Close to the surface, the relationship between $E$ and $\sigma$ is a fixed one, regardless of the geometry. The flea can't determine the size or shape of her world by comparing $E$ and $\sigma$.

The plates are a metal conductor, not an insulator, so the charge will tend to arrange itself more densely near the edges, rather than spreading itself uniformly on each plate. Furthermore, we have only calculated the on-axis field in example 15; in the off-axis region, each disk's contribution to the field will be weaker, and it will also point away from the axis a little. But if we are willing to ignore these complications for the sake of a rough analysis, then the fields superimpose as shown in figure f: the fields cancel the outside of the capacitor, but between the plates its value is double that contributed by a single plate. This cancellation on the outside is a very useful property for a practical capacitor. For instance, if you look at the printed circuit board in a typical piece of consumer electronics, there are many capacitors, often placed fairly close together. If their exterior fields didn't cancel out nicely, then each capacitor would interact with its neighbors in a complicated way, and the behavior of the circuit would depend on the exact physical layout, since the interaction would be stronger or weaker depending on distance. In reality, a capacitor does create weak external electric fields, but their effects are often negligible, and we can then use the lumped-circuit approximation, which states that each component's behavior depends only on the currents that flow in and out of it, not on the interaction of its fields with the other components.

10.3.2 The field near a charged surface

From a theoretical point of view, there is something even more intriguing about example 15: the magnitude of the field for small values of $z$ ($z \ll b$) is $E = 2\pi k \sigma$, which doesn't depend on $b$ at all for a fixed value of $\sigma$. If we made a disk with twice the radius, and covered it with the same number of coulombs per square meter (resulting in a total charge four times as great), the field close to the disk would be unchanged! That is, a flea living near the center of the disk, $h$, would have no way of determining the size of her flat "planet" by measuring the local field and charge density. (Only by leaping off the surface into outer space would she be able to measure fields that were dependent on $b$. If she traveled very far, to $z \gg b$, she would be in the region where the field is well approximated by $|E| \approx kQ/z^2 = k\pi b^2 \sigma/z^2$, which she could solve for $b$.)

What is the reason for this surprisingly simple behavior of the field? Is it a piece of mathematical trivia, true only in this particular case? What if the shape was a square rather than a circle? In other words, the flea gets no information about the size of the disk from measuring $E$, since $E = 2\pi k \sigma$, independent of $b$, but what if she didn't know the shape, either? If the result for a square had some other geometrical factor in front instead of $2\pi$, then she could tell which shape it was by measuring $E$. The surprising mathematical fact, however, is that the result for a square, indeed for any shape whatsoever, is $E = 2\pi \sigma k$. It doesn't even matter whether the sur-
face is flat or warped, or whether the density of charge is different at parts of the surface which are far away compared to the flea’s distance above the surface.

This universal $E_\perp = 2\pi k \sigma$ field perpendicular to a charged surface can be proved mathematically based on Gauss’s law\(^1\) (section 10.6), but we can understand what’s happening on qualitative grounds. Suppose on night, while the flea is asleep, someone adds more surface area, also positively charged, around the outside edge of her disk-shaped world, doubling its radius. The added charge, however, has very little effect on the field in her environment, as long as she stays at low altitudes above the surface. As shown in figure i, the new charge to her west contributes a field, $T$, that is almost purely “horizontal” (i.e., parallel to the surface) and to the east. It has a negligible upward component, since the angle is so shallow. This new eastward contribution to the field is exactly canceled out by the westward field, $S$, created by the new charge to her east. There is likewise almost perfect cancellation between any other pair of opposite compass directions.

A similar argument can be made as to the shape-independence of the result, as long as the shape is symmetric. For example, suppose that the next night, the tricky real estate developers decide to add corners to the disk and transform it into a square. Each corner’s contribution to the field measured at the center is canceled by the field due to the corner diagonally across from it.

What if the flea goes on a trip away from the center of the disk? The perfect cancellation of the “horizontal” fields contributed by distant charges will no longer occur, but the “vertical” field (i.e., the field perpendicular to the surface) will still be $E_\perp = 2\pi k \sigma$, where $\sigma$ is the local charge density, since the distant charges can’t contribute to the vertical field. The same result applies if the shape of the surface is asymmetric, and doesn’t even have any well-defined geometric center: the component perpendicular to the surface is $E_\perp = 2\pi k \sigma$, but we may have $E_\parallel \neq 0$. All of the above arguments can be made more rigorous by discussing mathematical limits rather than using words like “very small.” There is not much point in giving a rigorous proof here, however, since we will be able to demonstrate this fact as a corollary of Gauss’ Law in section 10.6. The result is as follows:

At a point lying a distance $z$ from a charged surface, the component of the electric field perpendicular to the surface obeys

$$\lim_{z \to 0} E_\perp = 2\pi k \sigma,$$

where $\sigma$ is the charge per unit area. This is true regardless of the shape or size of the surface.

---

\(^1\)rhymes with “mouse”
The field near a point, line, or surface charge example 16

Compare the variation of the electric field with distance, \(d\), for small values of \(d\) in the case of a point charge, an infinite line of charge, and an infinite charged surface.

For a point charge, we have already found \(E \propto d^{-2}\) for the magnitude of the field, where we are now using \(d\) for the quantity we would ordinarily notate as \(r\). This is true for all values of \(d\), not just for small \(d\) — it has to be that way, because the point charge has no size, so if \(E\) behaved differently for small and large \(d\), there would be no way to decide what \(d\) had to be small or large relative to.

For a line of charge, the result of example 13 is

\[ E = \frac{k\lambda L}{d^2 \sqrt{1 + L^2/4d^2}}. \]

In the limit of \(d \ll L\), the quantity inside the square root is dominated by the second term, and we have \(E \propto d^{-1}\).

Finally, in the case of a charged surface, the result is simply \(E = 2\pi\sigma k\), or \(E \propto d^0\).

Notice the lovely simplicity of the pattern, as shown in figure j. A point is zero-dimensional: it has no length, width, or breadth. A line is one-dimensional, and a surface is two-dimensional. As the dimensionality of the charged object changes from 0 to 1, and then to 2, the exponent in the near-field expression goes from 2 to 1 to 0.
10.4 Energy in fields

10.4.1 Electric field energy

Fields possess energy, as argued on page 579, but how much energy? The answer can be found using the following elegant approach. We assume that the electric energy contained in an infinitesimal volume of space $d\mathbf{v}$ is given by $dU_e = f(E)\, dv$, where $f$ is some function, which we wish to determine, of the field $E$. It might seem that we would have no easy way to determine the function $f$, but many of the functions we could cook up would violate the symmetry of space. For instance, we could imagine $f(E) = aE_y$, where $a$ is some constant with the appropriate units. However, this would violate the symmetry of space, because it would give the $y$ axis a different status from $x$ and $z$. As discussed on page 216, if we wish to calculate a scalar based on some vectors, the dot product is the only way to do it that has the correct symmetry properties. If all we have is one vector, $E$, then the only scalar we can form is $E \cdot E$, which is the square of the magnitude of the electric field vector.

In principle, the energy function we are seeking could be proportional to $E \cdot E$, or to any function computed from it, such as $\sqrt{E \cdot E}$ or $(E \cdot E)^7$. On physical grounds, however, the only possibility that works is $E \cdot E$. Suppose, for instance, that we pull apart two oppositely charged capacitor plates, as shown in figure a. We are doing work by pulling them apart against the force of their electrical attraction, and this quantity of mechanical work equals the increase in electrical energy, $U_e$. Using our previous approach to energy, we would have thought of $U_e$ as a quantity which depended on the distance of the positive and negative charges from each other, but now we’re going to imagine $U_e$ as being stored within the electric field that exists in the space between and around the charges. When the plates are touching, their fields cancel everywhere, and there is zero electrical energy. When they are separated, there is still approximately zero field on the outside, but the field between the plates is nonzero, and holds some energy.

Now suppose we carry out the whole process, but with the plates carrying double their previous charges. Since Coulomb’s law involves the product $q_1 q_2$ of two charges, we have quadrupled the force between any given pair of charged particles, and the total attractive force is therefore also four times greater than before. This means that the work done in separating the plates is four times greater, and so is the energy $U_e$ stored in the field. The field, however, has merely been doubled at any given location: the electric field $E_+$ due to the positively charged plate is doubled, and similarly for the contribution $E_-$ from the negative one, so the total electric field $E_+ + E_-$ is also doubled. Thus doubling the field results in an electrical energy which is four times greater, i.e., the energy density must be proportional to the square of the field, $dU_e \propto (E \cdot E)\, dv$. For ease
of notation, we write this as \( dU_e \propto E^2 \, dv \), or \( dU_e = aE^2 \, dv \), where \( a \) is a constant of proportionality. Note that we never really made use of any of the details of the geometry of figure a, so the reasoning is of general validity. In other words, not only is \( dU_e = aE^2 \, dv \) the function that works in this particular case, but there is every reason to believe that it would work in other cases as well.

It now remains only to find \( a \). Since the constant must be the same in all situations, we only need to find one example in which we can compute the field and the energy, and then we can determine \( a \). The situation shown in figure a is just about the easiest example to analyze. We let the square capacitor plates be uniformly covered with charge densities \( +\sigma \) and \( -\sigma \), and we write \( b \) for the lengths of their sides. Let \( h \) be the gap between the plates after they have been separated. We choose \( h \ll b \), so that the field experienced by the negative plate due to the positive plate is \( E_+ = 2\pi k\sigma \). The charge of the negative plate is \( -\sigma b^2 \), so the magnitude of the force attracting it back toward the positive plate is \( \text{(force)} = \text{(charge)(field)} = 2\pi k\sigma^2 b^2 \). The amount of work done in separating the plates is \( \text{(work)} = \text{(force)(distance)} = 2\pi k\sigma^2 b^2 h \). This is the amount of energy that has been stored in the field between the two plates, \( U_e = 2\pi k\sigma^2 b^2 h = 2\pi k\sigma^2 v \), where \( v \) is the volume of the region between the plates.

