \frac{dV}{dy} \), and relate these to the direction of the force that is pushing the current forward against the resistance of friction. (b) If you wanted to find a lot of electric charge on this map, where would you look? Answer, p. 1043
Two-dimensional field and potential patterns. Top: A uniformly charged rod. Bottom: A dipole. In each case, the diagram on the left shows the field vectors and constant-potential curves, while the one on the right shows the potential (up-down coordinate) as a function of x and y. Interpreting the field diagrams: Each arrow represents the field at the point where its tail has been positioned. For clarity, some of the arrows in regions of very strong field strength are not shown — they would be too long to show. Interpreting the constant-potential curves: In regions of very strong fields, the curves are not shown because they would merge together to make solid black regions. Interpreting the perspective plots: Keep in mind that even though we’re visualizing things in three dimensions, these are really two-dimensional potential patterns being represented. The third (up-down) dimension represents potential, not position.

22.7 ★ Field lines and Gauss’s law

When we look at the “sea of arrows” representation of a field, y/1, there is a natural visual tendency to imagine connecting the arrows as in y/2. The curves formed in this way are called field lines, and they have a direction, shown by the arrowheads.

y / Two different representations of an electric field.
An avalanche transceiver

The dipole field pattern is universal: it always has the same mathematical shape, and this holds regardless of whether the field is electric or magnetic, static or oscillating. One of its universal properties is that the field lines always pass through the source, and this is taken advantage of in a device called an avalanche transceiver used by backcountry skiers and hikers in winter. The device is worn on the body and is left in transmitting mode while the user is traveling. It creates a dipole pattern of electric and magnetic fields, oscillating at 457 kHz. If a member of the party is buried in an avalanche, his companions must find and locate him within about 30-90 minutes, or he will suffocate. The rescuers switch their own transceivers to receive mode, allowing them to determine the direction of the field line; when the antenna is parallel to the field line, the oscillating electric field drives electrons up and down it, creating the maximum possible current.

In the figure, the rescuer moves along the field line from A to B. At B, she verifies that the strength of the signal has increased. (If it hadn’t, she would have had to turn around and go in the opposite direction.) She redetermines the direction of the field and continues in the new direction. Continuing the process, she proceeds along an approximation to the field line. At D she finds that the field strength has fallen off again, so she knows that she has just passed over the victim’s position.

Electric field lines originate from positive charges and terminate on negative ones. We can choose a constant of proportionality that fixes how coarse or fine the “grain of the wood” is, but once this choice is made the strength of each charge is shown by the number of lines that begin or end on it. For example, figure y/2 shows eight lines at each charge, so we know that $q_1/q_2 = (-8)/8 = -1$. Because lines never begin or end except on a charge, we can always find the total charge inside any given region by subtracting the number of lines that go in from the number that come out and multiplying by the appropriate constant of proportionality. Ignoring the constant, we can apply this technique to figure aa to find $q_A = -8$, $q_B = 2 - 2 = 0$, and $q_C = 5 - 5 = 0$.

Figures y and aa are drawn in two dimensions, but we should really imagine them in three dimensions, so that the regions A, B, and C in figure aa are volumes bounded by surfaces. It is only because our universe happens to have three dimensions that the field line concept is as useful as it is. This is because the number of field lines per unit area perpendicularly piercing a surface can be interpreted as the strength of the electric field. To see this, consider an imaginary spherical surface of radius $r$, with a positive charge $q$ at its center. The field at the surface equals $kq/r^2$. The number of field lines piercing the surface is independent of $r$, but the surface’s area...
is proportional to \( r^2 \), so the number of piercings per unit area is proportional to \( 1/r^2 \), just like the electric field. With this interpretation, we arrive at Gauss’s law, which states that the field strength on a surface multiplied by the surface area equals the total charge enclosed by the surface. (This particular formulation only works in the case where the field pierces the surface perpendicularly and is constant in magnitude over the whole surface. It can be reformulated so as to eliminate these restrictions, but we won’t require that reformulation for our present purposes.)

\[ \text{The field of a line of charge} \]

In example 4 on p. 643, we found by simple scaling arguments that the electric field of a line of charge is proportional to \( 1/r \), where \( r \) is the distance from the line. Since electric fields are always proportional to the Coulomb constant \( k \) and to the amount of charge, clearly we must have something of the form \( E = (\ldots)k(q/L)/r \), where \( q/L \) is the charge per unit length and \( (\ldots) \) represents an unknown numerical constant (which you can easily verify is unitless). Using Gauss’s law we can fill in the final piece of the puzzle, which is the value of this constant.

Figure ab shows two surfaces, a cylindrical one A and a spherical one B. Every field line that passes through A also passes through B. By making A sufficiently small and B sufficiently large, we can make the field on each surface nearly constant and perpendicular to the surface, as required by our restricted form of Gauss’s law. If radius \( r_B \) is made large, the field at B made by the line of charge becomes indistinguishable from that of a point charge, \( kq/r_B^2 \). By Gauss’s law, the electric field at each surface is proportional to the number of field lines divided by its area. But the number of field lines is the same in both cases, so each field is inversely proportional to the corresponding surface area. We therefore have

\[
E_A = E_B \left( \frac{A_B}{A_A} \right)
= \left( \frac{kq}{r_B^2} \right) \left( \frac{4\pi r_A^2}{2\pi r_A L} \right)
= \frac{2kq}{Lr_A}.
\]

The unknown numerical constant equals 2.

\[ \text{No charge on the interior of a conductor} \]

I asserted on p. 593 that for a perfect conductor in equilibrium, excess charge is found only at the surface, never in the interior. This can be proved using Gauss’s law. Suppose that a charge \( q \) existed at some point in the interior, and it was in stable equilibrium. For concreteness, let’s say \( q \) is positive. If its equilibrium is to be stable, then we need an electric field everywhere around it that points inward like a pincushion, so that if the charge were to be
Example 17. Perturbed slightly, the field would bring it back to its equilibrium position. Since Newton’s third law forbids objects from making forces on themselves, this field would have to be the field contributed by all the other charges, not by \( q \) itself. But by Gauss’s law, an external set of charges cannot form this kind of inward-pointing pincushion pattern; for such a pattern, Gauss’s law would require there to be some negative charge inside the pincusion.

22.8 \( \int \) Electric field of a continuous charge distribution

Charge really comes in discrete chunks, but often it is mathematically convenient to treat a set of charges as if they were like a continuous fluid spread throughout a region of space. For example, a charged metal ball will have charge spread nearly uniformly all over its surface, and in for most purposes it will make sense to ignore the fact that this uniformity is broken at the atomic level. The electric field made by such a continuous charge distribution is the sum of the fields created by every part of it. If we let the “parts” become infinitesimally small, we have a sum of an infinite number of infinitesimal numbers, which is an integral. If it was a discrete sum, we would have a total electric field in the \( x \) direction that was the sum of all the \( x \) components of the individual fields, and similarly we’d have sums for the \( y \) and \( z \) components. In the continuous case, we have three integrals.

Field of a uniformly charged rod example 17

A rod of length \( L \) has charge \( Q \) spread uniformly along it. Find the electric field at a point a distance \( d \) from the center of the rod, along the rod’s axis. (This is different from examples 4 on p. 643 and 15 on p. 657, both because the point is on the axis of the rod and because the rod is of finite length.)

This is a one-dimensional situation, so we really only need to do a single integral representing the total field along the axis. We imagine breaking the rod down into short pieces of length \( dz \), each with charge \( dq \). Since charge is uniformly spread along the rod, we have \( dq = \lambda \, dz \), where \( \lambda = Q/L \) (Greek lambda) is the charge per unit length, in units of coulombs per meter. Since the pieces are infinitesimally short, we can treat them as point charges and use the expression \( k \, dq / r^2 \) for their contributions to the field, where \( r = d - z \) is the distance from the charge at \( z \) to
the point in which we are interested.

\[ E_z = \int \frac{k\,dq}{r^2} \]

\[ = \int_{-L/2}^{+L/2} \frac{k\lambda\,dz}{r^2} \]

\[ = k\lambda \int_{-L/2}^{+L/2} \frac{dz}{(d - z)^2} \]

The integral can be looked up in a table, or reduced to an elementary form by substituting a new variable for \( d - z \). The result is

\[ E_z = k\lambda \left( \frac{1}{d - z} \right)_{-L/2}^{+L/2} \]

\[ = k\lambda \frac{d - L/2}{d - L/2} - \frac{d + L/2}{d - L/2} \]

For large values of \( d \), this expression gets smaller for two reasons: (1) the denominators of the fractions become large, and (2) the two fractions become nearly the same, and tend to cancel out. This makes sense, since the field should get weaker as we get farther away from the charge. In fact, the field at large distances must approach \( kQ/d^2 \), since from a great distance, the rod looks like a point.

It's also interesting to note that the field becomes infinite at the ends of the rod, but is not infinite on the interior of the rod. Can you explain physically why this happens?
Summary

Selected vocabulary
field . . . . . . . . a property of a point in space describing the forces that would be exerted on a particle if it was there
sink . . . . . . . . a point at which field vectors converge
source . . . . . . . a point from which field vectors diverge; often used more inclusively to refer to points of either convergence or divergence
electric field . . . . the force per unit charge exerted on a test charge at a given point in space
gravitational field the force per unit mass exerted on a test mass at a given point in space
electric dipole . . an object that has an imbalance between positive charge on one side and negative charge on the other; an object that will experience a torque in an electric field

Notation
g . . . . . . . . . . the gravitational field
E . . . . . . . . . the electric field
D . . . . . . . . an electric dipole moment

c
Other terminology and notation
d, p, m . . . . . . other notations for the electric dipole moment

Summary

Experiments show that time is not absolute: it flows at different rates depending on an observer’s state of motion. This is an example of the strange effects predicted by Einstein’s theory of relativity. All of these effects, however, are very small when the relative velocities are small compared to \( c \). This makes sense, because Newton’s laws have already been thoroughly tested by experiments at such speeds, so a new theory like relativity must agree with the old one in their realm of common applicability. This requirement of backwards-compatibility is known as the correspondence principle.

