19. In vector diagrams such as those of Fig. 12-13, the distance from the center of the graph represents:

    A. Average amplitude.

    B. Frequency.

    C. Phase difference.

    D. Peak amplitude.

20. In diagrams like those of Fig. 12-13, the progression of time is sometimes depicted as:

    A. Movement to the right.

    B. Movement to the left.

    C. Rotation counterclockwise.

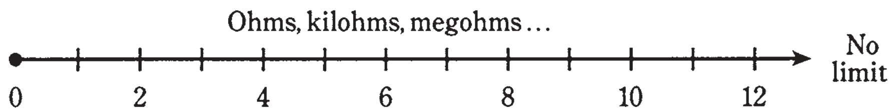
    D. Rotation clockwise.

13

CHAPTER

Inductive reactance

IN DC CIRCUITS, RESISTANCE IS A SIMPLE THING. IT CAN BE EXPRESSED AS A number, from zero (a perfect conductor) to extremely large values, increasing without limit through the millions, billions, and even trillions of ohms. Physicists call resistance a *scalar* quantity, because it can be expressed on a one-dimensional scale. In fact, dc resistance can be represented along a half line, or ray, as shown in Fig. 13-1.

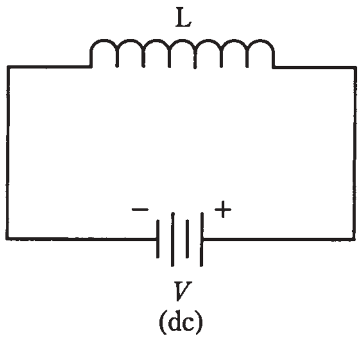


13-1 Resistance can be represented on a ray.

Given a certain dc voltage, the current decreases as the resistance increases, in accordance with Ohm's Law, as you already know. The same law holds for ac through a resistance, if the ac voltage and current are both specified as peak, pk-pk, or rms values.

Coils and direct current

Suppose that you have some wire that conducts electricity very well. What will happen if you wind a length of the wire into a coil and connect it to a source of dc, as shown in Fig. 13-2? The wire will draw a large amount of current, possibly blowing a fuse or overstressing a battery. It won't matter whether the wire is a single-turn loop, or whether it's lying haphazardly on the floor, or whether it's wrapped around a stick. The current will be large. In amperes, it will be equal to $I = E/R$, where I is the current, E is the dc voltage, and R is the resistance of the wire (a low resistance).



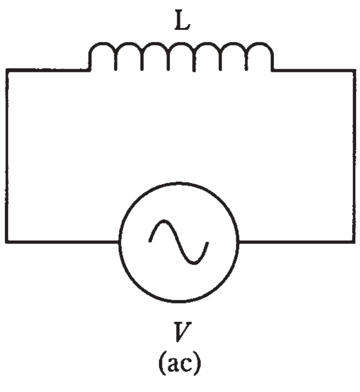
13-2 A coil connected across a source of dc.

You can make an electromagnet, as you've already seen, by passing dc through a coil wound around an iron rod. But there will still be a large, constant current in the coil. The coil will probably get more or less hot, as energy is dissipated in the resistance of the wire. The battery, too, or power supply components, will become warm or hot.

If the voltage of the battery or power supply is increased, the wire in the coil, iron core or not, will get hotter. Ultimately, if the supply can deliver the necessary current, the wire will melt.

Coils and alternating current

Suppose you change the voltage source, connected across the coil, from dc to ac (Fig. 13-3). Imagine that you can vary the frequency of the ac, from a few hertz to hundreds of hertz, then kilohertz, then megahertz.



13-3 A coil connected across a source of ac.

At first, the ac current will be high, just as it is with dc. But the coil has a certain amount of inductance, and it takes some time for current to establish itself in the coil. Depending on how many turns there are, and on whether the core is air or a ferromagnetic material, you'll reach a point, as the ac frequency increases, when the coil starts to get "sluggish." That is, the current won't have time to get established in the coil before the polarity of the voltage reverses.

At high ac frequencies, the *current* through the coil will have difficulty following the *voltage* placed across the coil. Just as the coil starts to "think" that it's making a good short circuit, the ac voltage wave will pass its peak, go back to zero, and then try to pull the electrons the other way.

This sluggishness in a coil for ac is, in effect, similar to dc resistance. As the frequency is raised higher and higher, the effect gets more and more pronounced. Eventually, if you keep on increasing the frequency of the ac source, the coil will not even begin to come near establishing a current with each cycle. It will act like a large resistance. Hardly any ac current will flow through it.

The opposition that the coil offers to ac is called *inductive reactance.* It, like resistance, is measured in ohms. It can vary just as resistance does, from near zero (a short piece of wire) to a few ohms (a small coil) to kilohms or megohms (bigger and bigger coils).

Like resistance, inductive reactance affects the current in an ac circuit. But, unlike simple resistance, reactance changes with frequency. And the effect is not just a decrease in the current, although in practice this will occur. It is a change in the way the current flows with respect to the voltage.

Reactance and frequency

Inductive reactance is one of two kinds of reactance; the other, called *capacitive reactance*, will be discussed in the next chapter. Reactance in general is symbolized by the capital letter X. Inductive reactance is indicated by the letter X with a subscript L: X_L.

If the frequency of an ac source is given, in hertz, as f, and the inductance of a coil in henrys is L, then the inductive reactance is

$$X_L = 6.28fL$$

This same formula applies if the frequency, f, is in kilohertz and the inductance, L, is in millihenrys. And it also applies if f is in megahertz and L is in microhenrys. Just remember that if frequency is in thousands, inductance must be in thousandths, and if frequency is in millions, inductance must be in millionths.

Inductive reactance increases *linearly* with increasing ac frequency. This means that the function of X_L vs f is a straight line when graphed.

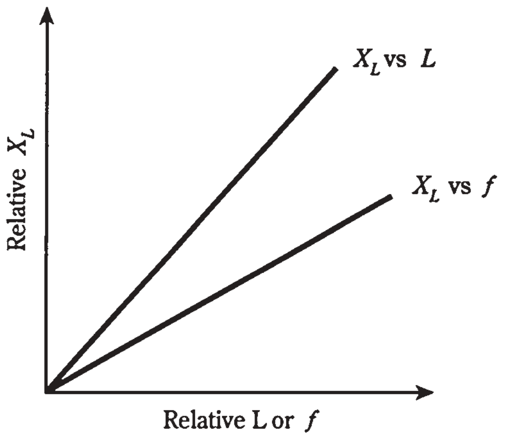
Inductive reactance also increases linearly with inductance. Therefore, the function of X_L vs L also appears as a straight line on a graph.

The value of X_L is directly proportional to f; X_L is also directly proportional to L. These relationships are graphed, in relative form, in Fig. 13-4.

Problem 13-1

Suppose a coil has an inductance of 0.50 H, and the frequency of the ac passing through it is 60 Hz. What is the inductive reactance?

Using the above formula, $X_L = 6.28 \times 60 \times 0.50 = 188.4$ ohms. You should round this off to two significant digits, or 190 ohms, because the inductance and frequency are both expressed to only two significant digits.



13-4 Inductive reactance is directly proportional to inductance and also to frequency.

Problem 13-2

What will be the inductive reactance of the above coil if the supply is a battery that supplies pure dc?

Because dc has a frequency of zero, $X_L = 6.28 \times 0 \times 0.50 = 0$. That is, there will be no inductive reactance. Inductance doesn't generally have any practical effect with pure dc.

Problem 13-3

If a coil has an inductive reactance of 100 Ω at a frequency of 5.00 MHz, what is its inductance?

In this case, you need to plug numbers into the formula and solve for the unknown L. Start out with the equation $100 = 6.28 \times 5.00 \times L = 31.4 \times L$. Then, recall that because the frequency is in megahertz, or millions of hertz, the inductance will come out in microhenrys, or millionths of a henry. You can divide both sides of the equation by 31.4, getting $L = 100/31.4 = 3.18\ \mu$H.

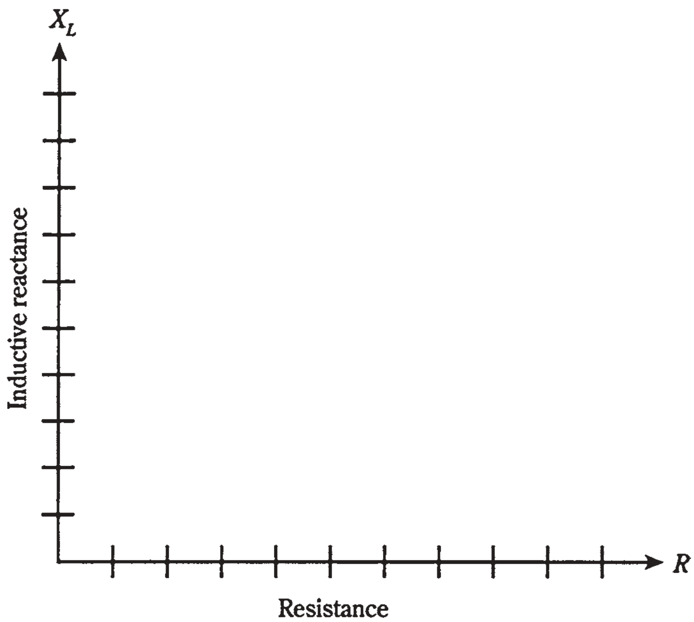
Points in the RL plane

Inductive reactance can be plotted along a half line, just as can resistance. In a circuit containing both resistance and inductance, the characteristics become two-dimensional. You can orient the resistance and reactance half lines perpendicular to each other to make a quarter-plane coordinate system, as shown in Fig. 13-5. Resistance is usually plotted horizontally, and inductive reactance is plotted vertically, going upwards.

In this scheme, RL combinations form *impedances.* You'll learn all about this in chapter 15. Each point on the *RL plane* corresponds to one unique impedance value. Conversely, each RL impedance value corresponds to one unique point on the plane.

For reasons made clear in chapter 15, impedances on the RL plane are written in the form $R + jX_L$, where R is the resistance and X_L is the inductive reactance.

If you have a pure resistance, say $R = 5\ \Omega$, then the *complex* impedance is $5 + j0$, and is at the point (5,0) on the RL plane. If you have a pure inductive reactance, such as



13-5 The RL quarter-plane.

$X_L = 3\ \Omega$, then the complex impedance is $0 + j3$, and is at the point (0, 3) on the RL plane. These points, and others, are shown in Fig. 13-6.

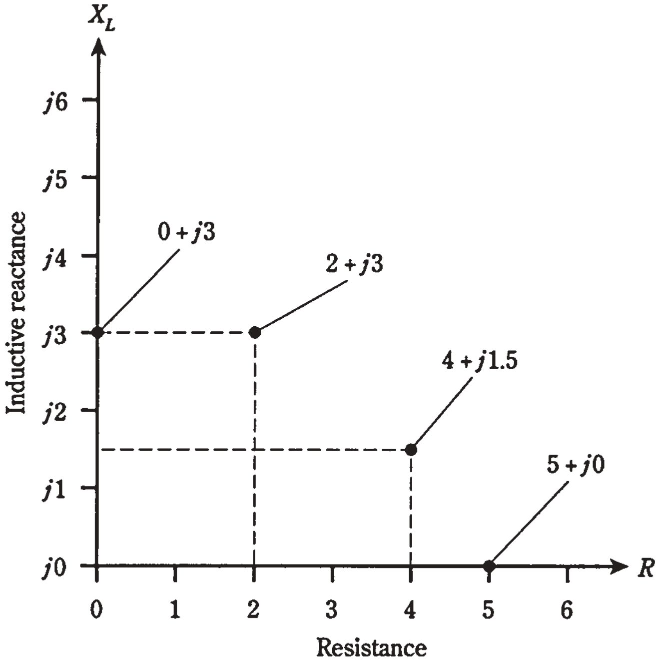
In real life, all coils have some resistance, because no wire is a perfect conductor. All resistors have at least a tiny bit of inductive reactance, because they take up some physical space. So there is really no such thing as a mathematically perfect pure resistance like $5 + j0$, or a mathematically perfect pure reactance like $0 + j3$. Sometimes you can get pretty close, but absolutely pure resistances or reactances never exist, if you want to get really theoretical.