We want to equate this to \( U_e = aE^2 v \). (We can write \( U_e \) and \( v \) rather than \( dU_e \) and \( dv \), since the field is constant in the region between the plates.) The field between the plates has contributions from both plates, \( E = E_+ + E_- = 4\pi k\sigma \). (We only used half this value in the computation of the work done on the moving plate, since the moving plate can’t make a force on itself. Mathematically, each plate is in a region where its own field is reversing directions, so we can think of its own contribution to the field as being zero within itself.) We then have \( aE^2 v = a \cdot 16\pi^2 k^2 \sigma^2 \cdot v \), and setting this equal to \( U_e = 2\pi k\sigma^2 v \) from the result of the work computation, we find \( a = 1/8\pi k \). Our final result is as follows:

The electric energy possessed by an electric field \( \mathbf{E} \) occupying an infinitesimal volume of space \( dv \) is given by

\[
dU_e = \frac{1}{8\pi k} E^2 \, dv,
\]

where \( E^2 = \mathbf{E} \cdot \mathbf{E} \) is the square of the magnitude of the electric field.

This is reminiscent of how waves behave: the energy content of a wave is typically proportional to the square of its amplitude.
We can think of the quantity $dU_e/dv$ as the energy density due to the electric field, i.e., the number of joules per cubic meter needed in order to create that field. (a) How does this quantity depend on the components of the field vector, $E_x$, $E_y$, and $E_z$? (b) Suppose we have a field with $E_x \neq 0$, $E_y=0$, and $E_z=0$. What would happen to the energy density if we reversed the sign of $E_x$?

A numerical example example 17

A capacitor has plates whose areas are $10^{-4} \text{ m}^2$, separated by a gap of $10^{-5} \text{ m}$. A 1.5-volt battery is connected across it. How much energy is sucked out of the battery and stored in the electric field between the plates? (A real capacitor typically has an insulating material between the plates whose molecules interact electrically with the charge in the plates. For this example, we’ll assume that there is just a vacuum in between the plates. The plates are also typically rolled up rather than flat.)

To connect this with our previous calculations, we need to find the charge density on the plates in terms of the voltage we were given. Our previous examples were based on the assumption that the gap between the plates was small compared to the size of the plates. Is this valid here? Well, if the plates were square, then the area of $10^{-4} \text{ m}^2$ would imply that their sides were $10^{-2} \text{ m}$ in length. This is indeed very large compared to the gap of $10^{-5} \text{ m}$, so this assumption appears to be valid (unless, perhaps, the plates have some very strange, long and skinny shape).

Based on this assumption, the field is relatively uniform in the whole volume between the plates, so we can use a single symbol, $E$, to represent its magnitude, and the relation $E = dV/dx$ is equivalent to $E = \Delta V/\Delta x = (1.5 \text{ V})/(\text{gap}) = 1.5 \times 10^5 \text{ V/m}$.

Since the field is uniform, we can dispense with the calculus, and replace $dU_e = (1/8\pi k)E^2 dv$ with $U_e = (1/8\pi k)E^2 v$. The volume equals the area multiplied by the gap, so we have

$$U_e = (1/8\pi k)E^2(\text{area})(\text{gap})$$

$$= \frac{1}{8\pi \times 9 \times 10^9 \text{ N.m}^2/\text{C}^2}(1.5 \times 10^5 \text{ V/m})^2(10^{-4} \text{ m}^2)(10^{-5} \text{ m})$$

$$= 1 \times 10^{-10} \text{ J}$$

self-check E

Show that the units in the preceding example really do work out to be joules.

Why $k$ is on the bottom example 18

It may also seem strange that the constant $k$ is in the denominator of the equation $dU_e = (1/8\pi k)E^2 dv$. The Coulomb constant $k$ tells us how strong electric forces are, so shouldn’t it be on top?
No. Consider, for instance, an alternative universe in which electric forces are twice as strong as in ours. The numerical value of \( k \) is doubled. Because \( k \) is doubled, all the electric field strengths are doubled as well, which quadruples the quantity \( E^2 \). In the expression \( E^2/8\pi k \), we've quadrupled something on top and doubled something on the bottom, which makes the energy twice as big. That makes perfect sense.

Example 19.

Imagine taking two opposite charges, \( b \), that were initially far apart and allowing them to come together under the influence of their electrical attraction.

According to our old approach, electrical energy is lost because the electric force did positive work as it brought the charges together. (This makes sense because as they come together and accelerate it is their electrical energy that is being lost and converted to kinetic energy.)

By the new method, we must ask how the energy stored in the electric field has changed. In the region indicated approximately by the shading in the figure, the superposing fields of the two charges undergo partial cancellation because they are in opposing directions. The energy in the shaded region is reduced by this effect. In the unshaded region, the fields reinforce, and the energy is increased.

It would be quite a project to do an actual numerical calculation of the energy gained and lost in the two regions (this is a case where the old method of finding energy gives greater ease of computation), but it is fairly easy to convince oneself that the energy is less when the charges are closer. This is because bringing the charges together shrinks the high-energy unshaded region and enlarges the low-energy shaded region.

Example 19.

A spherical capacitor example 20

A spherical capacitor, \( c \), consists of two concentric spheres of radii \( a \) and \( b \). Find the energy required to charge up the capacitor so that the plates hold charges \(+q\) and \(-q\).

On page 102, I proved that for gravitational forces, the interaction of a spherical shell of mass with other masses outside it is the same as if the shell’s mass was concentrated at its center. On the interior of such a shell, the forces cancel out exactly. Since gravity and the electric force both vary as \( 1/r^2 \), the same proof carries over immediately to electrical forces. The magnitude of the outward electric field contributed by the charge \(+q\) of the central sphere is therefore

\[
|E_+| = \begin{cases} 
0, & r < a \\
\frac{kq}{r^2}, & r > a 
\end{cases}
\]

where \( r \) is the distance from the center. Similarly, the magnitude of the inward electric field contributed by the charge \(-q\) of the inner sphere is

\[
|E_-| = \begin{cases} 
0, & r < a \\
\frac{kq}{r^2}, & r > a
\end{cases}
\]

For \( r < a \), the magnitude is zero because the field is not present. For \( r > a \), the magnitude is

\[
|E_-| = \frac{kq}{r^2}
\]

By superposing these fields, the net fields in the interior are

\[
E_i = E_+ + E_- = \frac{kq}{r^2}
\]

The total energy stored in the interior is

\[
\frac{1}{2} \kappa E_i^2 = \frac{1}{2} \frac{k^2q^2}{r^2}
\]

As \( r \) decreases, \( E_i^2 \) increases, and so does the energy as you approach the inner sphere.
of the inward field contributed by the outside sphere is

$$|E_-| = \begin{cases} 0, & r < b \\ kq/r^2, & r > b \end{cases}. $$

In the region outside the whole capacitor, the two fields are equal in magnitude, but opposite in direction, so they cancel. We then have for the total field

$$|E| = \begin{cases} 0, & r < a \\ kq/r^2, & a < r < b \\ 0, & r > b \end{cases} $$

so to calculate the energy, we only need to worry about the region $a < r < b$. The energy density in this region is

$$\frac{dU_e}{dV} = \frac{1}{8\pi k} E^2 = \frac{kq^2}{8\pi} r^{-4}. $$

This expression only depends on $r$, so the energy density is constant across any sphere of radius $r$. We can slice the region $a < r < b$ into concentric spherical layers, like an onion, and the energy within one such layer, extending from $r$ to $r + dr$ is

$$dU_e = \frac{dU_e}{dV} dV = \frac{dU_e}{dV} (\text{area of shell})(\text{thickness of shell}) = \left( \frac{kq^2}{8\pi} r^{-4} \right) (4\pi r^2) (dr) = \frac{kq^2}{2} r^{-2} dr. $$

Integrating over all the layers to find the total energy, we have

$$U_e = \int dU_e = \int_a^b \frac{kq^2}{2} r^{-2} dr = \left. -\frac{kq^2}{2} r^{-1} \right|_a^b = \frac{kq^2}{2} \left( \frac{1}{a} - \frac{1}{b} \right) $$
Discussion Questions

A The figure shows a positive charge in the gap between two capacitor plates. Compare the energy of the electric fields in the two cases. Does this agree with what you would have expected based on your knowledge of electrical forces?

B The figure shows a spherical capacitor. In the text, the energy stored in its electric field is shown to be

\[ U_e = \frac{kq^2}{2} \left( \frac{1}{a} - \frac{1}{b} \right). \]

What happens if the difference between \( b \) and \( a \) is very small? Does this make sense in terms of the mechanical work needed in order to separate the charges? Does it make sense in terms of the energy stored in the electric field? Should these two energies be added together?

Similarly, discuss the cases of \( b \to \infty \) and \( a \to 0 \).

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

D In example 19 on page 607, I argued that for the charges shown in the figure, the fields contain less energy when the charges are closer together, because the region of cancellation expanded, while the region of reinforcing fields shrank. Perhaps a simpler approach is to consider the two extreme possibilities: the case where the charges are infinitely far apart, and the one in which they are at zero distance from each other, i.e., right on top of each other. Carry out this reasoning for the case of (1) a positive charge and a negative charge of equal magnitude, (2) two positive charges of equal magnitude, (3) the gravitational energy of two equal masses.

10.4.2 Gravitational field energy

Example B depended on the close analogy between electric and gravitational forces. In fact, every argument, proof, and example discussed so far in this section is equally valid as a gravitational example, provided we take into account one fact: only positive mass exists, and the gravitational force between two masses is attractive. This is the opposite of what happens with electrical forces, which are repulsive in the case of two positive charges. As a consequence of this, we need to assign a negative energy density to the gravitational field! For a gravitational field, we have

\[ dU_g = \frac{1}{8\pi G} g^2 dv, \]

where \( g^2 = g \cdot g \) is the square of the magnitude of the gravitational field.