Since time is not absolute, simultaneity is not a well-defined concept, and therefore the universe cannot operate as Newton imagined, through instantaneous action at a distance. There is a delay in time before a change in the configuration of mass and charge in one corner of the universe will make itself felt as a change in the forces experienced far away. We imagine the outward spread of such a change as a ripple in an invisible universe-filling field of force.

We define the gravitational field at a given point as the force per unit mass exerted on objects inserted at that point, and likewise the electric field is defined as the force per unit charge. These fields are vectors, and the fields generated by multiple sources add according to the rules of vector addition.
When the electric field is constant, the $\Delta V$ between two points lying on a line parallel to the field is related to the field by the equation $\Delta V = -Ed$, where $d$ is the distance between the two points. In this context, one usually refers to $V$ as the electrical potential rather than “the voltage.”

Fields of force contain energy. The density of energy is proportional to the square of the magnitude of the field. In the case of static fields, we can calculate potential energy either using the previous definition in terms of mechanical work or by calculating the energy stored in the fields. If the fields are not static, the old method gives incorrect results and the new one must be used.
Problems

Key
✓ A computerized answer check is available online.
∫ A problem that requires calculus.
∗ A difficult problem.

1 In our by-now-familiar neuron, the voltage difference between the inner and outer surfaces of the cell membrane is about $V_{out} - V_{in} = -70$ mV in the resting state, and the thickness of the membrane is about 6.0 nm (i.e., only about a hundred atoms thick). What is the electric field inside the membrane? 

2 The gap between the electrodes in an automobile engine's spark plug is 0.060 cm. To produce an electric spark in a gasoline-air mixture, an electric field of $3.0 \times 10^6$ V/m must be achieved. On starting a car, what minimum voltage must be supplied by the ignition circuit? Assume the field is uniform. 

(b) The small size of the gap between the electrodes is inconvenient because it can get blocked easily, and special tools are needed to measure it. Why don’t they design spark plugs with a wider gap?

3 (a) At time $t = 0$, a positively charged particle is placed, at rest, in a vacuum, in which there is a uniform electric field of magnitude $E$. Write an equation giving the particle's speed, $v$, in terms of $t$, $E$, and its mass and charge $m$ and $q$. 

(b) If this is done with two different objects and they are observed to have the same motion, what can you conclude about their masses and charges? (For instance, when radioactivity was discovered, it was found that one form of it had the same motion as an electron in this type of experiment.)

4 Three charges are arranged on a square as shown. All three charges are positive. What value of $q_2/q_1$ will produce zero electric field at the center of the square? 

Solution, p. 1033

5 Show that the magnitude of the electric field produced by a simple two-charge dipole, at a distant point along the dipole’s axis, is to a good approximation proportional to $D/r^3$, where $r$ is the distance from the dipole. [Hint: Use the approximation $(1 + \epsilon)^p \approx 1 + p\epsilon$, which is valid for small $\epsilon$.]

6 Consider the electric field created by a uniform ring of total charge $q$ and radius $b$. (a) Show that the field at a point on the ring's axis at a distance $a$ from the plane of the ring is $kqa(a^2 + b^2)^{-3/2}$. (b) Show that this expression has the right behavior for $a = 0$ and for $a$ much greater than $b$.

7 Example 4 on p. 643 showed that the electric field of a long, uniform line of charge falls off with distance as $1/r$. By a similar technique, show that the electric field of a uniformly charged plane has no dependence on the distance from the plane.
8. Given that the field of a dipole is proportional to \( D/r^3 \) (problem 5), show that its potential varies as \( D/r^2 \). (Ignore positive and negative signs and numerical constants of proportionality.)

9. A carbon dioxide molecule is structured like O-C-O, with all three atoms along a line. The oxygen atoms grab a little bit of extra negative charge, leaving the carbon positive. The molecule’s symmetry, however, means that it has no overall dipole moment, unlike a V-shaped water molecule, for instance. Whereas the potential of a dipole of magnitude \( D \) is proportional to \( D/r^2 \), it turns out that the potential of a carbon dioxide molecule at a distant point along the molecule’s axis equals \( b/r^3 \), where \( r \) is the distance from the molecule and \( b \) is a constant. What would be the electric field of a carbon dioxide molecule at a point on the molecule’s axis, at a distance \( r \) from the molecule?

10. A proton is in a region in which the electric field is given by \( E = a + bx^3 \). If the proton starts at rest at \( x_1 = 0 \), find its speed, \( v \), when it reaches position \( x_2 \). Give your answer in terms of \( a \), \( b \), \( x_2 \), and \( e \) and \( m \), the charge and mass of the proton.

11. Consider the electric field created by a uniformly charged cylinder that extends to infinity in one direction. (a) Starting from the result of problem 8, show that the field at the center of the cylinder’s mouth is \( 2\pi k \sigma \), where \( \sigma \) is the density of charge on the cylinder, in units of coulombs per square meter. [Hint: You can use a method similar to the one in problem 9.] (b) This expression is independent of the radius of the cylinder. Explain why this should be so. For example, what would happen if you doubled the cylinder’s radius?

12. In an electrical storm, the cloud and the ground act like a parallel-plate capacitor, which typically charges up due to frictional electricity in collisions of ice particles in the cold upper atmosphere. Lightning occurs when the magnitude of the electric field reaches a critical value \( E_c \), at which air is ionized.
   (a) Treat the cloud as a flat square with sides of length \( L \). If it is at a height \( h \) above the ground, find the amount of energy released in the lightning strike.
   (b) Based on your answer from part a, which is more dangerous, a lightning strike from a high-altitude cloud or a low-altitude one?
   (c) Make an order-of-magnitude estimate of the energy released by a typical lightning bolt, assuming reasonable values for its size and altitude. \( E_c \) is about \( 10^6 \text{ N/C} \).

See problem 16 for a note on how recent research affects this estimate.

13. The neuron in the figure has been drawn fairly short, but some neurons in your spinal cord have tails (axons) up to a meter long. The inner and outer surfaces of the membrane act as the “plates”
of a capacitor. (The fact that it has been rolled up into a cylinder has very little effect.) In order to function, the neuron must create a voltage difference \( V \) between the inner and outer surfaces of the membrane. Let the membrane’s thickness, radius, and length be \( t \), \( r \), and \( L \). (a) Calculate the energy that must be stored in the electric field for the neuron to do its job. (In real life, the membrane is made out of a substance called a dielectric, whose electrical properties increase the amount of energy that must be stored. For the sake of this analysis, ignore this fact.)

\( \diamond \) Hint, p. 1032 \( \checkmark \)

(b) An organism’s evolutionary fitness should be better if it needs less energy to operate its nervous system. Based on your answer to part a, what would you expect evolution to do to the dimensions \( t \) and \( r \)? What other constraints would keep these evolutionary trends from going too far?

14 To do this problem, you need to understand how to do volume integrals in cylindrical and spherical coordinates. (a) Show that if you try to integrate the energy stored in the field of a long, straight wire, the resulting energy per unit length diverges both at \( r \to 0 \) and \( r \to \infty \). Taken at face value, this would imply that a certain real-life process, the initiation of a current in a wire, would be impossible, because it would require changing from a state of zero magnetic energy to a state of infinite magnetic energy. (b) Explain why the infinities at \( r \to 0 \) and \( r \to \infty \) don’t really happen in a realistic situation. (c) Show that the electric energy of a point charge diverges at \( r \to 0 \), but not at \( r \to \infty \).

A remark regarding part (c): Nature does seem to supply us with particles that are charged and pointlike, e.g., the electron, but one could argue that the infinite energy is not really a problem, because an electron traveling around and doing things neither gains nor loses infinite energy; only an infinite change in potential energy would be physically troublesome. However, there are real-life processes that create and destroy pointlike charged particles, e.g., the annihilation of an electron and antielectron with the emission of two gamma rays. Physicists have, in fact, been struggling with infinities like this since about 1950, and the issue is far from resolved. Some theorists propose that apparently pointlike particles are actually not pointlike: close up, an electron might be like a little circular loop of string.
15  The figure shows cross-sectional views of two cubical capacitors, and a cross-sectional view of the same two capacitors put together so that their interiors coincide. A capacitor with the plates close together has a nearly uniform electric field between the plates, and almost zero field outside; these capacitors don’t have their plates very close together compared to the dimensions of the plates, but for the purposes of this problem, assume that they still have approximately the kind of idealized field pattern shown in the figure. Each capacitor has an interior volume of 1.00 m$^3$, and is charged up to the point where its internal field is 1.00 V/m. (a) Calculate the energy stored in the electric field of each capacitor when they are separate. (b) Calculate the magnitude of the interior field when the two capacitors are put together in the manner shown. Ignore effects arising from the redistribution of each capacitor’s charge under the influence of the other capacitor. (c) Calculate the energy of the put-together configuration. Does assembling them like this release energy, consume energy, or neither? √

16  In problem 12 on p. 663, you estimated the energy released in a bolt of lightning, based on the energy stored in the electric field immediately before the lightning occurs. The assumption was that the field would build up to a certain value, which is what is necessary to ionize air. However, real-life measurements always seemed to show electric fields strengths roughly 10 times smaller than those required in that model. For a long time, it wasn’t clear whether the field measurements were wrong, or the model was wrong. Research carried out in 2003 seems to show that the model was wrong. It is now believed that the final triggering of the bolt of lightning comes from cosmic rays that enter the atmosphere and ionize some of the air. If the field is 10 times smaller than the value assumed in problem 12, what effect does this have on the final result of problem 12? √
Exercise 22: Field vectors

Apparatus:
3 solenoids
DC power supply
compass
ruler
cut-off plastic cup

At this point you’ve studied the gravitational field, \( \mathbf{g} \), and the electric field, \( \mathbf{E} \), but not the magnetic field, \( \mathbf{B} \). However, they all have some of the same mathematical behavior: they act like vectors. Furthermore, magnetic fields are the easiest to manipulate in the lab. Manipulating gravitational fields directly would require futuristic technology capable of moving planet-sized masses around! Playing with electric fields is not as ridiculously difficult, but static electric charges tend to leak off through your body to ground, and static electricity effects are hard to measure numerically. Magnetic fields, on the other hand, are easy to make and control. Any moving charge, i.e. any current, makes a magnetic field.