Often, resistance and inductive reactance are both deliberately placed in a circuit. Then you get impedances values such as $2 + j3$ or $4 + j1.5$. These are shown in Fig. 13-6 as points on the RL plane.

Remember that the values for X_L are *reactances*, not the actual inductances. Therefore, they vary with the frequency in the RL circuit. Changing the frequency has the effect of making the points move in the RL plane. They move vertically, going upwards as the ac frequency increases, and downwards as the ac frequency decreases. If the ac frequency goes down to zero, the inductive reactance vanishes. Then $X_L = 0$, and the point is along the resistance axis of the RL plane.

Vectors in the RL Plane

Engineers sometimes like to represent points in the RL plane (and in other types of coordinate planes, too) as vectors. This gives each point a definite magnitude and a precise direction.



13-6 Four points in the RL impedance plane. See text for discussion.

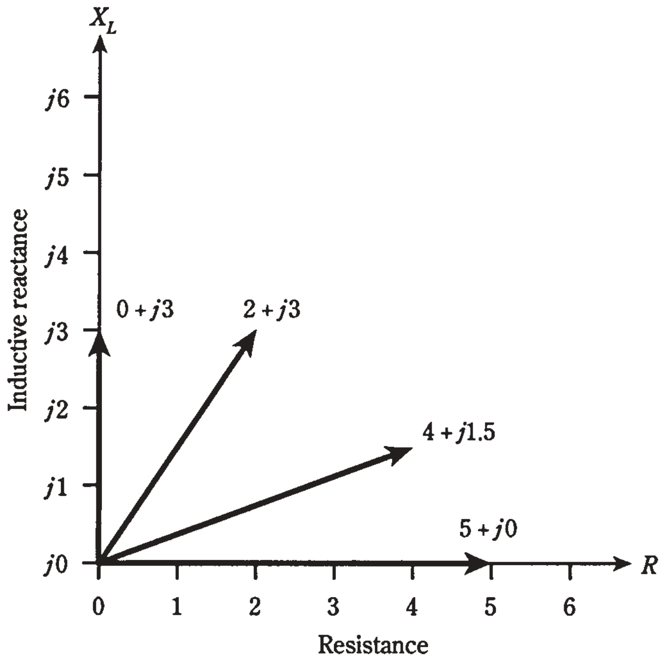
In Fig. 13-6, there are four different points shown. Each one is represented by a certain distance to the right of the origin (0,0), and a specific distance upwards from the origin. The first of these is the resistance, R, and the second is the inductive reactance, X_L. Thus, the RL combination is a two-dimensional quantity. There is no way to uniquely define RL combinations as single numbers, or *scalars*, because there are two different quantities that can vary independently.

Another way to think of these points is to draw lines from the origin out to them. Then you can think of the points as rays, each having a certain length, or magnitude, and a definite direction, or angle counterclockwise from the resistance axis. These rays, going out to the points, are vectors (Fig. 13-7). You've already been introduced to these things.

Vectors seem to engender apprehension in some people, as if they were invented by scientists for the perverse pleasure of befuddling ordinary folks. "What are you taking this semester?" asks Jane. "*Vector* analysis!" Joe shudders (if he's one of the timid types), or beams (if he wants to impress Jane).

This attitude is completely groundless. Just think of vectors as arrows that have a certain length, and that point in some direction.

In Fig. 13-7, the points of Fig. 13-6 are shown as vectors. The only difference is that there is some more ink on the paper.



13-7 Four vectors in the RL impedance plane.

Current lags voltage

Inductance, as you recall, stores electrical energy as a magnetic field. When a voltage is placed across a coil, it takes awhile for the current to build up to full value.

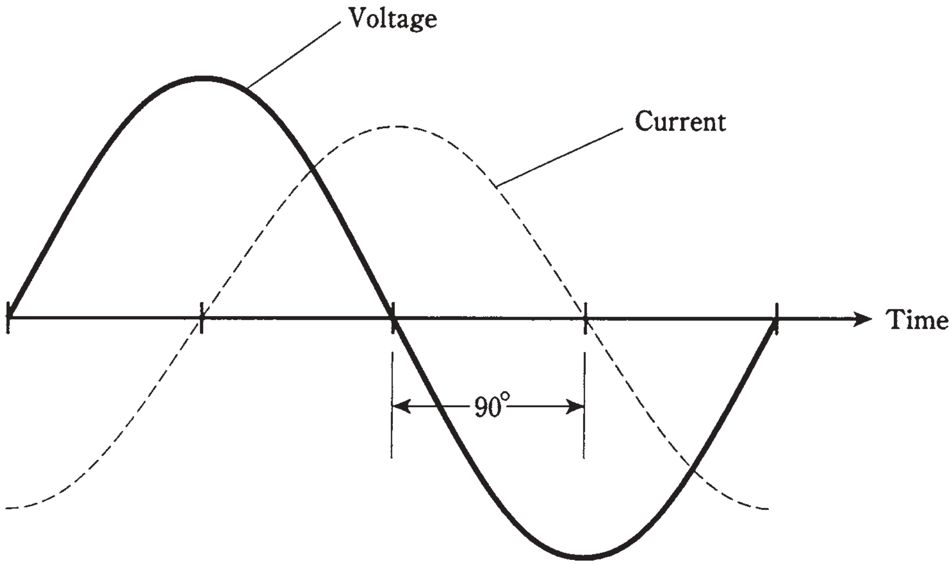
When ac is placed across a coil, the current *lags* the voltage in phase.

Pure inductance

Suppose that you place an ac voltage across a low-loss coil, with a frequency high enough so that the inductive reactance, X_L, is much larger than the resistance, R. In this situation, the current is one-quarter of a cycle behind the voltage. That is, the current lags the voltage by 90 degrees (Fig. 13-8).

At very low frequencies, large inductances are normally needed in order for this current lag to be a full ¼ cycle. This is because any coil has some resistance; no wire is a perfect conductor. If some wire were found that had a mathematically zero resistance, and if a coil of any size were wound from this wire, then the current would lag the voltage by 90 degrees in this inductor, no matter what the ac frequency.

When the value of X_L is very large compared with the value of R in a circuit—that is, when there is an essentially pure inductance—the vector in the RL plane points straight up along the X_L axis. Its angle is 90 degrees from the R axis, which is considered the *zero line* in the RL plane.



13-8 In a pure inductance, the current lags the voltage by 90 degrees.

Inductance and resistance

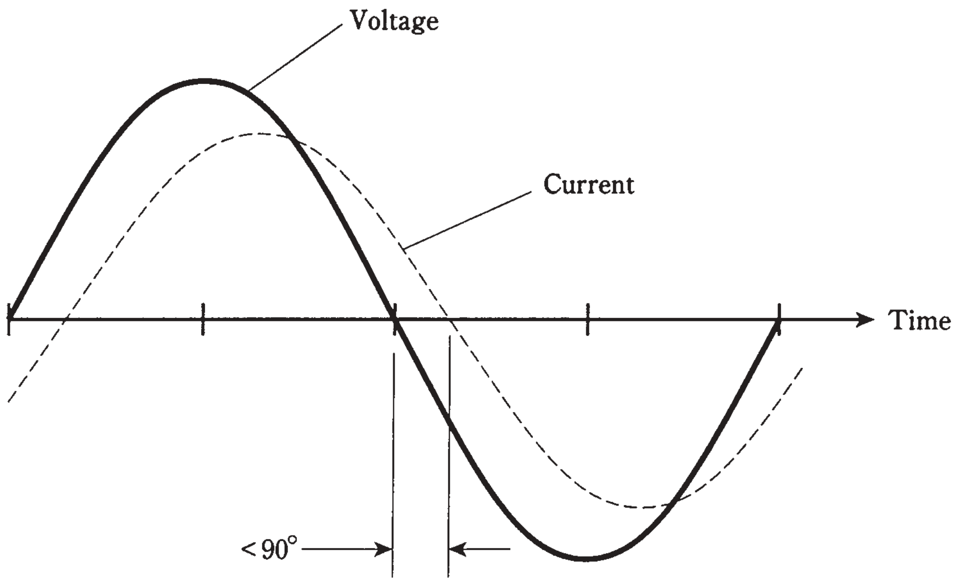
When the resistance in a resistance-inductance circuit is significant compared with the inductive reactance, the current lags the voltage by something less than 90 degrees (Fig. 13-9). If R is small compared with X_L, the current lag is almost 90 degrees; as R gets larger, the lag decreases. A circuit with resistance and inductance is called an *RL circuit*.

The value of R in an RL circuit might increase relative to X_L because resistance is deliberately placed in series with the inductance. Or, it might happen because the ac frequency gets so low that X_L decreases until it is in the same ball park with the loss resistance R in the coil winding. In either case, the situation can be schematically represented by a coil in series with a resistor (Fig. 13-10).

If you know the values of X_L and R, you can find the *angle of lag,* also called the *RL phase angle,* by plotting the point $R + jX_L$ on the RL plane, drawing the vector from the origin $0 + j0$ out to that point, and then measuring the angle of the vector, counterclockwise from the resistance axis. You can use a protractor to measure this angle, or you can compute its value using trigonometry.

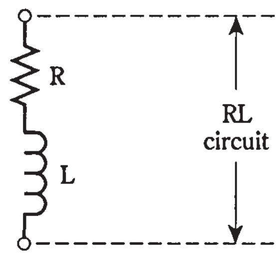
In fact, you don't need to know the actual values of X_L and R in order to find the angle of lag. All you need to know is their ratio. For example, if $L = 5\ \Omega$ and $R = 3\ \Omega$, you will get the same angle as you would get if $X_L = 50\ \Omega$ and $R = 30\ \Omega$, or if $X_L = 20\ \Omega$ and $R = 12\ \Omega$. The angle of lag will be the same for any values of X_L and R in the ratio of 5:3.

It's easy to find the angle of lag whenever you know the ratio of R to X_L. You'll see some examples shortly.



13-9 In a circuit with inductance and resistance, the current lags the voltage by less than 90 degrees.

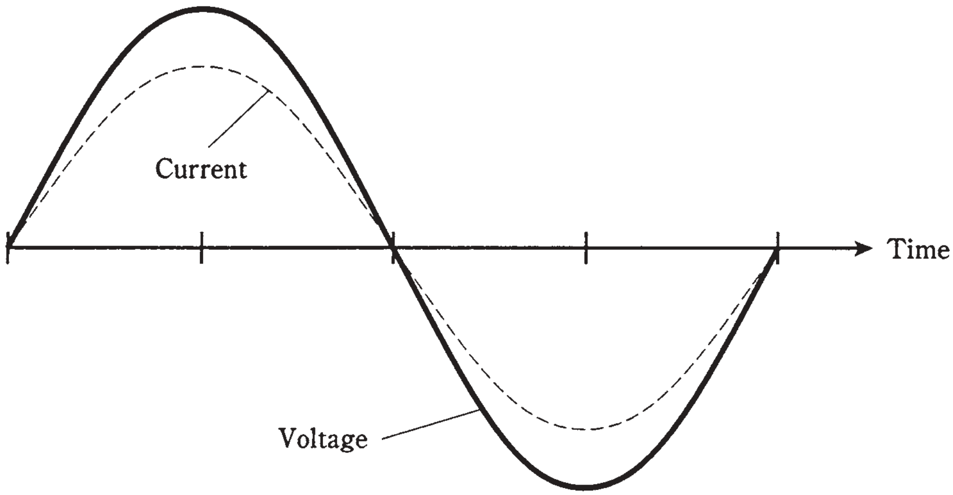
13-10 Schematic representation of an RL circuit.



Pure resistance

As the resistance in an RL circuit becomes large with respect to the inductive reactance, the angle of lag gets smaller and smaller. The same thing happens if the inductive reactance gets small compared with the resistance. When R is many times greater than X_L, whatever their actual magnitudes might be, the vector in the RL plane lies almost on the R axis, going "east" or to the right. The RL phase angle is nearly zero. The current comes into phase with the voltage.

In a pure resistance, with no inductance at all, the current is precisely in phase with the voltage (Fig. 13-11). A pure resistance doesn't store any energy, as an inductive circuit does, but sends the energy out as heat, light, electromagnetic waves, sound, or some other form that never comes back into the circuit.