10.4.3 Magnetic field energy

So far we’ve only touched in passing on the topic of magnetic fields, which will deal with in detail in chapter 11. Magnetism is an interaction between moving charge and moving charge, i.e., between currents and currents. Since a current has a direction in
space, while charge doesn’t, we can anticipate that the mathematical rule connecting a magnetic field to its source-currents will have to be completely different from the one relating the electric field to its source-charges. However, if you look carefully at the argument leading to the relation \( \frac{dU_e}{dv} = \frac{E^2}{8\pi k} \), you’ll see that these mathematical details were only necessary to the part of the argument in which we fixed the constant of proportionality. To establish \( \frac{dU_e}{dv} \propto E^2 \), we only had to use three simple facts:

- The field is proportional to the source.
- Forces are proportional to fields.
- Field contributed by multiple sources add like vectors.

All three of these statements are true for the magnetic field as well, so without knowing anything more specific about magnetic fields — not even what units are used to measure them! — we can state with certainty that the energy density in the magnetic field is proportional to the square of the magnitude of the magnetic field. The constant of proportionality is given on p. 691.

\[ \text{Current is a scalar, since the definition } I = \frac{dq}{dt} \text{ is the derivative of a scalar. However, there is a closely related quantity called the current density, } J, \text{ which is a vector, and } J \text{ is in fact the more fundamentally important quantity.} \]
10.5 LRC circuits

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 section 3.3 in the case of a mechanical system, although we didn’t call it that.

10.5.1 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 \(1/2C\),

\[U_C = \frac{1}{2C}q^2.\]

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 \(\mu\text{F}\), 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 parallel-plate capacitor example 21

Suppose a capacitor consists of two parallel metal plates with area \(A\), and the gap between them is \(h\). The gap is small compared to the dimensions of the plates. What is the capacitance?

Since the plates are metal, the charges on each plate are free to move, and will tend to cluster themselves more densely near the edges due to the mutual repulsion of the other charges in the same plate. However, it turns out that if the gap is small, this is a small effect, so we can get away with assuming uniform charge density on each plate. The result of example 17 then applies, and for the region between the plates, we have \(E = 4\pi kq/A\) and \(U_e = (1/8\pi k)E^2Ah\). Substituting the first expression into the second, we find \(U_e = 2\pi kq^2h/A\). Comparing this to the definition of capacitance, we end up with \(C = A/4\pi kh\).

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. 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

\[
U_L = \frac{L}{2} I^2.
\]

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 henry = 1 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!”

There is a lumped circuit approximation for inductors, just like the one for capacitors (p. 601). For a capacitor, this means assuming that the electric fields are completely internal, so that components only interact via currents that flow through wires, not due to the physical overlapping of their fields in space. Similarly for an inductor, the lumped circuit approximation is the assumption that the magnetic fields are completely internal.

**Identical inductances in series** example 22

If two inductors are placed in series, any current that passes through the combined double inductor must pass through both its parts. If we assume the lumped circuit approximation, the two inductors’ fields don’t interfere with each other, so the energy is doubled for a given current. 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. (This is only approximately true, because putting two solenoids end-to-end causes the fields just outside their mouths to overlap and add together in a complicated manner. In other words, the lumped-circuit approximation may not be very good.)

**Identical capacitances in parallel** example 23

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.

This is consistent with the result of example 21, which had the capacitance of a single parallel-plate capacitor 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, subject to any violation of the lumped-circuit approximation due to the interaction of the fields where the edges of the capacitors are joined together.

Inductances in parallel and capacitances in series are explored in homework problems 36 and 33.

Discussion Questions

A Suppose that two parallel-plate capacitors are wired in parallel, and are placed very close together, side by side, so that the lumped circuit approximation is not very accurate. 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.

10.5.2 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,

\[ U_{\text{spring}} = \frac{1}{2} k x^2 \]  
\[ K = \frac{1}{2} m v^2. \]  

In the case of a mechanical oscillator, we have usually assumed a 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 \( dt, \) divided by \( dt, \)

\[ P = \frac{W}{dt} = F \frac{dx}{dt} = Fv = -bv^2, \]  
so

\[ \text{rate of heat dissipation} = -bv^2. \]  

**self-check F**

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?  

Answer, p. 1059

In the circuit, the stored forms of energy are

\[ U_C = \frac{1}{2C} q^2 \]  
\[ U_L = \frac{1}{2} LI^2, \]  

and the rate of heat dissipation in the resistor is

\[ \text{rate of heat dissipation} = -RI^2. \]  

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 G

How is \( v \) related mathematically to \( x \)? How is \( I \) connected to \( q \)? Are the two relationships analogous? \( \triangleright \) Answer, p. 1059

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 frequency
\[ \omega \approx \sqrt{k/m} \]
we can immediately solve the electrical version by analogy, giving

\[ \omega \approx \frac{1}{\sqrt{LC}} \]

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 25

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 26

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.)

---

3 As in chapter 2, we use the word “frequency” to mean either \( f \) or \( \omega = 2\pi f \) when the context makes it clear which is being referred to.
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

\[
L = \frac{1}{C \omega^2}
\]

\[
= \frac{1}{(10^{-6} \text{ F})(2\pi \times 10^3 \text{ s}^{-1})^2}
\]

\[
= 2.5 \times 10^{-3} \text{ F}^{-1} \text{s}^2
\]

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

\[
F^{-1} \text{s}^2 = (C^2/J)^{-1} \text{s}^2
\]

\[
= J \cdot C^{-2} \text{s}^2
\]

\[
= J/A^2
\]

\[
= H
\]

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.

### 10.5.3 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 $q + dq$:

$$V_C = \frac{dU_C}{dq} = \frac{d}{dq} \left( \frac{1}{2C} q^2 \right) = \frac{q}{C}$$

Many books use this as the definition of capacitance. This equation, by the way, probably explains the historical reason why $C$ was de-
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 \( V_{\text{inductor}} + V_{\text{black box}} = 0 \), 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

\[
|IV| = \left| \frac{dU_L}{dt} \right| = \left| \frac{d}{dt} \left( \frac{1}{2} LI^2 \right) \right| = |LI \frac{dI}{dt}|,
\]

or

\[
|V| = \left| L \frac{dI}{dt} \right|,
\]

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 11, 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 and 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 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
Ballasts for fluorescent lights.

Top: a big, heavy inductor used as a ballast in an old-fashioned fluorescent bulb. Bottom: a small solid-state ballast, built into the base of a modern compact fluorescent bulb.

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, \( \frac{dI}{dt} \) is zero.

Example 27

In a gas discharge tube, such as a neon sign, enough voltage is applied to a tube full of gas to ionize some of the atoms in the gas. Once ions have been created, the voltage accelerates them, and they strike other atoms, ionizing them as well and resulting in a chain reaction. This is a spark, like a bolt of lightning. But once the spark starts up, the device begins to act as though it has no resistance: more and more current flows, without the need to apply any more voltage. The power, \( P = IV \), would grow without limit, and the tube would burn itself out.

The simplest solution is to connect an inductor, known as the “ballast,” in series with the tube, and run the whole thing on an AC voltage. During each cycle, as the voltage reaches the point where the chain reaction begins, there is a surge of current, but the inductor resists such a sudden change of current, and the energy that would otherwise have burned out the bulb is instead channeled into building a magnetic field.

A common household fluorescent light bulb consists of a gas discharge tube in which the glass is coated with a fluorescent material. The gas in the tube emits ultraviolet light, which is absorbed by the coating, and the coating then glows in the visible spectrum.

Until recently, it was common for a fluorescent light’s ballast to be a simple inductor, and for the whole device to be operated at the 60 Hz frequency of the electrical power lines. This caused the lights to flicker annoyingly at 120 Hz, and could also cause an audible hum, since the magnetic field surrounding the inductor could...
exert mechanical forces on things. Modern compact fluorescent bulbs have ballasts built into their bases that use a frequency in the kilohertz range, eliminating the flicker and hum.

**Discussion Question**

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

### 10.5.4 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.

We do not even have to reduce $k$ to exactly zero in order to get non-oscillatory behavior. There is actually a finite, critical value below which the behavior changes, so that the mass never even makes it through one cycle. This is the case of overdamping, discussed on page 190.

Electrical circuits can exhibit all the same behavior. For simplicity we will analyze only the cases where either the capacitor or the inductor is completely absent, giving $Q = 0$.

*The RC circuit*

We first analyze the RC circuit, o. 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:

\[
V_R = IR \\
V_C = q/C
\]

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 $dq/dt$, we have

$$\frac{dq}{dt} = -\frac{1}{RC}q.$$

We need a function $q(t)$ whose derivative equals itself, but multiplied by a negative constant. A function of the form $ae^t$, where $e = 2.718...$ is the base of natural logarithms, is the only one that has its derivative equal to itself, and $ae^{bt}$ has its derivative equal to itself multiplied by $b$. Thus our solution is

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

The RL circuit

The RL circuit, $q$, 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 $L/R$.

Death by solenoid; spark plugs example 28

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 \( V_L = dI/dt \): since the time constant is short, the time derivative \( dI/dt \) 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.

A spark-gap radio transmitter example 29

Figure r shows a primitive type of radio transmitter, called a spark gap transmitter, used to send Morse code around the turn of the twentieth century. The high voltage source, \( V \), is typically about 10,000 volts. When the telegraph switch, \( S \), is closed, the RC circuit on the left starts charging up. An increasing voltage difference develops between the electrodes of the spark gap, \( G \). When this voltage difference gets large enough, the electric field in the air between the electrodes causes a spark, partially discharging the RC circuit, but charging the LC circuit on the right. The LC circuit then oscillates at its resonant frequency (typically about 1 MHz), but the energy of these oscillations is rapidly radiated away by the antenna, \( A \), which sends out radio waves (chapter 11).

Discussion Questions

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?
10.5.5 Review of complex numbers

For a more detailed treatment of complex numbers, see ch. 3 of James Nearing’s free book at physics.miami.edu/~nearing.

We assume there is a number, i, such that \(i^2 = -1\). The square roots of \(-1\) are then \(i\) and \(-i\). (In electrical engineering work, where \(i\) stands for current, \(j\) is sometimes used instead.) This gives rise to a number system, called the complex numbers, containing the real numbers as a subset. Any complex number \(z\) can be written in the form \(z = a + bi\), where \(a\) and \(b\) are real, and \(a\) and \(b\) are then referred to as the real and imaginary parts of \(z\). A number with a zero real part is called an imaginary number. The complex numbers can be visualized as a plane, with the real number line placed horizontally like the \(x\) axis of the familiar \(x-y\) plane, and the imaginary numbers running along the \(y\) axis. The complex numbers are complete in a way that the real numbers aren’t: every nonzero complex number has two square roots. For example, 1 is a real number, so it is also a member of the complex numbers, and its square roots are \(-1\) and \(1\). Likewise, \(-1\) has square roots \(i\) and \(-i\), and the number \(i\) has square roots \(1/\sqrt{2} + i/\sqrt{2}\) and \(-1/\sqrt{2} - i/\sqrt{2}\).