A practical device for making a strong magnetic field is simply a coil of wire, formally known as a solenoid. The field pattern surrounding the solenoid gets stronger or weaker in proportion to the amount of current passing through the wire.

1. With a single solenoid connected to the power supply and laid with its axis horizontal, use a magnetic compass to explore the field pattern inside and outside it. The compass shows you the field vector’s direction, but not its magnitude, at any point you choose. Note that the field the compass experiences is a combination (vector sum) of the solenoid’s field and the earth’s field.

2. What happens when you bring the compass extremely far away from the solenoid?

What does this tell you about the way the solenoid’s field varies with distance?

Thus although the compass doesn’t tell you the field vector’s magnitude numerically, you can get at least some general feel for how it depends on distance.
3. The figure below is a cross-section of the solenoid in the plane containing its axis. Make a sea-of-arrows sketch of the magnetic field in this plane. The length of each arrow should at least approximately reflect the strength of the magnetic field at that point.

Does the field seem to have sources or sinks?

4. What do you think would happen to your sketch if you reversed the wires?

Try it.
5. Now hook up the two solenoids in parallel. You are going to measure what happens when their two fields combine at a certain point in space. As you’ve seen already, the solenoids’ nearby fields are much stronger than the earth’s field; so although we now theoretically have three fields involved (the earth’s plus the two solenoids’), it will be safe to ignore the earth’s field. The basic idea here is to place the solenoids with their axes at some angle to each other, and put the compass at the intersection of their axes, so that it is the same distance from each solenoid. Since the geometry doesn’t favor either solenoid, the only factor that would make one solenoid influence the compass more than the other is current. You can use the cut-off plastic cup as a little platform to bring the compass up to the same level as the solenoids’ axes.

a) What do you think will happen with the solenoids’ axes at 90 degrees to each other, and equal currents? Try it. Now represent the vector addition of the two magnetic fields with a diagram. Check your diagram with your instructor to make sure you’re on the right track.

b) Now try to make a similar diagram of what would happen if you switched the wires on one of the solenoids.

After predicting what the compass will do, try it and see if you were right.

C) Now suppose you were to go back to the arrangement you had in part a, but you changed one of the currents to half its former value. Make a vector addition diagram, and use trig to predict the angle.

Try it. To cut the current to one of the solenoids in half, an easy and accurate method is simply to put the third solenoid in series with it, and put that third solenoid so far away that its magnetic field doesn’t have any significant effect on the compass.
Chapter 23
Relativity and magnetism

Many people imagine Einstein’s theory of relativity as something exotic and speculative. It’s certainly not speculative at this point in history, since you use it every time you use a GPS receiver. But it’s even less exotic than that. Every time you stick a magnet to your refrigerator, you’re making use of relativity. Let’s dig a little deeper into relativity as preparation for understanding what magnetism is and where it comes from.

23.1 Spacetime

Let’s compare how Aristotle, Galileo, and Einstein would describe space and time.

Aristotle: All observers agree on whether or not two things happen at the same time, and also on whether they happen at the same place.

Galileo: Observers always agree on simultaneity, but not necessarily on whether things happen in the same place.

The reason for the disagreement is shown in figure a. Aristotle says that the only legitimate observers are those that are at rest relative to the ground, while Galileo is willing to accept any inertial frame of reference, such as the driver’s. Galileo ended up winning this argument because of experiments verifying the principle of inertia.

Einstein: Observers need not agree on whether two things happen at the same time or the same place.

We accept Einstein’s view because of evidence such as the atomic clock experiment described on p. 634. Such experiments rule out both the instantaneous transmission of signals (p. 635) and, as we will argue in more detail on p. 676, Galileo’s claim about universal agreement on simultaneity.

One of the reasons that nineteenth-century Europeans found Marxism alarming was because it was atheistic, and they felt that without the framework of religion, there could be no basis for morality. For similar reasons, I was deeply disoriented when I first encountered relativity. The idea had been firmly inculcated that the universe was described by mathematical functions, and the natural habitat of those functions was graph paper. The graph paper provided what seemed like a necessary framework. For a position-time graph,
vertical lines meant “same time,” and the horizontal ones “same place.” Somehow it didn’t bother me much when Galileo erased the same-place lines (or at least relegated them to subjectivity), but without the same-time lines I felt lost, as if I were wandering in a landscape of Hieronymus Bosch’s hell or Dali’s melting watches.

One of the disorienting things about this vision of the universe is that it takes away the notion that we can have a literal “vision” of the universe. We no longer have the idea of a snapshot of the landscape at a certain moment frozen in time. The sense of vision is merely a type of optical measurement, in which we receive signals that have traveled to our eyes at some finite speed (the speed of light). What relativity substitutes for the Galilean instantaneous snapshot is the concept of \textit{spacetime}, which is like the graph paper when its lines have been erased. Every point on the paper is called an \textit{event}. How can we even agree on the existence of an event, or define which one we are talking about, if we can’t necessarily agree on its time or position? The relativist’s attitude is that if a firecracker pops, that’s an event, everyone agrees that it’s an event, and $x$ and $t$ coordinates are just an optional and arbitrary name or label for the event. Labeling an event with coordinates is like God asking Adam to name all the birds and animals: the animals weren’t consulted and didn’t care.

My grandparents’ German shepherd lived for a certain amount of time, so he was not just a pointlike event in spacetime. Way back in ch. 2, we saw how to represent the motion of such things as curves on an $x$-$t$ graph. From the point of view of relativity, the curve is the thing — we make no distinction between the dog and the dog’s track through spacetime. Such a track is called a \textit{world-line}. A world-line is a set of events strung together continuously: the dog as a puppy in Walnut Creek in 1964, the dog dozing next to the TV in 1970, and so on. The strange terminology is translated from German, and is supposed to be a description of the idea that the line is the thing’s track \textit{through} the world, i.e., through spacetime.

Sometimes if we want to describe an event, we can describe it as the beginning or end of a world-line: the dog’s birth, or the firecracker’s self-destruction. More commonly, we pick out an event of interest as the intersection of two world-lines, as in figure c. In this figure, as is common in relativity, we omit any indications of the axes, since the idea is that events and world-lines are primary, and coordinates secondary. In this book, to be consistent with the familiar depiction of $x$-$t$ graphs, we will use the convention that later times on an object’s world-line are to the right, but it is actually more common in relativity to show time progressing from the bottom of the diagram to the top.

All observers agree on whether or not two world-lines intersect, and another aid in holding on to our sanity is that they agree on whether
or not world-lines are *straight*. A straight world-line is an object moving inertially, with no forces acting on it.

If we wish to, we are able to draw a graph-paper grid on our picture of spacetime, and assign $x$ and $t$ coordinates to events, but these are not built into the structure of spacetime, and they are observer-dependent — even more so than in Galilean spacetime. They are best thought of as the sophisticated results of a laborious process of collecting and analyzing data obtained by methods such as consulting clocks or exchanging signals between different places. Figure e outlines such a process in a cartoonish way. A fleet of rocket ships, carrying surveyors, is sent out from Earth and dispersed throughout a vast region of space. The surveyors look through their theodolites at images, which are formed by light rays (dashed lines) that have arrived after traveling at a finite speed. Such light rays carry old, stale information about various events. A nuclear war has broken out. Rock and roll music has arrived on Saturn. The resulting data are then transmitted by various means (passenger pigeon, Morse-coded radio, paper mail) and consolidated at the surveying office, where coordinates are charted.

![Diagram](image)

d / Normally we only try to represent one spatial dimension in a depiction of spacetime, as in panel 1, where a light bulb is momentarily turned on, producing flashes of light that spread out in both directions. In panel 2 we attempt to show a second spatial dimension, so that the world-line becomes a surface: a "world-sheet" shaped like a cone.

e / Coordinates like $x$ and $t$ are the after-the-fact result of a process analogous to surveying.

*self-check A*

Here is a spacetime graph for an empty object such as a house: [Image].

Explain why it looks like this. My grandparents had a dog door with a flap cut into their back door, so that their dog could come in and out. Draw a spacetime diagram showing the dog going out into the back yard. Can an observer using another frame of reference say that the dog didn’t go outside?

▷ Answer, p. 1043

In the next section we turn to a more quantitative treatment of how time and distance behave in relativity.
23.2 Relativistic distortion of space and time

Time distortion arising from motion and gravity

Let’s refer back to the results of the Hafele-Keating experiment described on p. 634. Hafele and Keating were testing specific quantitative predictions of relativity, and they verified them to within their experiment’s error bars. Let’s work backward instead, and inspect the empirical results for clues as to how time works.