13-11 In a circuit with only resistance, the current is in phase with the voltage.

How much lag?

If you know the ratio of the inductive reactance to the resistance, X_L/R, in an RL circuit, then you can find the phase angle. Of course, you can find the angle of lag if you know the actual values of X_L and R.

Pictorial method

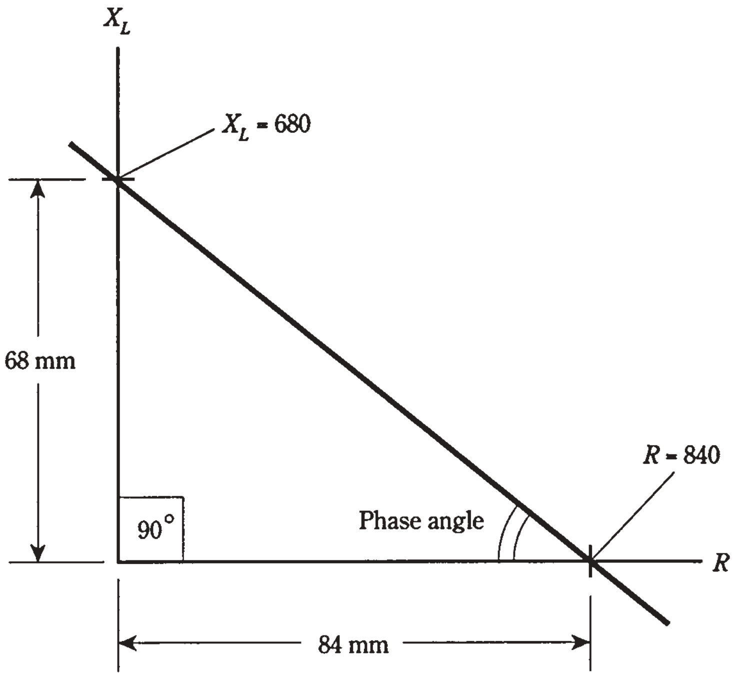
It isn't necessary to construct an entire RL plane to find phase angles. You can use a ruler that has centimeter (cm) and millimeter (mm) markings, and a protractor.

First, draw a line a little more than 10 cm long, going from left to right on a sheet of paper. Use the ruler and a sharp pencil. Then, with the protractor, construct a line off the left end of this first line, going vertically upwards. Make this line at least 10 cm long.

The horizontal line, or the one going to the right, is the R axis of a crude coordinate system. The vertical line, or the one going upwards, is the X_L axis.

If you know the values of X_L and R, divide them down, or multiply them up (in your head) so that they're both between 0 and 100. For example, if $X_L = 680\ \Omega$ and $R = 840\ \Omega$, you can divide them both by 10 to get $X_L = 68$ and $R = 84$. Plot these points lightly by making hash marks on the vertical and horizontal lines you've drawn. The R mark will be 84 mm to the right of the origin, or intersection of the original two perpendicular lines. The X_L mark will be 68 mm up from the origin.

Draw a line connecting the two hash marks, as shown in Fig. 13-12. This line will run at a slant, and will form a triangle along with the two axes. Your hash marks, and the origin of the coordinate system, form *vertices* of a *right triangle.* The triangle is called "right" because one of its angles is a right angle (90 degrees.)



13-12 Pictorial method of finding phase angle.

Measure the angle between the slanted line and the horizontal, or R, axis. Extend one or both of the lines if necessary in order to get a good reading on the protractor. This angle will be between 0 and 90 degrees, and represents the phase angle in the RL circuit.

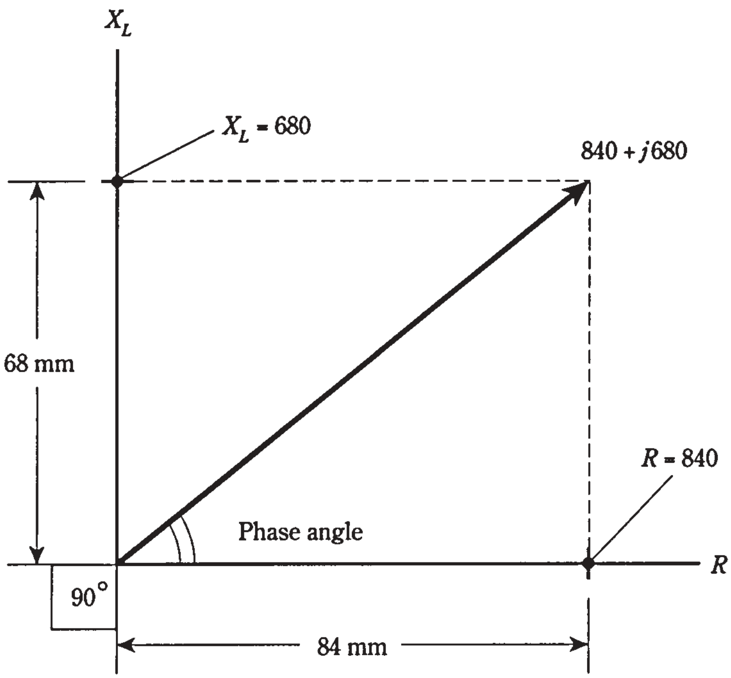
The actual vector, $R + jX_L$, is found by constructing a rectangle using the origin and your two hash marks as three of the four vertices, and drawing new horizontal and vertical lines to complete the figure. The vector is the diagonal of this rectangle, as shown in Fig. 13-13. The phase angle is the angle between this vector and the R axis. It will be the same as the angle of the slanted line in Fig. 13-12.

Trigonometric method

The pictorial method is inexact. It can be a bother, sometimes, to divide down and get numbers that are easy to work with. Drawing the pictures accurately requires care and patience. If the phase angle is very close to 0 degrees or 90 degrees—that is, if the ratio X_L/R is very small or large—the accuracy is especially poor in the drawing.

There's a better way. If you have a calculator that can find the *arctangent* of a number, you've got it made easy. Nowadays, if you intend to work with engineers, there's no excuse not to have a good calculator. If you don't have one, I suggest you go out and buy one right now. It should have "trig" functions and their inverses, "log" functions, exponential functions, and others that engineers use often.

If you know the values X_L and R, then the phase angle is simply the arctangent of their ratio, or arctan (X_L/R). This might also be written $\tan^{-1}(X_L/R)$. Punch a few buttons on the calculator, and you have it.



13-13 Another pictorial way of finding phase angle. This method shows the actual impedance vector.

Problem 13-4

The inductive reactance in an RL circuit is 680 Ω, and the resistance is 840 Ω. What is the phase angle?

Find the ratio $X_L/R = 680/840$. The calculator will display something like 0.809523809. Find the arctangent, or $\tan^{-1}$, getting a phase angle of 38.99099404 degrees (as shown on the calculator). Round this off to 39.0 degrees.

Problem 13-5

An RL circuit works at a frequency of 1.0 MHz. It has a resistance of 10 Ω, and an inductance of 90 μH. What is the phase angle?

This is a rather complicated problem, because it requires several steps. But each step is straightforward. You need to do them carefully, one at a time, and then recheck the whole problem once or twice when you're done calculating.

First, find the inductive reactance. This is found by the formula $X_L = 6.28fL = 6.28 \times 1.0 \times 90 = 565\ \Omega$. Then find the ratio $X_L/R = 565/10 = 56.5$. The phase angle is arctan $56.5 = 89$ degrees. This indicates that the circuit is almost purely reactive. The resistance contributes little to the behavior of this RL circuit at this frequency.

Problem 13-6

What is the phase angle of the above circuit at 10 kHz?

This requires that X_L be found over again, for the new frequency. Suppose you decide to use megahertz, so it will go nicely in the formula with microhenrys. A frequency of 10 kHz is the same as 0.010 MHz. Calculating, you get $X_L = 6.28fL = 6.28 \times 0.010 \times 90 = 6.28 \times 0.90 = 5.65\ \Omega$. The ratio X_L/R is then $5.65/10 = 0.565$. The phase angle is $\arctan 0.565 = 29$ degrees. At this frequency, the resistance and inductance both play significant roles in the RL circuit.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. As the number of turns in a coil increases, the current in the coil will eventually:
   A. Become very large.
   B. Stay the same.
   C. Decrease to near zero.
   D. Be stored in the core material.

2. As the number of turns in a coil increases, the reactance:
   A. Increases.
   B. Decreases.
   C. Stays the same.
   D. Is stored in the core material.

3. As the frequency of an ac wave gets lower, the value of X_L for a particular coil:
   A. Increases.
   B. Decreases.
   C. Stays the same.
   D. Depends on the voltage.

4. A coil has an inductance of 100 mH. What is the reactance at a frequency of 1000 Hz?
   A. 0.628 Ω.
   B. 6.28 Ω.
   C. 62.8 Ω.
   D. 628 Ω.

5. A coil shows an inductive reactance of 200 Ω at 500 Hz. What is its inductance?
   A. 0.637 H.
   B. 628 H.
   C. 63.7 mH.
   D. 628 mH.

6. A coil has an inductance of 400 μH. Its reactance is 33 Ω. What is the frequency?
   A. 13 kHz.
   B. 0.013 kHz.
   C. 83 kHz.
   D. 83 MHz.
7. An inductor has $X_L = 555\ \Omega$ at $f = 132$ kHz. What is L?
   A. 670 mH.
   B. 670 μH.
   C. 460 mH.
   D. 460 μH.
8. A coil has $L = 689\ \mu$H at $f = 990$ kHz. What is X_L?
   A. 682 Ω.
   B. 4.28 Ω.
   C. 4.28 KΩ.
   D. 4.28 MΩ.
9. An inductor has $L = 88$ mH with $X_L = 100\ \Omega$. What is f?
   A. 55.3 kHz.
   B. 55.3 Hz.
   C. 181 kHz.
   D. 181 Hz.
10. Each point in the RL plane:
    A. Corresponds to a unique resistance.
    B. Corresponds to a unique inductance.
    C. Corresponds to a unique combination of resistance and inductive reactance.
    D. Corresponds to a unique combination of resistance and inductance.
11. If the resistance R and the inductive reactance X_L both vary from zero to unlimited values, but are always in the ratio 3:1, the points in the RL plane for all the resulting impedances will fall along:
    A. A vector pointing straight up.
    B. A vector pointing "east."
    C. A circle.
    D. A ray of unlimited length.
12. Each impedance $R + jX_L$:
    A. Corresponds to a unique point in the RL plane.
    B. Corresponds to a unique inductive reactance.

C. Corresponds to a unique resistance.

D. All of the above.

13. A vector is a quantity that has:

A. Magnitude and direction.

B. Resistance and inductance.

C. Resistance and reactance.

D. Inductance and reactance.

14. In an RL circuit, as the ratio of inductive reactance to resistance, X_L/R, decreases, the phase angle:

A. Increases.

B. Decreases.

C. Stays the same.

D. Cannot be found.

15. In a purely reactive circuit, the phase angle is:

A. Increasing.

B. Decreasing.

C. 0 degrees.

D. 90 degrees.

16. If the inductive reactance is the same as the resistance in an RL circuit, the phase angle is:

A. 0 degrees.

B. 45 degrees.

C. 90 degrees.

D. Impossible to find; there's not enough data given.

17. In Fig. 13-14, the impedance shown is:

A. 8.0.

B. 90.

C. $90 + j8.0$.

D. $8.0 + j90$.

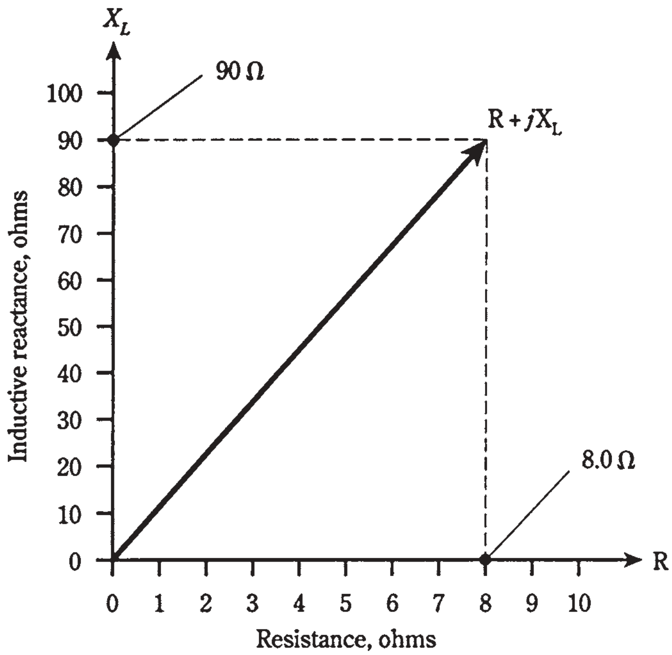
18. In Fig. 13-14, note that the R and X_L scale divisions are of different sizes. The phase angle is:

A. About 50 degrees, from the looks of it.

B. 48 degrees, as measured with a protractor.

C. 85 degrees, as calculated trigonometrically.

D. 6.5 degrees, as calculated trigonometrically.



13-14 Illustration for quiz questions 17 and 18.