Complex numbers can be added and subtracted by adding or subtracting their real and imaginary parts. Geometrically, this is the same as vector addition.

The complex numbers \(a + bi\) and \(a - bi\), lying at equal distances above and below the real axis, are called complex conjugates. The results of the quadratic formula are either both real, or complex conjugates of each other. The complex conjugate of a number \(z\) is notated as \(\bar{z}\) or \(z^*\).

The complex numbers obey all the same rules of arithmetic as the reals, except that they can’t be ordered along a single line. That is, it’s not possible to say whether one complex number is greater than another. We can compare them in terms of their magnitudes (their distances from the origin), but two distinct complex numbers may have the same magnitude, so, for example, we can’t say whether 1 is greater than \(i\) or \(i\) is greater than 1.

\[1\]A square root of \(i\)  
\[\text{example 30}\]

\[\text{Prove that } 1/\sqrt{2} + i/\sqrt{2} \text{ is a square root of } i.\]

\[\text{Our proof can use any ordinary rules of arithmetic, except for ordering.}\]

\[(1/\sqrt{2} + i/\sqrt{2})^2 = 1/\sqrt{2} \cdot 1/\sqrt{2} + 1/\sqrt{2} \cdot i/\sqrt{2} + i/\sqrt{2} \cdot 1/\sqrt{2} + i/\sqrt{2} \cdot i/\sqrt{2}\]

\[= 1/2 (1 + i + i - 1)\]

\[= i\]
A complex number can be described in terms of its magnitude and argument.

Example 30 showed one method of multiplying complex numbers. However, there is another nice interpretation of complex multiplication. We define the argument of a complex number as its angle in the complex plane, measured counterclockwise from the positive real axis. Multiplying two complex numbers then corresponds to multiplying their magnitudes, and adding their arguments.

**self-check H**
Using this interpretation of multiplication, how could you find the square roots of a complex number?

An identity example 31
The magnitude \(|z|\) of a complex number \(z\) obeys the identity \(|z|^2 = z\overline{z}\). To prove this, we first note that \(\bar{z}\) has the same magnitude as \(z\), since flipping it to the other side of the real axis doesn’t change its distance from the origin. Multiplying \(z\) by \(\bar{z}\) gives a result whose magnitude is found by multiplying their magnitudes, so the magnitude of \(z\bar{z}\) must therefore equal \(|z|^2\). Now we just have to prove that \(z\bar{z}\) is a positive real number. But if, for example, \(z\) lies counterclockwise from the real axis, then \(\bar{z}\) lies clockwise from it. If \(z\) has a positive argument, then \(\bar{z}\) has a negative one, or vice-versa. The sum of their arguments is therefore zero, so the result has an argument of zero, and is on the positive real axis.

This whole system was built up in order to make every number have square roots. What about cube roots, fourth roots, and so on? Does it get even more weird when you want to do those as well? No. The complex number system we’ve already discussed is sufficient to handle all of them. The nicest way of thinking about it is in terms of roots of polynomials. In the real number system, the polynomial \(x^2 - 1\) has two roots, i.e., two values of \(x\) (plus and minus one) that we can plug in to the polynomial and get zero. Because it has these two real roots, we can rewrite the polynomial as \((x - 1)(x + 1)\). However, the polynomial \(x^2 + 1\) has no real roots. It’s ugly that in the real number system, some second-order polynomials have two roots, and can be factored, while others can’t. In the complex number system, they all can. For instance, \(x^2 + 1\) has roots \(i\) and \(-i\), and can be factored as \((x - i)(x + i)\). In general, the fundamental theorem of algebra states that in the complex number system, any nth-order polynomial can be factored completely into \(n\) linear factors, and we can also say that it has \(n\) complex roots, with the understanding that some of the roots may be the same. For instance, the fourth-order polynomial \(x^4 + 1\) can be factored as \((x - i)(x + i)(x - 0)(x - 0)\), and we say that it has four roots, \(i\), \(-i\), 0, and 0, two of which happen to be the same. This is a sensible way to think about it, because

---

\(^4\)I cheated a little. If \(z\)’s argument is 30 degrees, then we could say \(\bar{z}\)’s was -30, but we could also call it 330. That’s OK, because 330+30 gives 360, and an argument of 360 is the same as an argument of zero.
in real life, numbers are always approximations anyway, and if we make tiny, random changes to the coefficients of this polynomial, it will have four distinct roots, of which two just happen to be very close to zero.

**Discussion Questions**

**A** Find \( \arg i \), \( \arg (-i) \), and \( \arg 37 \), where \( \arg z \) denotes the argument of the complex number \( z \).

**B** Visualize the following multiplications in the complex plane using the interpretation of multiplication in terms of multiplying magnitudes and adding arguments: \((i)(i) = -1, (i)(-i) = 1, (-i)(-i) = -1\).

**C** If we visualize \( z \) as a point in the complex plane, how should we visualize \(-z\)? What does this mean in terms of arguments? Give similar interpretations for \( z^2 \) and \( \sqrt{z} \).

**D** Find four different complex numbers \( z \) such that \( z^4 = 1 \).

**E** Compute the following. For the final two, use the magnitude and argument, not the real and imaginary parts.

\[
|1 + i|, \quad \arg(1 + i), \quad \left| \frac{1}{1 + i} \right|, \quad \arg \left( \frac{1}{1 + i} \right).
\]

From these, find the real and imaginary parts of \( 1/(1 + i) \).

### 10.5.6 Euler’s formula

Having expanded our horizons to include the complex numbers, it’s natural to want to extend functions we knew and loved from the world of real numbers so that they can also operate on complex numbers. The only really natural way to do this in general is to use Taylor series. A particularly beautiful thing happens with the functions \( e^x \), \( \sin x \), and \( \cos x \):

\[
e^x = 1 + \frac{1}{2!} x^2 + \frac{1}{3!} x^3 + \ldots
\]

\[
\cos x = 1 - \frac{1}{2!} x^2 + \frac{1}{4!} x^4 - \ldots
\]

\[
\sin x = x - \frac{1}{3!} x^3 + \frac{1}{5!} x^5 - \ldots
\]

If \( x = i\phi \) is an imaginary number, we have

\[
e^{i\phi} = \cos \phi + i \sin \phi,
\]

a result known as Euler’s formula. The geometrical interpretation in the complex plane is shown in figure x.

Although the result may seem like something out of a freak show at first, applying the definition of the exponential function makes it clear how natural it is:

\[
e^x = \lim_{n \to \infty} \left( 1 + \frac{x}{n} \right)^n.
\]
When \( x = i\phi \) is imaginary, the quantity \( (1 + i\phi/n) \) represents a number lying just above 1 in the complex plane. For large \( n \), \( (1 + i\phi/n) \) becomes very close to the unit circle, and its argument is the small angle \( \phi/n \). Raising this number to the \( n \)th power multiplies its argument by \( n \), giving a number with an argument of \( \phi \).

Euler’s formula is used frequently in physics and engineering.

*Trig functions in terms of complex exponentials*  
**example 32**  
> Write the sine and cosine functions in terms of exponentials.

> Euler’s formula for \( x = -i\phi \) gives \( \cos \phi - i \sin \phi \), since \( \cos(-\theta) = \cos \theta \), and \( \sin(-\theta) = -\sin \theta \).

\[
\begin{align*}
\cos x &= \frac{e^{ix} + e^{-ix}}{2} \\
\sin x &= \frac{e^{ix} - e^{-ix}}{2i}
\end{align*}
\]

*A hard integral made easy*  
**example 33**  
> Evaluate \( \int e^{x} \cos x \, dx \)

> This seemingly impossible integral becomes easy if we rewrite the cosine in terms of exponentials:

\[
\begin{align*}
\int e^{x} \cos x \, dx &= \int e^{x} \left( \frac{e^{ix} + e^{-ix}}{2} \right) \, dx \\
&= \frac{1}{2} \int (e^{(1+i)x} + e^{(1-i)x}) \, dx \\
&= \frac{1}{2} \left( \frac{e^{(1+i)x}}{1+i} + \frac{e^{(1-i)x}}{1-i} \right) + c
\end{align*}
\]

Since this result is the integral of a real-valued function, we’d like it to be real, and in fact it is, since the first and second terms are complex conjugates of one another. If we wanted to, we could use Euler’s theorem to convert it back to a manifestly real result.\(^5\)

10.5.7 **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

\(^5\)In general, the use of complex number techniques to do an integral could result in a complex number, but that complex number would be a constant, which could be subsumed within the usual constant of integration.
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

\[ I = \frac{dq}{dt} = C\frac{dV}{dt}. \]

This is partly analogous to Ohm’s law. For example, if we double the amplitude of a sinusoidally varying AC voltage, the derivative \( dV/dt \) will also double, and the amplitude of the sinusoidally varying current will also double. However, it is not true that \( I = V/R \), because taking the derivative of a sinusoidal function shifts its phase by 90 degrees. If the voltage varies as, for example, \( V(t) = V_o \sin(\omega t) \), then the current will be \( I(t) = \omega CV_o \cos(\omega t) \). The amplitude of the current is \( \omega CV_o \), which is proportional to \( V_o \), but it’s not true that \( I(t) = V(t)/R \) for some constant \( R \).

A second problem that crops up is that our entire analysis of DC resistive circuits was built on the foundation of the loop rule and the junction rule, both of which are statements about sums. To apply the junction rule to an AC circuit, for example, we would say that the sum of the sine waves describing the currents coming into the junction is equal (at every moment in time) to the sum of the sine waves going out. Now sinusoidal functions have a remarkable property, which is that if you add two different sinusoidal functions having the same frequency, the result is also a sinusoid with that frequency. For example, \( \cos \omega t + \sin \omega t = \sqrt{2} \sin(\omega t + \pi/4) \), which can be proved using trig identities. The trig identities can get very cumbersome, however, and there is a much easier technique involving complex numbers.