The east-going clock lost time, ending up off by \(-59 \pm 10\) nanoseconds, while the west-going one gained \(273 \pm 7\) ns. Since two traveling clocks experienced effects in opposite directions, we can tell that the rate at which time flows depends on the motion of the observer. The east-going clock was moving in the same direction as the earth’s rotation, so its velocity relative to the earth’s center was greater than that of the clock that remained in Washington, while the west-going clock’s velocity was correspondingly reduced. The fact that the east-going clock fell behind, and the west-going one got ahead, shows that the effect of motion is to make time go more slowly. This effect of motion on time was predicted by Einstein in his original 1905 paper on relativity, written when he was 26.

If this had been the only effect in the Hafele-Keating experiment, then we would have expected to see effects on the two flying clocks that were equal in size. Making up some simple numbers to keep the arithmetic transparent, suppose that the earth rotates from west to east at 1000 km/hr, and that the planes fly at 300 km/hr. Then the speed of the clock on the ground is 1000 km/hr, the speed of the clock on the east-going plane is 1300 km/hr, and that of the west-going clock 700 km/hr. Since the speeds of 700, 1000, and 1300 km/hr have equal spacing on either side of 1000, we would expect the discrepancies of the moving clocks relative to the one in the lab to be equal in size but opposite in sign.

In fact, the two effects are unequal in size: \(-59\) ns and \(273\) ns. This implies that there is a second effect involved, a speeding up of time simply due to the planes’ being up in the air. This was verified more directly in a 1978 experiment by Iijima and Fujiwara, figure g, in which identical atomic clocks were kept at rest at the top and bottom of a mountain near Tokyo. This experiment, unlike the Hafele-Keating one, isolates one effect on time, the gravitational one: time’s rate of flow increases with height in a gravitational field. Einstein didn’t figure out how to incorporate gravity into relativity until 1915, after much frustration and many false starts. The simpler version of the theory without gravity is known as special relativity, the full version as general relativity. We’ll restrict ourselves to special relativity until chapter 27, and that means that what we want to focus on right now is the distortion of time due to motion, not gravity.
The correspondence principle requires that the relativistic distortion of time become small for small velocities. A graph showing the time difference between two atomic clocks. One clock was kept at Mitaka Observatory, at 58 m above sea level. The other was moved back and forth to a second observatory, Norikura Corona Station, at the peak of the Norikura volcano, 2876 m above sea level. The plateaus on the graph are data from the periods when the clocks were compared side by side at Mitaka. The difference between one plateau and the next shows a gravitational effect on the rate of flow of time, accumulated during the period when the mobile clock was at the top of Norikura. Cf. problem 4, p. 833.

We can now see in more detail how to apply the correspondence principle. The behavior of the three clocks in the Hafele-Keating experiment shows that the amount of time distortion increases as the speed of the clock’s motion increases. Newton lived in an era when the fastest mode of transportation was a galloping horse, and the best pendulum clocks would accumulate errors of perhaps a minute over the course of several days. A horse is much slower than a jet plane, so the distortion of time would have had a relative size of only $\sim 10^{-15}$ — much smaller than the clocks were capable of detecting. At the speed of a passenger jet, the effect is about $10^{-12}$, and state-of-the-art atomic clocks in 1971 were capable of measuring that. A GPS satellite travels much faster than a jet airplane, and the effect on the satellite turns out to be $\sim 10^{-10}$. The general idea here is that all physical laws are approximations, and approximations aren’t simply right or wrong in different situations. Approximations are better or worse in different situations, and the question is whether a particular approximation is good enough in a given situation to serve a particular purpose. The faster the motion, the worse the Newtonian approximation of absolute time. Whether the approximation is good enough depends on what you’re trying to accomplish. The correspondence principle says that the approxi-
Two events are given as points on a graph of position versus time. Joan of Arc helps to restore Charles VII to the throne. At a later time and a different position, Joan of Arc is sentenced to death.

A change of units distorts an x-t graph. This graph depicts exactly the same events as figure i. The only change is that the x and t coordinates are measured using different units, so the grid is compressed in t and expanded in x.

A convention we’ll use to represent a distortion of time and space.

Relativity says that when two observers are in different frames of reference, each observer considers the other one’s perception of time to be distorted. We’ll also see that something similar happens to their observations of distances, so both space and time are distorted. What exactly is this distortion? How do we even conceptualize it?

The idea isn’t really as radical as it might seem at first. We can visualize the structure of space and time using a graph with position and time on its axes. These graphs are familiar by now, but we’re going to look at them in a slightly different way. Before, we used them to describe the motion of objects. The grid underlying the graph was merely the stage on which the actors played their parts. Now the background comes to the foreground: it’s time and space themselves that we’re studying. We don’t necessarily need to have a line or a curve drawn on top of the grid to represent a particular object. We may, for example, just want to talk about events, depicted as points on the graph as in figure i. A distortion of the Cartesian grid underlying the graph can arise for perfectly ordinary reasons that Newton would have readily accepted. For example, we can simply change the units used to measure time and position, as in figure j.

We’re going to have quite a few examples of this type, so I’ll adopt the convention shown in figure k for depicting them. Figure k summarizes the relationship between figures i and j in a more compact form. The gray rectangle represents the original coordinate grid of figure i, while the grid of black lines represents the new version from figure j. Omitting the grid from the gray rectangle makes the diagram easier to decode visually.

Our goal of unraveling the mysteries of special relativity amounts
to nothing more than finding out how to draw a diagram like k in the case where the two different sets of coordinates represent measurements of time and space made by two different observers, each in motion relative to the other. Galileo and Newton thought they knew the answer to this question, but their answer turned out to be only approximately right. To avoid repeating the same mistakes, we need to clearly spell out what we think are the basic properties of time and space that will be a reliable foundation for our reasoning. I want to emphasize that there is no purely logical way of deciding on this list of properties. The ones I’ll list are simply a summary of the patterns observed in the results from a large body of experiments. Furthermore, some of them are only approximate. For example, property 1 below is only a good approximation when the gravitational field is weak, so it is a property that applies to special relativity, not to general relativity.

Experiments show that:

1. No point in time or space has properties that make it different from any other point.
2. Likewise, all directions in space have the same properties.
3. Motion is relative, i.e., all inertial frames of reference are equally valid.
4. Causality holds, in the sense described on page 635.
5. Time depends on the state of motion of the observer.

Most of these are not very subversive. Properties 1 and 2 date back to the time when Galileo and Newton started applying the same universal laws of motion to the solar system and to the earth; this contradicted Aristotle, who believed that, for example, a rock would naturally want to move in a certain special direction (down) in order to reach a certain special location (the earth’s surface). Property 3 is the reason that Einstein called his theory “relativity,” but Galileo and Newton believed exactly the same thing to be true, as dramatized by Galileo’s run-in with the Church over the question of whether the earth could really be in motion around the sun. Example 3 on p. 247 describes a modern, high-precision experiment that can be interpreted as a test of this principle. Property 4 would probably surprise most people only because it asserts in such a weak and specialized way something that they feel deeply must be true. The only really strange item on the list is 5, but the Hafele-Keating experiment forces it upon us.
A Galilean version of the relationship between two frames of reference. As in all such graphs in this chapter, the original coordinates, represented by the gray rectangle, have a time axis that goes to the right, and a position axis that goes straight up.

A transformation that leads to disagreements about whether two events occur at the same time and place. This is not just a matter of opinion. Either the arrow hit the bull’s-eye or it didn’t. If it were not for property 5, the relativity of time, we could imagine that figure 1 would give the correct transformation between frames of reference in motion relative to one another. Let’s say that observer 1, whose grid coincides with the gray rectangle, is a hitch-hiker standing by the side of a road. Event A is a raindrop hitting his head, and event B is another raindrop hitting his head. He says that A and B occur at the same location in space. Observer 2 is a motorist who drives by without stopping; to him, the passenger compartment of his car is at rest, while the asphalt slides by underneath. He says that A and B occur at different points in space, because during the time between the first raindrop and the second, the hitch-hiker has moved backward. On the other hand, observer 2 says that events A and C occur in the same place, while the hitch-hiker disagrees. The slope of the grid-lines is simply the velocity of the relative motion of each observer relative to the other.

Figure 1 has familiar, comforting, and eminently sensible behavior, but it also happens to be wrong, because it violates property 5. The distortion of the coordinate grid has only moved the vertical lines up and down, so both observers agree that events like B and C are simultaneous. If this was really the way things worked, then all observers could synchronize all their clocks with one another forever, and the clocks would never get out of sync. This contradicts the results of the Hafele-Keating experiment, in which all three clocks were initially synchronized in Washington, but later went out of sync because of their different states of motion.

It might seem as though we still had a huge amount of wiggle room available for the correct form of the distortion. It turns out, however, that properties 1-5 are sufficient to prove that there is only one answer, which is the one found by Einstein in 1905. To see why this is, let’s work by a process of elimination.

Figure m shows a transformation that might seem at first glance to be as good a candidate as any other, but it violates property 3, that motion is relative, for the following reason. In observer 2’s frame of reference, some of the grid lines cross one another. This means that observers 1 and 2 disagree on whether or not certain events are the same. But as described on p. 671, the intersection of world-lines is supposed to be something that all observers agree on, and this means that they must be able to agree whether two events are the same or different. For instance, suppose that event A marks the arrival of an arrow at the bull’s-eye of a target, and event B is the location and time when the bull’s-eye is punctured. Events A and B occur at the same location and at the same time. If one observer says that A and B coincide, but another says that they don’t, we have a direct contradiction. Since the two frames of reference in figure m give contradictory results, one of them is right and one is wrong. This violates property 3, because all inertial frames of reference are supposed to be equally valid. To avoid problems like
this, we clearly need to make sure that none of the grid lines ever cross one another.