19 An RL circuit consists of a 100-μH inductor and a 100-Ω resistor. What is the phase angle at a frequency of 200 kHz?

A. 45.0 degrees.

B. 51.5 degrees.

C. 38.5 degrees.

D. There isn't enough data to know.

20. An RL circuit has an inductance of 88 mH. The resistance is 95 Ω. What is the phase angle at 800 Hz?

A. 78 degrees.

B. 12 degrees.

C. 43 degrees.

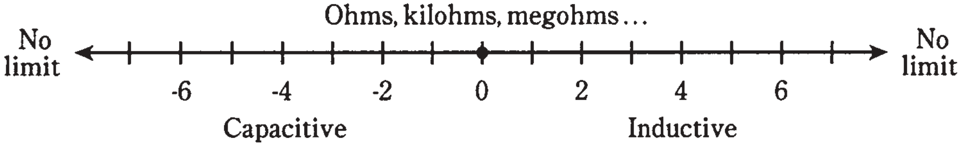
D. 47 degrees.

14
CHAPTER

Capacitive reactance

INDUCTIVE REACTANCE IS SOMETHING LIKE RESISTANCE, IN THE SENSE THAT IT is a one-dimensional, or scalar, quantity that can vary from zero upwards without limit. Inductive reactance, like resistance, can be represented by a ray, and is measured in ohms.

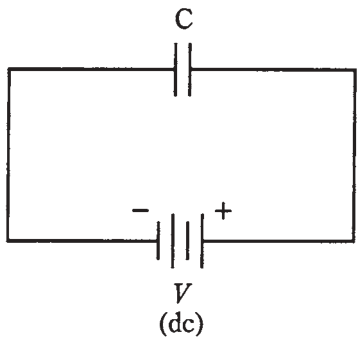
Inductive reactance has its counterpart in the form of *capacitive reactance.* This too can be represented as a ray, starting at the same zero point as inductive reactance, but running off in the opposite direction, having *negative* ohmic values (Fig. 14-1). When the ray for capacitive reactance is combined with the ray for inductive reactance, a number line is the result, with ohmic values that range from the huge negative numbers, through zero, to huge positive numbers.



14-1 Inductive and capacitive reactance can be represented on a complete ohmic number line.

Capacitors and direct current

Suppose that you have two big, flat metal plates, both of which are excellent electrical conductors. Imagine that you stack them one on top of the other, with only air in between. What will take place if you connect a source of dc across the plates (Fig. 14-2)? The plates will become electrically charged, and will reach a potential difference equal to the dc source voltage. It won't matter how big or small the plates are; their mutual voltage will always be the same as that of the source, although, if the plates are monstrously large, it might take awhile for them to become fully charged. The current, once the plates are charged, will be zero.



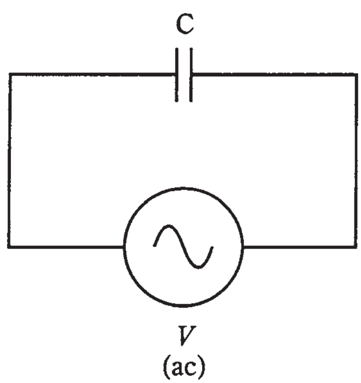
14-2 A capacitor connected across a source of dc.

If you put some insulating material, such as glass, between the plates, their mutual voltage will not change, although the charging time might increase.

If you increase the source voltage, the potential difference between the plates will follow along, more or less rapidly, depending on how large the plates are and on what is between them. If the voltage is increased without limit, arcing will eventually take place. That is, sparks will begin to jump between the plates.

Capacitors and alternating current

Suppose that the source is changed from direct to alternating current (Fig. 14-3). Imagine that you can adjust the frequency of this ac from a low value of a few hertz, to hundreds of hertz, to many kilohertz, megahertz, and gigahertz.



14-3 A capacitor across a source of ac.

At first, the voltage between the plates will follow just about exactly along as the ac source polarity reverses over and over. But the set of plates has a certain amount of capacitance, as you have learned. Perhaps they can charge up fast, if they are small and if the space between them is large, but they can't charge instantaneously.

As you increase the frequency of the ac voltage source, there will come a point at which the plates do not get charged up very much before the source polarity reverses. The set of plates will be "sluggish." The charge won't have time to get established with each ac cycle.

At high ac frequencies, the voltage between the plates will have trouble following the current that is charging and discharging them. Just as the plates begin to get a good charge, the ac current will pass its peak and start to discharge them, pulling electrons out of the negative plate and pumping electrons into the positive plate.

As the frequency is raised, the set of plates starts to act more and more like a short circuit. When the frequency is low, there is a small charging current, but this quickly tails off and drops to zero as the plates become fully charged. As the frequency becomes high, the current flows for more and more of every cycle before dropping off; the charging time remains constant while the *period* of the charging/discharging wave is getting shorter. Eventually, if you keep on increasing the frequency, the period of the wave will be much shorter than the charging/discharging time, and current will flow in and out of the plates in just about the same way as it would flow if the plates were shorted out.

The opposition that the set of plates offers to ac is the capacitive reactance. It is measured in ohms, just like inductive reactance, and just like resistance. But it is, by convention, assigned negative values rather than positive ones. Capacitive reactance, denoted X_{C}, can vary, just as resistance and inductive reactance do, from near zero (when the plates are huge and close together, and/or the frequency is very high) to a few negative ohms, to many negative kilohms or megohms.

Capacitive reactance varies with frequency. But X_C gets larger (negatively) as the frequency goes *down.* This is the opposite of what happens with inductive reactance, which gets larger (positively) as the frequency goes *up.*

Sometimes, capacitive reactance is talked about in terms of its *absolute value,* with the minus sign removed. Then you might say that X_C is increasing as the frequency decreases, or that X_C is decreasing as the frequency is raised. It's best, however, if you learn to work with negative X_C values right from the start. This will be important later, when you need to work with inductive and capacitive reactances together in the same circuits.

Reactance and frequency

In many ways capacitive reactance behaves like a mirror image of inductive reactance. But in another sense, X_C is an extension of X_L into negative values—below zero.

If the frequency of an ac source is given in hertz as f and the capacitance of a capacitor in farads as C, then the capacitive reactance is

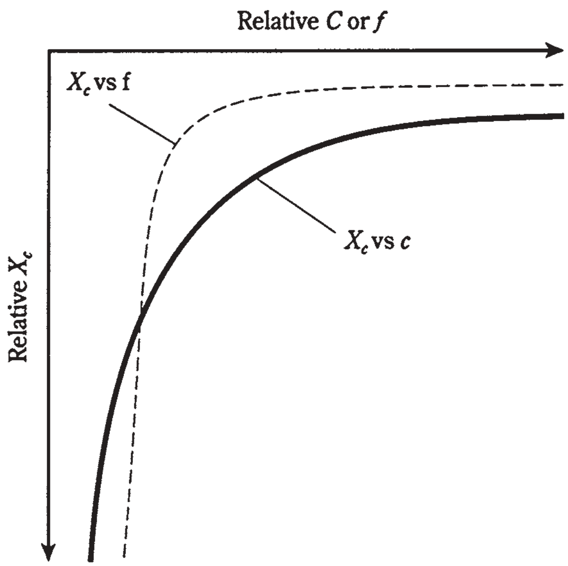
$$X_C = -1/(6.28fC)$$

This same formula applies if the frequency, f, is in megahertz and the capacitance, C, is in microfarads (μF). Just remember that if the frequency is in millions, the capacitance must be in million*ths*.

Capacitive reactance varies *inversely* with the frequency. This means that the function X_C vs f appears as a curve when graphed, and this curve "blows up" as the frequency nears zero.

Capacitive reactance also varies inversely with the actual value of capacitance, given a fixed frequency. Therefore, the function of X_C vs C also appears as a curve that "blows up" as the capacitance approaches zero.

The negative of X_c is *inversely proportional* to frequency, and also to capacitance. Relative graphs of these functions are shown in Fig. 14-4.



14-4 Capacitive reactance is negatively, and inversely, proportional to frequency (f) and also to capacitance (C).

Problem 14-1

A capacitor has a value of 0.00100 μF at a frequency of 1.00 MHz. What is the capacitive reactance?

Use the formula and plug in the numbers. You can do this directly, since the data is specified in *micro*farads (millionths) and in *mega*hertz (millions):

$$X_C = -1/(6.28 \times 1.0 \times 0.00100) = -1/(0.00628) = -159\ \Omega$$

This is rounded to three significant figures, since all the data is given to this many digits.

Problem 14-2

What will be the capacitive reactance of the above capacitor, if the frequency decreases to zero? That is, if the source is dc?

In this case, if you plug the numbers into the formula, you'll get zero in the denominator. Mathematicians will tell you that this is a no-no. But in reality, you can say that the reactance will be "extremely large negatively, and for practical purposes, negative infinity."

Problem 14-3

Suppose a capacitor has a reactance of $-100\ \Omega$ at a frequency of 10.0 MHz. What is its capacitance?

In this problem, you need to put the numbers in the formula and solve for the unknown C. Begin with the equation

$$-100 = -1/(6.28 \times 10.0 \times C)$$

Dividing through by -100, you get:

$$1 = 1/(628 \times 10.0 \times C)$$

Multiply each side of this by C, and you obtain:

$$C = 1/(628 \times 10.0)$$

This can be solved easily enough. Divide out $C = 1/6280$ on your calculator, and you'll get $C = 0.000159$. Because the frequency is given in megahertz, this capacitance comes out in microfarads, so that C = 0.000159 μF. You might rather say that this is 159 pf (remember that 1 pF = 0.000001 μF).

Admittedly, the arithmetic for dealing with capacitive reactance is a little messier than that for inductive reactance. This is the case for two reasons. First, you have to work with reciprocals, and therefore the numbers sometimes get awkward. Second, you have to watch those negative signs. It's easy to leave them out. But they're important when looking at reactances in the coordinate plane, because the minus sign tells you that the reactance is capacitive, rather than inductive.

Points in the RC plane

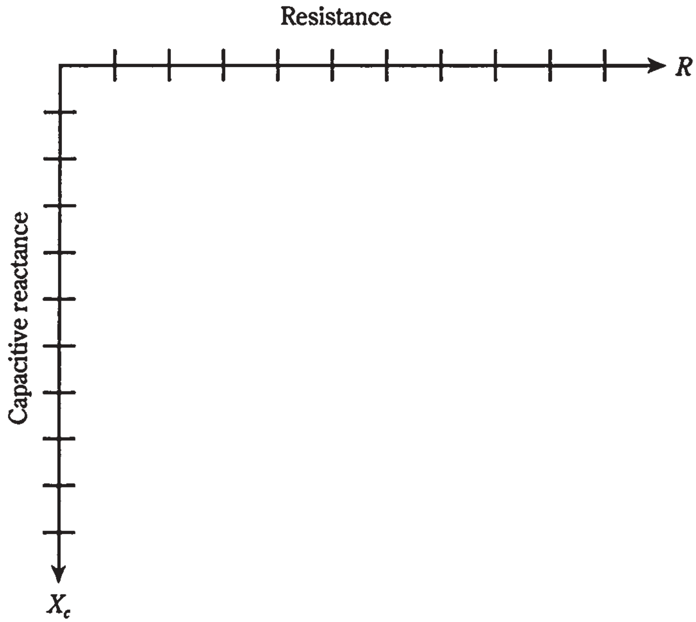
Capacitive reactance can be plotted along a half line, or ray, just as can inductive reactance. In fact, capacitive and inductive reactance, considered as one, form a whole line that is made of two half lines stuck together and pointing in opposite directions. The point where they join is the zero-reactance point. This was shown back in Fig. 14-1.