Figure aa shows a useful way to visualize what’s going on. When a circuit is oscillating at a frequency \( \omega \), we use points in the plane to represent sinusoidal functions with various phases and amplitudes.

**Self-check l**

Which of the following functions can be represented in this way? \( \cos(6t - 4) \), \( \cos^2 t \), \( \tan t \)

> Answer, p. 1060
The simplest examples of how to visualize this in polar coordinates are ones like \( \cos \omega t + \cos \omega t = 2 \cos \omega t \), where everything has the same phase, so all the points lie along a single line in the polar plot, and addition is just like adding numbers on the number line. The less trivial example \( \cos \omega t + \sin \omega t = \sqrt{2} \sin(\omega t + \pi/4) \), can be visualized as in figure ab.

Figure ab suggests that all of this can be tied together nicely if we identify our plane with the plane of complex numbers. For example, the complex numbers 1 and \( i \) represent the functions \( \sin \omega t \) and \( \cos \omega t \). In figure z, for example, the voltage across the capacitor is a sine wave multiplied by a number that gives its amplitude, so we associate that function with a number \( V \) lying on the real axis. Its magnitude, \( |V| \), gives the amplitude in units of volts, while its argument \( \arg V \), gives its phase angle, which is zero. The current is a multiple of the cosine, so we identify it with a number \( I \) lying on the imaginary axis. We have \( \arg I = 90^\circ \), and \( |I| \) is the amplitude of the current, in units of amperes. But comparing with our result above, we have \( |I| = \omega C |V| \). Bringing together the phase and magnitude information, we have \( I = i \omega CV \). This looks very much like Ohm’s law, so we write

\[
\tilde{I} = \frac{\tilde{V}}{Z_C},
\]

where the quantity

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

having units of ohms, is called the impedance of the capacitor at this frequency.

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.

Note that we have two types of complex numbers: those that represent sinusoidal functions of time, and those that represent impedances. The ones that represent sinusoidal functions have tildes on top, which look like little sine waves.

**self-check J**

Why can’t a capacitor have its impedance printed on it along with its capacitance?  

\( \triangleright \) Answer, p. 1060
Similar math (but this time with an integral instead of a derivative) gives

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

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 z and ac 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 (see p. 713).

Summarizing, the impedances of resistors, capacitors, and inductors are

\[
\begin{align*}
Z_R &= R \\
Z_C &= -\frac{i}{\omega C} \\
Z_L &= i\omega L.
\end{align*}
\]

\[\text{Low-pass and high-pass filters example 34}\]

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.

**Discussion Question**

A Figure z on page 629 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 Figure ac on page 630 shows the voltage and current for an inductor. The power is considered to be positive when energy is being put into the inductor’s magnetic field. Sketch the graph of the power, and then the graph of \( U \), the energy stored in the magnetic field, and use it to give a physical explanation of the \( P-t \) graph. In particular, discuss why the frequency is doubled on the \( P-t \) graph.

C Relate the features of the graph in figure ac on page 630 to the story told in cartoons in figure m/2-3 on page 620.

**10.5.8 Power**

How much power is delivered when an oscillating voltage is applied to an impedance? The equation \( P = IV \) is generally true, since voltage is defined as energy per unit charge, and current is defined as charge per unit time: multiplying them gives energy per unit time. In a DC circuit, all three quantities were constant, but in an oscillating (AC) circuit, all three display time variation.

A resistor

First let’s examine the case of a resistor. For instance, you’re probably reading this book from a piece of paper illuminated by a glowing lightbulb, which is driven by an oscillating voltage with amplitude \( V_0 \). In the special case of a resistor, we know that \( I \) and \( V \) are in phase. For example, if \( V \) varies as \( V_0 \cos \omega t \), then \( I \) will be a cosine as well, \( I_0 \cos \omega t \). The power is then \( I_0 V_0 \cos^2 \omega t \), which is always positive, and varies between 0 and \( I_0 V_0 \). Even if the time variation was \( \cos \omega t \) or \( \sin(\omega t + \pi/4) \), we would still have a maximum power of \( I_0 V_0 \), because both the voltage and the current would reach their maxima at the same time. In a lightbulb, the moment of maximum power is when the circuit is most rapidly heating the filament. At the instant when \( P = 0 \), a quarter of a cycle later, no current is flowing, and no electrical energy is being turned into heat. Throughout the whole cycle, the filament is getting rid of energy by

\[ \text{A resistor always turns electrical energy into heat. It never turns heat into electrical energy!} \]
radiating light.\(^7\) Since the circuit oscillates at a frequency\(^8\) of 60 Hz, the temperature doesn’t really have time to cycle up or down very much over the 1/60 s period of the oscillation, and we don’t notice any significant variation in the brightness of the light, even with a short-exposure photograph.

Thus, what we really want to know is the average power, “average” meaning the average over one full cycle. Since we’re covering a whole cycle with our average, it doesn’t matter what phase we assume. Let’s use a cosine. The total amount of energy transferred over one cycle is

\[
E = \int \frac{dE}{dt} dt,
\]

where \(T = 2\pi/\omega\) is the period.

\[
E = \int_0^T P dt
= \int_0^T I_o V_o \cos^2 \omega t dt
= I_o V_o \int_0^T \cos^2 \omega t dt
= I_o V_o \int_0^T \frac{1}{2} (1 + \cos 2\omega t) dt
\]

The reason for using the trig identity \(\cos^2 x = (1 + \cos 2x)/2\) in the last step is that it lets us get the answer without doing a hard integral. Over the course of one full cycle, the quantity \(\cos 2\omega t\) goes positive, negative, positive, and negative again, so the integral of it is zero. We then have

\[
E = I_o V_o \int_0^T \frac{1}{2} dt
= \frac{I_o V_o T}{2}
\]

\(^7\)To many people, the word “radiation” implies nuclear contamination. Actually, the word simply means something that “radiates” outward. Natural sunlight is “radiation.” So is the light from a lightbulb, or the infrared light being emitted by your skin right now.

\(^8\)Note that this time “frequency” means \(f\), not \(\omega\)! Physicists and engineers generally use \(\omega\) because it simplifies the equations, but electricians and technicians always use \(f\). The 60 Hz frequency is for the U.S.
The average power is

\[ P_{av} = \frac{\text{energy transferred in one full cycle}}{\text{time for one full cycle}} = \frac{I_o V_o T/2}{T} = \frac{I_o V_o}{2}, \]

i.e., the average is half the maximum. The power varies from 0 to \( I_o V_o \), and it spends equal amounts of time above and below the maximum, so it isn’t surprising that the average power is half-way in between zero and the maximum. Summarizing, we have

\[ P_{av} = \frac{I_o V_o}{2} \quad \text{[average power in a resistor]} \]

for a resistor.

RMS quantities

Suppose one day the electric company decided to start supplying your electricity as DC rather than AC. How would the DC voltage have to be related to the amplitude \( V_o \) of the AC voltage previously used if they wanted your lightbulbs to have the same brightness as before? The resistance of the bulb, \( R \), is a fixed value, so we need to relate the power to the voltage and the resistance, eliminating the current. In the DC case, this gives \( P = IV = (V/R)V = V^2/R \).

(For DC, \( P \) and \( P_{av} \) are the same.) In the AC case, \( P_{av} = I_o V_o/2 = V_o^2/2R \). Since there is no factor of 1/2 in the DC case, the same power could be provided with a DC voltage that was smaller by a factor of \( 1/\sqrt{2} \). Although you will hear people say that household voltage in the U.S. is 110 V, its amplitude is actually \((110 \text{ V}) \times \sqrt{2} \approx 160 \text{ V}\). The reason for referring to \( V_o/\sqrt{2} \) as “the” voltage is that people who are naive about AC circuits can plug \( V_o/\sqrt{2} \) into a familiar DC equation like \( P = V^2/R \) and get the right average answer. The quantity \( V_o/\sqrt{2} \) is called the “RMS” voltage, which stands for “root mean square.” The idea is that if you square the function \( V(t) \), take its average (mean) over one cycle, and then take the square root of that average, you get \( V_o/\sqrt{2} \). Many digital meters provide RMS readouts for measuring AC voltages and currents.

A capacitor

For a capacitor, the calculation starts out the same, but ends up with a twist. If the voltage varies as a cosine, \( V_o \cos \omega t \), then the relation \( I = C \frac{dV}{dt} \) tells us that the current will be some constant multiplied by minus the sine, \(-V_o \sin \omega t \). The integral we did in the case of a resistor now becomes

\[ E = \int_0^T -I_o V_o \sin \omega t \cos \omega t \, dt, \]
and based on figure ae, you can easily convince yourself that over the course of one full cycle, the power spends two quarter-cycles being negative and two being positive. In other words, the average power is zero!

Why is this? It makes sense if you think in terms of energy. A resistor converts electrical energy to heat, never the other way around. A capacitor, however, merely stores electrical energy in an electric field and then gives it back. For a capacitor,

\[
P_{av} = 0 \quad \text{[average power in a capacitor]}
\]

Notice that although the average power is zero, the power at any given instant is not typically zero, as shown in figure ae. The capacitor does transfer energy: it’s just that after borrowing some energy, it always pays it back in the next quarter-cycle.

**An inductor**

The analysis for an inductor is similar to that for a capacitor: the power averaged over one cycle is zero. Again, we’re merely storing energy temporarily in a field (this time a magnetic field) and getting it back later.

### 10.5.9 Impedance matching

Figure af shows a commonly encountered situation: we wish to maximize the average power, \(P_{av}\), delivered to the load for a fixed value of \(V_o\), the amplitude of the oscillating driving voltage. We assume that the impedance of the transmission line, \(Z_T\) is a fixed value, over which we have no control, but we are able to design the load, \(Z_o\), with any impedance we like. For now, we’ll also assume that both impedances are resistive. For example, \(Z_T\) could be the resistance of a long extension cord, and \(Z_o\) could be a lamp at the end of it. The result generalizes immediately, however, to any kind of impedance. For example, the load could be a stereo speaker’s magnet coil, which is displays both inductance and resistance. (For a purely inductive or capacitive load, \(P_{av}\) equals zero, so the problem isn’t very interesting!)