The next type of transformation we want to kill off is shown in figure n, in which the grid lines curve, but never cross one another. The trouble with this one is that it violates property 1, the uniformity of time and space. The transformation is unusually “twisty” at A, whereas at B it’s much more smooth. This can’t be correct, because the transformation is only supposed to depend on the relative state of motion of the two frames of reference, and that given information doesn’t single out a special role for any particular point in spacetime. If, for example, we had one frame of reference rotating relative to the other, then there would be something special about the axis of rotation. But we’re only talking about inertial frames of reference here, as specified in property 3, so we can’t have rotation; each frame of reference has to be moving in a straight line at constant speed. For frames related in this way, there is nothing that could single out an event like A for special treatment compared to B, so transformation n violates property 1.

The examples in figures m and n show that the transformation we’re looking for must be linear, meaning that it must transform lines into lines, and furthermore that it has to take parallel lines to parallel lines. Einstein wrote in his 1905 paper that “...on account of the property of homogeneity [property 1] which we ascribe to time and space, the [transformation] must be linear.”¹ Applying this to our diagrams, the original gray rectangle, which is a special type of parallelogram containing right angles, must be transformed into another parallelogram. There are three types of transformations, figure o, that have this property. Case I is the Galilean transformation of figure l on page 676, which we’ve already ruled out.

Case II can also be discarded. Here every point on the grid rotates counterclockwise. What physical parameter would determine the

---

amount of rotation? The only thing that could be relevant would be \( v \), the relative velocity of the motion of the two frames of reference with respect to one another. But if the angle of rotation was proportional to \( v \), then for large enough velocities the grid would have left and right reversed, and this would violate property 4, causality: one observer would say that event A caused a later event B, but another observer would say that B came first and caused A.

The only remaining possibility is case III, which I’ve redrawn in figure p with a couple of changes. This is the one that Einstein predicted in 1905. The transformation is known as the Lorentz transformation, after Hendrik Lorentz (1853-1928), who partially anticipated Einstein’s work, without arriving at the correct interpretation. The distortion is a kind of smooshing and stretching, as suggested by the hands. Also, we’ve already seen in figures i-k on page 674 that we’re free to stretch or compress everything as much as we like in the horizontal and vertical directions, because this simply corresponds to changing the units of measurement for time and distance. In figure p I’ve chosen units that give the whole drawing a convenient symmetry about a 45-degree diagonal line. Ordinarily it wouldn’t make sense to talk about a 45-degree angle on a graph whose axes had different units. But in relativity, the symmetric appearance of the transformation tells us that space and time ought to be treated on the same footing, and measured in the same units.
As in our discussion of the Galilean transformation, slopes are interpreted as velocities, and the slope of the near-horizontal lines in figure q is interpreted as the relative velocity of the two observers. The difference between the Galilean version and the relativistic one is that now there is smooshing happening from the other side as well. Lines that were vertical in the original grid, representing simultaneous events, now slant over to the right. This tells us that, as required by property 5, different observers do not agree on whether events that occur in different places are simultaneous. The Hafele-Keating experiment tells us that this non-simultaneity effect is fairly small, even when the velocity is as big as that of a passenger jet, and this is what we would have anticipated by the correspondence principle. The way that this is expressed in the graph is that if we pick the time unit to be the second, then the distance unit turns out to be hundreds of thousands of miles. In these units, the velocity of a passenger jet is an extremely small number, so the slope $v$ in a figure like q is extremely small, and the amount of distortion is tiny — it would be much too small to see on this scale.

The only thing left to determine about the Lorentz transformation is the size of the transformed parallelogram relative to the size of the original one. Although the drawing of the hands in figure p may suggest that the grid deforms like a framework made of rigid coat-hanger wire, that is not the case. If you look carefully at the figure, you’ll see that the edges of the smooshed parallelogram are actually a little longer than the edges of the original rectangle. In fact what stays the same is not lengths but areas, as proved in the caption to figure r.
The $\gamma$ factor

Figure q showed us that observers in different frames disagree on whether different events are simultaneous. This is an indication that time is not absolute, so we shouldn’t be surprised that time’s rate of flow is also different for different observers. We use the symbol $\gamma$ (Greek letter gamma) defined in the figure s to measure this unequal rate of flow. With a little algebra and geometry (homework problem 2, page 694), one can use the equal-area property to show that this ratio is given by

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

If you’ve had good training in physics, the first thing you probably think when you look at this equation is that it must be nonsense, because its units don’t make sense. How can we take something with units of velocity squared, and subtract it from a unitless 1? But remember that this is expressed in our new relativistic units, in which the same units are used for distance and time. We refer to these as natural units. In this system, velocities are always unitless. This sort of thing happens frequently in physics. For instance, before James Joule discovered conservation of energy, nobody knew that heat and mechanical energy were different forms of the same thing, so instead of measuring them both in units of joules as we would do now, they measured heat in one unit (such as calories) and mechanical energy in another (such as foot-pounds). In ordinary metric units, we just need an extra conversion factor, called $c$, and the equation becomes

$$\gamma = \frac{1}{\sqrt{1 - (\frac{v}{c})^2}}.$$ 

When we say, “It’s five hours from LA to Vegas,” we’re using a unit of time as a unit of distance. This works because there is a standard speed implied: the speed of a car on the freeway. Similarly, the conversion factor $c$ can be interpreted as a speed, so that $v/c$ is the unitless ratio of two speeds.

As argued on p. 635, cause and effect can never be propagated instantaneously; $c$ turns out to be the specific numerical speed limit on cause and effect. In particular, we’ll see in section 24.3 that light travels at $c$, which has a numerical value in SI units of $3.0 \times 10^8$ m/s.
Because $\gamma$ is always greater than 1, we have the following interpretation:

**Time dilation**
A clock runs fastest in the frame of reference of an observer who is at rest relative to the clock. An observer in motion relative to the clock at speed $v$ perceives the clock as running more slowly by a factor of $\gamma$.

As proved in figures u and v, lengths are also distorted:

**Length contraction**
A meter-stick appears longest to an observer who is at rest relative to it. An observer moving relative to the meter-stick at $v$ observes the stick to be shortened by a factor of $\gamma$.

$v / This figure proves, as claimed in figure u, that the length contraction is $x = 1/\gamma$. First we slice the parallelogram vertically like a salami and slide the slices down, making the top and bottom edges horizontal. Then we do the same in the horizontal direction, forming a rectangle with sides $\gamma$ and $x$. Since both the Lorentz transformation and the slicing processes leave areas unchanged, the area $\gamma x$ of the rectangle must equal the area of the original square, which is 1.

**self-check B**
What is $G$ when $v = 0$? What does this mean? $\Rightarrow$ Answer, p. 1044

---

**Changing an equation from natural units to SI example 1**
Often it is easier to do all of our algebra in natural units, which are simpler because $c = 1$, and all factors of $c$ can therefore be omitted. For example, suppose we want to solve for $v$ in terms of $\gamma$. In natural units, we have $\gamma = 1/\sqrt{1-v^2}$, so $\gamma^2 = 1-v^2$, and $v = \sqrt{1-\gamma^2}$.

This form of the result might be fine for many purposes, but if we wanted to find a value of $v$ in SI units, we would need to reinsert factors of $c$ in the final result. There is no need to do this throughout the whole derivation. By looking at the final result, we see that there is only one possible way to do this so that the results make sense in SI, which is to write $v = c\sqrt{1-\gamma^2}$.
Example 2

The motion of a certain ray of light is given by the equation \( x = -t \). Is this expressed in natural units, or in SI units? Convert to the other system.

The equation is in natural units. It wouldn’t make sense in SI units, because we would have meters on the left and seconds on the right. To convert to SI units, we insert a factor of \( c \) in the only possible place that will cause the equation to make sense: \( x = -ct \).

---

Example 3

An interstellar road trip

Alice stays on earth while her twin Betty heads off in a spaceship for Tau Ceti, a nearby star. Tau Ceti is 12 light-years away, so even though Betty travels at 87% of the speed of light, it will take her a long time to get there: 14 years, according to Alice.

Betty experiences time dilation. At this speed, her \( \gamma \) is 2.0, so that the voyage will only seem to her to last 7 years. But there is perfect symmetry between Alice’s and Betty’s frames of reference, so Betty agrees with Alice on their relative speed; Betty sees herself as being at rest, while the sun and Tau Ceti both move backward at 87% of the speed of light. How, then, can she observe Tau Ceti to get to her in only 7 years, when it should take 14 years to travel 12 light-years at this speed?

We need to take into account length contraction. Betty sees the distance between the sun and Tau Ceti to be shrunk by a factor of 2. The same thing occurs for Alice, who observes Betty and her spaceship to be foreshortened.
The correspondence principle

The correspondence principle requires that $\gamma$ be close to 1 for the velocities much less than $c$ encountered in everyday life. In natural units, $\gamma = (1 - v^2)^{-1/2}$. For small values of $v$, the approximation $(1 + e)^p \approx 1 + pe$ holds (see p. 1061). Applying this approximation, we find $\gamma \approx 1 + v^2/2$.

As expected, this gives approximately 1 when $v$ is small compared to 1 (i.e., compared to $c$, which equals 1 in natural units).

In problem 7 on p. 695 we rewrite this in SI units.

Figure t on p. 680 shows that the approximation is not valid for large values of $v/c$. In fact, $\gamma$ blows up to infinity as $v$ gets closer and closer to $c$.

