

Figure f shows the equations for some of the more commonly encountered configurations, with illustrations of their field patterns. They all have a factor of k/c^2 in front, which shows that magnetism is just electricity (k) seen through the lens of relativity ($1/c^2$). A convenient feature of SI units is that k/c^2 has a numerical value of exactly 10^{-7} , with units of N/A^2 .

Field created by a long, straight wire carrying current I :

$$B = \frac{k}{c^2} \cdot \frac{2I}{r}$$

Here r is the distance from the center of the wire. The field vectors trace circles in planes perpendicular to the wire, going clockwise when viewed from along the direction of the current.

Field created by a single circular loop of current:

The field vectors form a dipole-like pattern, coming through the loop and back around on the outside. Each oval path traced out by the field vectors appears clockwise if viewed from along the direction the current is going when it punches through it. There is no simple equation for a field at an arbitrary point in space, but for a point lying *along the central axis* perpendicular to the loop, the field is

$$B = \frac{k}{c^2} \cdot 2\pi I b^2 (b^2 + z^2)^{-3/2},$$

where b is the radius of the loop and z is the distance of the point from the plane of the loop.

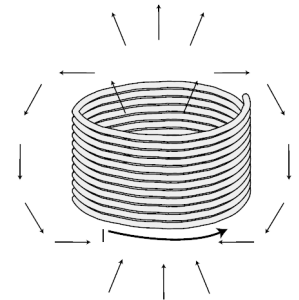
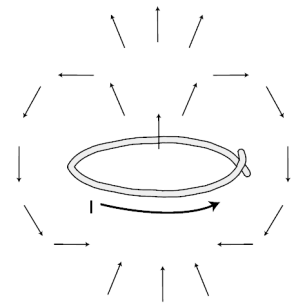
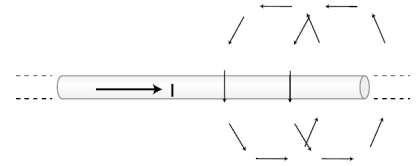
Field created by a solenoid (cylindrical coil):

The field pattern is similar to that of a single loop, but for a long solenoid the paths of the field vectors become very straight on the inside of the coil and on the outside immediately next to the coil. For a sufficiently long solenoid, the interior field also becomes very nearly uniform, with a magnitude of

$$B = \frac{k}{c^2} \cdot 4\pi I N / \ell,$$

where N is the number of turns of wire and ℓ is the length of the solenoid. The field near the mouths or outside the coil is not constant, and is more difficult to calculate. For a long solenoid, the exterior field is much smaller than the interior field.

Don't memorize the equations!



f / Some magnetic fields.

Force on a charge moving through a magnetic field

We now know how to calculate magnetic fields in some typical situations, but one might also like to be able to calculate magnetic forces, such as the force of a solenoid on a moving charged particle, or the force between two parallel current-carrying wires.

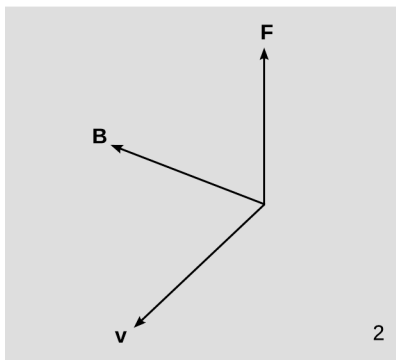
We will restrict ourselves to the case of the force on a charged particle moving through a magnetic field, which allows us to calculate the force between two objects when one is a moving charged particle and the other is one whose magnetic field we know how to find. An example is the use of solenoids inside a TV tube to guide the electron beam as it paints a picture.

Experiments show that the magnetic force on a moving charged particle has a magnitude given by

$$|\mathbf{F}| = q|\mathbf{v}||\mathbf{B}|\sin\theta,$$

where \mathbf{v} is the velocity vector of the particle, and θ is the angle between the \mathbf{v} and \mathbf{B} vectors. Unlike electric and gravitational forces, magnetic forces do not lie along the same line as the field vector. The force is always *perpendicular* to both \mathbf{v} and \mathbf{B} . Given two vectors, there is only one line perpendicular to both of them, so the force vector points in one of the two possible directions along this line. For a positively charged particle, the direction of the force vector can be found as follows. First, position the \mathbf{v} and \mathbf{B} vectors with their tails together. The direction of \mathbf{F} is such that if you sight along it, the \mathbf{B} vector is clockwise from the \mathbf{v} vector; for a negatively charged particle the direction of the force is reversed. Note that since the force is perpendicular to the particle's motion, the magnetic field never does work on it.

If we place a moving test charge in a magnetic field, we can use the equation $|\mathbf{F}| = q|\mathbf{v}||\mathbf{B}|\sin\theta$ and the geometrical relationship discussed above to indirectly determine \mathbf{B} . (More than one measurement will in general be required.) This can also serve as a definition of the magnetic field, analogous to the one on p. 641 for the electric field.



Example 1.

Magnetic levitation

example 1

In figure 24.2.2, a small, disk-shaped permanent magnet is stuck on the side of a battery, and a wire is clasped loosely around the battery, shorting it. A large current flows through the wire. The electrons moving through the wire feel a force from the magnetic field made by the permanent magnet, and this force levitates the wire.

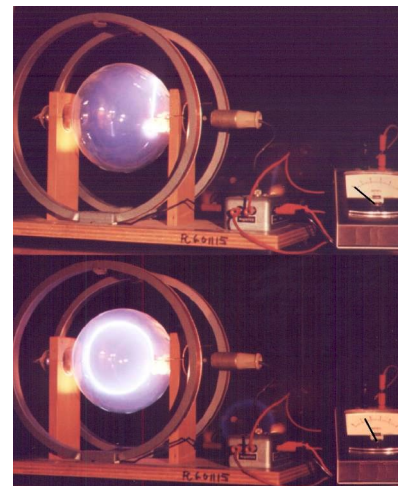
From the photo, it's possible to find the direction of the magnetic field made by the permanent magnet. The electrons in the copper wire are negatively charged, so they flow from the negative (flat) terminal of the battery to the positive terminal (the one with the

bump, in front). As the electrons pass by the permanent magnet, we can imagine that they would experience a field either toward the magnet, or away from it, depending on which way the magnet was flipped when it was stuck onto the battery. Imagine sighting along the upward force vector, which you could do if you were a tiny bug lying on your back underneath the wire. Since the electrons are negatively charged, the \mathbf{B} vector must be counter-clockwise from the \mathbf{v} vector, which means toward the magnet.

A circular orbit

example 2

Magnetic forces cause a beam of electrons to move in a circle. The beam is created in a vacuum tube, in which a small amount of hydrogen gas has been left. A few of the electrons strike hydrogen molecules, creating light and letting us see the beam. A magnetic field is produced by passing a current (meter) through the circular coils of wire in front of and behind the tube. In the bottom figure, with the magnetic field turned on, the force perpendicular to the electrons' direction of motion causes them to move in a circle.



Example 2.

Nervous-system effects during an MRI scan

example 3

During an MRI scan of the head, the patient's nervous system is exposed to intense magnetic fields, and there are ions moving around in the nerves. The resulting forces on the ions can cause symptoms such as vertigo.

Energy in the magnetic field

On p. 649 I gave equations for the energy stored in the gravitational and electric fields. Since a magnetic field is essentially an electric field seen in a different frame of reference, we expect the magnetic-field equation to be closely analogous to the electric version, and it is:

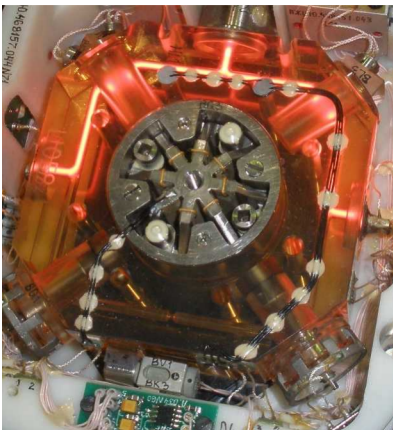
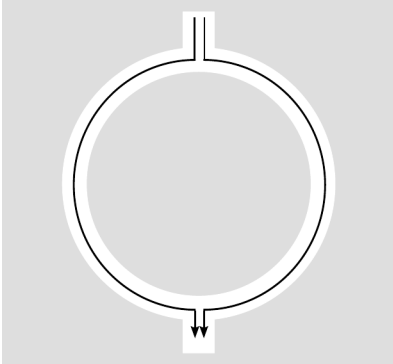
$$(\text{energy stored in the gravitational field per m}^3) = -\frac{1}{8\pi G}|\mathbf{g}|^2$$

$$(\text{energy stored in the electric field per m}^3) = \frac{1}{8\pi k}|\mathbf{E}|^2$$

$$(\text{energy stored in the magnetic field per m}^3) = \frac{c^2}{8\pi k}|\mathbf{B}|^2$$

The idea here is that k/c^2 is the magnetic version of the electric quantity k , the $1/c^2$ representing the fact that magnetism is a relativistic effect.

Solenoids are very common electrical devices, but they can be a hazard to someone who is working on them. Imagine a solenoid that initially has a DC current passing through it. The current creates a magnetic field inside and around it, which contains energy. Now suppose that we break the circuit. Since there is no longer a complete circuit, current will quickly stop flowing, and the magnetic field will collapse very quickly. The field had energy stored in it, and even a small amount of energy can create a dangerous power surge if released over a short enough time interval. It is prudent not to fiddle with a solenoid that has current flowing through it, since breaking the circuit could be hazardous to your health.



As a typical numerical estimate, let's assume a 40 cm × 40 cm × 40 cm solenoid with an interior magnetic field of 1.0 T (quite a strong field). For the sake of this rough estimate, we ignore the exterior field, which is weak, and assume that the solenoid is cubical in shape. The energy stored in the field is

$$\begin{aligned} (\text{energy per unit volume})(\text{volume}) &= \frac{c^2}{8\pi k} |\mathbf{B}|^2 V \\ &= 3 \times 10^4 \text{ J} \end{aligned}$$

That's a lot of energy!

24.3 The universal speed c

Let's think a little more about the role of the 45-degree diagonal in the Lorentz transformation. Slopes on these graphs are interpreted as velocities. This line has a slope of 1 in relativistic units, but that slope corresponds to c in ordinary metric units. We already know that the relativistic distance unit must be extremely large compared to the relativistic time unit, so c must be extremely large. Now note what happens when we perform a Lorentz transformation: this particular line gets stretched, but the new version of the line lies right on top of the old one, and its slope stays the same. In other words, if one observer says that something has a velocity equal to c , every other observer will agree on that velocity as well. (The same thing happens with $-c$.)

Velocities don't simply add and subtract.

This is counterintuitive, since we expect velocities to add and subtract in relative motion. If a dog is running away from me at 5 m/s relative to the sidewalk, and I run after it at 3 m/s, the dog's velocity in my frame of reference is 2 m/s. According to everything we have learned about motion (p. 82), the dog must have different speeds in the two frames: 5 m/s in the sidewalk's frame and 2 m/s

g/If you've flown in a jet plane, you can thank relativity for helping you to avoid crashing into a mountain or an ocean. The figure shows a standard piece of navigational equipment called a ring laser gyroscope. A beam of light is split into two parts, sent around the perimeter of the device, and reunited. Since light travels at the universal speed c , which is constant, we expect the two parts to come back together at the same time. If they don't, it's evidence that the device has been rotating. The plane's computer senses this and notes how much rotation has accumulated.

in mine. But velocities are measured by dividing a distance by a time, and both distance and time are distorted by relativistic effects, so we actually shouldn't expect the ordinary arithmetic addition of velocities to hold in relativity; it's an approximation that's valid at velocities that are small compared to c .

A universal speed limit

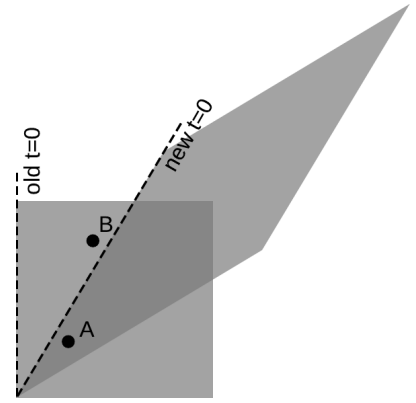
For example, suppose Janet takes a trip in a spaceship, and accelerates until she is moving at $0.6c$ relative to the earth. She then launches a space probe in the forward direction at a speed relative to her ship of $0.6c$. We might think that the probe was then moving at a velocity of $1.2c$, but in fact the answer is still less than c (problem 1, page 722). This is an example of a more general fact about relativity, which is that c represents a universal speed limit. This is required by causality, as shown in figure h.

Light travels at c .

Now consider a beam of light. We're used to talking casually about the "speed of light," but what does that really mean? Motion is relative, so normally if we want to talk about a velocity, we have to specify what it's measured relative to. A sound wave has a certain speed relative to the air, and a water wave has its own speed relative to the water. If we want to measure the speed of an ocean wave, for example, we should make sure to measure it in a frame of reference at rest relative to the water. But light isn't a vibration of a physical medium; it can propagate through the near-perfect vacuum of outer space, as when rays of sunlight travel to earth. This seems like a paradox: light is supposed to have a specific speed, but there is no way to decide what frame of reference to measure it in. The way out of the paradox is that light must travel at a velocity equal to c . Since all observers agree on a velocity of c , regardless of their frame of reference, everything is consistent.

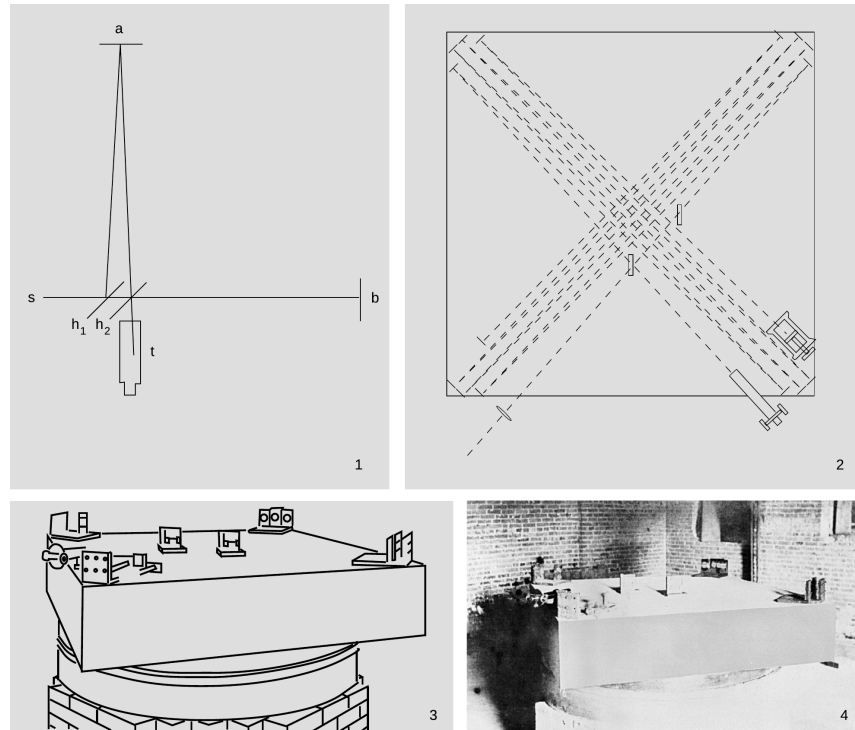
The Michelson-Morley experiment

The constancy of the speed of light had in fact already been observed when Einstein was an 8-year-old boy, but because nobody could figure out how to interpret it, the result was largely ignored. In 1887 Michelson and Morley set up a clever apparatus to measure any difference in the speed of light beams traveling east-west and north-south. The motion of the earth around the sun at 110,000 km/hour (about 0.01% of the speed of light) is to our west during the day. Michelson and Morley believed that light was a vibration of a mysterious medium called the ether, so they expected that the speed of light would be a fixed value relative to the ether. As the earth moved through the ether, they thought they would observe an effect on the velocity of light along an east-west line. For instance, if they released a beam of light in a westward direction during the day, they expected that it would move away from them at less than the normal



h / A proof that causality imposes a universal speed limit. In the original frame of reference, represented by the square, event A happens a little before event B. In the new frame, shown by the parallelogram, A happens after $t = 0$, but B happens before $t = 0$; that is, B happens before A. The time ordering of the two events has been reversed. This can only happen because events A and B are very close together in time and fairly far apart in space. The line segment connecting A and B has a slope greater than 1, meaning that if we wanted to be present at both events, we would have to travel at a speed greater than c (which equals 1 in the units used on this graph). You will find that if you pick any two points for which the slope of the line segment connecting them is less than 1, you can never get them to straddle the new $t = 0$ line in this funny, time-reversed way. Since different observers disagree on the time order of events like A and B, causality requires that information never travel from A to B or from B to A; if it did, then we would have time-travel paradoxes. The conclusion is that c is the maximum speed of cause and effect in relativity.