In a circuit containing resistance and capacitive reactance, the characteristics are two-dimensional, in a way that is analogous to the situation with the RL plane from the previous chapter. The resistance ray and the capacitive-reactance ray can be placed end to end at right angles to make a quarter plane called the *RC plane* (Fig. 14-5). Resistance is plotted horizontally, with increasing values toward the right. Capacitive reactance is plotted *downwards,* with increasingly negative values as you go down.

The combinations of R and X_C in this RC plane form impedances. You'll learn about impedance in greater detail in the next chapter. Each point on the RC plane corresponds to one and only one impedance. Conversely, each specific impedance coincides with one and only one point on the plane.

Impedances that contain resistance and capacitance are written in the form $R + jX_C$. Remember that X_C is *never positive,* that is, it is always negative or zero. Because of this, engineers will often write $R - jX_C$, dropping the minus sign from X_C and replacing addition with subtraction in the complex rendition of impedance.

If the resistance is pure, say R = 3 Ω, then the complex impedance is $3 - j0$ and this corresponds to the point (3,0) on the RC plane. You might at this point suspect that $3 - j0$ is the same as $3 + j0$, and that you really need not even write the "$j0$" part at all. In theory, both of these notions are indeed correct. But writing the "$j0$" part indicates that you are open to the



14-5 The RC quarter-plane.

possibility that there might be reactance in the circuit, and that you're working in two dimensions. It also underscores the fact that the impedance is a pure resistance.

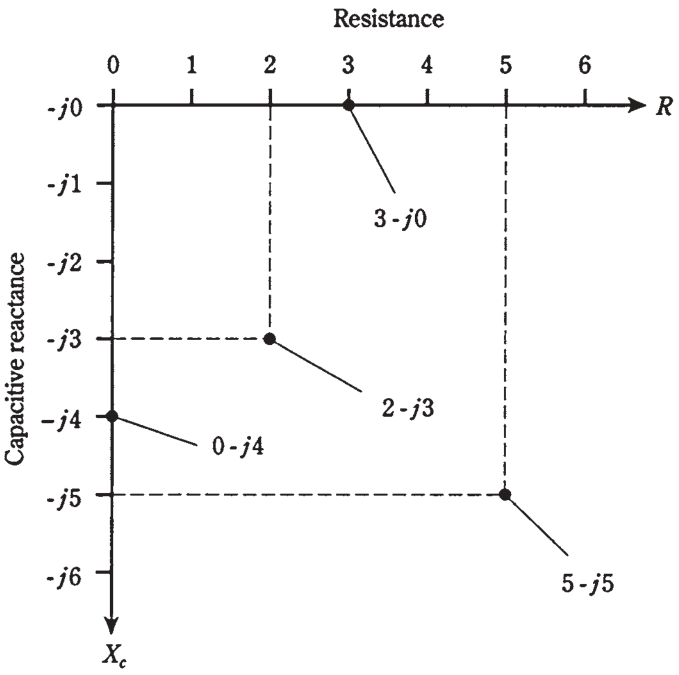
If you have a pure capacitive reactance, say $X_C = -4\ \Omega$, then the complex impedance is $0 - j4$, and this is at the point $(0, -4)$ on the RC plane. Again, it's important, for completeness, to write the "0" and not just the " $-j4$."

The points for $3 - j0$ and $0 - j4$, and two others, are plotted on the RC plane in Fig. 14-6.

In practical circuits, all capacitors have some *leakage resistance.* If the frequency goes to zero, that is, if the source is dc, a tiny current will flow because no insulator is perfect. Some capacitors have almost no leakage resistance, and come close to being perfect. But none are mathematically flawless. All resistors have a little bit of capacitive reactance, just because they occupy physical space. So there is no such thing as a mathematically pure resistance, either. The points $3 - j0$ and $0 - j4$ are idealized.

Often, resistance and capacitive reactance are both placed in a circuit deliberately. Then you get impedances such as $2 - j3$ and $5 - j5$, both shown in Fig. 14-6.

Remember that the values for X_C are *reactances,* and not the actual capacitances. They vary with the frequency in an RC circuit. If you raise or lower the frequency, the value of X_C will change. A higher frequency causes X_C to get smaller and smaller negatively (closer to zero). A lower frequency causes X_C to get larger and larger negatively (farther from zero, or lower down on the RC plane). If the frequency goes to zero, then the capacitive reactance drops off the bottom of the plane, out of sight. In that case you have two oppositely charged plates or sets of plates, and no "action."



14-6 Four points in the RC plane. See text for discussion.

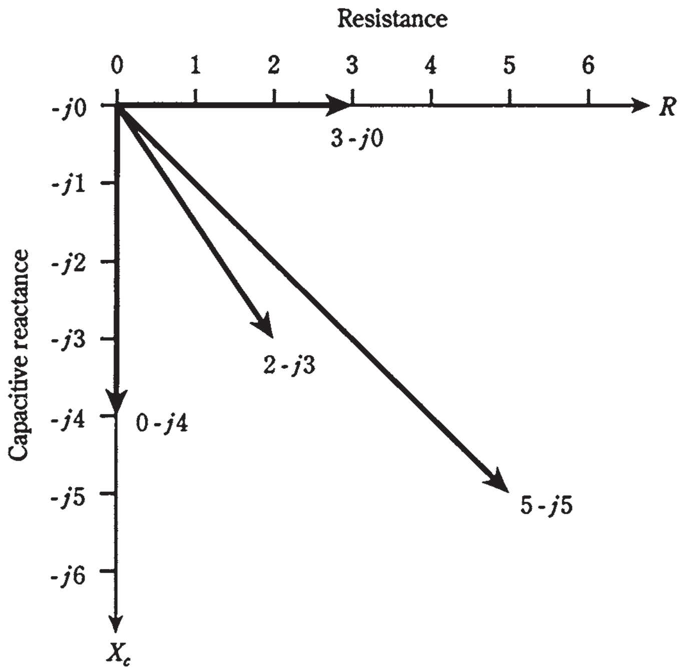
Vectors in the RC plane

If you work much with engineers, or if you plan to become one, you'll get familiar with the RC plane, just as you will with the RL plane. Recall from the last chapter that RL impedances can be represented as vectors. The same is true for RC impedances.

In Fig. 14-6, there are four different impedance points. Each one is represented by a certain distance to the right of the origin (0,0), and a certain displacement downwards. The first of these is the resistance, R, and the second is the capacitive reactance, X_C. Therefore, the RC impedance is a two-dimensional quantity.

Doesn't this look like a mirror-image reflection of RL impedances? You could almost imagine that we're looking at an RL plane reflected in a pool of still water. This is, in fact, an excellent way to envision this situation.

The impedance points in the RC plane can be rendered as vectors, just as this can be done in the RL plane. Then the points become rays, each with a certain length and direction. The magnitude and direction for a vector, and the coordinates for the point, both uniquely define the same impedance value. The length of the vector is the distance of the point from the origin, and the direction is the angle measured *clockwise* from the resistance (R) line, and specified in *negative* degrees. The equivalent vectors, for the points in Fig. 14-6, are illustrated in Fig. 14-7.



14-7 Four vectors in the RC impedance plane.

Current leads voltage

Capacitance stores energy in the form of an electric field. When a current is driven through a capacitor, it takes a little time before the plates can fully charge to the full potential difference of the source voltage.

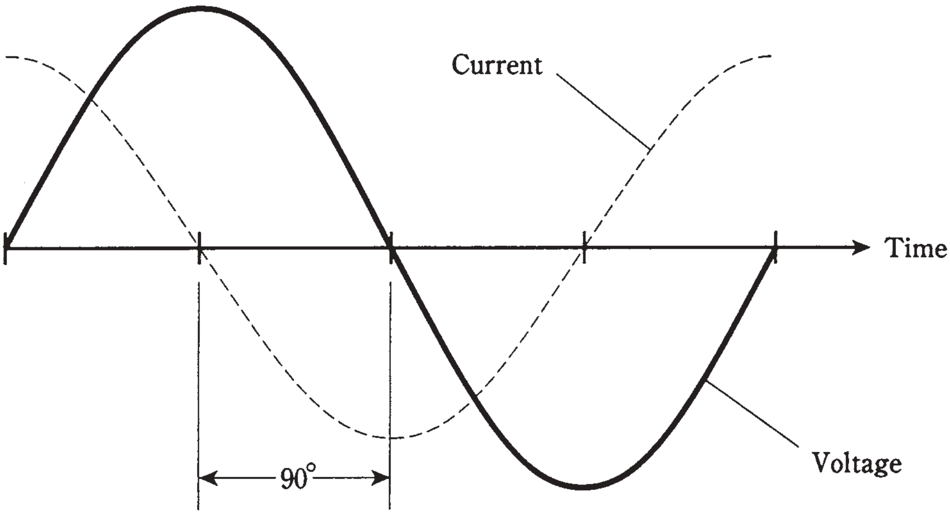
When an ac voltage source is placed across a capacitor, the voltage in the capacitor lags the current in phase. Another way of saying this is that the current *leads* the voltage. The phase difference can range from zero, to a small part of a cycle, to a quarter of a cycle (90 degrees).

Pure capacitance

Imagine placing an ac voltage source across a capacitor. Suppose that the frequency is high enough, and/or the capacitance large enough, so that the capacitive reactance, X_C, is extremely small compared with the resistance, R. Then the current leads the voltage by a full 90 degrees (Fig. 14-8).

At very high frequencies, it doesn't take very much capacitance for this to happen. Small capacitors usually have less leakage resistance than large ones. At lower frequencies, the capacitance must be larger, although high-quality, low-loss capacitors are not too difficult to manufacture except at audio frequencies and at the 60-Hz utility frequency.

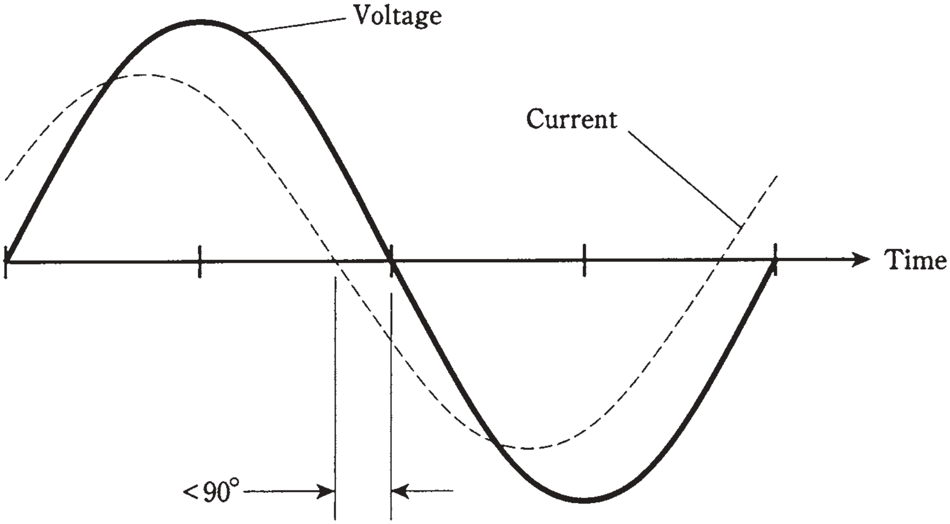
The situation depicted in Fig. 14-8 represents an essentially pure capacitive reactance. The vector in the RC plane points just about straight down. Its angle is -90 degrees from the R axis or zero line.



14-8 In a pure capacitance, the current leads the voltage by 90 degrees.