A moving atomic clock

Example 4 shows that when $v$ is small, relativistic effects are approximately proportional to $v^2$, so it is very difficult to observe them at low speeds. For example, a car on the freeway travels at about 1/10 the speed of a passenger jet, so the resulting time dilation is only 1/100 as much. For this reason, it was not until four decades after Hafele and Keating that anyone did a conceptually simple atomic clock experiment in which the only effect was motion, not gravity; it is difficult to move a clock at a high enough velocity without putting it in some kind of aircraft, which then has to fly at some altitude. In 2010, however, Chou et al. succeeded in building an atomic clock accurate enough to detect time dilation at speeds as low as 10 m/s. Figure x shows their results. Since it was not practical to move the entire clock, the experimenters only moved the aluminum atoms inside the clock that actually made it “tick.”

Large time dilation

The time dilation effect in the Hafele-Keating experiment was very small. If we want to see a large time dilation effect, we can’t do it with something the size of the atomic clocks they used; the kinetic energy would be greater than the total megatonnage of all the world’s nuclear arsenals. We can, however, accelerate subatomic particles to speeds at which $\gamma$ is large. For experimental particle physicists, relativity is something you do all day before heading home and stopping off at the store for milk. An early, low-precision experiment of this kind was performed by Rossi and Hall in 1941, using naturally occurring cosmic rays.

Science 329 (2010) 1630
Example 6: Muons accelerated to nearly \( c \) undergo radioactive decay much more slowly than they would according to an observer at rest with respect to the muons. The first two data-points (unfilled circles) were subject to large systematic errors.

Figure y shows a 1974 experiment\(^3\) of a similar type which verified the time dilation predicted by relativity to a precision of about one part per thousand. Particles called muons (named after the Greek letter \( \mu \), “myoo”) were produced by an accelerator at CERN, near Geneva. A muon is essentially a heavier version of the electron. Muons undergo radioactive decay, lasting an average of only 2.197 \( \mu \)s before they evaporate into an electron and two neutrinos. The 1974 experiment was actually built in order to measure the magnetic properties of muons, but it produced a high-precision test of time dilation as a byproduct. Because muons have the same electric charge as electrons, they can be trapped using magnetic fields. Muons were injected into the ring shown in figure y, circling around it until they underwent radioactive decay. At the speed at which these muons were traveling, they had \( \gamma = 29.33 \), so on the average they lasted 29.33 times longer than the normal lifetime. In other words, they were like tiny alarm clocks that self-destructed at a randomly selected time. Figure z shows the number of radioactive decays counted, as a function of the time elapsed after a given stream of muons was injected into the storage ring. The two dashed lines show the rates of decay predicted with and without relativity. The relativistic line is the one that agrees with experiment.

\(^3\) Bailey at al., Nucl. Phys. B150(1979) 1
Example 7: In the garage’s frame of reference, the bus is moving, and fits in the garage due to length contraction. In the bus’s frame, the garage is moving, and can’t hold the bus due to its length contraction.

The garage paradox example 7

Suppose we take a schoolbus and drive it at relativistic speeds into a garage of ordinary size, in which it normally would not fit. Because of the length contraction, it fits. But the driver will perceive the garage as being contracted and thus even less able to contain the bus.

The paradox is resolved when we recognize that the concept of fitting the bus in the garage “all at once” contains a hidden assumption, the assumption that it makes sense to ask whether the front and back of the bus can simultaneously be in the garage. Observers in different frames of reference moving at high relative speeds do not necessarily agree on whether things happen simultaneously. As shown in figure aa, the person in the garage’s frame can shut the door at an instant B he perceives to be simultaneous with the front bumper’s arrival A at the back wall of the garage, but the driver would not agree about the simultaneity of these two events, and would perceive the door as having shut long after she plowed through the back wall.
Example 8: Colliding nuclei show relativistic length contraction.

Discussion questions

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

B A question that students often struggle with is whether time and space can really be distorted, or whether it just seems that way. Compare with optical illusions or magic tricks. How could you verify, for instance, that the lines in the figure are actually parallel? Are relativistic effects the same, or not?

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

D Mechanical clocks can be affected by motion. For example, it was a significant technological achievement to build a clock that could sail aboard a ship and still keep accurate time, allowing longitude to be determined. How is this similar to or different from relativistic time dilation?
Discussion question E: colliding nuclei show relativistic length contraction.

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

Discussion question G: The rod in the figure is perfectly rigid. At event A, the hammer strikes one end of the rod. At event B, the other end moves. Since the rod is perfectly rigid, it can’t compress, so A and B are simultaneous. In frame 2, B happens before A. Did the motion at the right end cause the person on the left to decide to pick up the hammer and use it?
ad / 1. When the circuit is incomplete, no current flows through the wire, and the magnet is unaffected. It points in the direction of the Earth's magnetic field. 2. The circuit is completed, and current flows through the wire. The wire has a strong effect on the magnet, which turns almost perpendicular to it. If the earth's field could be removed entirely, the compass would point exactly perpendicular to the wire; this is the direction of the wire's field.

23.3 Magnetic interactions

Think not that I am come to destroy the law, or the prophets: I am not come to destroy, but to fulfill.  
Matthew 5:17

At this stage, you understand roughly as much about the classification of interactions as physicists understood around the year 1800. There appear to be three fundamentally different types of interactions: gravitational, electrical, and magnetic. Many types of interactions that appear superficially to be distinct — stickiness, chemical interactions, the energy an archer stores in a bow — are really the same: they're manifestations of electrical interactions between atoms. Is there any way to shorten the list any further? The prospects seem dim at first. For instance, we find that if we rub a piece of fur on a rubber rod, the fur does not attract or repel a magnet. The fur has an electric field, and the magnet has a magnetic field. The two are completely separate, and don't seem to affect one another. Likewise we can test whether magnetizing a piece of iron changes its weight. The weight doesn't seem to change by any measurable amount, so magnetism and gravity seem to be unrelated.

That was where things stood until 1820, when the Danish physicist Hans Christian Oersted was delivering a lecture at the University of Copenhagen, and he wanted to give his students a demonstration that would illustrate the cutting edge of research. He generated a current in a wire by making a short circuit across a battery, and held the wire near a magnetic compass. The idea was to give an example of how one could search for a previously undiscovered link between electricity (the electric current in the wire) and magnetism. One never knows how much to believe from these dramatic legends, but the story is\(^4\) that the experiment he'd expected to turn out negative instead turned out positive: when he held the wire near the compass, the current in the wire caused the compass to twist!

People had tried similar experiments before, but only with static electricity, not with a moving electric current. For instance, they had hung batteries so that they were free to rotate in the earth's magnetic field, and found no effect; since the battery was not connected to a complete circuit, there was no current flowing. With Oersted's own setup, ad, the effect was only produced when the "circuit was closed, but not when open, as certain very celebrated physicists in vain attempted several years ago."\(^5\)

Oersted was eventually led to the conclusion that magnetism was an interaction between moving charges and other moving charges, i.e., between one current and another. A permanent magnet, he in-

---

\(^4\)Oersted's paper describing the phenomenon says that "The first experiments on the subject . . . were set on foot in the classes for electricity, galvanism, and magnetism, which were held by me in the winter just past," but that doesn't tell us whether the result was really a surprise that occurred in front of his students.

\(^5\)All quotes are from the 1876 translation by J.E. Kempe.
Magnetism is an interaction between moving charges and moving charges. The moving charges in the wire attract the moving charges in the beam of charged particles in the vacuum tube.

Figure af shows an example that is conceptually simple, but not very practical. If you try this with a typical vacuum tube, like a TV or computer monitor, the current in the wire probably won’t be enough to produce a visible effect. A more practical method is to hold a magnet near the screen. We still have an interaction between moving charges and moving charges, but the swirling electrons in the atoms in the magnet are now playing the role played by the moving charges in the wire in figure af. Warning: if you do this, make sure your monitor has a demagnetizing button! If not, then your monitor may be permanently ruined.

Relativity requires magnetism

So magnetism is an interaction between moving charges and moving charges. But how can that be? Relativity tells us that motion is a matter of opinion. Consider figure ag. In this figure and in figure ah, the dark and light coloring of the particles represents the fact that one particle has positive charge and the other negative. Observer ag/2 sees the two particles as flying through space side by side, so they would interact both electrically (simply because they’re charged) and magnetically (because they’re charges in motion). But an observer moving along with them, ag/1, would say they were both at rest, and would expect only an electrical interaction. This seems like a paradox. Magnetism, however, comes not to destroy relativity but to fulfill it. Magnetic interactions must exist according to the theory of relativity. To understand how this can be, consider how time and space behave in relativity. Observers in different frames of reference disagree about the lengths of measuring sticks and the speeds of clocks, but the laws of physics are valid and self-consistent in either frame of reference. Similarly, observers in different frames of reference disagree about what electric and magnetic fields there are, but they agree about concrete physical events. An observer in frame of reference ag/1 says there are electric fields around the particles, and predicts that as time goes on, the particles will begin to accelerate towards one another, eventually colliding. She explains the collision as being due to the electrical attraction between the particles. A different observer, ag/2, says the particles are moving. This observer also predicts that the particles will collide, but explains their motion in terms of both an electric field and a magnetic field. As we’ll see shortly, the magnetic field is required in order...
One observer sees an electric field, while the other sees both an electric field and a magnetic one.

A model of a charged particle and a current-carrying wire, seen in two different frames of reference. The relativistic length contraction is highly exaggerated. The force on the lone particle is purely magnetic in 1, and purely electric in 2.

To maintain consistency between the predictions made in the two frames of reference.