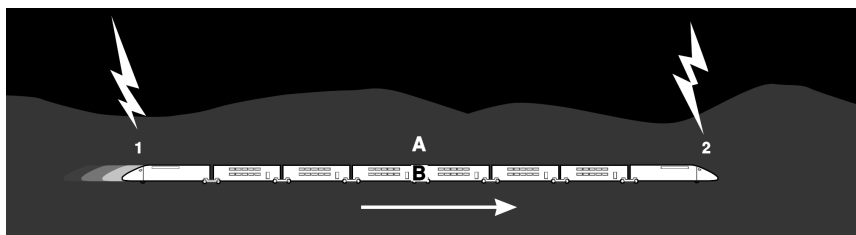
speed because the earth was chasing it through the ether. They were surprised when they found that the expected 0.01% change in the speed of light did not occur.



The Michelson-Morley experiment, shown in photographs, and drawings from the original 1887 paper. 1. A simplified drawing of the apparatus. A beam of light from the source, s , is partially reflected and partially transmitted by the half-silvered mirror h_1 . The two half-intensity parts of the beam are reflected by the mirrors at a and b , reunited, and observed in the telescope, t . If the earth's surface was supposed to be moving through the ether, then the times taken by the two light waves to pass through the moving ether would be unequal, and the resulting time lag would be detectable by observing the interference between the waves when they were reunited. 2. In the real apparatus, the light beams were reflected multiple times. The effective length of each arm was increased to 11 meters, which greatly improved its sensitivity to the small expected difference in the speed of light. 3. In an earlier version of the experiment, they had run into problems with its "extreme sensitiveness to vibration," which was "so great that it was impossible to see the interference fringes except at brief intervals . . . even at two o'clock in the morning." They therefore mounted the whole thing on a massive stone floating in a pool of mercury, which also made it possible to rotate it easily. 4. A photo of the apparatus.

Discussion questions

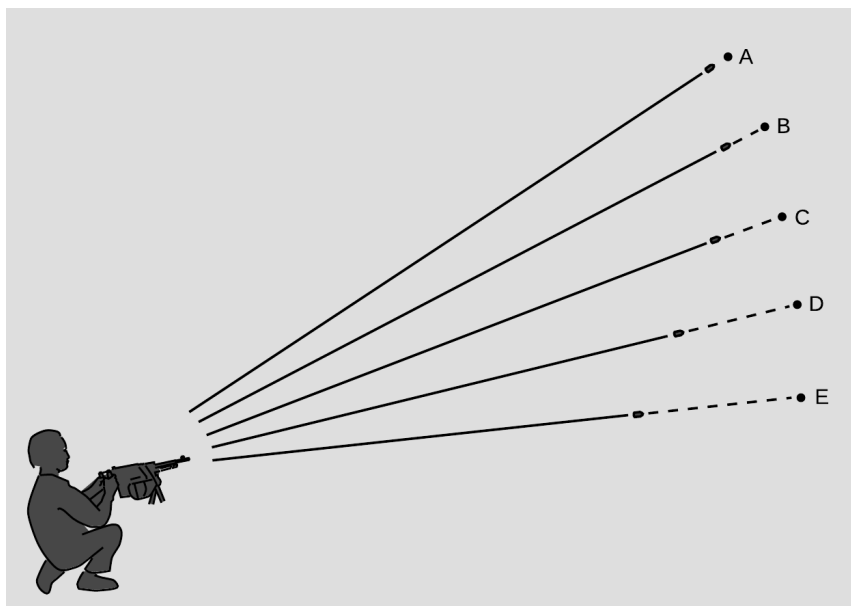
A The figure shows a famous thought experiment devised by Einstein. A train is moving at constant velocity to the right when bolts of lightning strike the ground near its front and back. Alice, standing on the dirt at the midpoint of the flashes, observes that the light from the two flashes arrives simultaneously, so she says the two strikes must have occurred simultaneously. Bob, meanwhile, is sitting aboard the train, at its middle. He passes by Alice at the moment when Alice later figures out that the flashes happened. Later, he receives flash 2, and then flash 1. He infers that since both flashes traveled half the length of the train, flash 2 must have occurred first. How can this be reconciled with Alice's belief that the flashes were simultaneous? Explain using a graph.



B Use a graph to resolve the following relativity paradox. Relativity says that in one frame of reference, event A could happen before event B, but in someone else's frame B would come before A. How can this be? Obviously the two people could meet up at A and talk as they cruised past each other. Wouldn't they have to agree on whether B had already happened?

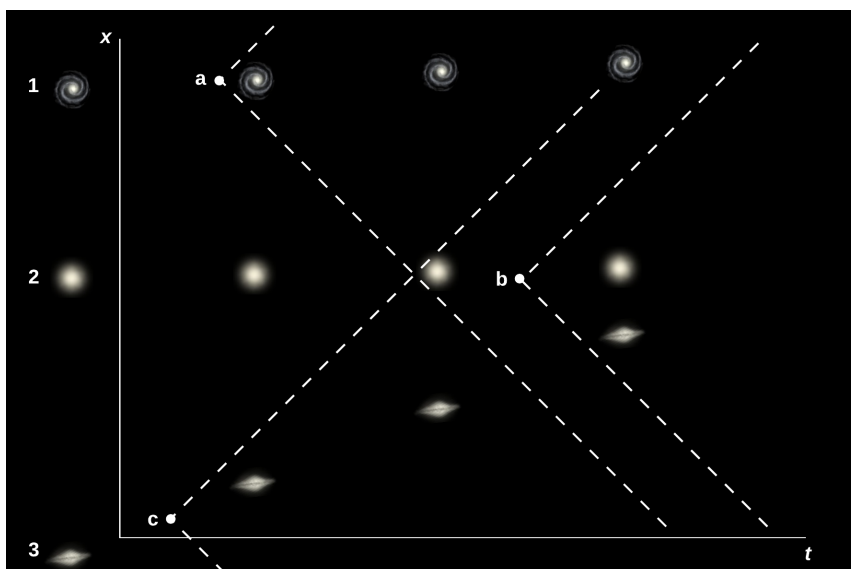
C The machine-gunner in the figure sends out a spray of bullets. Suppose that the bullets are being shot into outer space, and that the distances traveled are trillions of miles (so that the human figure in the diagram is not to scale). After a long time, the bullets reach the points shown with dots which are all equally far from the gun. Their arrivals at those points are events A through E, which happen at different times. Sketch these events on a position-time graph. The chain of impacts extends across space at a speed greater than c . Does this violate special relativity?

Discussion question C.



D The graph shows three galaxies. The axes are drawn according to an observer at rest relative to the galaxy 2, so that that galaxy is always at the same x coordinate. Intelligent species in the three different galaxies develop radio technology independently, and at some point each begins to actively send out signals in an attempt to communicate with other civilizations. Events a , b , and c mark the points at which these signals begin spreading out across the universe at the speed of light. Find the events at which the inhabitants of galaxy 2 detect the signals from galaxies 1 and 3. According to 2, who developed radio first, 1 or 3? On top of the graph, draw a new pair of position and time axes, for the frame in which galaxy 3 is at rest. According to 3, in what order did events a , b , and c happen?

Discussion question D.



24.4 Induction

The principle of induction

Physicists of Michelson and Morley's generation thought that light was a mechanical vibration of the ether, but we now know that it is a ripple in the electric and magnetic fields. With hindsight, relativity essentially requires this:

1. Relativity requires that changes in any field propagate as waves at a finite speed (p. 635).
2. Relativity says that if a wave has a fixed speed but is not a mechanical disturbance in a physical medium, then it must travel at the universal velocity c (p. 705).

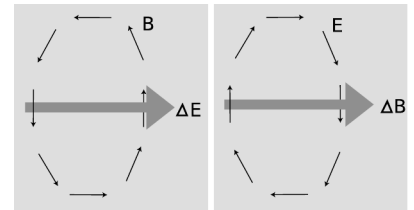
What is less obvious is that there are not two separate kinds of waves, electric and magnetic. In fact an electric wave can't exist without a magnetic one, or a magnetic one without an electric one. This new fact follows from the principle of induction, which was discovered experimentally by Faraday in 1831, seventy-five years before Einstein. Let's state Faraday's idea first, and then see how something like it must follow inevitably from relativity:

the principle of induction

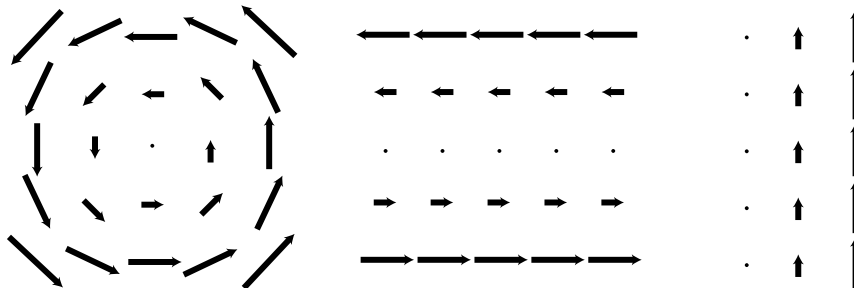
Any electric field that changes over time will produce a magnetic field in the space around it.

Any magnetic field that changes over time will produce an electric field in the space around it.

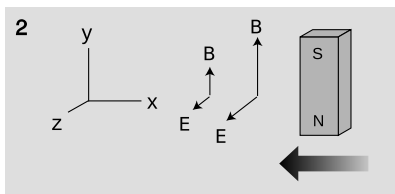
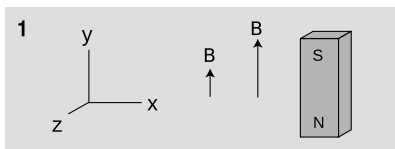
The induced field tends to have a whirlpool pattern, as shown in figure l, but the whirlpool image is not to be taken too literally; the principle of induction really just requires a field pattern such that, if one inserted a paddlewheel in it, the paddlewheel would spin. All of the field patterns shown in figure m are ones that could be created by induction; all have a counterclockwise "curl" to them.



l / The geometry of induced fields. The induced field tends to form a whirlpool pattern around the change in the vector producing it. Note how they circulate in opposite directions.



m / Three fields with counterclockwise "curls."

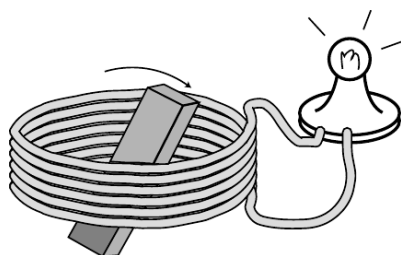


n / Observer 1 is at rest with respect to the bar magnet, and observes magnetic fields that have different strengths at different distances from the magnet. Observer 2, hanging out in the region to the left of the magnet, sees the magnet moving toward her, and detects that the magnetic field in that region is getting stronger as time passes.

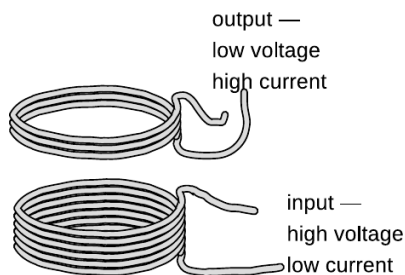
Figure n shows an example of the fundamental reason why a changing \mathbf{B} field must create an \mathbf{E} field. In section 23.3 we established that according to relativity, what one observer describes as a purely magnetic field, an observer in a different state of motion describes as a mixture of magnetic and electric fields. This is why there must be both an \mathbf{E} and a \mathbf{B} in observer 2's frame. Observer 2 cannot explain the electric field as coming from any charges. In frame 2, the \mathbf{E} can only be explained as an effect caused by the changing \mathbf{B} .

Observer 1 says, "2 feels a changing \mathbf{B} field because he's moving through a static field." Observer 2 says, "I feel a changing \mathbf{B} because the magnet is getting closer."

Although this argument doesn't prove the "whirlpool" geometry, we can verify that the fields I've drawn in figure n are consistent with it. The $\Delta\mathbf{B}$ vector is upward, and the electric field has a curliness to it: a paddlewheel inserted in the electric field would spin clockwise as seen from above, since the clockwise torque made by the strong electric field on the right is greater than the counterclockwise torque made by the weaker electric field on the left.



o / A generator.



p / A transformer.

The generator

example 5

A generator, o, consists of a permanent magnet that rotates within a coil of wire. The magnet is turned by a motor or crank, (not shown). As it spins, the nearby magnetic field changes. According to the principle of induction, this changing magnetic field results in an electric field, which has a whirlpool pattern. This electric field pattern creates a current that whips around the coils of wire, and we can tap this current to light the lightbulb.

self-check A

When you're driving a car, the engine recharges the battery continuously using a device called an alternator, which is really just a generator like the one shown on the previous page, except that the coil rotates while the permanent magnet is fixed in place. Why can't you use the alternator to start the engine if your car's battery is dead? ▷ Answer, p. 1044

The transformer

example 6

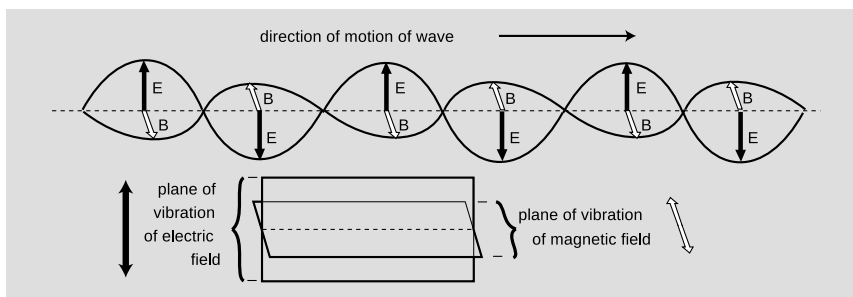
In example 18 on p. 607 we discussed the advantages of transmitting power over electrical lines using high voltages and low currents. However, we don't want our wall sockets to operate at 10000 volts! For this reason, the electric company uses a device called a transformer, p, to convert to lower voltages and higher currents inside your house. The coil on the input side creates a magnetic field. Transformers work with alternating current, so the

magnetic field surrounding the input coil is always changing. This induces an electric field, which drives a current around the output coil.

If both coils were the same, the arrangement would be symmetric, and the output would be the same as the input, but an output coil with a smaller number of coils gives the electric forces a smaller distance through which to push the electrons. Less mechanical work per unit charge means a lower voltage. Conservation of energy, however, guarantees that the amount of power on the output side must equal the amount put in originally, $I_{in} V_{in} = I_{out} V_{out}$, so this reduced voltage must be accompanied by an increased current.

24.5 Electromagnetic waves

The most important consequence of induction is the existence of electromagnetic waves. Whereas a gravitational wave would consist of nothing more than a rippling of gravitational fields, the principle of induction tells us that there can be no purely electrical or purely magnetic waves. Instead, we have waves in which there are both electric and magnetic fields, such as the sinusoidal one shown in the figure. Maxwell proved that such waves were a direct consequence of his equations, and derived their properties mathematically. The derivation would be beyond the mathematical level of this book, so we will just state the results.