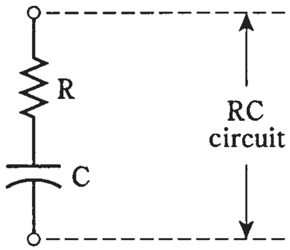
Capacitance and resistance

When the resistance in a resistance-capacitance circuit is significant compared with the capacitive reactance, the current leads the voltage by something less than 90 degrees (Fig. 14-9). If R is small compared with X_C, the difference is almost a quarter of a cycle. As R gets larger, or as X_C becomes smaller, the phase difference gets less. A circuit containing resistance and capacitance is called an *RC circuit.*



14-9 In a circuit with capacitance and resistance, the current leads the voltage by less than 90 degrees.

The value of R in an RC circuit might increase relative to X_C because resistance is deliberately put into a circuit. Or, it might happen because the frequency becomes so low that X_C rises to a value comparable with the leakage resistance of the capacitor. In either case, the situation can be represented by a resistor, R, in series with a capacitor, C (Fig. 14-10).



14-10 Schematic representation of an RC circuit.

If you know the values of X_C and R, you can find the *angle of lead,* also called the *RC phase angle,* by plotting the point $R - jX_C$ on the RC plane, drawing the vector from the origin $0 - j0$ out to that point, and then measuring the angle of the vector, clockwise from the resistance axis. You can use a protractor to measure this angle, as you did in the previous chapter for RL phase angles. Or you can use trigonometry to calculate the angle.

As with RL circuits, you only need to know the ratio of X_C to R to determine the phase angle. For example, if $X_C = -4\ \Omega$ and $R = 7\ \Omega$, you'll get the same angle as with $X_C = -400\ \Omega$ and $R = 700\ \Omega$, or with $X_C = -16\ \Omega$ and $R = 28\ \Omega$. The phase angle will be the same for any ratio of X_C:$R = -4$:7.

Pure resistance

As the resistance in an RC circuit gets large compared with the capacitive reactance, the angle of lead becomes smaller. The same thing happens if the value of X_C gets small compared with the value of R. When you call X_C "large," you mean large *negatively.* When you say that X_C is "small," you mean that it is close to zero, or small *negatively.*

When R is many times larger than X_C, whatever their actual values, the vector in the RC plane will be almost right along the R axis. Then the RC phase angle will be nearly zero, that is, just a little bit negative. The voltage will come nearly into phase with the current. The plates of the capacitor will not come anywhere near getting fully charged with each cycle. The capacitor will be said to "pass the ac" with very little loss, as if it were shorted out. But it will still have an extremely high X_C for any ac signals at much lower frequencies that might exist across it at the same time. (This property of capacitors can be put to use in electronic circuits, for example when an engineer wants to let radio signals get through while blocking audio frequencies.)

Ultimately, if the capacitive reactance gets small enough, the circuit will act as a pure resistance, and the current will be in phase with the voltage.

How much lead?

If you know the ratio of capacitive reactance to resistance, or X_C/R, in an RC circuit, then you can find the phase angle. Of course, you can find this angle of lead if you know the precise values.

Pictorial method

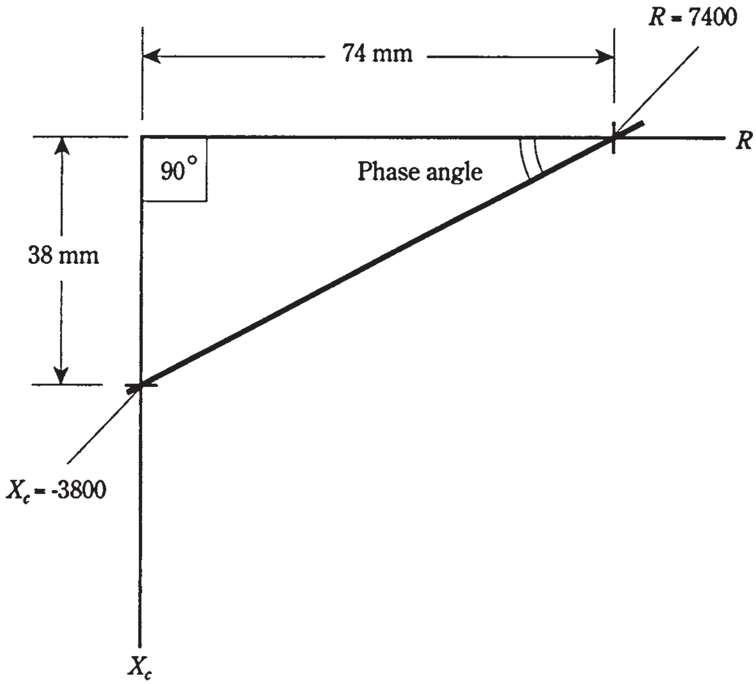
You can use a protractor and a ruler to find phase angles for RC circuits, just as you did with RL circuits, as long as the angles aren't too close to 0 or 90 degrees.

First, draw a line somewhat longer than 10 cm, going from left to right on the paper. Then, use the protractor to construct a line going somewhat more than 10 cm vertically downwards, starting at the left end of the horizontal line.

The horizontal line is the R axis of a crude RC plane. The line going down is the X_C axis.

If you know the values of X_C and R, divide or multiply them by a constant, chosen to make both values fall between -100 and 100. For example, if $X_C = -3800\ \Omega$ and $R = 7400\ \Omega$, divide them both by 100, getting -38 and 74. Plot these points on the lines as hash marks. The X_C mark goes 38 mm down from the intersection point between your two axes (*negative* 38 mm *up* from the intersection point). The R mark goes 74 mm to the right of the intersection point.

Now, draw a line connecting the two hash marks, as shown in Fig. 14-11. This line will be at a slant, and will form a triangle along with the two axes. This is a right triangle, with the right angle at the origin of the RC plane.

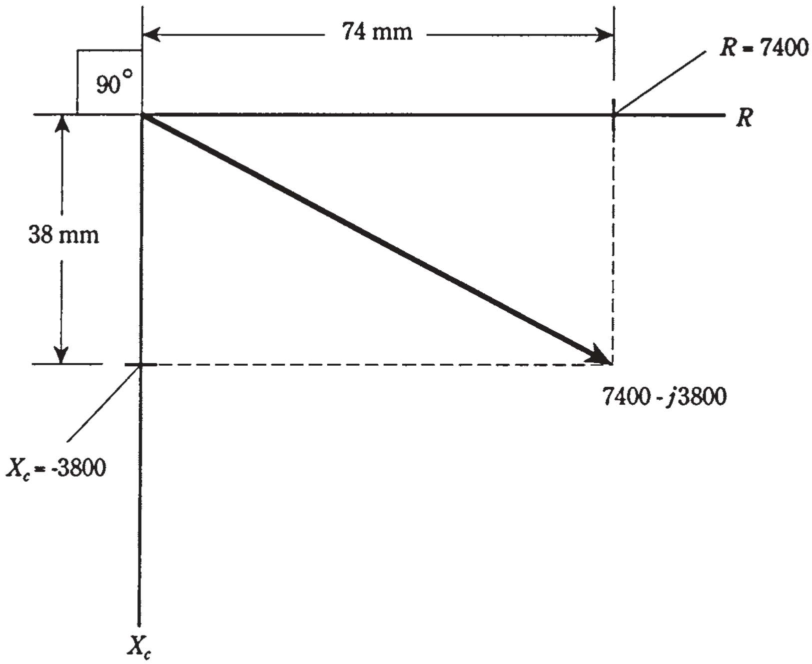


14-11 Pictorial method of finding phase angle.

Measure the angle between the slanted line and the R, or horizontal, axis. Use the protractor for this. Extend the lines, if necessary, using the ruler, to get a good reading on the protractor. This angle will be between 0 and 90 degrees. Multiply this reading by

−1 to get the RC phase angle. That is, if the protractor shows 27 degrees, the RC phase angle is −27 degrees.

The actual vector is found by constructing a rectangle using the origin and your two hash marks, making new perpendicular lines to complete the figure. The vector is the diagonal of this rectangle, running out from the origin (Fig. 14-12). The phase angle is the angle between the R axis and this vector, multiplied by −1. It will have the same measure as the angle of the slanted line you constructed in Fig. 14-11.



14-12 Another pictorial way of finding phase angle. This method shows the actual impedance vector.

Trigonometric method

The more accurate way to find RC phase angles is to use trigonometry. Then you don't have to draw a figure and use a protractor. You only need to punch a few buttons on your calculator. (Just make sure they're the right ones, in the right order.)

If you know the values X_C and R, find the ratio X_C/R. If you know the ratio only, call it X_C/R and enter it into the calculator. Be sure not to make the mistake of getting the ratio upside down (R/X_C). This ratio will be a negative number or zero, because X_C is always negative or zero, and R is always positive. Find the arctangent (arctan or $\tan^{-1}$) of this number. This is the RC phase angle.

Always remember, when doing problems of this kind, to use the *capacitive reactance* for X_C, and not the capacitance. This means that, if you are given the capacitance, you must use the formula for X_C and then calculate the RC phase angle.

Problem 14-4

The capacitive reactance in an RC circuit is $-3800\ \Omega$, and the resistance is $7400\ \Omega$. What is the phase angle?

Find the ratio $X_C/R = -3800/7400$. The calculator win display something like 0.513513513. Find the arctangent, or $\tan^{-1}$, getting a phase angle of 27.18111109 degrees on the calculator display. Round this off to 27.18 degrees.

Problem 14-5

An RC circuit works at a frequency of 3.50 MHz. It has a resistance of 130 Ω and a capacitance of 150 pF. What is the phase angle?

This problem is a little more involved. First, you must find the capacitive reactance for a capacitor of 150 pF. Convert this to microfarads, getting $C = 0.000150\ \mu$F. Remember that *micro*farads go with *mega*hertz (millionths go with minions to cancel each other out). Then

$$\begin{aligned} X_C &= -1/(6.28 \times 3.50 \times 0.000150) \\ &= -1/0.003297 = -303\ \Omega \end{aligned}$$

Now you can find the ratio $X_C/R = -303/130\ -2.33$; the phase angle is arctan (-2.33) $= -66.8$ degrees.

Problem 14-6

What is the phase angle in the above circuit if the frequency is raised to 7.10 MHz?

You need to find the new value for X_C, because it will change as a result of the frequency change. Calculating,

$$\begin{aligned} X_C &= -1/(6.28 \times 7.10 \times 0.000150) \\ &= -1/0.006688 = -150\ \Omega \end{aligned}$$

The ratio $X_c/R = -150/130 = -1.15$; the phase angle is therefore arctan $(-1.15) =$ -49.0 degrees.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. As the size of the plates in a capacitor increases, all other things being equal:
   A. The value of X_C increases negatively.
   B. The value of X_C decreases negatively.
   C. The value of X_C does not change.
   D. You can't say what happens to X_C without more data.

2. If the dielectric material between the plates of a capacitor is changed, all other things being equal:
   A. The value of X_C increases negatively.

B. The value of X_C decreases negatively.

C. The value of X_C does not change.

D. You can't say what happens to X_C without more data.

3. As the frequency of a wave gets lower, all other things being equal, the value of X_C for a capacitor:

A. Increases negatively.

B. Decreases negatively.

C. Does not change.

D. Depends on the current.

4. A capacitor has a value of 330 pF. What is its capacitive reactance at a frequency of 800 kHz?

A. $-1.66\ \Omega$.

B. $-0.00166\ \Omega$.

C. $-603\ \Omega$.

D. $-603\ \mathrm{K}\Omega$.

5. A capacitor has a reactance of $-4.50\ \Omega$ at 377 Hz. What is its capacitance?

A. 9.39 μF.

B. 93.9 μF.

C. 7.42 μF.

D. 74.2 μF.

6. A capacitor has a value of 47 μF. Its reactance is $-47\ \Omega$. What is the frequency?

A. 72 Hz.

B. 7.2 MHz.

C. 0.000072 Hz.

D. 7.2 Hz.

7. A capacitor has $X_C = -8800\ \Omega$ at $f = 830$ kHz. What is C?

A. 2.18 μF.

B. 21.8 pF.

C. 0.00218 μF.

D. 2.18 pF.

8. A capacitor has $C = 166$ pF at $f = 400$ kHz. What is X_C?

A. $-2.4\ \mathrm{K}\ \Omega$.

B. $-2.4\ \Omega$.

C. $-2.4 \times 10^{-6}\ \Omega$.

D. $-2.4\ \mathrm{M}\ \Omega$.

9. A capacitor has $C = 4700\ \mu F$ and $X_C = -33\ \Omega$. What is f?
   A. 1.0 Hz.
   B. 10 Hz.
   C. 1.0 kHz.
   D. 10 kHz.