To see how this really works out, we need to find a nice simple example. An example like figure ag is not easy to handle, because in the second frame of reference, the moving charges create fields that change over time at any given location, like when the V-shaped wake of a speedboat washes over a buoy. Examples like figure af are easier, because there is a steady flow of charges, and all the fields stay the same over time. Figure ah/1 shows a simplified and idealized model of figure af. The charge by itself is like one of the charged particles in the vacuum tube beam of figure af, and instead of the wire, we have two long lines of charges moving in opposite directions. Note that, as discussed in discussion question C on page 589, the currents of the two lines of charges do not cancel out. The dark and light balls represent particles with opposite charges. Because of this, the total current in the “wire” is double what it would be if we took away one line.

As a model of figure af, figure ah/1 is partly realistic and partly unrealistic. In a real piece of copper wire, there are indeed charged particles of both types, but it turns out that the particles of one type (the protons) are locked in place, while only some of the other type (the electrons) are free to move. The model also shows the particles moving in a simple and orderly way, like cars on a two-lane road, whereas in reality most of the particles are organized into copper atoms, and there is also a great deal of random thermal motion. The model’s unrealistic features aren’t a problem, because the point of this exercise is only to find one particular situation that shows magnetic effects must exist based on relativity.

What electrical force does the lone particle in figure ah/1 feel? Since the density of “traffic” on the two sides of the “road” is equal, there is zero overall electrical force on the lone particle. Each “car” that attracts the lone particle is paired with a partner on the other side of the road that repels it. If we didn’t know about magnetism, we’d think this was the whole story: the lone particle feels no force at all from the wire.

Figure ah/2 shows what we’d see if we were observing all this from a frame of reference moving along with the lone charge. Here’s where the relativity comes in. Relativity tells us that moving objects appear contracted to an observer who is not moving along with them. Both lines of charge are in motion in both frames of reference, but in frame 1 they were moving at equal speeds, so their contractions were equal. In frame 2, however, their speeds are unequal. The dark charges are moving more slowly than in frame 1, so in frame 2 they are less contracted. The light-colored charges are moving more quickly, so their contraction is greater now. The “cars” on the two sides of the “road” are no longer paired off, so the electrical forces
on the lone particle no longer cancel out as they did in ah/1. The lone particle is attracted to the wire, because the particles attracting it are more dense than the ones repelling it. Furthermore, the attraction felt by the lone charge must be purely electrical, since the lone charge is at rest in this frame of reference, and magnetic effects occur only between moving charges and other moving charges.

Now observers in frames 1 and 2 disagree about many things, but they do agree on concrete events. Observer 2 is going to see the lone particle drift toward the wire due to the wire’s electrical attraction, gradually speeding up, and eventually hit the wire. If 2 sees this collision, then 1 must as well. But 1 knows that the total electrical force on the lone particle is exactly zero. There must be some new type of force. She invents a name for this new type of force: magnetism. This was a particularly simple example, because the force was purely magnetic in one frame of reference, and purely electrical in another. In general, an observer in a certain frame of reference will measure a mixture of electric and magnetic fields, while an observer in another frame, in motion with respect to the first, says that the same volume of space contains a different mixture.

We therefore arrive at the conclusion that electric and magnetic phenomena aren’t separate. They’re different sides of the same coin. We refer to electric and magnetic interactions collectively as electromagnetic interactions. Our list of the fundamental interactions of nature now has two items on it instead of three: gravity and electromagnetism.

Oersted found that magnetism was an interaction between moving charges and other moving charges. We can see this in the situation described in figure ah/1, in which the result of the argument depended on the fact that both the lone charge and the charges in the wire were moving. To see this in a different way, we can apply the result of example 4 on p. 683, that for small velocities the \( \gamma \) factor differs from 1 by about \( \frac{v^2}{2c^2} \). Let the lone charge in figure ah/1 have velocity \( u \), the ones in the wire \( \pm v \). As we’ll see on p. 704, velocities in relative motion don’t exactly add and subtract relativistically, but as long as we assume that \( u \) and \( v \) are small, the correspondence principle guarantees that they will approximately add and subtract. Then the velocities in the lone charge’s rest frame, ah/2, are approximately 0, \( v - u \), and \( -v - u \). The nonzero charge density of the wire in frame ah/2 is then proportional to the difference in the length contractions \( \gamma_{v-u} - \gamma_{v+u} \approx \frac{2uv}{c^2} \). This depends on the product of the velocities \( u \) and \( v \), which is as expected if magnetism is an interaction of moving charges with moving charges.

The basic rules for magnetic attractions and repulsions, shown in figure ai, aren’t quite as simple as the ones for gravity and electricity. Rules ai/1 and ai/2 follow directly from our previous analysis of figure ah. Rules 3 and 4 are obtained by flipping the type of
charge that the bottom particle has. For instance, rule 3 is like rule 1, except that the bottom charge is now the opposite type. This turns the attraction into a repulsion. (We know that flipping the charge reverses the interaction, because that’s the way it works for electric forces, and magnetic forces are just electric forces viewed in a different frame of reference.)

“A magnetic weathervane placed near a current.”

Figure 23.3.1 shows a magnetic weathervane, consisting of two charges that spin in circles around the axis of the arrow. (The magnetic field doesn’t cause them to spin; a motor is needed to get them to spin in the first place.) Just like the magnetic compass in figure ad, the weathervane’s arrow tends to align itself in the direction perpendicular to the wire. This is its preferred orientation because the charge close to the wire is attracted to the wire, while the charge far from the wire is repelled by it.

Discussion questions

A. In the situation shown in figure ah, is there a frame in which the force \( \mathbf{F} \) is a purely electric one, \( \mathbf{F}_E \)? Pure \( \mathbf{F}_B \)? Is there a frame in which the electromagnetic field is a pure \( \mathbf{E} \)? Pure \( \mathbf{B} \)? Is there zero net charge in both frames? One? Neither? What about the current?

B. For the situation shown in figure ah, draw a spacetime diagram showing the positive charges as black world-lines and the negative as red, in the wire’s rest frame. Use a ruler, and draw the spacing fairly accurately. Interpret this in the frame of the lone charge.

C. Resolve the following paradox concerning the argument given in this section. We would expect that at any given time, electrons in a solid would be associated with protons in a definite way. For simplicity, let’s imagine that the solid is made out of hydrogen (which actually does become a metal under conditions of very high pressure). A hydrogen atom consists of a single proton and a single electron. Even if the electrons are moving and forming an electric current, we would imagine that this would be like a game of musical chairs, with the protons as chairs and the electrons as people. Each electron has a proton that is its “friend,” at least for the moment. This is the situation shown in figure ah/1. How, then, can an observer in a different frame see the electrons and protons as not being paired up, as in ah/2?
Summary

Notation

\[ y = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \]

an abbreviation for \( 1/\sqrt{1 - v^2/c^2} \)

Summary

Experiments show that space and time do not have the properties claimed by Galileo and Newton. Time and space as seen by one observer are distorted compared to another observer’s perceptions if they are moving relative to each other. This distortion is quantified by the factor

\[ y = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \]

where \( v \) is the relative velocity of the two observers, and \( c \) is a universal velocity that is the same in all frames of reference. Light travels at \( c \). A clock appears to run fastest to an observer who is not in motion relative to it, and appears to run too slowly by a factor of \( y \) to an observer who has a velocity \( v \) relative to the clock. Similarly, a meter-stick appears longest to an observer who sees it at rest, and appears shorter to other observers. Time and space are relative, not absolute.

As a consequence of relativity, we must have not just electrical interactions of charges with charges, but also an additional magnetic interaction of moving charges with other moving charges.
Problems

Key
√ A computerized answer check is available online.
∫ A problem that requires calculus.
⋆ A difficult problem.

1 What happens in the equation for $\gamma$ when you put in a negative number for $v$? Explain what this means physically, and why it makes sense.

2 In this homework problem, you’ll fill in the steps of the algebra required in order to find the equation for $\gamma$ on page 680. To keep the algebra simple, let the time $t$ in figure s equal 1, as suggested in the figure accompanying this homework problem. The original square then has an area of 1, and the transformed parallelogram must also have an area of 1. (a) Prove that point P is at $x = v\gamma$, so that its $(t, x)$ coordinates are $(\gamma, v\gamma)$. (b) Find the $(t, x)$ coordinates of point Q. (c) Find the length of the short diagonal connecting P and Q. (d) Average the coordinates of P and Q to find the coordinates of the midpoint C of the parallelogram, and then find distance OC. (e) Find the area of the parallelogram by computing twice the area of triangle PQO. [Hint: You can take PQ to be the base of the triangle.] (f) Set this area equal to 1 and solve for $\gamma$ to prove $\gamma = 1/\sqrt{1 - v^2}$.

3 Astronauts in three different spaceships are communicating with each other. Those aboard ships A and B agree on the rate at which time is passing, but they disagree with the ones on ship C. (a) Alice is aboard ship A. How does she describe the motion of her own ship, in its frame of reference? (b) Describe the motion of the other two ships according to Alice. (c) Give the description according to Betty, whose frame of reference is ship B. (d) Do the same for Cathy, aboard ship C.

4 The Voyager 1 space probe, launched in 1977, is moving faster relative to the earth than any other human-made object, at 17,000 meters per second. (a) Calculate the probe’s $\gamma$. (b) Over the course of one year on earth, slightly less than one year passes on the probe. How much less? (There are 31 million seconds in a year.)

5 The earth is orbiting the sun, and therefore is contracted relativistically in the direction of its motion. Compute the amount by which its diameter shrinks in this direction.
Page 675 lists five observed properties of space and time that sufficed to derive relativity. There are other sets of postulates that could also have been used, and in fact this was not the set that was first used by Einstein. Suppose that our first inkling that spacetime isn’t Galilean consists of an experiment like the one in figure af on p. 689, and that we interpret it, correctly, as being evidence of a relativistic length contraction as in figure ah on p. 690. Besides this fact, what else is needed in order to get the full structure of relativity?