An electromagnetic wave.

A sinusoidal electromagnetic wave has the geometry shown above. The \mathbf{E} and \mathbf{B} fields are perpendicular to the direction of motion, and are also perpendicular to each other. If you look along the direction of motion of the wave, the \mathbf{B} vector is always 90 degrees clockwise from the \mathbf{E} vector. In a plane wave, the magnitudes of the two fields are related by $|\mathbf{E}| = c|\mathbf{B}|$.

How is an electromagnetic wave created? It could be emitted, for example, by an electron orbiting an atom or currents going back and forth in a transmitting antenna. In general any accelerating charge will create an electromagnetic wave, although only a current that

varies sinusoidally with time will create a sinusoidal wave. Once created, the wave spreads out through space without any need for charges or currents along the way to keep it going. As the electric field oscillates back and forth, it induces the magnetic field, and the oscillating magnetic field in turn creates the electric field. The whole wave pattern propagates through empty space at the velocity c .

Einstein's motorcycle

example 7

As a teenage physics student, Einstein imagined the following paradox. (See p. 513.) What if he could get on a motorcycle and ride at speed c , alongside a beam of light? In his frame of reference, he observes constant electric and magnetic fields. But only a *changing* electric field can induce a magnetic field, and only a *changing* magnetic field can induce an electric field. The laws of physics are violated in his frame, and this seems to violate the principle that all frames of reference are equally valid.

The resolution of the paradox is that c is a universal speed limit, so the motorcycle can't be accelerated to c . Observers can never be at rest relative to a light wave, so no observer can have a frame of reference in which a light wave is observed to be at rest.

Reflection

example 8

The wave in figure r hits a silvered mirror. The metal is a good conductor, so it has constant voltage throughout, and the electric field equals zero inside it: the wave doesn't penetrate and is 100% reflected. If the electric field is to be zero at the surface as well, the reflected wave must have its electric field inverted (p. 525), so that the incident and reflected fields cancel there.

But the magnetic field of the reflected wave is *not* inverted. This is because the reflected wave, when viewed along its leftward direction of propagation, needs to have its \mathbf{B} vector 90 degrees clockwise from its \mathbf{E} vector.

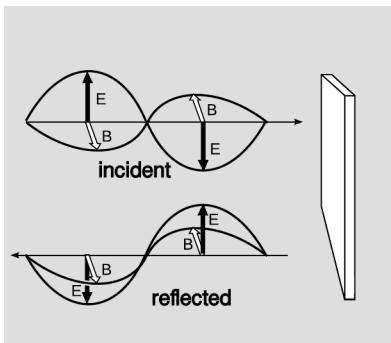
Polarization

Two electromagnetic waves traveling in the same direction through space can differ by having their electric and magnetic fields in different directions, a property of the wave called its polarization.

Light is an electromagnetic wave

Once Maxwell had derived the existence of electromagnetic waves, he became certain that they were the same phenomenon as light. Both are transverse waves (i.e., the vibration is perpendicular to the direction the wave is moving), and the velocity is the same.

Heinrich Hertz (for whom the unit of frequency is named) verified Maxwell's ideas experimentally. Hertz was the first to succeed in producing, detecting, and studying electromagnetic waves in detail



r / Example 8. The incident and reflected waves are drawn offset from each other for clarity, but are actually on top of each other so that their fields superpose.

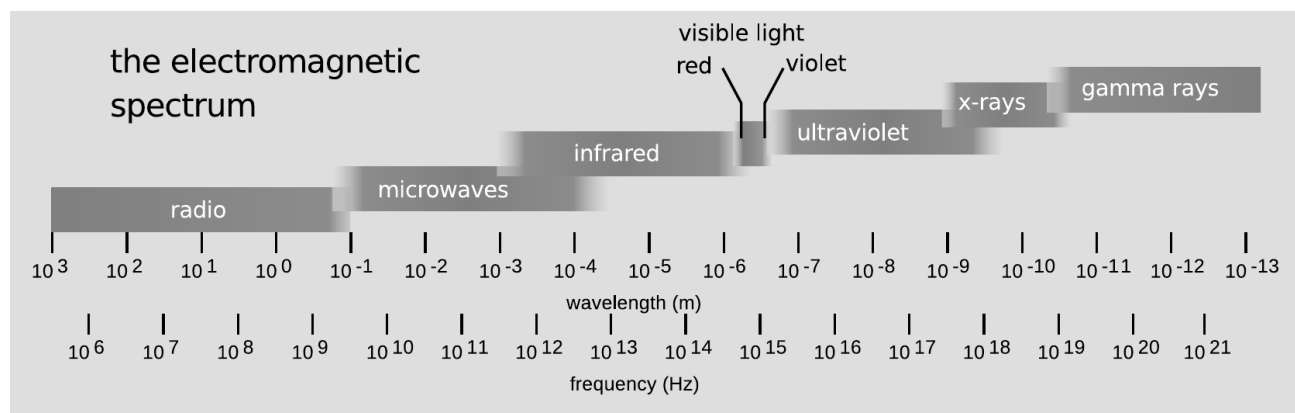


Heinrich Hertz (1857-1894).

using antennas and electric circuits. To produce the waves, he had to make electric currents oscillate very rapidly in a circuit. In fact, there was really no hope of making the current reverse directions at the frequencies of 10^{15} Hz possessed by visible light. The fastest electrical oscillations he could produce were 10^9 Hz, which would give a wavelength of about 30 cm. He succeeded in showing that, just like light, the waves he produced were polarizable, and could be reflected and refracted (i.e., bent, as by a lens), and he built devices such as parabolic mirrors that worked according to the same optical principles as those employing light. Hertz's results were convincing evidence that light and electromagnetic waves were one and the same.

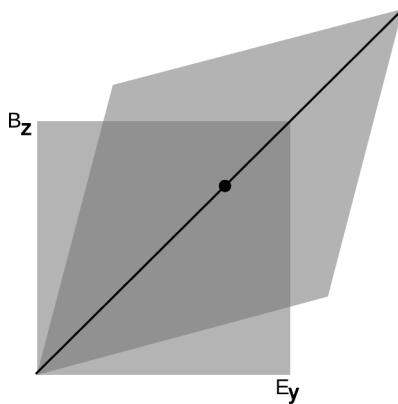
The electromagnetic spectrum

Today, electromagnetic waves with frequencies in the range employed by Hertz are known as radio waves. Any remaining doubts that the “Hertzian waves,” as they were then called, were the same type of wave as light waves were soon dispelled by experiments in the whole range of frequencies in between, as well as the frequencies outside that range. In analogy to the spectrum of visible light, we speak of the entire electromagnetic spectrum, of which the visible spectrum is one segment.



The terminology for the various parts of the spectrum is worth memorizing, and is most easily learned by recognizing the logical relationships between the wavelengths and the properties of the waves with which you are already familiar. Radio waves have wavelengths that are comparable to the size of a radio antenna, i.e., meters to tens of meters. Microwaves were named that because they have much shorter wavelengths than radio waves; when food heats unevenly in a microwave oven, the small distances between neighboring hot and cold spots is half of one wavelength of the standing wave the oven creates. The infrared, visible, and ultraviolet obviously have

much shorter wavelengths, because otherwise the wave nature of light would have been as obvious to humans as the wave nature of ocean waves. To remember that ultraviolet, x-rays, and gamma rays all lie on the short-wavelength side of visible, recall that all three of these can cause cancer. (As we'll discuss later in the course, there is a basic physical reason why the cancer-causing disruption of DNA can only be caused by very short-wavelength electromagnetic waves. Contrary to popular belief, microwaves cannot cause cancer, which is why we have microwave ovens and not x-ray ovens!)



Example 9: an electromagnetic wave that is legal in one frame of reference is legal in another. As in figure e on p. 699, the each frame of reference is in motion relative to the other along the x axis. If the wave's electric field is aligned with the y axis, and its magnetic field with z , then x is also the direction in which the wave is moving, as required for our example.

Switching frames of reference

example 9

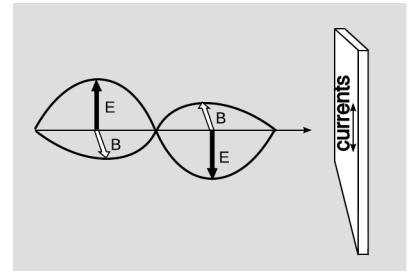
If we switch to a different frame of reference, a legal light wave should still be legal. Consider the requirement $\mathbf{E} = c\mathbf{B}$, in the case where observer 1 says observer 2 is trying to run away from the wave. In figure e on p. 699, we saw that the familiar parallelogram graphs described the transformation of electric and magnetic fields from one frame of reference to another. These pictures are intended to be used in units where $c = 1$, so the requirement for the fields becomes $\mathbf{E} = \mathbf{B}$, and such a combination of fields is represented by a dot on the diagonal, which is the same line in both frames.

Why the sky is blue

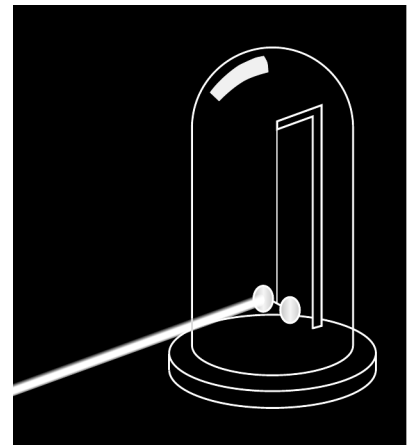
example 10

When sunlight enters the upper atmosphere, a particular air molecule finds itself being washed over by an electromagnetic wave of frequency f . The molecule's charged particles (nuclei and electrons) act like oscillators being driven by an oscillating force, and respond by vibrating at the same frequency f . Energy is sucked out of the incoming beam of sunlight and converted into the kinetic energy of the oscillating particles. However, these particles are accelerating, so they act like little radio antennas that put the energy back out as spherical waves of light that spread out in all directions. An object oscillating at a frequency f has an acceleration proportional to f^2 , and an accelerating charged particle creates an electromagnetic wave whose fields are proportional to its acceleration, so the field of the reradiated spherical wave is proportional to f^2 . The energy of a field is proportional to the square of the field, so the energy of the reradiated wave is proportional to f^4 . Since blue light has about twice the frequency of red light, this process is about $2^4 = 16$ times as strong for blue as for red, and that's why the sky is blue.

s / An electromagnetic wave strikes an ohmic surface. The wave's electric field causes currents to flow up and down. The wave's magnetic field then acts on these currents, producing a force in the direction of the wave's propagation. This is a pre-relativistic argument that light must possess inertia.



t / A simplified drawing of the 1903 experiment by Nichols and Hull that verified the predicted momentum of light waves. Two circular mirrors were hung from a fine quartz fiber, inside an evacuated bell jar. A 150 mW beam of light was shone on one of the mirrors for 6 s, producing a tiny rotation, which was measurable by an optical lever (not shown). The force was within 0.6% of the theoretically predicted value (problem 12 on p. 813) of $0.001 \mu\text{N}$. For comparison, a short clipping of a single human hair weighs $\sim 1 \mu\text{N}$.



Momentum of light

Newton defined momentum as mv , which would imply that light, which has no mass, should have no momentum. But Newton's laws only work at speeds small compared to the speed of light, and light travels *at* the speed of light. In fact, it's straightforward to show that electromagnetic waves have momentum. If a light wave strikes an ohmic surface, as in figure s, the wave's electric field causes charges to vibrate back and forth in the surface. These currents then experience a magnetic force from the wave's magnetic field, and application of the geometrical rule on p. 702 shows that the resulting force is in the direction of propagation of the wave. A light wave has momentum and inertia. This is explored further in problem 12 on p. 813. Figure t shows an experimental confirmation.

24.6 ★ Symmetry and handedness

Imagine that you establish radio contact with an alien on another planet. Neither of you even knows where the other one's planet is, and you aren't able to establish any landmarks that you both recognize. You manage to learn quite a bit of each other's languages, but you're stumped when you try to establish the definitions of left and right (or, equivalently, clockwise and counterclockwise). Is there any way to do it?

If there was any way to do it without reference to external landmarks, then it would imply that the laws of physics themselves were asymmetric, which would be strange. Why should they distinguish left from right? The gravitational field pattern surrounding a star or planet looks the same in a mirror, and the same goes for electric fields. The field patterns shown in section 24.2 seem to violate this



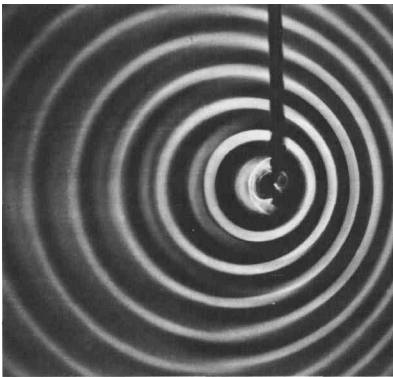
u / In this scene from *Swan Lake*, the choreography has a symmetry with respect to left and right.



C.S. Wu



v/A graphical representation of the Lorentz transformation for a velocity of $(3/5)c$. The long diagonal is stretched by a factor of two, the short one is half its former length, and the area is the same as before.



$w/$ The pattern of waves made by a point source moving to the right across the water. Note the shorter wavelength of the forward-emitted waves and the longer wavelength of the backward-going ones.

principle, but do they really? Could you use these patterns to explain left and right to the alien? In fact, the answer is no. If you look back at the definition of the magnetic field in section 24.1, it also contains a reference to handedness: the counterclockwise direction of the loop's current as viewed along the magnetic field. The aliens might have reversed their definition of the magnetic field, in which case their drawings of field patterns would look like mirror images of ours.

Until the middle of the twentieth century, physicists assumed that any reasonable set of physical laws would have to have this kind of symmetry between left and right. An asymmetry would be grotesque. Whatever their aesthetic feelings, they had to change their opinions about reality when experiments by C.S. Wu et al. showed that the weak nuclear force (section 22.4) violates right-left symmetry! It is still a mystery why right-left symmetry is observed so scrupulously in general, but is violated by one particular type of physical process.

24.7 ★ Doppler shifts and clock time

Figure v shows our now-familiar method of visualizing a Lorentz transformation, in a case where the numbers come out to be particularly simple. This diagram has two geometrical features that we have referred to before without digging into their physical significance: the *stretch factor* of the diagonals, and the *area*. In this section we'll see that the former can be related to the Doppler effect, and the latter to clock time.

Doppler shifts of light

When Doppler shifts happen to ripples on a pond or the sound waves from an airplane, they can depend on the relative motion of three different objects: the source, the receiver, and the medium. But light waves don't have a medium. Therefore Doppler shifts of light can only depend on the relative motion of the source and observer.

One simple case is the one in which the relative motion of the source and the receiver is perpendicular to the line connecting them. That is, the motion is transverse. Nonrelativistic Doppler shifts happen because the distance between the source and receiver is changing, so in nonrelativistic physics we don't expect any Doppler shift at all when the motion is transverse, and this is what is in fact observed to high precision. For example, the photo shows shortened and lengthened wavelengths to the right and left, along the source's line of motion, but an observer above or below the source measures just the normal, unshifted wavelength and frequency. But relativistically, we have a time dilation effect, so for light waves emitted transversely, there is a Doppler shift of $1/\gamma$ in frequency (or γ in wavelength).

The other simple case is the one in which the relative motion of the

source and receiver is longitudinal, i.e., they are either approaching or receding from one another. For example, distant galaxies are receding from our galaxy due to the expansion of the universe, and this expansion was originally detected because Doppler shifts toward the red (low-frequency) end of the spectrum were observed.