10. Each point in the RC plane:
    A. Corresponds to a unique inductance.
    B. Corresponds to a unique capacitance.
    C. Corresponds to a unique combination of resistance and capacitance.
    D. Corresponds to a unique combination of resistance and reactance.

11 If R increases in an RC circuit, but X_C is always zero, then the vector in the RC plane will:
    A. Rotate clockwise.
    B. Rotate counterclockwise.
    C. Always point straight towards the right.
    D. Always point straight down.

12. If the resistance R increases in an RC circuit, but the capacitance and the frequency are nonzero and constant, then the vector in the RC plane will:
    A. Get longer and rotate clockwise.
    B. Get longer and rotate counterclockwise.
    C. Get shorter and rotate clockwise.
    D. Get shorter and rotate counterclockwise.

13. Each impedance $R - jX_C$:
    A. Represents a unique combination of resistance and capacitance.
    B. Represents a unique combination of resistance and reactance.
    C. Represents a unique combination of resistance and frequency.
    D. All of the above.

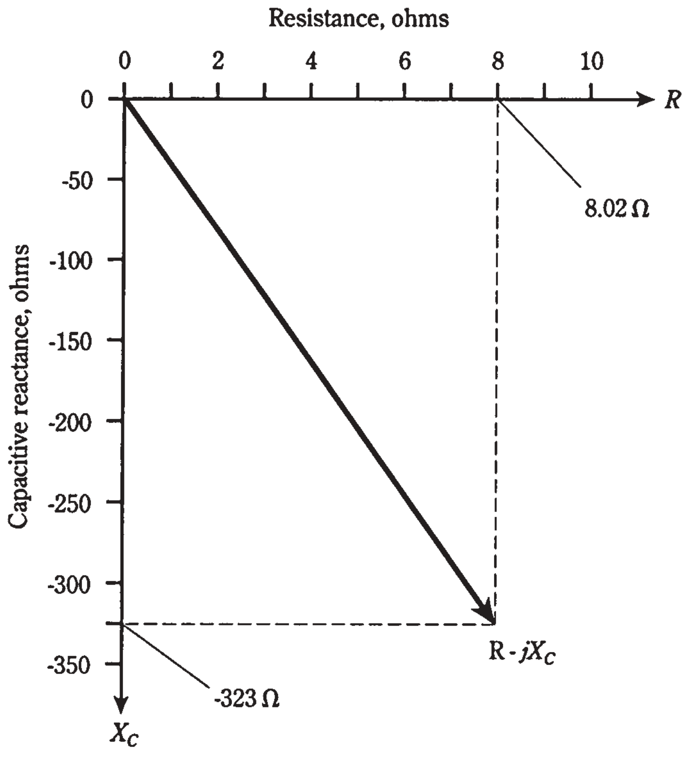
14. In an RC circuit, as the ratio of capacitive reactance to resistance, $-X_C/R$, gets closer to zero, the phase angle:
    A. Gets closer to -90 degrees.
    B. Gets closer to 0 degrees.
    C. Stays the same.
    D. Cannot be found.

15. In a purely resistive circuit, the phase angle is:
    A. Increasing.
    B. Decreasing.

C. 0 degrees.
D. −90 degrees.

16. If the ratio of X_C/R is 1, the phase angle is:
A. 0 degrees.
B. −45 degrees.
C. −90 degrees.
D. Impossible to find; there's not enough data given.

17. In Fig. 14-13, the impedance shown is:
A. $8.02 + j323$.
B. $323 + j8.02$.
C. $8.02 - j323$.
D. $323 - j8.02$.



14-13 Illustration for quiz questions 17 and 18.

18. In Fig. 14-13, note that the R and X_C scale divisions are not the same size. The phase angle is
A. 1.42 degrees.

B. About −60 degress, from the looks of it.

C.−58.9 degrees.

D. −88.6 degrees.

19. An RC circuit consists of a 150-pF capacitor and a 330 Ω resisitor in series. What is the phase angle at a frequency of 1.34 MHz?

A. –67.4 degrees.

B. –22.6 degrees.

C. –24.4 degrees.

D. –65.6 degrees.

20. An RC circuit has a capitance of 0.015 μF. The resistance is 52 Ω. What is the phase angle at 90 kHz?

A. –24 degrees.

B. –0.017 degrees.

C. –66 degrees.

D. None of the above.

15

CHAPTER

Impedance and admittance

YOU'VE SEEN HOW INDUCTIVE AND CAPACITIVE REACTANCE CAN BE REPRESENTED along a line perpendicular to resistance. In this chapter, you'll put all three of these quantities—R, X_L, and X_C—together, forming a complete, working definition of impedance. You'll also get acquainted with *admittance*, impedance's evil twin.

To express the behavior of alternating-current (ac) circuits, you need two dimensions, because ac has variable frequency along with variable current. One dimension (resistance) will suffice for dc, but not for ac.

In this chapter and the two that follow, the presentation is rather mathematical. You can get a grasp of the general nature of the subject matter without learning how to do all of the calculations presented. The mathematics is given for those of you who wish to gain a firm understanding of how ac circuits behave.

Imaginary numbers

What does the lowercase j actually mean in expressions of impedance such as $4 + j7$ and $45 - j83$? This was briefly discussed earlier in this book, but what is this thing, really?

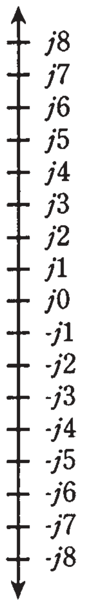
Mathematicians use the lowercase letter i to represent j. (Mathematicians and physicists/engineers often differ in notation as well as in philosophy.) This *imaginary number* is the square root of -1. It is the number that, when multiplied by itself, gives -1. So $i = j$, and $j \times j = -1$.

The entire set of imaginary numbers derives from this single unit. The square of an imaginary number is negative—always. No real *number* has this property. Whether a real number is positive, zero, or negative, its square can never be negative—never.

The notion of j (or i, if you're a mathematician) came about simply because some mathematicians wondered what the square root of -1 would behave like, if there were such a thing. So the mathematicians "imagined" the existence of this animal, and found that it had certain properties. Eventually, the number i was granted a place among the realm of numbers. Mathematically, it's as real as the real numbers. But the original term "imaginary" stuck, so that this number carries with it a mysterious aura.

It's not important, in this context, to debate the reality of the abstract, but to reassure you that imaginary numbers are not particularly special, and are not intended or reserved for just a few eccentric geniuses. "Imaginary" numbers are as real as the "real" ones. And just as unreal, in that neither kind are concrete; you can hold neither type of number in your hand, nor eat them, nor throw them in a wastebasket.

The *unit imaginary number j* can be multiplied by any real number, getting an infinitude of imaginary numbers forming an *imaginary number line* (Fig. 15-1). This is a duplicate of the *real number line* you learned about in school. It must be at a right angle to the real number line when you think of real and imaginary numbers at the same time.



15-1 The imaginary number line.

Complex numbers

When you add a real number and an imaginary number, such as $4 + j7$ or $45 - j83$, you have a *complex number.* This term doesn't mean complicated; it would better be called composite. But again, the original name stuck, even if it wasn't the best possible thing to call it.

Real numbers are one dimensional. They can be depicted on a line. Imaginary numbers are also one dimensional for the same reason. But complex numbers need two dimensions to be completely defined.

Adding and subtracting complex numbers

Adding complex numbers is just a matter of adding the real parts and the complex parts separately. The sum of $4 + j7$ and $45 - j83$ is therefore $(4 + 45) + j(7 - 83) = 49 + j(-76) = 49 - j76$.

Subtracting complex numbers works similarly. The difference $(4 + j7) - (45 - j83)$ is found by multiplying the second complex number by -1 and then adding the result, getting $(4 + j7) + (-1(45 - j83) = (4 + j7) + (-45 + j83) = -41 + j90$.

The general formula for the sum of two complex numbers $(a + jb)$ and $(c + jd)$ is

$$(a + jb) + (c + jd) = (a + c) + j(b + d)$$

The plus and minus number signs get tricky when working with sums and differences of complex numbers. Just remember that any difference can be treated as a sum: multiply the second number by -1 and then add. You might want to do some exercises to get yourself acquainted with the way these numbers behave, but in working with engineers, you will not often be called upon to wrestle with complex numbers at the level of "nitty-gritty."

If you plan to become an engineer, you'll need to practice adding and subtracting complex numbers. But it's not difficult once you get used to it by doing a few sample problems.

Multiplying complex numbers

You should know how complex numbers are multiplied, to have a full understanding of their behavior. When you multiply these numbers, you only need to treat them as sums of number pairs, that is, as binomials.

It's easier to give the general formula than to work with specifics here. The product of $(a + jb)$ and $(c + jd)$ is equal to $ac + jad + jbc + jjbd$. Simplifying, remember that $jj = -1$, so you get the final formula:

$$(a + jb)(c + jd) = (ac - bd) + j(ad + bc)$$

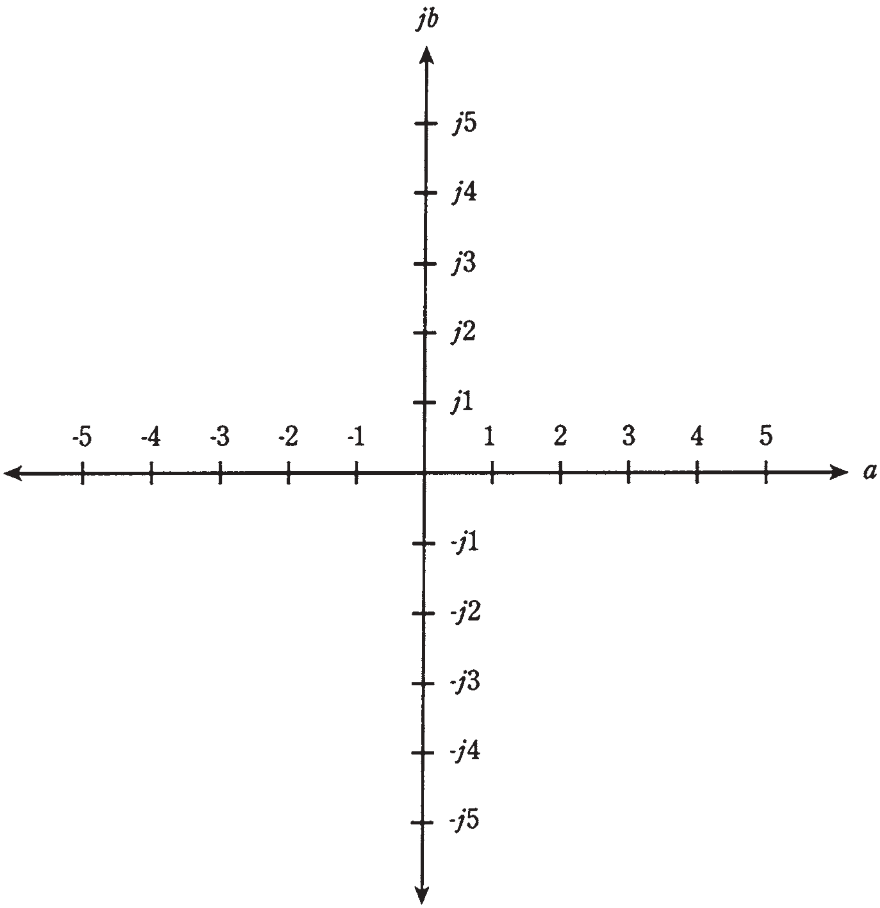
As with the addition and subtraction of complex numbers, you must be careful with signs (plus and minus). And also, as with addition and subtraction, you can get used to doing these problems with a little practice. Engineers sometimes (but not too often) have to multiply complex numbers.