Since we’re assuming both the load and the transmission line are resistive, their impedances add in series, and the amplitude of the current is given by

\[
I_o = \frac{V_o}{Z_o + Z_T},
\]
\[ P_{av} = \frac{I_o V_o}{2} \]
\[ = \frac{I_o^2 Z_o}{2} \]
\[ = \frac{V_o^2 Z_o}{(Z_o + Z_T)^2}/2. \]

The maximum of this expression occurs where the derivative is zero,

\[ 0 = \frac{1}{2} \frac{d}{dZ_o} \left[ \frac{V_o^2 Z_o}{(Z_o + Z_T)^2} \right] \]
\[ 0 = \frac{1}{2} \frac{d}{dZ_o} \left[ \frac{Z_o}{(Z_o + Z_T)^2} \right] \]
\[ 0 = (Z_o + Z_T)^{-2} - 2Z_o (Z_o + Z_T)^{-3} \]
\[ 0 = (Z_o + Z_T) - 2Z_o \]
\[ Z_o = Z_T \]

In other words, to maximize the power delivered to the load, we should make the load’s impedance the same as the transmission line’s. This result may seem surprising at first, but it makes sense if you think about it. If the load’s impedance is too high, it’s like opening a switch and breaking the circuit; no power is delivered. On the other hand, it doesn’t pay to make the load’s impedance too small. Making it smaller does give more current, but no matter how small we make it, the current will still be limited by the transmission line’s impedance. As the load’s impedance approaches zero, the current approaches this fixed value, and the the power delivered, \( I_o^2 Z_o \), decreases in proportion to \( Z_o \).

Maximizing the power transmission by matching \( Z_T \) to \( Z_o \) is called *impedance matching*. For example, an 8-ohm home stereo speaker will be correctly matched to a home stereo amplifier with an internal impedance of 8 ohms, and 4-ohm car speakers will be correctly matched to a car stereo with a 4-ohm internal impedance. You might think impedance matching would be unimportant because even if, for example, we used a car stereo to drive 8-ohm speakers, we could compensate for the mismatch simply by turning the volume knob higher. This is indeed one way to compensate for any impedance mismatch, but there is always a price to pay. When the impedances are matched, half the power is dissipated in the transmission line and half in the load. By connecting a 4-ohm amplifier to an 8-ohm speaker, however, you would be setting up a situation in two watts were being dissipated as heat inside the amp for every amp being delivered to the speaker. In other words, you would be wasting energy, and perhaps burning out your amp when you turned up the volume to compensate for the mismatch.


10.5.10 **Impedances in series and parallel**

How do impedances combine in series and parallel? The beauty of treating them as complex numbers is that they simply combine according to the same rules you’ve already learned as resistances.

**Series impedance**

A capacitor and an inductor in series with each other are driven by a sinusoidally oscillating voltage. At what frequency is the current maximized?

Impedances in series, like resistances in series, add. The capacitor and inductor act as if they were a single circuit element with an impedance

$$Z = Z_L + Z_C = i\omega L - \frac{i}{\omega C}.$$  

The current is then

$$\tilde{I} = \frac{\tilde{V}}{i\omega L - i/\omega C}.$$  

We don’t care about the phase of the current, only its amplitude, which is represented by the absolute value of the complex number $\tilde{I}$, and this can be maximized by making $|i\omega L - i/\omega C|$ as small as possible. But there is some frequency at which this quantity is zero —

$$0 = i\omega L - \frac{i}{\omega C}$$

$$\frac{1}{\omega C} = \omega L$$

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

At this frequency, the current is infinite! What is going on physically? This is an LRC circuit with $R = 0$. It has a resonance at this frequency, and because there is no damping, the response at resonance is infinite. Of course, any real LRC circuit will have some damping, however small (cf. figure j on page 185).

**Resonance with damping**

What is the amplitude of the current in a series LRC circuit?

Generalizing from example 35, we add a third, real impedance:

$$|\tilde{I}| = \frac{|\tilde{V}|}{|Z|} = \frac{|\tilde{V}|}{|R + i\omega L - i/\omega C|} = \frac{|\tilde{V}|}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}}$$

This result would have taken pages of algebra without the complex number technique!
A second-order stereo crossover filter example 37

A stereo crossover filter ensures that the high frequencies go to the tweeter and the lows to the woofer. This can be accomplished simply by putting a single capacitor in series with the tweeter and a single inductor in series with the woofer. However, such a filter does not cut off very sharply. Suppose we model the speakers as resistors. (They really have inductance as well, since they have coils in them that serve as electromagnets to move the diaphragm that makes the sound.) Then the power they draw is \( \mathcal{P} \). Putting an inductor in series with the woofer, \( \mathcal{L} \), gives a total impedance that at high frequencies is dominated by the inductor's, so the current is proportional to \( \omega^{-1} \), and the power drawn by the woofer is proportional to \( \omega^{-2} \).

A second-order filter, like \( \mathcal{L} \), is one that cuts off more sharply: at high frequencies, the power goes like \( \omega^{-4} \). To analyze this circuit, we first calculate the total impedance:

\[
Z = Z_L + \left( \frac{1}{Z_C} + \frac{1}{Z_R} \right)^{-1}
\]

All the current passes through the inductor, so if the driving voltage being supplied on the left is \( \tilde{V}_d \), we have

\[
\tilde{V}_d = \tilde{I}_L Z
\]

and we also have

\[
\tilde{V}_L = \tilde{I}_L Z_L.
\]

The loop rule, applied to the outer perimeter of the circuit, gives

\[
\tilde{V}_d = \tilde{V}_L + \tilde{V}_R.
\]

Straightforward algebra now results in

\[
\tilde{V}_R = \frac{\tilde{V}_d}{1 + Z_L/Z_C + Z_L/Z_R}.
\]

At high frequencies, the \( Z_L/Z_C \) term, which varies as \( \omega^2 \), dominates, so \( \tilde{V}_R \) and \( \tilde{I}_R \) are proportional to \( \omega^{-2} \), and the power is proportional to \( \omega^{-4} \).
10.6 Fields by Gauss’ law

10.6.1 Gauss’ law

The flea of subsection 10.3.2 had a long and illustrious scientific career, and we’re now going to pick up her story where we left off. This flea, whose name is Gauss\(^9\), has derived the equation \(E_{\perp} = 2\pi k\sigma\) for the electric field very close to a charged surface with charge density \(\sigma\). Next we will describe two improvements she is going to make to that equation.

First, she realizes that the equation is not as useful as it could be, because it only gives the part of the field due to the surface. If other charges are nearby, then their fields will add to this field as vectors, and the equation will not be true unless we carefully subtract out the field from the other charges. This is especially problematic for her because the planet on which she lives, known for obscure reasons as planet Flatcat, is itself electrically charged, and so are all the fleas — the only thing that keeps them from floating off into outer space is that they are negatively charged, while Flatcat carries a positive charge, so they are electrically attracted to it. When Gauss found the original version of her equation, she wanted to demonstrate it to her skeptical colleagues in the laboratory, using electric field meters and charged pieces of metal foil. Even if she set up the measurements by remote control, so that her the charge on her own body would be too far away to have any effect, they would be disrupted by the ambient field of planet Flatcat. Finally, however, she realized that she could improve her equation by rewriting it as follows:

\[
E_{\text{outward, on side 1}} + E_{\text{outward, on side 2}} = 4\pi k\sigma.
\]

The tricky thing here is that “outward” means a different thing, depending on which side of the foil we’re on. On the left side, “outward” means to the left, while on the right side, “outward” is right. A positively charged piece of metal foil has a field that points leftward on the left side, and rightward on its right side, so the two contributions of \(2\pi k\sigma\) are both positive, and we get \(4\pi k\sigma\). On the other hand, suppose there is a field created by other charges, not by the charged foil, that happens to point to the right. On the right side, this externally created field is in the same direction as the foil’s field, but on the left side, the it reduces the strength of the leftward field created by the foil. The increase in one term of the equation balances the decrease in the other term. This new version of the equation is thus exactly correct regardless of what externally generated fields are present!

Her next innovation starts by multiplying the equation on both sides by the area, \(A\), of one side of the foil:

\[
(E_{\text{outward, on side 1}} + E_{\text{outward, on side 2}}) A = 4\pi k\sigma A.
\]

\(^9\)no relation to the human mathematician of the same name
The area vector is defined to be perpendicular to the surface, in the outward direction. Its magnitude tells how much the area is.

\[ E_{\text{outward, on side 1}} A_1 + E_{\text{outward, on side 2}} A_2 = 4\pi kq, \]

where \( q \) is the charge of the foil. The reason for this modification is that she can now make the whole thing more attractive by defining a new vector, the area vector \( A \). As shown in figure a, she defines an area vector for side 1 which has magnitude \( A \) and points outward from side 1, and an area vector for side 2 which has the same magnitude and points outward from that side, which is in the opposite direction. The dot product of two vectors, \( \mathbf{u} \cdot \mathbf{v} \), can be interpreted as \( u_{\parallel} \frac{1}{v} |v| \), and she can therefore rewrite her equation as

\[ E_1 \cdot A_1 + E_2 \cdot A_2 = 4\pi kq. \]

The quantity on the left side of this equation is called the flux through the surface, written \( \Phi \).

Gauss now writes a grant proposal to her favorite funding agency, the BSGS (Blood-Suckers’ Geological Survey), and it is quickly approved. Her audacious plan is to send out exploring teams to chart the electric fields of the whole planet of Flatcat, and thereby determine the total electric charge of the planet. The fleas’ world is commonly assumed to be a flat disk, and its size is known to be finite, since the sun passes behind it at sunset and comes back around on the other side at dawn. The most daring part of the plan is that it requires surveying not just the known side of the planet but the uncharted Far Side as well. No flea has ever actually gone around the edge and returned to tell the tale, but Gauss assures them that they won’t fall off — their negatively charged bodies will be attracted to the disk no matter which side they are on.