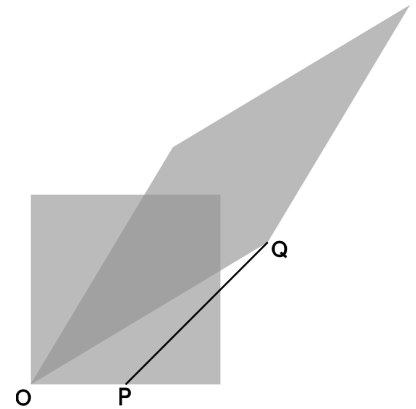
Nonrelativistically, we would expect the light from such a galaxy to be Doppler shifted down in frequency by some factor, which would depend on the relative velocities of three different objects: the source, the wave's medium, and the receiver. Relativistically, things get simpler, because light isn't a vibration of a physical medium, so the Doppler shift can only depend on a single velocity v , which is the rate at which the separation between the source and the receiver is increasing.

The square in figure x is the “graph paper” used by someone who considers the source to be at rest, while the parallelogram plays a similar role for the receiver. The figure is drawn for the case where $v = 3/5$ (in units where $c = 1$), and in this case the stretch factor of the long diagonal is 2. To keep the area the same, the short diagonal has to be squished to half its original size. But now it's a matter of simple geometry to show that OP equals half the width of the square, and this tells us that the Doppler shift is a factor of $1/2$ in frequency. That is, the squish factor of the short diagonal is interpreted as the Doppler shift. To get this as a general equation for velocities other than $3/5$, one can show by straightforward fiddling with the result of part c of problem 2 on p. 694 that the Doppler shift is

$$D(v) = \sqrt{\frac{1-v}{1+v}}.$$

Here $v > 0$ is the case where the source and receiver are getting farther apart, $v < 0$ the case where they are approaching. (This is the opposite of the sign convention used in section 19.5. It is convenient to change conventions here so that we can use positive values of v in the case of cosmological red-shifts, which are the most important application.)

Suppose that Alice stays at home on earth while her twin Betty takes off in her rocket ship at $3/5$ of the speed of light. When I first learned relativity, the thing that caused me the most pain was understanding how each observer could say that the other was the one whose time was slow. It seemed to me that if I could take a pill that would speed up my mind and my body, then naturally I would see everybody *else* as being *slow*. Shouldn't the same apply to relativity? But suppose Alice and Betty get on the radio and try to settle who is the fast one and who is the slow one. Each twin's voice sounds slooooooowed doooowwwwn to the other. If Alice claps her hands twice, at a time interval of one second by her clock, Betty hears the hand-claps coming over the radio two seconds apart, but the situation is exactly symmetric, and Alice hears the same thing if Betty claps.



x / At event O, the source and the receiver are on top of each other, so as the source emits a wave crest, it is received without any time delay. At P, the source emits another wave crest, and at Q the receiver receives it.

Each twin analyzes the situation using a diagram identical to x, and attributes her sister's observations to a complicated combination of time distortion, the time taken by the radio signals to propagate, and the motion of her twin relative to her.

self-check B

Turn your book upside-down and reinterpret figure x. ▷ Answer, p. 1044

A symmetry property of the Doppler effect example 11

Suppose that A and B are at rest relative to one another, but C is moving along the line between A and B. A transmits a signal to C, who then retransmits it to B. The signal accumulates two Doppler shifts, and the result is their product $D(v)D(-v)$. But this product must equal 1, so we must have $D(-v)D(v) = 1$, which can be verified directly from the equation.

The Ives-Stilwell experiment example 12

The result of example 11 was the basis of one of the earliest laboratory tests of special relativity, by Ives and Stilwell in 1938. They observed the light emitted by excited by a beam of H_2^+ and H_3^+ ions with speeds of a few tenths of a percent of c . Measuring the light from both ahead of and behind the beams, they found that the product of the Doppler shifts $D(v)D(-v)$ was equal to 1, as predicted by relativity. If relativity had been false, then one would have expected the product to differ from 1 by an amount that would have been detectable in their experiment. In 2003, Saathoff et al. carried out an extremely precise version of the Ives-Stilwell technique with Li^+ ions moving at 6.4% of c . The frequencies observed, in units of MHz, were:

$$\begin{aligned} f_0 &= 546466918.8 \pm 0.4 && \text{(unshifted frequency)} \\ f_0 D(-v) &= 582490203.44 \pm .09 && \text{(shifted frequency, forward)} \\ f_0 D(v) &= 512671442.9 \pm 0.5 && \text{(shifted frequency, backward)} \\ \sqrt{f_0 D(-v) \cdot f_0 D(v)} &= 546466918.6 \pm 0.3 \end{aligned}$$

The results show incredibly precise agreement between f_0 and $\sqrt{f_0 D(-v) \cdot f_0 D(v)}$, as expected relativistically because $D(v)D(-v)$ is supposed to equal 1. The agreement extends to 9 significant figures, whereas if relativity had been false there should have been a relative disagreement of about $v^2 = .004$, i.e., a discrepancy in the third significant figure. The spectacular agreement with theory has made this experiment a lightning rod for anti-relativity kooks.

We saw on p. 704 that relativistic velocities should not be expected to be exactly additive, and problem 1 on p. 722 verifies this in the special case where A moves relative to B at $0.6c$ and B relative to

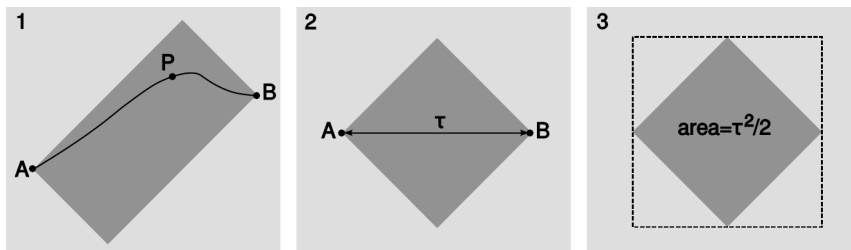
C at $0.6c$ — the result *not* being $1.2c$. The relativistic Doppler shift provides a simple way of deriving a general equation for the relativistic combination of velocities; problem 21 on p. 727 guides you through the steps of this derivation.

★ **Clock time**

On p. 679 we proved that the Lorentz transformation doesn't change the area of a shape in the $x-t$ plane. We used this only as a stepping stone toward the Lorentz transformation, but it is natural to wonder whether this kind of area has any physical interest of its own.

The equal-area result is not relativistic, since the proof never appeals to property 5 on page 675. Cases I and II on page 677 also have the equal-area property. We can see this clearly in a Galilean transformation like figure 1 on p. 676, where the distortion of the rectangle could be accomplished by cutting it into vertical slices and then displacing the slices upward without changing their areas.

But the area does have a nice interpretation in the relativistic case. Suppose that we have events A (Charles VII is restored to the throne) and B (Joan of Arc is executed). Now imagine that technologically advanced aliens want to be present at both A and B, but in the interim they wish to fly away in their spaceship, be present at some other event P (perhaps a news conference at which they give an update on the events taking place on earth), but get back in time for B. Since nothing can go faster than c (which we take to equal 1 in appropriate units), P cannot be too far away. The set of all possible events P forms a rectangle, figure $y/1$, in the $x-t$ plane that has A and B at opposite corners and whose edges have slopes equal to ± 1 . We call this type of rectangle a light-rectangle, because its sides could represent the motion of rays of light.



$y/1$. The gray light-rectangle represents the set of all events such as P that could be visited after A and before B.

2. The rectangle becomes a square in the frame in which A and B occur at the same location in space.

3. The area of the dashed square is τ^2 , so the area of the gray square is $\tau^2/2$.

The area of this rectangle will be the same regardless of one's frame of reference. In particular, we could choose a special frame of reference, panel 2 of the figure, such that A and B occur in the same place. (They do not occur at the same place, for example, in the sun's frame, because the earth is spinning and going around the sun.) Since the speed c , which equals 1 in our units, is the same in all frames of reference, and the sides of the rectangle had slopes ± 1 in frame 1, they must still have slopes ± 1 in frame 2. The rectangle

becomes a square with its diagonals parallel to the x and t axes, and the length of these diagonals equals the time τ elapsed on a clock that is at rest in frame 2, i.e., a clock that glides through space at constant velocity from A to B, meeting up with the planet earth at the appointed time. As shown in panel 3 of the figure, the area of the gray regions can be interpreted as half the square of this gliding-clock time. If events A and B are separated by a distance x and a time t , then in general $t^2 - x^2$ gives the square of the gliding-clock time.¹

When $|x|$ is greater than $|t|$, events A and B are so far apart in space and so close together in time that it would be impossible to have a cause and effect relationship between them, since $c = 1$ is the maximum speed of cause and effect. In this situation $t^2 - x^2$ is negative and cannot be interpreted as a clock time, but it can be interpreted as minus the square of the distance between A and B as measured by rulers at rest in a frame in which A and B are simultaneous.

No matter what, $t^2 - x^2$ is the same as measured in all frames of reference. Geometrically, it plays the same role in the $x-t$ plane that ruler measurements play in the Euclidean plane. In Euclidean geometry, the ruler-distance between any two points stays the same regardless of rotation, i.e., regardless of the angle from which we view the scene; according to the Pythagorean theorem, the square of this distance is $x^2 + y^2$. In the $x-t$ plane, $t^2 - x^2$ stays the same regardless of the frame of reference.

To avoid overloading the reader with terms to memorize, some commonly used terminology is relegated to problem 22 on p. 728.

¹Proof: Based on units, the expression must have the form $(\dots)t^2 + (\dots)tx + (\dots)x^2$, where each (\dots) represents a unitless constant. The tx coefficient must be zero by property 2 on p. 675. For consistency with figure y/3, the t^2 coefficient must equal 1. Since the area vanishes for $x = t$, the x^2 coefficient must equal -1 .

Summary

Selected vocabulary

- magnetic field . . . a field of force, defined in terms of the torque exerted on a test dipole
- magnetic dipole . . . an object, such as a current loop, an atom, or a bar magnet, that experiences torques due to magnetic forces; the strength of magnetic dipoles is measured by comparison with a standard dipole consisting of a square loop of wire of a given size and carrying a given amount of current
- induction the production of an electric field by a changing magnetic field, or vice-versa

Notation

- \mathbf{B} the magnetic field
- D_m magnetic dipole moment

Summary

The magnetic field is defined in terms of the torque on a magnetic test dipole. It has no sources or sinks; magnetic field patterns never converge on or diverge from a point.

Relativity dictates a maximum speed limit c for cause and effect. This speed is the same in all frames of reference.

Relativity requires that the magnetic and electric fields be intimately related. The principle of induction states that any changing electric field produces a magnetic field in the surrounding space, and vice-versa. These induced fields tend to form whirlpool patterns.

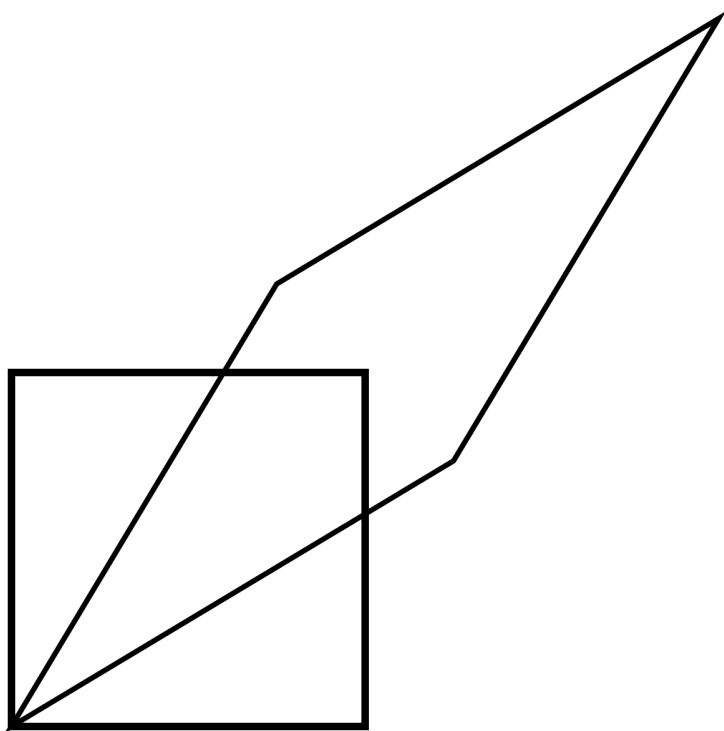
The most important consequence of the principle of induction is that there are no purely magnetic or purely electric waves. Electromagnetic disturbances propagate outward at c as combined magnetic and electric waves, with a well-defined relationship between the magnitudes and directions of the electric and magnetic fields. These electromagnetic waves are what light is made of, but other forms of electromagnetic waves exist besides visible light, including radio waves, x-rays, and gamma rays.

Problems

Key

- ✓ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ★ A difficult problem.

1 The figure illustrates a Lorentz transformation using the conventions employed in section 23.2. For simplicity, the transformation chosen is one that lengthens one diagonal by a factor of 2. Since Lorentz transformations preserve area, the other diagonal is shortened by a factor of 2. Let the original frame of reference, depicted with the square, be A, and the new one B. (a) By measuring with a ruler on the figure, show that the velocity of frame B relative to frame A is $0.6c$. (b) Print out a copy of the page. With a ruler, draw a third parallelogram that represents a second successive Lorentz transformation, one that lengthens the long diagonal by another factor of 2. Call this third frame C. Use measurements with a ruler to determine frame C's velocity relative to frame A. Does it equal double the velocity found in part a? Explain why it should be expected to turn out the way it does. ✓



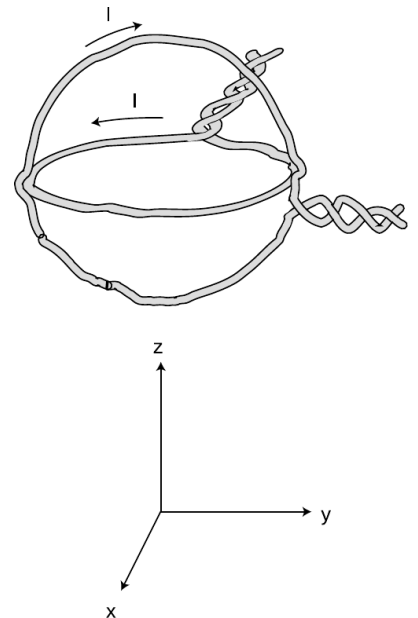
2 (a) In this chapter we've represented Lorentz transformations as distortions of a square into various parallelograms, with the degree of distortion depending on the velocity of one frame of reference relative to another. Suppose that one frame of reference was moving at c relative to another. Discuss what would happen in terms of distortion of a square, and show that this is impossible by using an argument similar to the one used to rule out transformations like the one in figure m on page 676.

(b) Resolve the following paradox. Two pen-pointer lasers are placed side by side and aimed in parallel directions. Their beams both travel at c relative to the hardware, but each beam has a velocity of zero relative to the neighboring beam. But the speed of light can't be zero; it's supposed to be the same in all frames of reference.

3 Consider two solenoids, one of which is smaller so that it can be put inside the other. Assume they are long enough so that each one only contributes significantly to the field inside itself, and the interior fields are nearly uniform. Consider the configuration where the small one is inside the big one with their currents circulating in the same direction, and a second configuration in which the currents circulate in opposite directions. Compare the energies of these configurations with the energy when the solenoids are far apart. Based on this reasoning, which configuration is stable, and in which configuration will the little solenoid tend to get twisted around or spit out?
 ▷ Hint, p. 1032

4 The figure shows a nested pair of circular wire loops used to create magnetic fields. (The twisting of the leads is a practical trick for reducing the magnetic fields they contribute, so the fields are very nearly what we would expect for an ideal circular current loop.) The coordinate system below is to make it easier to discuss directions in space. One loop is in the $y - z$ plane, the other in the $x - y$ plane. Each of the loops has a radius of 1.0 cm, and carries 1.0 A in the direction indicated by the arrow.