Natural relativistic units were introduced on p. 680, and examples 1 and 2 on pp. 681 and 682 gave examples of how to convert an equation from natural units to SI units. In example 4 on p. 683, we derived the approximation

\[ \gamma \approx 1 + \frac{v^2}{2} \]

for values of \( v \) that are small compared to 1 (i.e., small compared to the speed of light in natural units). As in the other examples, convert this equation to SI units. > Solution, p. 1033

We want to throw a ball of diameter \( b \) through a hole of diameter \( h \) in a thin wall. Clearly this is possible if \( b < h \), but consider the case where \( b > h \). If the motion is relativistic, then is it unambiguous whether the ball fits through the hole, or is this frame-dependent, as in example 7 on p. 685? If the former, then is there some velocity \( v \) that is required, expressible in terms of \( b \) and \( h \)?
Chapter 24
Electromagnetism

24.1 The magnetic field

No magnetic monopoles

If you could play with a handful of electric dipoles and a handful of bar magnets, they would appear very similar. For instance, a pair of bar magnets wants to align themselves head-to-tail, and a pair of electric dipoles does the same thing. (It is unfortunately not that easy to make a permanent electric dipole that can be handled like this, since the charge tends to leak.)

You would eventually notice an important difference between the two types of objects, however. The electric dipoles can be broken apart to form isolated positive charges and negative charges. The two-ended device can be broken into parts that are not two-ended. But if you break a bar magnet in half, a, you will find that you have simply made two smaller two-ended objects.

The reason for this behavior is not hard to divine from our microscopic picture of permanent iron magnets. An electric dipole has extra positive “stuff” concentrated in one end and extra negative in the other. The bar magnet, on the other hand, gets its magnetic properties not from an imbalance of magnetic “stuff” at the two ends but from the orientation of the rotation of its electrons. One end is the one from which we could look down the axis and see the electrons rotating clockwise, and the other is the one from which they would appear to go counterclockwise. There is no difference between the “stuff” in one end of the magnet and the other, b.

Nobody has ever succeeded in isolating a single magnetic pole. In technical language, we say that magnetic monopoles do not seem to exist. Electric monopoles do exist — that’s what charges are.

Electric and magnetic forces seem similar in many ways. Both act at a distance, both can be either attractive or repulsive, and both are intimately related to the property of matter called charge. (Recall that magnetism is an interaction between moving charges.) Physicists’s aesthetic senses have been offended for a long time because this seeming symmetry is broken by the existence of electric monopoles and the absence of magnetic ones. Perhaps some exotic form of matter exists, composed of particles that are magnetic monopoles. If such particles could be found in cosmic rays...
The unit of magnetic field, the tesla, is named after Serbian-American inventor Nikola Tesla.

A standard dipole made from a square loop of wire shorting across a battery. It acts very much like a bar magnet, but its strength is more easily quantified.

A dipole tends to align itself to the surrounding magnetic field.

or moon rocks, it would be evidence that the apparent asymmetry was only an asymmetry in the composition of the universe, not in the laws of physics. For these admittedly subjective reasons, there have been several searches for magnetic monopoles. Experiments have been performed, with negative results, to look for magnetic monopoles embedded in ordinary matter. Soviet physicists in the 1960's made exciting claims that they had created and detected magnetic monopoles in particle accelerators, but there was no success in attempts to reproduce the results there or at other accelerators. The most recent search for magnetic monopoles, done by reanalyzing data from the search for the top quark at Fermilab, turned up no candidates, which shows that either monopoles don’t exist in nature or they are extremely massive and thus hard to create in accelerators.

**Definition of the magnetic field**

Since magnetic monopoles don’t seem to exist, it would not make much sense to define a magnetic field in terms of the force on a test monopole. Instead, we follow the philosophy of the alternative definition of the electric field, and define the field in terms of the torque on a magnetic test dipole. This is exactly what a magnetic compass does: the needle is a little iron magnet which acts like a magnetic dipole and shows us the direction of the earth’s magnetic field.

To define the strength of a magnetic field, however, we need some way of defining the strength of a test dipole, i.e., we need a definition of the magnetic dipole moment. We could use an iron permanent magnet constructed according to certain specifications, but such an object is really an extremely complex system consisting of many iron atoms, only some of which are aligned. A more fundamental standard dipole is a square current loop. This could be a little resistive circuit consisting of a square of wire shorting across a battery.

We will find that such a loop, when placed in a magnetic field, experiences a torque that tends to align plane so that its face points in a certain direction. (Since the loop is symmetric, it doesn’t care if we rotate it like a wheel without changing the plane in which it lies.) It is this preferred facing direction that we will end up defining as the direction of the magnetic field.

Experiments show if the loop is out of alignment with the field, the torque on it is proportional to the amount of current, and also to the interior area of the loop. The proportionality to current makes sense, since magnetic forces are interactions between moving charges, and current is a measure of the motion of charge. The proportionality to the loop’s area is also not hard to understand, because increasing the length of the sides of the square increases both the amount of charge contained in this circular “river” and
the amount of leverage supplied for making torque. Two separate physical reasons for a proportionality to length result in an overall proportionality to length squared, which is the same as the area of the loop. For these reasons, we define the magnetic dipole moment of a square current loop as

\[ D_m = IA, \]

[definition of the magnetic dipole moment of a square current loop]

We now define the magnetic field in a manner entirely analogous to the alternative definition of the electric field on p. 646:

**definition of the magnetic field**

The magnetic field vector, \( \mathbf{B} \), at any location in space is defined by observing the torque exerted on a magnetic test dipole \( D_{mt} \) consisting of a square current loop. The field’s magnitude is \( |\mathbf{B}| = \tau/D_{mt} \sin \theta \), where \( \theta \) is the angle by which the loop is misaligned. The direction of the field is perpendicular to the loop; of the two perpendiculars, we choose the one such that if we look along it, the loop’s current is counterclockwise.

We find from this definition that the magnetic field has units of N·m/A·m² = N/A·m. This unwieldy combination of units is abbreviated as the tesla, 1 T = 1 N/A·m. Refrain from memorizing the part about the counterclockwise direction at the end; in section 24.5 we’ll see how to understand this in terms of more basic principles.

The nonexistence of magnetic monopoles means that unlike an electric field, d/1, a magnetic one, d/2, can never have sources or sinks. The magnetic field vectors lead in paths that loop back on themselves, without ever converging or diverging at a point.

**Relativity**

The definition of the tesla as 1 N/A·m looks messy, but relativity suggests a simple explanation. We saw in section 23.3 that a particular mixture of electric and magnetic fields appears to be a different mixture in a different frame of reference. In a system of units designed for relativity, \( \mathbf{E} \) and \( \mathbf{B} \) have the same units. The SI, which predates relativity, wasn’t designed this way, which is also why c is some number with units of m/s rather than simply equaling 1. The SI units of the \( \mathbf{E} \) and \( \mathbf{B} \) fields are almost the same: they differ only by a factor of m/s.

Figure e shows that something similar to the parallelogram diagrams developed in ch. 23 also works as a way of representing the transformation of the \( \mathbf{E} \) and \( \mathbf{B} \) fields from one frame of reference to another. One frame is moving relative to the other in the \( x \) direction, and this mixes certain components of the fields. A dot on the graph represents a particular set of fields, which are seen by each observer according to her own coordinate grid.
24.2 Calculating magnetic fields and forces

Magnetostatics

Our study of the electric field built on our previous understanding of electric forces, which was ultimately based on Coulomb's law for the electric force between two point charges. Since magnetism is ultimately an interaction between currents, i.e., between moving charges, it is reasonable to wish for a magnetic analog of Coulomb's law, an equation that would tell us the magnetic force between any two moving point charges.

Such a law, unfortunately, does not exist. Coulomb's law describes the special case of electrostatics: if a set of charges is sitting around and not moving, it tells us the interactions among them. Coulomb's law fails if the charges are in motion, since it does not incorporate any allowance for the time delay in the outward propagation of a change in the locations of the charges.

A pair of moving point charges will certainly exert magnetic forces on one another, but their magnetic fields are like the v-shaped bow waves left by boats. Each point charge experiences a magnetic field that originated from the other charge when it was at some previous position. There is no way to construct a force law that tells us the force between them based only on their current positions in space.

There is, however, a science of magnetostatics that covers a great many important cases. Magnetostatics describes magnetic forces among currents in the special case where the currents are steady and continuous, leading to magnetic fields throughout space that do not change over time.

The magnetic field of a long, straight wire is one example that we can say something about without resorting to fancy mathematics. We saw in examples 4 on p. 643 and 15 on p. 657 that the electric field of a uniform line of charge is \( E = \frac{2kq}{Lr} \), where \( r \) is the distance from the line and \( q/L \) is the charge per unit length. In a frame of reference moving at velocity \( v \) parallel to the line, this electric field will be observed as a combination of electric and magnetic fields. It therefore follows that the magnetic field of a long, straight, current-carrying wire must be proportional to \( 1/r \). We also expect that it will be proportional to the Coulomb constant, which sets the strength of electric and magnetic interactions, and to the current \( I \) in the wire. The complete expression turns out to be \( B = \left( \frac{k}{c^2} \right) \left( \frac{2I}{r} \right) \). This is identical to the expression for \( E \) except for replacement of \( q/L \) with \( I \) and an additional factor of \( 1/c^2 \). The latter occurs because magnetism is a purely relativistic effect, and the relativistic length contraction depends on \( v^2/c^2 \).