Of course it is possible that the electric charge of planet Flatcat is not perfectly uniform, but that isn’t a problem. As discussed in subsection 10.3.2, as long as one is very close to the surface, the field only depends on the local charge density. In fact, a side-benefit of Gauss’s program of exploration is that any such local irregularities will be mapped out. But what the newspapers find exciting is the idea that once all the teams get back from their voyages and tabulate their data, the total charge of the planet will have been determined for the first time. Each surveying team is assigned to visit a certain list of republics, duchies, city-states, and so on. They are to record each territory’s electric field vector, as well as its area. Because the electric field may be nonuniform, the final equation for determining the planet’s electric charge will have many terms, not just one for each side of the planet:

\[ \Phi = \sum E_j \cdot A_j = 4\pi kq_{\text{total}} \]
Gauss herself leads one of the expeditions, which heads due east, toward the distant Tail Kingdom, known only from fables and the occasional account from a caravan of traders. A strange thing happens, however. Gauss embarks from her college town in the wetlands of the Tongue Republic, travels straight east, passes right through the Tail Kingdom, and one day finds herself right back at home, all without ever seeing the edge of the world! What can have happened? All at once she realizes that the world isn’t flat.

Now what? The surveying teams all return, the data are tabulated, and the result for the total charge of Flatcat is \( \frac{1}{4\pi k} \sum E_j \cdot A_j = 37 \text{ nC} \). But the equation was derived under the assumption that Flatcat was a disk. If Flatcat is really round, then the result may be completely wrong. Gauss and two of her grad students go to their favorite bar, and decide to keep on ordering Bloody Marys until they either solve their problems or forget them. One student suggests that perhaps Flatcat really is a disk, but the edges are rounded. Maybe the surveying teams really did flip over the edge at some point, but just didn’t realize it. Under this assumption, the original equation will be approximately valid, and 37 nC really is the total charge of Flatcat.

A second student, named Newton, suggests that they take seriously the possibility that Flatcat is a sphere. In this scenario, their planet’s surface is really curved, but the surveying teams just didn’t notice the curvature, since they were close to the surface, and the surface was so big compared to them. They divided up the surface into a patchwork, and each patch was fairly small compared to the whole planet, so each patch was nearly flat. Since the patch is nearly flat, it makes sense to define an area vector that is perpendicular to it. In general, this is how we define the direction of an area vector, as shown in figure d. This only works if the areas are small. For instance, there would be no way to define an area vector for an entire sphere, since “outward” is in more than one direction.

If Flatcat is a sphere, then the inside of the sphere must be vast, and there is no way of knowing exactly how the charge is arranged below the surface. However, the survey teams all found that the electric field was approximately perpendicular to the surface everywhere, and that its strength didn’t change very much from one location to another. The simplest explanation is that the charge is all concentrated in one small lump at the center of the sphere. They have no way of knowing if this is really the case, but it’s a hypothesis that allows them to see how much their 37 nC result would change if they assumed a different geometry. Making this assumption, Newton performs the following simple computation on a napkin. The field at the surface is related to the charge at the center by
\[ |E| = \frac{kq_{\text{total}}}{r^2}, \]

where \( r \) is the radius of Flatcat. The flux is then

\[ \Phi = \sum E_j \cdot A_j, \]

and since the \( E_j \) and \( A_j \) vectors are parallel, the dot product equals \( |E_j||A_j| \), so

\[ \Phi = \sum \frac{kq_{\text{total}}}{r^2} |A_j|. \]

But the field strength is always the same, so we can take it outside the sum, giving

\[
\begin{align*}
\Phi &= \frac{kq_{\text{total}}}{r^2} \sum |A_j| \\
&= \frac{kq_{\text{total}}}{r^2} A_{\text{total}} \\
&= \frac{kq_{\text{total}}}{r^2} 4\pi r^2 \\
&= 4\pi kq_{\text{total}}.
\end{align*}
\]

Not only have all the factors of \( r \) canceled out, but the result is the same as for a disk!

Everyone is pleasantly surprised by this apparent mathematical coincidence, but is it anything more than that? For instance, what if the charge wasn’t concentrated at the center, but instead was evenly distributed throughout Flatcat’s interior volume? Newton, however, is familiar with a result called the shell theorem (page 102), which states that the field of a uniformly charged sphere is the same as if all the charge had been concentrated at its center.\(^\text{10}\) We now have three different assumptions about the shape of Flatcat and the arrangement of the charges inside it, and all three lead to exactly the same mathematical result, \( \Phi = 4\pi kq_{\text{total}} \). This is starting to look like more than a coincidence. In fact, there is a general mathematical theorem, called Gauss’ theorem, which states the following:

For any region of space, the flux through the surface equals \( 4\pi kq_{\text{in}} \), where \( q_{\text{in}} \) is the total charge in that region.

Don’t memorize the factor of \( 4\pi \) in front — you can rederive it any time you need to, by considering a spherical surface centered on a point charge.

\(^{10}\)Newton’s human namesake actually proved this for gravity, not electricity, but they’re both \( 1/r^2 \) forces, so the proof works equally well in both cases.
Note that although region and its surface had a definite physical existence in our story — they are the planet Flatcat and the surface of planet Flatcat — Gauss’ law is true for any region and surface we choose, and in general, the Gaussian surface has no direct physical significance. It’s simply a computational tool.

Rather than proving Gauss’ theorem and then presenting some examples and applications, it turns out to be easier to show some examples that demonstrate its salient properties. Having understood these properties, the proof becomes quite simple.

self-check K

Suppose we have a negative point charge, whose field points inward, and we pick a Gaussian surface which is a sphere centered on that charge. How does Gauss’ theorem apply here?

Answer, p. 1060

10.6.2 Additivity of flux

Figure e shows two different ways in which flux is additive. Figure e/1, additivity by charge, shows that we can break down a charge distribution into two or more parts, and the flux equals the sum of the fluxes due to the individual charges. This follows directly from the fact that the flux is defined in terms of a dot product, \( \mathbf{E} \cdot \mathbf{A} \), and the dot product has the additive property \((a+b) \cdot c = a \cdot c + b \cdot c\).

To understand additivity of flux by region, e/2, we have to consider the parts of the two surfaces that were eliminated when they were joined together, like knocking out a wall to make two small apartments into one big one. Although the two regions shared this wall before it was removed, the area vectors were opposite: the direction that is outward from one region is inward with respect to the other. Thus if the field on the wall contributes positive flux to one region, it contributes an equal amount of negative flux to the other region, and we can therefore eliminate the wall to join the two regions, without changing the total flux.

10.6.3 Zero flux from outside charges

A third important property of Gauss’ theorem is that it only refers to the charge inside the region we choose to discuss. In other words, it asserts that any charge outside the region contributes zero to the flux. This makes at least some sense, because a charge outside the region will have field vectors pointing into the surface on one side, and out of the surface on the other. Certainly there should be at least partial cancellation between the negative (inward) flux on one side and the positive (outward) flux on the other. But why should this cancellation be exact?

To see the reason for this perfect cancellation, we can imagine space as being built out of tiny cubes, and we can think of any charge distribution as being composed of point charges. The additivity-by-charge property tells us that any charge distribution can be handled.
by considering its point charges individually, and the additivity-by-region property tells us that if we have a single point charge outside a big region, we can break the region down into tiny cubes. If we can prove that the flux through such a tiny cube really does cancel exactly, then the same must be true for any region, which we could build out of such cubes, and any charge distribution, which we can build out of point charges.

For simplicity, we will carry out this calculation only in the special case shown in figure f, where the charge lies along one axis of the cube. Let the sides of the cube have length $2b$, so that the area of each side is $(2b)^2 = 4b^2$. The cube extends a distance $b$ above, below, in front of, and behind the horizontal $x$ axis. There is a distance $d - b$ from the charge to the left side, and $d + b$ to the right side.

There will be one negative flux, through the left side, and five positive ones. Of these positive ones, the one through the right side is very nearly the same in magnitude as the negative flux through the left side, but just a little less because the field is weaker on the right, due to the greater distance from the charge. The fluxes through the other four sides are very small, since the field is nearly perpendicular to their area vectors, and the dot product $E_j \cdot A_j$ is zero if the two vectors are perpendicular. In the limit where $b$ is very small, we can approximate the flux by evaluating the field at the center of each of the cube’s six sides, giving

$$
\Phi = \Phi_{\text{left}} + 4\Phi_{\text{side}} + \Phi_{\text{right}}
= |E_{\text{left}}||A_{\text{left}}| \cos 180^\circ + 4|E_{\text{side}}||A_{\text{side}}| \cos \theta_{\text{side}}
+ |E_{\text{right}}||A_{\text{right}}| \cos 0^\circ,
$$

and a little trig gives $\cos \theta_{\text{side}} \approx b/d$, so

$$
\Phi = -|E_{\text{left}}||A_{\text{left}}| + 4|E_{\text{side}}||A_{\text{side}}| \frac{b}{d} + |E_{\text{right}}||A_{\text{right}}|
= (4b^2) \left( -|E_{\text{left}}| + 4|E_{\text{side}}| \frac{b}{d} + |E_{\text{right}}| \right)
= (4b^2) \left( -\frac{kq}{(d-b)^2} + 4\frac{kq}{d^2} \frac{b}{d} + \frac{kq}{(d+b)^2} \right)
= \left( \frac{4kqb^2}{d^2} \right) \left( -\frac{1}{(1-b/d)^2} + \frac{4b}{d} + \frac{1}{(1+b/d)^2} \right).
$$

Using the approximation $(1+\epsilon)^{-2} \approx 1 - 2\epsilon$ for small $\epsilon$, this becomes

$$
\Phi = \left( \frac{4kqb^2}{d^2} \right) \left( -1 - \frac{2b}{d} + \frac{4b}{d} + 1 - \frac{2b}{d} \right)
= 0.
$$

Thus in the limit of a very small cube, $b \ll d$, we have proved that the flux due to this exterior charge is zero. The proof can be
extended to the case where the charge is not along any axis of the
cube,\textsuperscript{11} and based on additivity we then have a proof that the flux
due to an outside charge is always zero.