- (a) Using the equation in optional section 24.2, calculate the magnetic field that would be produced by *one* such loop, at its center. ✓
- (b) Describe the direction of the magnetic field that would be produced, at its center, by the loop in the $x - y$ plane alone.
- (c) Do the same for the other loop.
- (d) Calculate the magnitude of the magnetic field produced by the two loops in combination, at their common center. Describe its direction. ✓



Problem 4.

5 One model of the hydrogen atom has the electron circling around the proton at a speed of 2.2×10^6 m/s, in an orbit with a radius of 0.05 nm. (Although the electron and proton really orbit around their common center of mass, the center of mass is very close to the proton, since it is 2000 times more massive. For this problem, assume the proton is stationary.) In homework problem 19, p. 619, you calculated the electric current created.

(a) Now estimate the magnetic field created at the center of the atom by the electron. We are treating the circling electron as a current loop, even though it's only a single particle. ✓

(b) Does the proton experience a nonzero force from the electron's magnetic field? Explain.

(c) Does the electron experience a magnetic field from the proton? Explain.

(d) Does the electron experience a magnetic field created by its own current? Explain.

(e) Is there an electric force acting between the proton and electron? If so, calculate it. ✓

(f) Is there a gravitational force acting between the proton and electron? If so, calculate it.

(g) An inward force is required to keep the electron in its orbit – otherwise it would obey Newton's first law and go straight, leaving the atom. Based on your answers to the previous parts, which force or forces (electric, magnetic and gravitational) contributes significantly to this inward force?

[Based on a problem by Arnold Arons.]

6 Suppose a charged particle is moving through a region of space in which there is an electric field perpendicular to its velocity vector, and also a magnetic field perpendicular to both the particle's velocity vector and the electric field. Show that there will be one particular velocity at which the particle can be moving that results in a total force of zero on it; this requires that you analyze both the magnitudes and the directions of the forces compared to one another. Relate this velocity to the magnitudes of the electric and magnetic fields. (Such an arrangement, called a velocity filter, is one way of determining the speed of an unknown particle.)

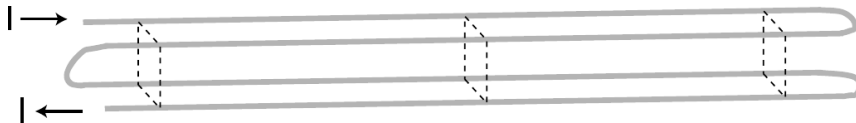
7 If you put four times more current through a solenoid, how many times more energy is stored in its magnetic field? ✓

8 Suppose we are given a permanent magnet with a complicated, asymmetric shape. Describe how a series of measurements with a magnetic compass could be used to determine the strength and direction of its magnetic field at some point of interest. Assume that you are only able to see the direction to which the compass needle settles; you cannot measure the torque acting on it. ★

9 Consider two solenoids, one of which is smaller so that it can be put inside the other. Assume they are long enough to act like ideal solenoids, so that each one only contributes significantly to the field inside itself, and the interior fields are nearly uniform. Consider the configuration where the small one is partly inside and partly hanging out of the big one, with their currents circulating in the same direction. Their axes are constrained to coincide.

(a) Find the difference in the magnetic energy between the configuration where the solenoids are separate and the configuration where the small one is inserted into the big one. Your equation will include the length x of the part of the small solenoid that is inside the big one, as well as other relevant variables describing the two solenoids. \checkmark

(b) Based on your answer to part a, find the force acting



Problem 10.

10 Four long wires are arranged, as shown, so that their cross-section forms a square, with connections at the ends so that current flows through all four before exiting. Note that the current is to the right in the two back wires, but to the left in the front wires. If the dimensions of the cross-sectional square (height and front-to-back) are b , find the magnetic field (magnitude and direction) along the long central axis. \checkmark

11 The purpose of this problem is to find the force experienced by a straight, current-carrying wire running perpendicular to a uniform magnetic field. (a) Let A be the cross-sectional area of the wire, n the number of free charged particles per unit volume, q the charge per particle, and v the average velocity of the particles. Show that the current is $I = Avnq$. (b) Show that the magnetic force per unit length is $AvnqB$. (c) Combining these results, show that the force on the wire per unit length is equal to IB . \triangleright Solution, p. 1033

12 Suppose two long, parallel wires are carrying current I_1 and I_2 . The currents may be either in the same direction or in opposite directions. (a) Using the information from section 24.2, determine under what conditions the force is attractive, and under what conditions it is repulsive. Note that, because of the difficulties explored in problem 14, it's possible to get yourself tied up in knots if you use the energy approach of section 22.4. (b) Starting from the result of problem 11, calculate the force per unit length.

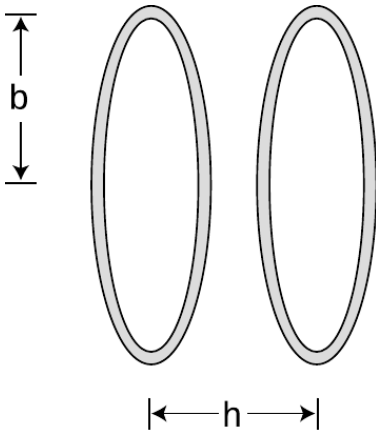
\triangleright Solution, p. 1033

13 Section 24.2 states the following rule:

For a positively charged particle, the direction of the F vector is the one such that if you sight along it, the \mathbf{B} vector is clockwise from the v vector.

Make a three-dimensional model of the three vectors using pencils or rolled-up pieces of paper to represent the vectors assembled with their tails together. Now write down every possible way in which the rule could be rewritten by scrambling up the three symbols F , \mathbf{B} , and v . Referring to your model, which are correct and which are incorrect?

14 Prove that any two planar current loops with the same value of IA will experience the same torque in a magnetic field, regardless of their shapes. In other words, the dipole moment of a current loop can be defined as IA , regardless of whether its shape is a square. ★



Problem 15.

15 A Helmholtz coil is defined as a pair of identical circular coils lying in parallel planes and separated by a distance, h , equal to their radius, b . (Each coil may have more than one turn of wire.) Current circulates in the same direction in each coil, so the fields tend to reinforce each other in the interior region. This configuration has the advantage of being fairly open, so that other apparatus can be easily placed inside and subjected to the field while remaining visible from the outside. The choice of $h = b$ results in the most uniform possible field near the center. A photograph of a Helmholtz coil is shown in example 2 on page 703.

- (a) Find the percentage drop in the field at the center of one coil, compared to the full strength at the center of the whole apparatus. ✓
- (b) What value of h (not equal to b) would make this difference equal to zero? ✓

16 (a) In the photo of the vacuum tube apparatus in figure 24.2.2 on p. 703, infer the direction of the magnetic field from the motion of the electron beam. (b) Based on your answer to a, find the direction of the currents in the coils. (c) What direction are the electrons in the coils going? (d) Are the currents in the coils repelling or attracting the currents consisting of the beam inside the tube? Compare with figure ai on p. 691. ▷ Solution, p. 1034

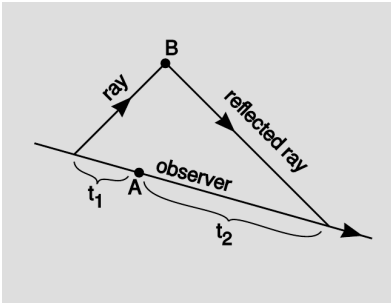
17 In the photo of the vacuum tube apparatus in section 24.2, an approximately uniform magnetic field caused circular motion. Is there any other possibility besides a circle? What can happen in general? ★

18 This problem is now problem 16 on p. 665.

19 In section 24.2 I gave an equation for the magnetic field in the interior of a solenoid, but that equation doesn't give the right answer near the mouths or on the outside. Although in general the computation of the field in these other regions is complicated, it is possible to find a precise, simple result for the field at the center of one of the mouths, using only symmetry and vector addition. What is it?
▷ Solution, p. 1034 ★

20 Prove that in an electromagnetic plane wave, half the energy is in the electric field and half in the magnetic field.

21 As promised in section 24.7, this problem will lead you through the steps of finding an equation for the combination of velocities in relativity, generalizing the numerical result found in problem 1. Suppose that A moves relative to B at velocity u , and B relative to C at v . We want to find A's velocity w relative to C, in terms of u and v . Suppose that A emits light with a certain frequency. This will be observed by B with a Doppler shift $D(u)$. C detects a further shift of $D(v)$ relative to B. We therefore expect the Doppler shifts to multiply, $D(w) = D(u)D(v)$, and this provides an implicit rule for determining w if u and v are known. (a) Using the expression for D given in section 24.7.1, write down an equation relating u , v , and w . (b) Solve for w in terms of u and v . (c) Show that your answer to part b satisfies the correspondence principle.
▷ Solution, p. 1034



Problem 22.

22 On p. 719, we defined a quantity $t^2 - x^2$, which is often referred to as the spacetime interval. Let's notate it as \mathcal{I} (cursive letter "I"). The only reason this quantity is interesting is that it stays the same in all frames of reference, but to define it, we first had to pick a frame of reference in order to define an $x - t$ plane, and then turn around and prove that it didn't matter what frame had been chosen. It might thus be nicer simply to define it as the square of the gliding-clock time, in the case where B can be reached from A. Since this definition never refers to any coordinates or frame of reference, we know automatically that it is frame-independent. In this case where $\mathcal{I} > 0$, we say that the relationship between A and B is timelike; there is enough time for cause and effect to propagate between A and B. An interval $\mathcal{I} < 0$ is called spacelike.

In the spacelike case, we can define \mathcal{I} using rulers, as on p. 720, but it's awkward to have to introduce an entirely new measuring instrument in order to complete the definition. Geroch² suggests a cute alternative in which this case as well can be treated using clocks. Let observer O move inertially (i.e., without accelerating), and let her initial position and state of motion be chosen such that she will be present at event A. Before A, she emits a ray of light, choosing to emit it at the correct time and in the correct direction so that it will reach B. At B, we arrange to have the ray reflected so that O can receive the reflection at some later time. Let t_1 be the time elapsed on O's clock from emission of the first ray until event A, and let t_2 be the time from A until she receives the second ray. The goal of this problem is to show that if we define \mathcal{I} as $-t_1 t_2$, we obtain the same result as with the previous definition. Since t_1 and t_2 are simply clock readings, not coordinates defined in an arbitrary frame of reference, this definition is automatically frame-independent.

- Show that \mathcal{I} , as originally defined on p. 719, has the same units as the expression $-t_1 t_2$.
- Pick an event in the $x - t$ plane, and sketch the regions that are timelike and spacelike in relation to it.
- The special case of $\mathcal{I} = 0$ is called a lightlike interval. Such events lie on a cone in the diagram drawn in part b, and this cone is called the light cone. Verify that the two definitions of \mathcal{I} agree on the light cone.
- Prove that the two definitions agree on \mathcal{I} in the spacelike case.
- What goes wrong if O doesn't move inertially?

23 Can a field that is purely electrical in one frame of reference be purely magnetic in some other frame? Use figure e on p. 699.

²Robert Geroch, *General Relativity from A to B*, University of Chicago Press, 1978

Exercise 24A: Polarization

Apparatus:

calcite (Iceland spar) crystal

polaroid film

1. Lay the crystal on a piece of paper that has print on it. You will observe a double image. See what happens if you rotate the crystal.

Evidently the crystal does something to the light that passes through it on the way from the page to your eye. One beam of light enters the crystal from underneath, but two emerge from the top; by conservation of energy the energy of the original beam must be shared between them. Consider the following three possible interpretations of what you have observed:

(a) The two new beams differ from each other, and from the original beam, only in energy. Their other properties are the same.

(b) The crystal adds to the light some mysterious new property (not energy), which comes in two flavors, X and Y. Ordinary light doesn't have any of either. One beam that emerges from the crystal has some X added to it, and the other beam has Y.

(c) There is some mysterious new property that is possessed by all light. It comes in two flavors, X and Y, and most ordinary light sources make an equal mixture of type X and type Y light. The original beam is an even mixture of both types, and this mixture is then split up by the crystal into the two purified forms.

In parts 2 and 3 you'll make observations that will allow you to figure out which of these is correct.

2. Now place a polaroid film over the crystal and see what you observe. What happens when you rotate the film in the horizontal plane? Does this observation allow you to rule out any of the three interpretations?

3. Now put the polaroid film under the crystal and try the same thing. Putting together all your observations, which interpretation do you think is correct?

4. Look at an overhead light fixture through the polaroid, and try rotating it. What do you observe? What does this tell you about the light emitted by the lightbulb?

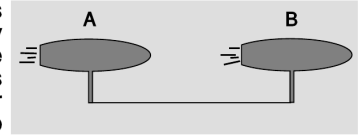
5. Now position yourself with your head under a light fixture and directly over a shiny surface, such as a glossy tabletop. You'll see the lamp's reflection, and the light coming from the lamp

to your eye will have undergone a reflection through roughly a 180-degree angle (i.e. it very nearly reversed its direction). Observe this reflection through the polaroid, and try rotating it. Finally, position yourself so that you are seeing glancing reflections, and try the same thing. Summarize what happens to light with properties X and Y when it is reflected. (This is the principle behind polarizing sunglasses.)

Exercise 24B: Events and spacetime

Bell's spaceship paradox

A difficult philosophical question is whether the time dilation and length contractions predicted by relativity are "real." This depends, of course, on what one means by "real." They are frame-dependent, i.e., observers in different frames of reference disagree about them. But this doesn't tell us much about their reality, since velocities are frame-dependent in Newtonian mechanics, but nobody worries about whether velocities are real. John Bell (1928-1990) proposed the following thought experiment to physicists in the CERN cafeteria, and found that nearly all of them got it wrong. He took this as evidence that their intuitions had been misguided by the standard way of approaching this question of the reality of Lorentz contractions.



Let spaceships A and B accelerate as shown in the figure, along a straight line. Observer C, whose frame of reference is indicated by the square, does not accelerate. The accelerations, as judged by C, are constant, and equal for the two ships. Each ship is equipped with a yard-arm, and a thread is tied between the two arms. Does the thread break, due to length contraction?

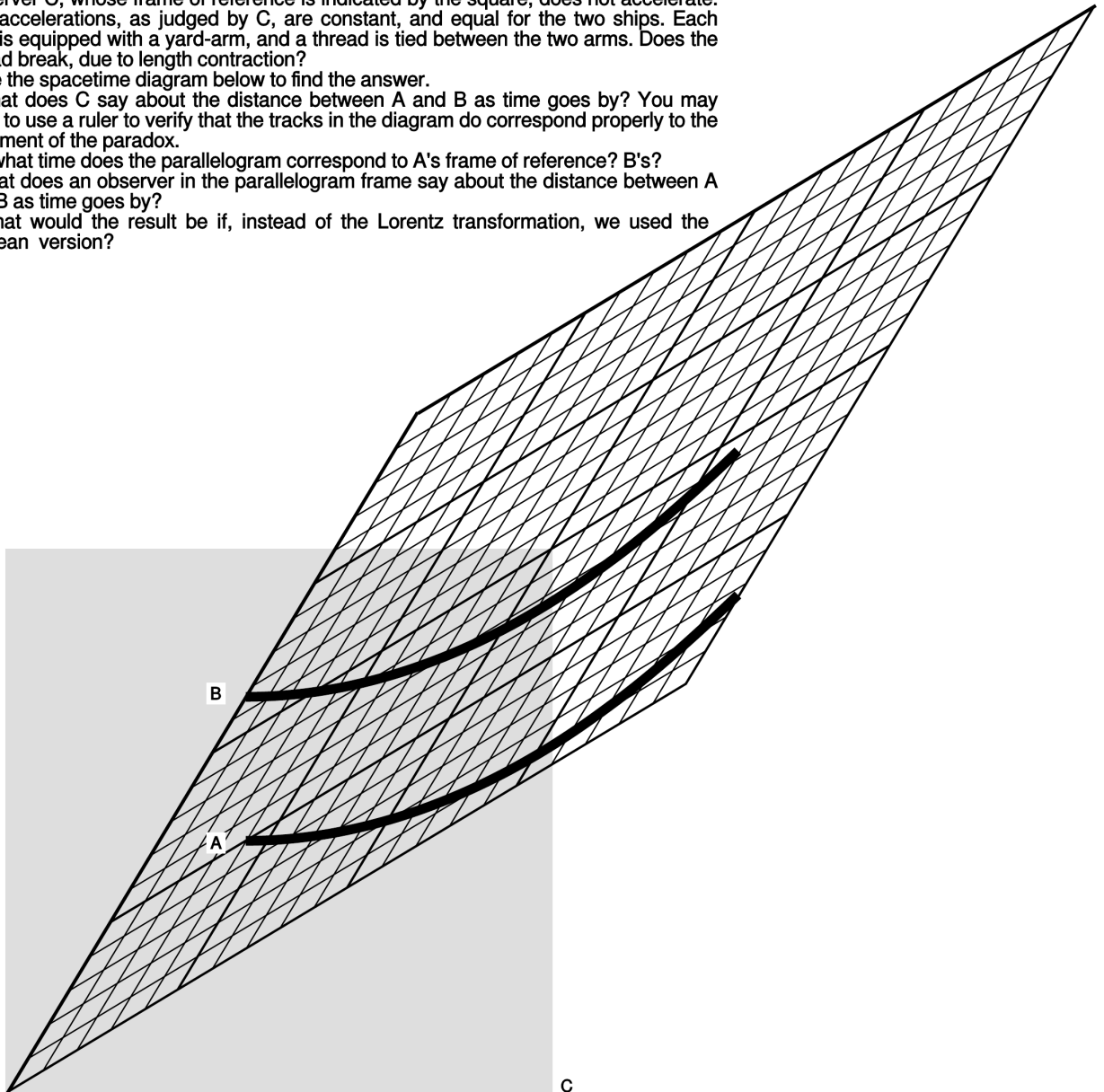
Use the spacetime diagram below to find the answer.

What does C say about the distance between A and B as time goes by? You may want to use a ruler to verify that the tracks in the diagram do correspond properly to the statement of the paradox.

At what time does the parallelogram correspond to A's frame of reference? B's?

What does an observer in the parallelogram frame say about the distance between A and B as time goes by?

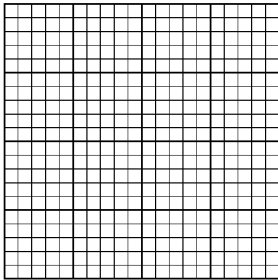
What would the result be if, instead of the Lorentz transformation, we used the Galilean version?



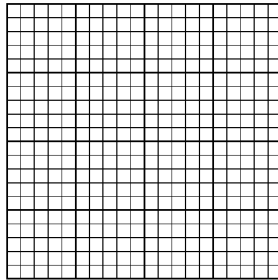
The following paper-and-pencil exercises involve graphs of position versus time. These are often referred to as spacetime diagrams, and tracks on them are called world-lines. As in the other examples in this book, a time axis is always closer to horizontal, and a position axis closer to vertical, and the units are such that the universal speed is a slope of +1 or -1.

Draw spacetime diagrams of: (1) a box with a particle bouncing back and forth inside it, (2) a ray of light being absorbed by a leaf on a tree, (3) an atomic nucleus splitting up into two parts (fissioning), (4) a cloud of gas collapsing gravitationally to form our solar system

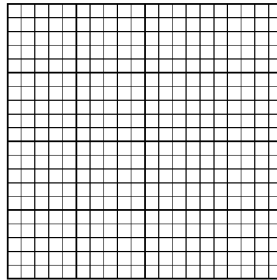
1



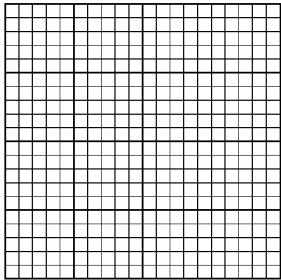
2



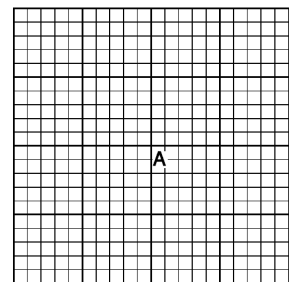
3



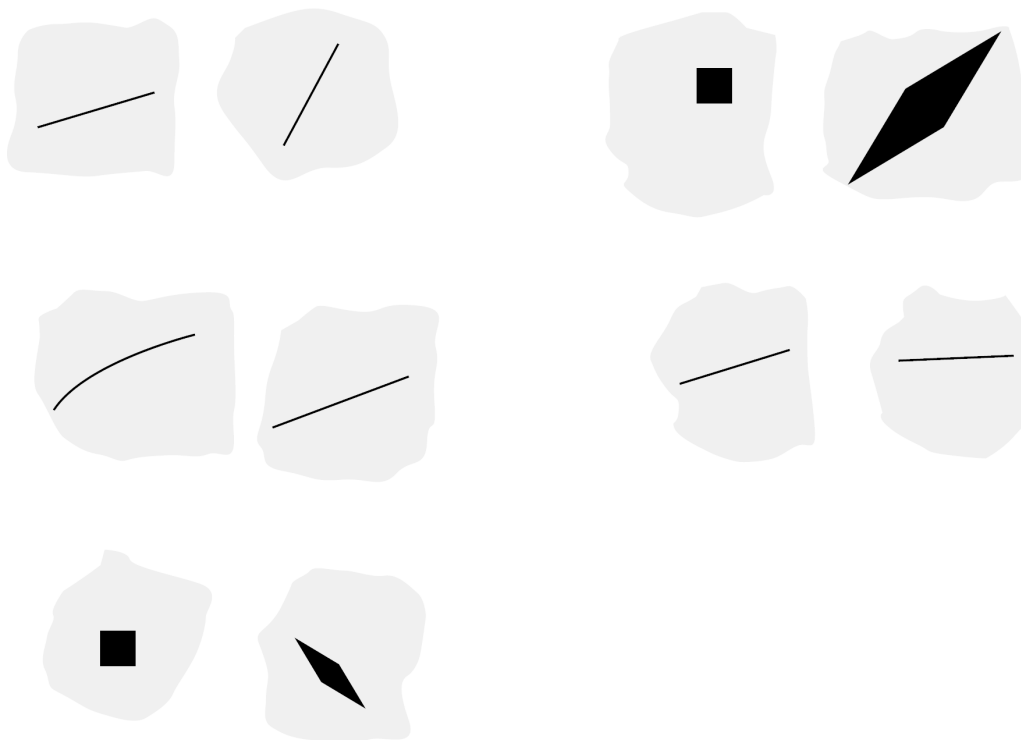
4



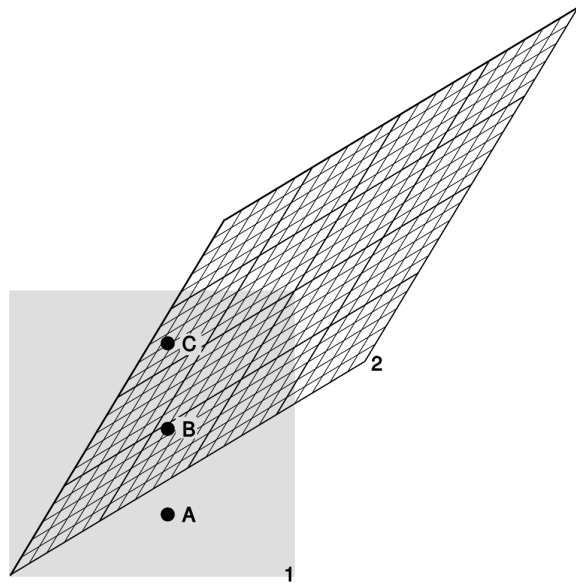
5. Event A is given. Mark examples of events B-F that satisfy these criteria. A material object travels from A to B. An object traveled from C to A. A ray of light emitted at A is received at D. A ray of light emitted at E is received at A. F can have no possible cause-and-effect relationship with A.



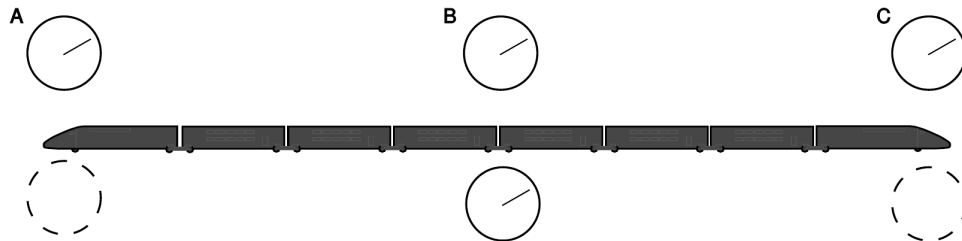
6. These are five pairs of spacetime diagrams, organized so that there is some resemblance between the left and right of each pair. In each case, decide whether the two could actually be the same thing represented in a different inertial frame of reference. Write yes or no, and explain why.



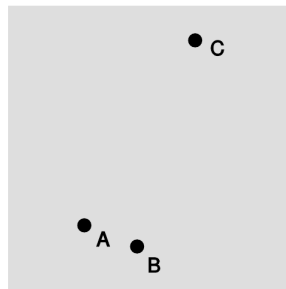
7. The diagram shows three events and two frames of reference. Describe the time-ordering of the events in these frames. Is any other time-ordering possible in any other frame?



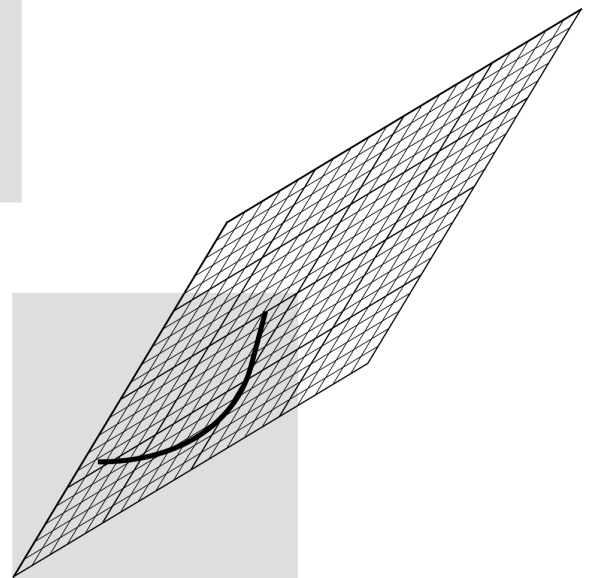
8. The top row shows three clocks located in three different places. They have been synchronized in the frame of reference of the earth, represented by the paper. This synchronization is carried out by exchanging light signals. For example, if the front and back clocks both send out flashes of light when they think it's 2 o'clock, the one in the middle will receive them both at the same time. Event A is the one at which the back clock A reads 2 o'clock, etc. The second row represents clocks that are synchronized aboard the train, which is moving to the right at a substantial fraction of the speed of light. How should the clocks shown with dashed outlines compare with the one at the middle of the train?



9 The figure shows three events and a square representing the t and x axes of a frame of reference. In this frame, the time-ordering of the events is ABC. If we switch to another frame, what other orderings, if any, are possible?



10. The figure shows the motion of an object, with two different frames of reference superimposed. Is anything strange going on?



11. Suppose that the train in 8 is hooked up in a circle, like a chain necklace. What happens?

Chapter 25

Capacitance and Inductance

The long road leading from the light bulb to the computer started with one very important step: the introduction of feedback into electronic circuits. Although the principle of feedback has been understood and applied to mechanical systems for centuries, and to electrical ones since the early twentieth century, for most of us the word evokes an image of Jimi Hendrix intentionally creating earsplitting screeches, or of the school principal doing the same inadvertently in the auditorium. In the guitar example, the musician stands in front of the amp and turns it up so high that the sound waves coming from the speaker come back to the guitar string and make it shake harder. This is an example of *positive* feedback: the harder the string vibrates, the stronger the sound waves, and the stronger the sound waves, the harder the string vibrates. The only limit is the power-handling ability of the amplifier.

Negative feedback is equally important. Your thermostat, for example, provides negative feedback by kicking the heater off when the house gets warm enough, and by firing it up again when it gets too cold. This causes the house's temperature to oscillate back and forth within a certain range. Just as out-of-control exponential freak-outs are a characteristic behavior of positive-feedback systems, oscillation is typical in cases of negative feedback. You have already studied negative feedback extensively in ch. 17 in the case of a mechanical system, although we didn't call it that.

25.1 Capacitance and inductance

In a mechanical oscillation, energy is exchanged repetitively between potential and kinetic forms, and may also be siphoned off in the form of heat dissipated by friction. In an electrical circuit, resistors are the circuit elements that dissipate heat. What are the electrical analogs of storing and releasing the potential and kinetic energy of a vibrating object? When you think of energy storage in an electrical circuit, you are likely to imagine a battery, but even rechargeable batteries can only go through 10 or 100 cycles before they wear out. In addition, batteries are not able to exchange energy on a short enough time scale for most applications. The circuit in a musical synthesizer may be called upon to oscillate thousands of times a

second, and your microwave oven operates at gigahertz frequencies. Instead of batteries, we generally use capacitors and inductors to store energy in oscillating circuits. Capacitors, which you've already encountered, store energy in electric fields. An inductor does the same with magnetic fields.

Capacitors

A capacitor's energy exists in its surrounding electric fields. It is proportional to the square of the field strength, which is proportional to the charges on the plates. If we assume the plates carry charges that are the same in magnitude, $+q$ and $-q$, then the energy stored in the capacitor must be proportional to q^2 . For historical reasons, we write the constant of proportionality as $1/2C$,

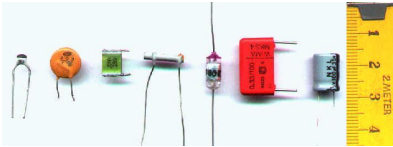
$$E_C = \frac{1}{2C}q^2.$$

The constant C is a geometrical property of the capacitor, called its capacitance.

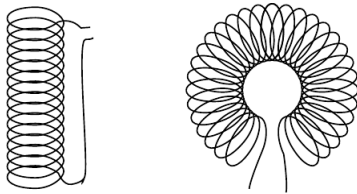
Based on this definition, the units of capacitance must be coulombs squared per joule, and this combination is more conveniently abbreviated as the farad, $1 \text{ F} = 1 \text{ C}^2/\text{J}$. "Condenser" is a less formal term for a capacitor. Note that the labels printed on capacitors often use MF to mean μF , even though MF should really be the symbol for megafarads, not microfarads. Confusion doesn't result from this nonstandard notation, since picofarad and microfarad values are the most common, and it wasn't until the 1990's that even millifarad and farad values became available in practical physical sizes. Figure a shows the symbol used in schematics to represent a capacitor.



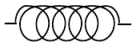
a / The symbol for a capacitor.



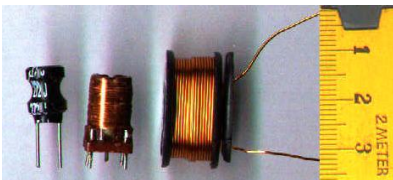
b / Some capacitors.



c / Two common geometries for inductors. The cylindrical shape on the left is called a solenoid.



d / The symbol for an inductor.



e / Some inductors.

Inductors

Any current will create a magnetic field, so in fact every current-carrying wire in a circuit acts as an inductor! However, this type of "stray" inductance is typically negligible, just as we can usually ignore the stray resistance of our wires and only take into account the actual resistors. To store any appreciable amount of magnetic energy, one usually uses a coil of wire designed specifically to be an inductor. All the loops' contribution to the magnetic field add together to make a stronger field. Unlike capacitors and resistors, practical inductors are easy to make by hand. One can for instance spool some wire around a short wooden dowel, put the spool inside a plastic aspirin bottle with the leads hanging out, and fill the bottle with epoxy to make the whole thing rugged. An inductor like this, in the form cylindrical coil of wire, is called a solenoid, c, and a stylized solenoid, d, is the symbol used to represent an inductor in a circuit regardless of its actual geometry.