\textit{No charge on the interior of a conductor example 38}

I asserted on p. 543 that for a perfect conductor in equilibrium, ex-
cess charge is found only at the surface, never in the interior. This
can be proved using Gauss’s theorem. Suppose that a charge \( q \)
existed at some point in the interior, and it was in stable equilib-
rium. For concreteness, let’s say \( q \) is positive. If its equilibrium
is to be stable, then we need an electric field everywhere around
it that points inward like a pincushion, so that if the charge were
to be perturbed slightly, the field would bring it back to its equi-
librium position. Since Newton’s third law forbids objects from
making forces on themselves, this field would have to be the field
contributed by all the other charges, not by \( q \) itself. But this is im-
possible, because this kind of inward-pointing pincushion pattern
would have a nonzero (negative) flux through the pincushion, but
Gauss’s theorem says we can’t have flux from outside charges.

\textsuperscript{11}The math gets messy for the off-axis case. This part of the proof can be
completed more easily and transparently using the techniques of section 10.7,
and that is exactly we’ll do in example 40 on page 653.
Discussion Questions

A One question that might naturally occur to you about Gauss’s law is what happens for charge that is exactly on the surface — should it be counted toward the enclosed charge, or not? If charges can be perfect, infinitesimal points, then this could be a physically meaningful question. Suppose we approach this question by way of a limit: start with charge $q$ spread out over a sphere of finite size, and then make the size of the sphere approach zero. The figure shows a uniformly charged sphere that’s exactly half-way in and half-way out of the cubical Gaussian surface. What is the flux through the cube, compared to what it would be if the charge was entirely enclosed? (There are at least three ways to find this flux: by direct integration, by Gauss’s law, or by the additivity of flux by region.)

B The dipole is completely enclosed in the cube. What does Gauss’s law say about the flux through the cube? If you imagine the dipole’s field pattern, can you verify that this makes sense?

C The wire passes in through one side of the cube and out through the other. If the current through the wire is increasing, then the wire will act like an inductor, and there will be a voltage difference between its ends. (The inductance will be relatively small, since the wire isn’t coiled up, and the $\Delta V$ will therefore also be fairly small, but still not zero.) The $\Delta V$ implies the existence of electric fields, and yet Gauss’s law says the flux must be zero, since there is no charge inside the cube. Why isn’t Gauss’s law violated?

D The charge has been loitering near the edge of the cube, but is then suddenly hit with a mallet, causing it to fly off toward the left side of the cube. We haven’t yet discussed in detail how disturbances in the electric and magnetic fields ripple outward through space, but it turns out that they do so at the speed of light. (In fact, that’s what light is: ripples in the electric and magnetic fields.) Because the charge is closer to the left side of the cube, the change in the electric field occurs there before
the information reaches the right side. This would seem certain to lead to a violation of Gauss’s law. How can the ideas explored in discussion question C show the resolution to this paradox?

10.6.4 Proof of Gauss’ theorem

With the computational machinery we’ve developed, it is now simple to prove Gauss’ theorem. Based on additivity by charge, it suffices to prove the law for a point charge. We have already proved Gauss’ law for a point charge in the case where the point charge is outside the region. If we can prove it for the inside case, then we’re all done.

If the charge is inside, we reason as follows. First, we forget about the actual Gaussian surface of interest, and instead construct a spherical one, centered on the charge. For the case of a sphere, we’ve already seen the proof written on a napkin by the flea named Newton (page 641). Now wherever the actual surface sticks out beyond the sphere, we glue appropriately shaped pieces onto the sphere. In the example shown in figure h, we have to add two Mickey Mouse ears. Since these added pieces do not contain the point charge, the flux through them is zero, and additivity of flux by region therefore tells us that the total flux is not changed when we make this alteration. Likewise, we need to chisel out any regions where the sphere sticks out beyond the actual surface. Again, there is no change in flux, since the region being altered doesn’t contain the point charge. This proves that the flux through the Gaussian surface of interest is the same as the flux through the sphere, and since we’ve already proved that that flux equals \( 4\pi kq \), our proof of Gauss’ theorem is complete.

Discussion Questions

A critical part of the proof of Gauss’ theorem was the proof that a tiny cube has zero flux through it due to an external charge. Discuss qualitatively why this proof would fail if Coulomb’s law was a \( 1/r \) or \( 1/r^3 \) law.

10.6.5 Gauss’ law as a fundamental law of physics

Note that the proof of Gauss’ theorem depended on the computation on the napkin discussed on page 10.6.1. The crucial point in this computation was that the electric field of a point charge falls off like \( 1/r^2 \), and since the area of a sphere is proportional to \( r^2 \), the result is independent of \( r \). The \( 1/r^2 \) variation of the field also came into play on page 644 in the proof that the flux due to an outside charge is zero. In other words, if we discover some other force of nature which is proportional to \( 1/r^3 \) or \( r \), then Gauss’ theorem will not apply to that force. Gauss’ theorem is not true for nuclear forces, which fall off exponentially with distance. However, this is the only assumption we had to make about the nature of the field. Since gravity, for instance, also has fields that fall off as \( 1/r^2 \), Gauss’ theorem is equally valid for gravity — we just have to replace mass...
with charge, change the Coulomb constant $k$ to the gravitational constant $G$, and insert a minus sign because the gravitational fields around a (positive) mass point inward.

Gauss' theorem can only be proved if we assume a $1/r^2$ field, and the converse is also true: any field that satisfies Gauss' theorem must be a $1/r^2$ field. Thus although we previously thought of Coulomb's law as the fundamental law of nature describing electric forces, it is equally valid to think of Gauss' theorem as the basic law of nature for electricity. From this point of view, Gauss' theorem is not a mathematical fact but an experimentally testable statement about nature, so we'll refer to it as Gauss' law, just as we speak of Coulomb's law or Newton's law of gravity.

If Gauss' law is equivalent to Coulomb's law, why not just use Coulomb’s law? First, there are some cases where calculating a field is easy with Gauss’ law, and hard with Coulomb’s law. More importantly, Gauss' law and Coulomb's law are only mathematically equivalent under the assumption that all our charges are standing still, and all our fields are constant over time, i.e., in the study of electrostatics, as opposed to electrodynamics. As we broaden our scope to study generators, inductors, transformers, and radio antennas, we will encounter cases where Gauss’ law is valid, but Coulomb’s law is not.

10.6.6 Applications

Often we encounter situations where we have a static charge distribution, and we wish to determine the field. Although superposition is a generic strategy for solving this type of problem, if the charge distribution is symmetric in some way, then Gauss' law is often a far easier way to carry out the computation.

Field of a long line of charge

Consider the field of an infinitely long line of charge, holding a uniform charge per unit length $\lambda$. Computing this field by brute-force superposition was fairly laborious (examples 13 on page 596 and 16 on page 603). With Gauss' law it becomes a very simple calculation.

The problem has two types of symmetry. The line of charge, and therefore the resulting field pattern, look the same if we rotate them about the line. The second symmetry occurs because the line is infinite: if we slide the line along its own length, nothing changes. This sliding symmetry, known as a translation symmetry, tells us that the field must point directly away from the line at any given point.

Based on these symmetries, we choose the Gaussian surface shown in figure i. If we want to know the field at a distance $R$ from the line, then we choose this surface to have a radius $R$, as
shown in the figure. The length, $L$, of the surface is irrelevant.

The field is parallel to the surface on the end caps, and therefore perpendicular to the end caps’ area vectors, so there is no contribution to the flux. On the long, thin strips that make up the rest of the surface, the field is perpendicular to the surface, and therefore parallel to the area vector of each strip, so that the dot product occurring in the definition of the flux is $E_j \cdot A_j = |E_j||A_j|\cos 0^\circ = |E_j||A_j|$. Gauss’ law gives

$$4\pi k q_{in} = \sum E_j \cdot A_j$$
$$4\pi k \lambda L = \sum |E_j||A_j|.$$ The magnitude of the field is the same on every strip, so we can take it outside the sum.

$$4\pi k \lambda L = |E| \sum |A_j|$$

In the limit where the strips are infinitely narrow, the surface becomes a cylinder, with (area)= (circumference)(length)= $2\pi RL$.

$$4\pi k \lambda L = |E| \times 2\pi RL$$
$$|E| = \frac{2k\lambda}{R}$$

Field near a surface charge

As claimed earlier, the result $E = 2\pi k \sigma$ for the field near a charged surface is a special case of Gauss’ law. We choose a Gaussian surface of the shape shown in figure j, known as a Gaussian pillbox. The exact shape of the flat end caps is unimportant.

The symmetry of the charge distribution tells us that the field points directly away from the surface, and is equally strong on both sides of the surface. This means that the end caps contribute equally to the flux, and the curved sides have zero flux through them. If the area of each end cap is $A$, then

$$4\pi k q_{in} = E_1 \cdot A_1 + E_2 \cdot A_2,$$ where the subscripts 1 and 2 refer to the two end caps. We have $A_2 = -A_1$, so

$$4\pi k q_{in} = E_1 \cdot A_1 - E_2 \cdot A_1$$
$$4\pi k q_{in} = (E_1 - E_2) \cdot A_1,$$ and by symmetry the magnitudes of the two fields are equal, so

$$2|E|A = 4\pi k \sigma A$$
$$|E| = 2\pi k \sigma$$
The symmetry between the two sides could be broken by the existence of other charges nearby, whose fields would add onto the field of the surface itself. Even then, Gauss’s law still guarantees

\[ 4\pi k q_{in} = (\mathbf{E}_1 - \mathbf{E}_2) \cdot \mathbf{A}_1, \]

or

\[ |\mathbf{E}_{\perp,1} - \mathbf{E}_{\perp,2}| = 4\pi k \sigma, \]

where the subscript \( \perp \) indicates the component of the field parallel to the surface (i.e., parallel to the area vectors). In other words, the electric field changes discontinuously when we pass through a charged surface; the discontinuity occurs in the component of the field perpendicular to the surface, and the amount of discontinuous change is \( 4\pi k \sigma \). This is a completely general statement that is true near any charged surface, regardless of the existence of other charges nearby.