How much energy does an inductor store? The energy density is proportional to the square of the magnetic field strength, which is

in turn proportional to the current flowing through the coiled wire, so the energy stored in the inductor must be proportional to I^2 . We write $L/2$ for the constant of proportionality, giving

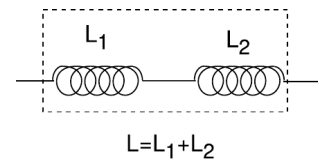
$$E_L = \frac{L}{2} I^2.$$

As in the definition of capacitance, we have a factor of $1/2$, which is purely a matter of definition. The quantity L is called the *inductance* of the inductor, and we see that its units must be joules per ampere squared. This clumsy combination of units is more commonly abbreviated as the henry, $1 \text{ henry} = 1 \text{ J/A}^2$. Rather than memorizing this definition, it makes more sense to derive it when needed from the definition of inductance. Many people know inductors simply as “coils,” or “chokes,” and will not understand you if you refer to an “inductor,” but they will still refer to L as the “inductance,” not the “coilance” or “chokeance!”

Identical inductances in series

example 1

If two inductors are placed in series, any current that passes through the combined double inductor must pass through both its parts. Thus by the definition of inductance, the inductance is doubled as well. In general, inductances in series add, just like resistances. The same kind of reasoning also shows that the inductance of a solenoid is approximately proportional to its length, assuming the number of turns per unit length is kept constant.

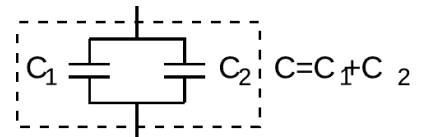


f / Inductances in series add.

Identical capacitances in parallel

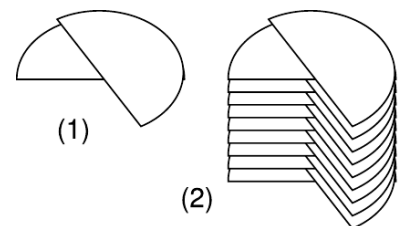
example 2

When two identical capacitances are placed in parallel, any charge deposited at the terminals of the combined double capacitor will divide itself evenly between the two parts. The electric fields surrounding each capacitor will be half the intensity, and therefore store one quarter the energy. Two capacitors, each storing one quarter the energy, give half the total energy storage. Since capacitance is inversely related to energy storage, this implies that identical capacitances in parallel give double the capacitance. In general, capacitances in parallel add. This is unlike the behavior of inductors and resistors, for which series configurations give addition.



g / Capacitances in parallel add.

This is consistent with the fact that the capacitance of a single parallel-plate capacitor is proportional to the area of the plates. If we have two parallel-plate capacitors, and we combine them in parallel and bring them very close together side by side, we have produced a single capacitor with plates of double the area, and it has approximately double the capacitance.



h / A variable capacitor.

Inductances in parallel and capacitances in series are explored in homework problems 4 and 6.

A variable capacitor

example 3

Figure h/1 shows the construction of a variable capacitor out of

two parallel semicircles of metal. One plate is fixed, while the other can be rotated about their common axis with a knob. The opposite charges on the two plates are attracted to one another, and therefore tend to gather in the overlapping area. This overlapping area, then, is the only area that effectively contributes to the capacitance, and turning the knob changes the capacitance. The simple design can only provide very small capacitance values, so in practice one usually uses a bank of capacitors, wired in parallel, with all the moving parts on the same shaft.

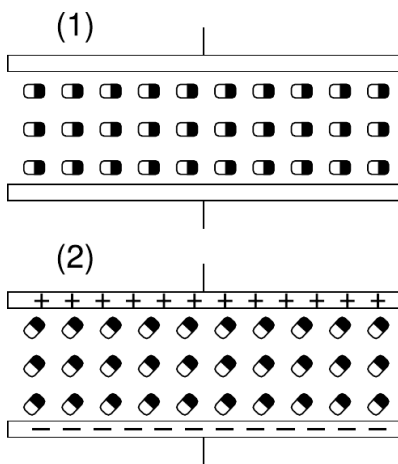
Discussion questions

A Suppose that two parallel-plate capacitors are wired in parallel, and are placed very close together, side by side, so that their fields overlap. Will the resulting capacitance be too small, or too big? Could you twist the circuit into a different shape and make the effect be the other way around, or make the effect vanish? How about the case of two inductors in series?

B Most practical capacitors do not have an air gap or vacuum gap between the plates; instead, they have an insulating substance called a dielectric. We can think of the molecules in this substance as dipoles that are free to rotate (at least a little), but that are not free to move around, since it is a solid. The figure shows a highly stylized and unrealistic way of visualizing this. We imagine that all the dipoles are initially turned sideways, (1), and that as the capacitor is charged, they all respond by turning through a certain angle, (2). (In reality, the scene might be much more random, and the alignment effect much weaker.)

For simplicity, imagine inserting just one electric dipole into the vacuum gap. For a given amount of charge on the plates, how does this affect the amount of energy stored in the electric field? How does this affect the capacitance?

Now redo the analysis in terms of the mechanical work needed in order to charge up the plates.



i / Discussion question B.

25.2 Oscillations

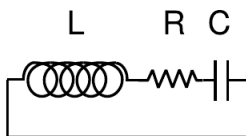
Figure j shows the simplest possible oscillating circuit. For any useful application it would actually need to include more components. For example, if it was a radio tuner, it would need to be connected to an antenna and an amplifier. Nevertheless, all the essential physics is there.

We can analyze it without any sweat or tears whatsoever, simply by constructing an analogy with a mechanical system. In a mechanical oscillator, k , we have two forms of stored energy,

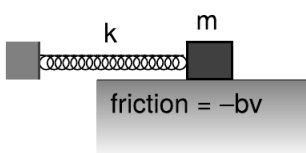
$$E_{spring} = \frac{1}{2}kx^2 \quad (1)$$

$$K = \frac{1}{2}mv^2. \quad (2)$$

In the case of a mechanical oscillator, we have usually assumed a



j / A series LRC circuit.



k / A mechanical analogy for the LRC circuit.

friction force of the form that turns out to give the nicest mathematical results, $F = -bv$. In the circuit, the dissipation of energy into heat occurs via the resistor, with no mechanical force involved, so in order to make the analogy, we need to restate the role of the friction force in terms of energy. The power dissipated by friction equals the mechanical work it does in a time interval Δt , divided by Δt , $P = W/\Delta t = F\Delta x/\Delta t = Fv = -bv^2$, so

$$\text{rate of heat dissipation} = -bv^2. \quad (3)$$

self-check A

Equation (1) has x squared, and equations (2) and (3) have v squared. Because they're squared, the results don't depend on whether these variables are positive or negative. Does this make physical sense? \triangleright
 Answer, p. 1044

In the circuit, the stored forms of energy are

$$E_C = \frac{1}{2C}q^2 \quad (1')$$

$$E_L = \frac{1}{2}LI^2, \quad (2')$$

and the rate of heat dissipation in the resistor is

$$\text{rate of heat dissipation} = -RI^2. \quad (3')$$

Comparing the two sets of equations, we first form analogies between quantities that represent the state of the system at some moment in time:

$$x \leftrightarrow q$$

$$v \leftrightarrow I$$

self-check B

How is v related mathematically to x ? How is I connected to q ? Are the two relationships analogous? \triangleright Answer, p. 1044

Next we relate the ones that describe the system's permanent characteristics:

$$k \leftrightarrow 1/C$$

$$m \leftrightarrow L$$

$$b \leftrightarrow R$$

Since the mechanical system naturally oscillates with a period $T = 2\pi\sqrt{m/k}$, we can immediately solve the electrical version by analogy, giving

$$T = 2\pi\sqrt{LC}.$$

Rather than period, T , and frequency, f , it turns out to be more convenient if we work with the quantity $\omega = 2\pi f$, which can be interpreted as the number of radians per second. Then

$$\omega = \frac{1}{\sqrt{LC}}.$$

Since the resistance R is analogous to b in the mechanical case, we find that the Q (quality factor, not charge) of the resonance is inversely proportional to R , and the width of the resonance is directly proportional to R .

Tuning a radio receiver

example 4

A radio receiver uses this kind of circuit to pick out the desired station. Since the receiver resonates at a particular frequency, stations whose frequencies are far off will not excite any response in the circuit. The value of R has to be small enough so that only one station at a time is picked up, but big enough so that the tuner isn't too touchy. The resonant frequency can be tuned by adjusting either L or C , but variable capacitors are easier to build than variable inductors.

A numerical calculation

example 5

The phone company sends more than one conversation at a time over the same wire, which is accomplished by shifting each voice signal into different range of frequencies during transmission. The number of signals per wire can be maximized by making each range of frequencies (known as a bandwidth) as small as possible. It turns out that only a relatively narrow range of frequencies is necessary in order to make a human voice intelligible, so the phone company filters out all the extreme highs and lows. (This is why your phone voice sounds different from your normal voice.)

▷ If the filter consists of an LRC circuit with a broad resonance centered around 1.0 kHz, and the capacitor is 1 μF (microfarad), what inductance value must be used?

▷ Solving for L , we have

$$\begin{aligned} L &= \frac{1}{C\omega^2} \\ &= \frac{1}{(10^{-6} \text{ F})(2\pi \times 10^3 \text{ s}^{-1})^2} \\ &= 2.5 \times 10^{-3} \text{ F}^{-1} \text{ s}^2 \end{aligned}$$

Checking that these really are the same units as henries is a little tedious, but it builds character:

$$\begin{aligned} \text{F}^{-1} \text{s}^2 &= (\text{C}^2/\text{J})^{-1} \text{s}^2 \\ &= \text{J} \cdot \text{C}^{-2} \text{s}^2 \\ &= \text{J}/\text{A}^2 \\ &= \text{H} \end{aligned}$$

The result is 25 mH (millihenries).

This is actually quite a large inductance value, and would require a big, heavy, expensive coil. In fact, there is a trick for making this kind of circuit small and cheap. There is a kind of silicon chip called an op-amp, which, among other things, can be used to simulate the behavior of an inductor. The main limitation of the op-amp is that it is restricted to low-power applications.

25.3 Voltage and current

What is physically happening in one of these oscillating circuits? Let's first look at the mechanical case, and then draw the analogy to the circuit. For simplicity, let's ignore the existence of damping, so there is no friction in the mechanical oscillator, and no resistance in the electrical one.

Suppose we take the mechanical oscillator and pull the mass away from equilibrium, then release it. Since friction tends to resist the spring's force, we might naively expect that having zero friction would allow the mass to leap instantaneously to the equilibrium position. This can't happen, however, because the mass would have to have infinite velocity in order to make such an instantaneous leap. Infinite velocity would require infinite kinetic energy, but the only kind of energy that is available for conversion to kinetic is the energy stored in the spring, and that is finite, not infinite. At each step on its way back to equilibrium, the mass's velocity is controlled exactly by the amount of the spring's energy that has so far been converted into kinetic energy. After the mass reaches equilibrium, it overshoots due to its own momentum. It performs identical oscillations on both sides of equilibrium, and it never loses amplitude because friction is not available to convert mechanical energy into heat.

Now with the electrical oscillator, the analog of position is charge. Pulling the mass away from equilibrium is like depositing charges $+q$ and $-q$ on the plates of the capacitor. Since resistance tends to resist the flow of charge, we might imagine that with no friction present, the charge would instantly flow through the inductor (which is, after all, just a piece of wire), and the capacitor would discharge instantly. However, such an instant discharge is impossible, because it would require infinite current for one instant. Infinite current would create infinite magnetic fields surrounding the inductor, and these fields would have infinite energy. Instead, the rate of flow of current is controlled at each instant by the relationship between the amount of energy stored in the magnetic field and the amount of current that must exist in order to have that strong a field. After the capacitor reaches $q = 0$, it overshoots. The circuit has its own kind of electrical "inertia," because if charge was to stop flowing, there would have to be zero current through the inductor. But the current in the inductor must be related to the amount of energy stored in

its magnetic fields. When the capacitor is at $q = 0$, all the circuit's energy is in the inductor, so it must therefore have strong magnetic fields surrounding it and quite a bit of current going through it.

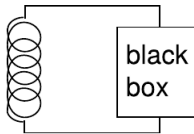
The only thing that might seem spooky here is that we used to speak as if the current in the inductor caused the magnetic field, but now it sounds as if the field causes the current. Actually this is symptomatic of the elusive nature of cause and effect in physics. It's equally valid to think of the cause and effect relationship in either way. This may seem unsatisfying, however, and for example does not really get at the question of what brings about a voltage difference across the resistor (in the case where the resistance is finite); there must be such a voltage difference, because without one, Ohm's law would predict zero current through the resistor.

Voltage, then, is what is really missing from our story so far.

Let's start by studying the voltage across a capacitor. Voltage is electrical potential energy per unit charge, so the voltage difference between the two plates of the capacitor is related to the amount by which its energy would increase if we increased the absolute values of the charges on the plates from q to $q + \Delta q$:

$$\begin{aligned} V_C &= (E_{q+\Delta q} - E_q) / \Delta q \\ &= \frac{\Delta E_C}{\Delta q} \\ &= \frac{\Delta}{\Delta q} \left(\frac{1}{2C} q^2 \right) \\ &= \frac{q}{C} \end{aligned}$$

Many books use this as the definition of capacitance. This equation, by the way, probably explains the historical reason why C was defined so that the energy was *inversely* proportional to C for a given value of q : the people who invented the definition were thinking of a capacitor as a device for storing charge rather than energy, and the amount of charge stored for a fixed voltage (the charge "capacity") is proportional to C .



1 / The inductor releases energy and gives it to the black box.

In the case of an inductor, we know that if there is a steady, constant current flowing through it, then the magnetic field is constant, and so is the amount of energy stored; no energy is being exchanged between the inductor and any other circuit element. But what if the current is changing? The magnetic field is proportional to the current, so a change in one implies a change in the other. For concreteness, let's imagine that the magnetic field and the current are both decreasing. The energy stored in the magnetic field is therefore decreasing, and by conservation of energy, this energy can't just go away — some other circuit element must be taking energy from the inductor. The simplest example, shown in figure 1, is a series circuit consisting of the inductor plus one other circuit element. It

doesn't matter what this other circuit element is, so we just call it a black box, but if you like, we can think of it as a resistor, in which case the energy lost by the inductor is being turned into heat by the resistor. The junction rule tells us that both circuit elements have the same current through them, so I could refer to either one, and likewise the loop rule tells us $V_{inductor} + V_{black\ box} = 0$, so the two voltage drops have the same absolute value, which we can refer to as V . Whatever the black box is, the rate at which it is taking energy from the inductor is given by $|P| = |IV|$, so

$$\begin{aligned} |IV| &= \left| \frac{\Delta E_L}{\Delta t} \right| \\ &= \left| \frac{\Delta}{\Delta t} \left(\frac{1}{2} LI^2 \right) \right| \\ &= \left| LI \frac{\Delta I}{\Delta t} \right|, \end{aligned}$$

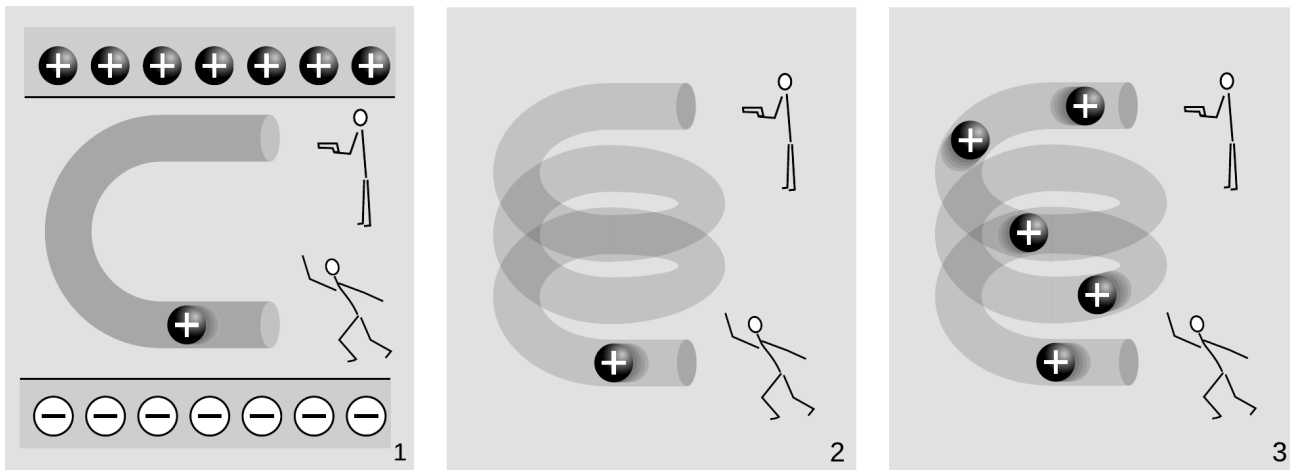
or

$$|V| = \left| L \frac{\Delta I}{\Delta t} \right|,$$

which in many books is taken to be the definition of inductance. The direction of the voltage drop (plus or minus sign) is such that the inductor resists the change in current.

There's one very intriguing thing about this result. Suppose, for concreteness, that the black box in figure 1 is a resistor, and that the inductor's energy is decreasing, and being converted into heat in the resistor. The voltage drop across the resistor indicates that it has an electric field across it, which is driving the current. But where is this electric field coming from? There are no charges anywhere that could be creating it! What we've discovered is one special case of a more general principle, the principle of induction: a changing magnetic field creates an electric field, which is in addition to any electric field created by charges. (The reverse is also true: any electric field that changes over time creates a magnetic field.) Induction forms the basis for such technologies as the generator and the transformer, and ultimately it leads to the existence of light, which is a wave pattern in the electric and magnetic fields. These are all topics for chapter 24, but it's truly remarkable that we could come to this conclusion without yet having learned any details about magnetism.

The cartoons in figure m compares electric fields made by charges, 1, to electric fields made by changing magnetic fields, 2-3. In m/1, two physicists are in a room whose ceiling is positively charged and whose floor is negatively charged. The physicist on the bottom



m / Electric fields made by charges, 1, and by changing magnetic fields, 2 and 3.

throws a positively charged bowling ball into the curved pipe. The physicist at the top uses a radar gun to measure the speed of the ball as it comes out of the pipe. They find that the ball has slowed down by the time it gets to the top. By measuring the change in the ball's kinetic energy, the two physicists are acting just like a voltmeter. They conclude that the top of the tube is at a higher voltage than the bottom of the pipe. A difference in voltage indicates an electric field, and this field is clearly being caused by the charges in the floor and ceiling.

In m/2, there are no charges anywhere in the room except for the charged bowling ball. Moving charges make magnetic fields, so there is a magnetic field surrounding the helical pipe while the ball is moving through it. A magnetic field has been created where there was none before, and that field has energy. Where could the energy have come from? It can only have come from the ball itself, so the ball must be losing kinetic energy. The two physicists working together are again acting as a voltmeter, and again they conclude that there is a voltage difference between the top and bottom of the pipe. This indicates an electric field, but this electric field can't have been created by any charges, because there aren't any in the room. This electric field was created by the change in the magnetic field.

The bottom physicist keeps on throwing balls into the pipe, until the pipe is full of balls, m/3, and finally a steady current is established. While the pipe was filling up with balls, the energy in the magnetic field was steadily increasing, and that energy was being stolen from the balls' kinetic energy. But once a steady current is established, the energy in the magnetic field is no longer changing. The balls no longer have to give up energy in order to build up the field, and the physicist at the top finds that the balls are exiting the pipe at full speed again. There is no voltage difference any more. Although

there is a current, $\Delta I/\Delta t$ is zero.

Discussion question

A What happens when the physicist at the bottom in figure m/3 starts getting tired, and decreases the current?

25.4 Decay

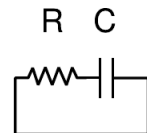
Up until now I've soft-pedaled the fact that by changing the characteristics of an oscillator, it is possible to produce non-oscillatory behavior. For example, imagine taking the mass-on-a-spring system and making the spring weaker and weaker. In the limit of small k , it's as though there was no spring whatsoever, and the behavior of the system is that if you kick the mass, it simply starts slowing down. For friction proportional to v , as we've been assuming, the result is that the velocity approaches zero, but never actually reaches zero. This is unrealistic for the mechanical oscillator, which will not have vanishing friction at low velocities, but it is quite realistic in the case of an electrical circuit, for which the voltage drop across the resistor really does approach zero as the current approaches zero.

Electrical circuits can exhibit all the same behavior. For simplicity we will analyze only the cases of LRC circuits with $L = 0$ or $C = 0$.

The RC circuit

We first analyze the RC circuit, n. In reality one would have to "kick" the circuit, for example by briefly inserting a battery, in order to get any interesting behavior. We start with Ohm's law and the equation for the voltage across a capacitor:

$$V_R = IR$$
$$V_C = q/C$$



n / An RC circuit.

The loop rule tells us

$$V_R + V_C = 0,$$

and combining the three equations results in a relationship between q and I :

$$I = -\frac{1}{RC}q$$

The negative sign tells us that the current tends to reduce the charge on the capacitor, i.e. to discharge it. It makes sense that the current is proportional to q : if q is large, then the attractive forces between the $+q$ and $-q$ charges on the plates of the capacitor are large, and charges will flow more quickly through the resistor in order to reunite. If there was zero charge on the capacitor plates, there would be no reason for current to flow. Since amperes, the unit of current, are the same as coulombs per second, it appears that the quantity RC must have units of seconds, and you can check for yourself that

this is correct. RC is therefore referred to as the time constant of the circuit.

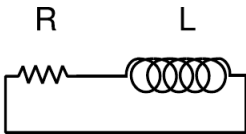
How exactly do I and q vary with time? Rewriting I as $\Delta q/\Delta t$, we have

$$\frac{\Delta q}{\Delta t} = -\frac{1}{RC}q.$$

This equation describes a function $q(t)$ that always gets smaller over time, and whose rate of decrease is big at first, when q is big, but gets smaller and smaller as q approaches zero. As an example of this type of mathematical behavior, we could imagine a man who has 1024 weeds in his backyard, and resolves to pull out half of them every day. On the first day, he pulls out half, and has 512 left. The next day, he pulls out half of the remaining ones, leaving 256. The sequence continues exponentially: 128, 64, 32, 16, 8, 4, 2, 1. Returning to our electrical example, the function $q(t)$ apparently needs to be an exponential, which we can write in the form ae^{bt} , where $e = 2.718\dots$ is the base of natural logarithms. We could have written it with base 2, as in the story of the weeds, rather than base e , but the math later on turns out simpler if we use e . It doesn't make sense to plug a number that has units into a function like an exponential, so bt must be unitless, and b must therefore have units of inverse seconds. The number b quantifies how fast the exponential decay is. The only physical parameters of the circuit on which b could possibly depend are R and C , and the only way to put units of ohms and farads together to make units of inverse seconds is by computing $1/RC$. Well, actually we could use $7/RC$ or $3\pi/RC$, or any other unitless number divided by RC , but this is where the use of base e comes in handy: for base e , it turns out that the correct unitless constant is 1. Thus our solution is

$$q = q_0 \exp\left(-\frac{t}{RC}\right).$$

The number RC , with units of seconds, is called the RC time constant of the circuit, and it tells us how long we have to wait if we want the charge to fall off by a factor of $1/e$.



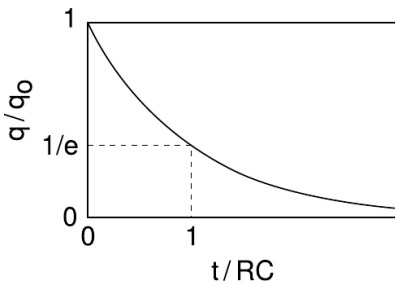
p / An RL circuit.

The RL circuit

The RL circuit, p, can be attacked by similar methods, and it can easily be shown that it gives

$$I = I_0 \exp\left(-\frac{R}{L}t\right).$$

The RL time constant equals L/R .



o / Over a time interval RC , the charge on the capacitor is reduced by a factor of e .

When we suddenly break an RL circuit, what will happen? It might seem that we're faced with a paradox, since we only have two forms of energy, magnetic energy and heat, and if the current stops suddenly, the magnetic field must collapse suddenly. But where does the lost magnetic energy go? It can't go into resistive heating of the resistor, because the circuit has now been broken, and current can't flow!

The way out of this conundrum is to recognize that the open gap in the circuit has a resistance which is large, but not infinite. This large resistance causes the RL time constant L/R to be very small. The current thus continues to flow for a very brief time, and flows straight across the air gap where the circuit has been opened. In other words, there is a spark!

We can determine based on several different lines of reasoning that the voltage drop from one end of the spark to the other must be very large. First, the air's resistance is large, so $V = IR$ requires a large voltage. We can also reason that all the energy in the magnetic field is being dissipated in a short time, so the power dissipated in the spark, $P = IV$, is large, and this requires a large value of V . (I isn't large — it is decreasing from its initial value.) Yet a third way to reach the same result is to consider the equation $V_L = \Delta I/\Delta t$: since the time constant is short, the time derivative $\Delta I/\Delta t$ is large.

This is exactly how a car's spark plugs work. Another application is to electrical safety: it can be dangerous to break an inductive circuit suddenly, because so much energy is released in a short time. There is also no guarantee that the spark will discharge across the air gap; it might go through your body instead, since your body might have a lower resistance.

Discussion question

A A gopher gnaws through one of the wires in the DC lighting system in your front yard, and the lights turn off. At the instant when the circuit becomes open, we can consider the bare ends of the wire to be like the plates of a capacitor, with an air gap (or gopher gap) between them. What kind of capacitance value are we talking about here? What would this tell you about the RC time constant?

25.5 Impedance

So far we have been thinking in terms of the free oscillations of a circuit. This is like a mechanical oscillator that has been kicked but then left to oscillate on its own without any external force to keep the vibrations from dying out. Suppose an LRC circuit is driven with a sinusoidally varying voltage, such as will occur when a radio tuner is hooked up to a receiving antenna. We know that a current will flow in the circuit, and we know that there will be resonant behavior, but it is not necessarily simple to relate current to voltage in the most general case. Let's start instead with the special cases of LRC circuits consisting of only a resistance, only a capacitance, or only an inductance. We are interested only in the steady-state response.

The purely resistive case is easy. Ohm's law gives

$$I = \frac{V}{R}.$$

In the purely capacitive case, the relation $V = q/C$ lets us calculate

$$\begin{aligned} I &= \frac{\Delta q}{\Delta t} \\ &= C \frac{\Delta V}{\Delta t}. \end{aligned}$$

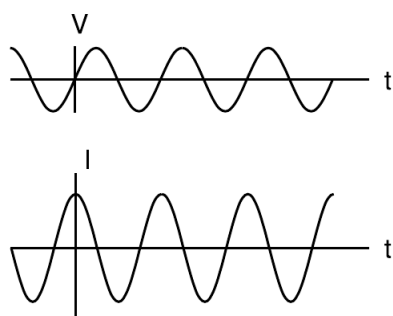
If the voltage varies as, for example, $V(t) = \tilde{V} \sin(\omega t)$, then it can be shown using calculus that the current will be $I(t) = \omega C \tilde{V} \cos(\omega t)$, so the maximum current is $\tilde{I} = \omega C \tilde{V}$. By analogy with Ohm's law, we can then write

$$\tilde{I} = \frac{\tilde{V}}{Z_C},$$

where the quantity

$$Z_C = \frac{1}{\omega C}, \quad [\text{impedance of a capacitor}]$$

having units of ohms, is called the *impedance* of the capacitor at this frequency. Note that it is only the *maximum* current, \tilde{I} , that is proportional to the *maximum* voltage, \tilde{V} , so the capacitor is not behaving like a resistor. The maxima of V and I occur at different times, as shown in figure q. It makes sense that the impedance becomes infinite at zero frequency. Zero frequency means that it would take an infinite time before the voltage would change by any amount. In other words, this is like a situation where the capacitor has been connected across the terminals of a battery and been allowed to settle down to a state where there is constant charge on both terminals. Since the electric fields between the plates are constant, there is no energy being added to or taken out of the field. A capacitor that can't exchange energy with any other circuit component is nothing more than a broken (open) circuit.



q / In a capacitor, the current is 90° ahead of the voltage in phase.

self-check C

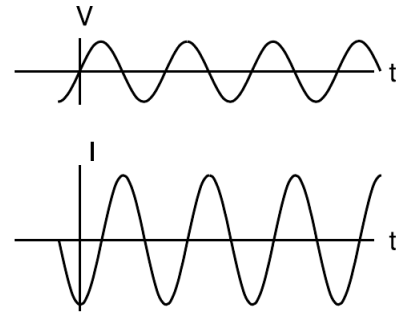
Why can't a capacitor have its impedance printed on it along with its capacitance? ▷ Answer, p. 1044

Similar math gives

$$Z_L = \omega L \quad [\text{impedance of an inductor}]$$

for an inductor. It makes sense that the inductor has lower impedance at lower frequencies, since at zero frequency there is no change in the magnetic field over time. No energy is added to or released from the magnetic field, so there are no induction effects, and the inductor acts just like a piece of wire with negligible resistance. The term “choke” for an inductor refers to its ability to “choke out” high frequencies.

The phase relationships shown in figures q and r can be remembered using my own mnemonic, “eVIL,” which shows that the voltage (V) leads the current (I) in an inductive circuit, while the opposite is true in a capacitive one. A more traditional mnemonic is “ELI the ICE man,” which uses the notation E for emf, a concept closely related to voltage.



r / The current through an inductor lags behind the voltage by a phase angle of 90° .

Low-pass and high-pass filters *example 7*

An LRC circuit only responds to a certain range (band) of frequencies centered around its resonant frequency. As a filter, this is known as a bandpass filter. If you turn down both the bass and the treble on your stereo, you have created a bandpass filter.

To create a high-pass or low-pass filter, we only need to insert a capacitor or inductor, respectively, in series. For instance, a very basic surge protector for a computer could be constructed by inserting an inductor in series with the computer. The desired 60 Hz power from the wall is relatively low in frequency, while the surges that can damage your computer show much more rapid time variation. Even if the surges are not sinusoidal signals, we can think of a rapid “spike” qualitatively as if it was very high in frequency — like a high-frequency sine wave, it changes very rapidly.

Inductors tend to be big, heavy, expensive circuit elements, so a simple surge protector would be more likely to consist of a capacitor in *parallel* with the computer. (In fact one would normally just connect one side of the power circuit to ground via a capacitor.) The capacitor has a very high impedance at the low frequency of the desired 60 Hz signal, so it siphons off very little of the current. But for a high-frequency signal, the capacitor’s impedance is very small, and it acts like a zero-impedance, easy path into which the current is diverted.

The main things to be careful about with impedance are that (1) the concept only applies to a circuit that is being driven sinusoidally, (2)

the impedance of an inductor or capacitor is frequency-dependent, and (3) impedances in parallel and series don't combine according to the same rules as resistances. It is possible, however, to get around the third limitation, as discussed in subsection .

Discussion questions

A Figure q on page 748 shows the voltage and current for a capacitor. Sketch the $q-t$ graph, and use it to give a physical explanation of the phase relationship between the voltage and current. For example, why is the current zero when the voltage is at a maximum or minimum?

B Relate the features of the graph in figure r on page 749 to the story told in cartoons in figure m/2-3 on page 744.