The complex number plane

Real and imaginary numbers can be thought of as points on a line. Complex numbers lend themselves to the notion of points on a plane. This plane is made by taking the real and imaginary number lines and placing them together, at right angles, so that they intersect at the zero points, 0 and $j0$. This is shown in Fig. 15-2. The result is a *Cartesian* coordinate plane, just like the ones you use to make graphs of everyday things like bank-account balance versus time.

Notational neuroses

On this plane, a complex number might be represented as $a + jb$ (in engineering or physicists' notation), or as $a + bi$ (in mathematicians' notation), or as an *ordered pair* (a, b). "Wait," you ask. "Is there a misprint here? Why does b go after the j, but in front of the i?" The answer is as follows: Mathematicians and engineers/physicists just don't think alike, and this is but one of myriad ways in which this is apparent. In other words, it's a matter of



15-2 The complex number plane.

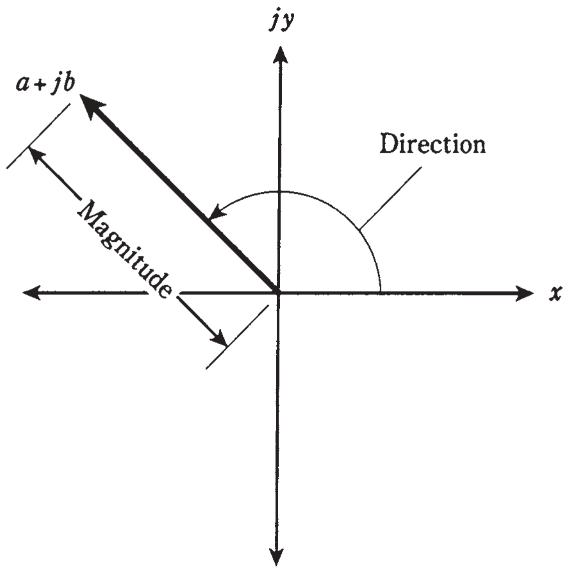
notational convention, and that is all. (It's also a somewhat humorous illustration of the different angle that an engineer takes in approaching a problem, as opposed to a mathematician.)

Complex number vectors

Complex numbers can also be represented as vectors in the complex plane. This gives each complex number a unique magnitude and direction. The magnitude is the distance of the point $a + jb$ from the origin $0 + j0$. The direction is the angle of the vector, measured counterclockwise from the $+a$ axis. This is shown in Fig. 15-3.

Absolute value

The *absolute value* of a complex number $a + jb$ is the length, or magnitude, of its vector in the complex plane, measured from the origin (0,0) to the point (a,b).

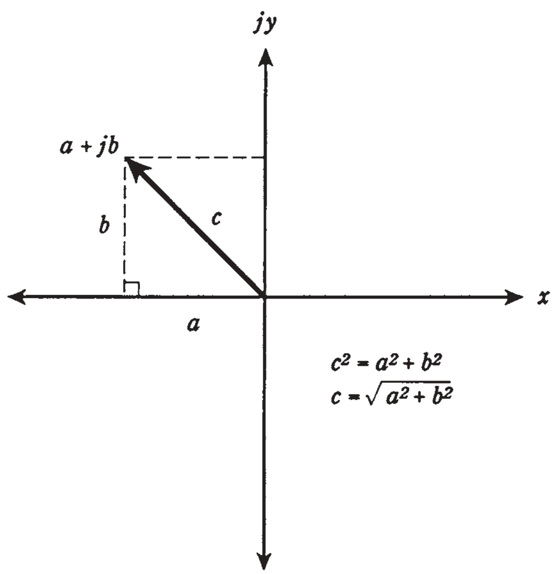


15-3 Magnitude and direction of a vector in the complex number plane.

In the case of a *pure real* number $a + j0$, the absolute value is simply the number itself, a, if it is positive, and $-a$ if a is negative.

In the case of a *pure imaginary* number $0 + jb$, the absolute value is equal to b if b (which is a real number) is positive, and $-b$ if b is negative.

If the number is neither pure real or pure imaginary, the absolute value must be found by using a formula. First, square both a and b. Then add them. Finally, take the square root. This is the length of the vector $a + jb$. The situation is illustrated in Fig. 15-4.



15-4 Calculation of absolute value, or vector length.

Problem 15-1

Find the absolute value of the complex number $-22 - j0$.

Note that this is a pure real. Actually, it is the same as $-22 + j0$, because $j0 = 0$. Therefore, the absolute value of this complex number is $-(-22) = 22$.

Problem 15-2

Find the absolute value of $0 - j34$.

This is a pure imaginary number. The value of b in this case is -34, because $0 - j34 = 0 + j(-34)$. Therefore, the absolute value is $-(-34) = 34$.

Problem 15-3

Find the absolute value of $3 - j4$.

In this number, $a = 3$ and $b = -4$, because $3 - j4$ can be rewritten as $3 + j(-4)$. Squaring both of these, and adding the results, gives $3^2 + (-4)^2 = 9 + 16 = 25$. The square root of 25 is 5; therefore, the absolute value of this complex number is 5.

You might notice this "3, 4, 5" relationship and recall the Pythagorean theorem for finding the length of the *hypotenuse* of a right triangle. The formula for finding the length of a vector in the complex-number plane comes directly from this theorem.

If you don't remember the Pythagorean theorem, don't worry; just remember the formula for the length of a vector.

The RX plane

Recall the planes for resistance (R) and inductive reactance (X_L) from chapter 13. This is the same as the upper-right quadrant of the complex-number plane shown in Fig. 15-2.

Similarly, the plane for resistance and capacitive reactance (X_C) is the same as the lower-right quadrant of the complex number plane.

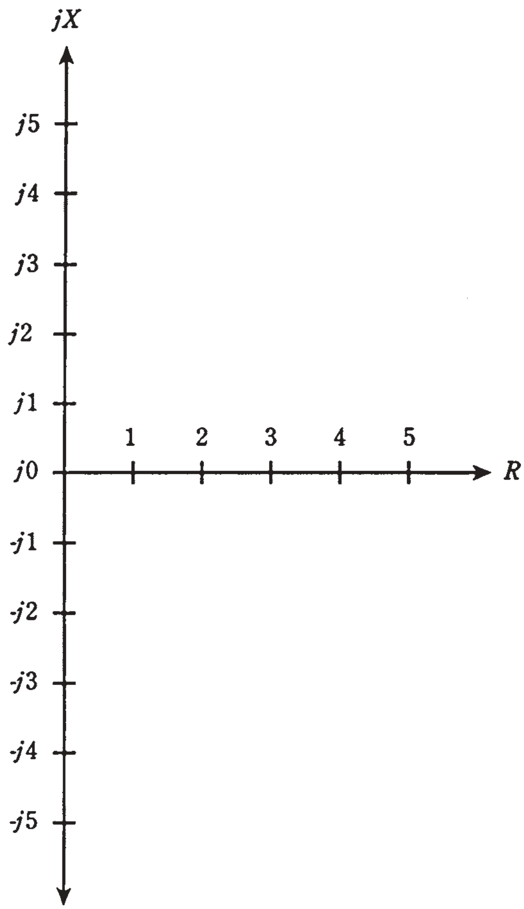
Resistances are represented by nonnegative real numbers. Reactances, whether they are inductive (positive) or capacitive (negative), correspond to imaginary numbers.

No negative resistance

There is no such thing, strictly speaking, as negative resistance. That is to say, one cannot have anything better than a perfect conductor. In some cases, a supply of direct current, such as a battery, can be treated as a negative resistance; in other cases, you can have a device that acts as if its resistance were negative under certain changing conditions. But generally, in the *RX (resistance-reactance) plane,* the resistance value is always positive. This means that you can remove the negative axis, along with the upper-left and lower-left quadrants, of the complex-number plane, obtaining a half plane as shown in Fig. 15-5.

Reactance in general

Now you should get a better idea of why capacitive reactance, X_C, is considered negative. In a sense, it is an extension of inductive reactance, X_L, into the realm of negatives, in a



15-5 The complex impedance RX plane.

way that cannot generally occur with resistance. Capacitors act like "negative inductors." Interesting things happen when capacitors and inductors are combined, which is discussed in the next couple of chapters.

Reactance can vary from extremely large negative values, through zero, to extremely large positive values. Engineers and physicists always consider reactance to be imaginary. In the mathematical model of impedance, capacitances and inductances manifest themselves "perpendicularly" to resistance.

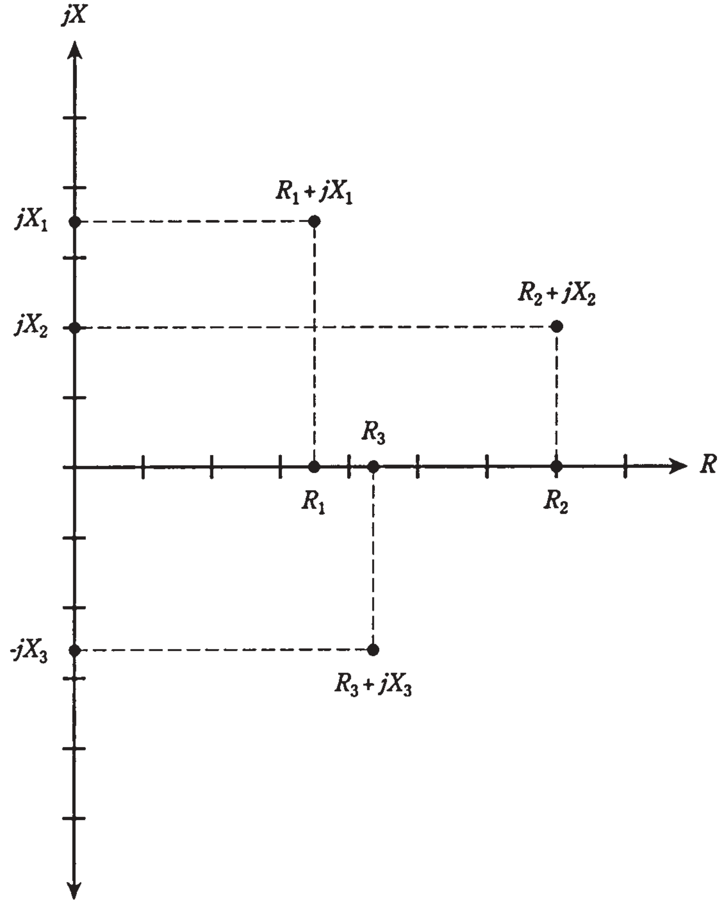
The general symbol for reactance is X; this encompasses both inductive reactance (X_L) and capacitive reactance (X_C).

Vector representation of impedance

Any impedance $R + jX$ can be represented by a complex number of the form $a + jb$. Just let $R = a$ and $X = b$.

It should be easy to visualize, now, how the impedance vector changes as either R *or* X, or both, are varied. If X remains constant, an increase in R will cause the vector to get longer. If R remains constant and X_L gets larger, the vector will also grow longer. If R stays the same as X_C gets larger (negatively), the vector will grow longer yet again.

Think of the point $a + jb$, or $R + jX$, moving around in the plane, and imagine where the corresponding points on the axes lie. These points are found by drawing dotted lines from the point $R + jX$ to the R and X axes, so that the lines intersect the axes at right angles (Fig. 15-6).



15-6 Some points in the RX plane, and their components on the R and X axes.

Now think of the points for R and X moving toward the right and left, or up and down, on their axes. Imagine what happens to the point $R + jX$ in various scenarios. This is how impedance changes as the resistance and reactance in a circuit are varied.

Resistance is one-dimensional. Reactance is also one-dimensional. But impedance is two-dimensional. To fully define impedance, you must render it on a half plane, specifying the resistance and the reactance, which are independent.

Absolute-value impedance

There will be times when you'll hear that the "impedance" of some device or component is a certain number of ohms. For example, in audio electronics, there are "8-Ω" speakers and "600-Ω" amplifier inputs. How can manufacturers quote a single number for a quantity that is two-dimensional, and needs two numbers to be completely expressed?

There are two answers to this. First, figures like this generally refer to devices that have purely resistive impedances. Thus, the "8-Ω" speaker really has a complex impedance of $8 + j0$, and the "600-Ω" input circuit is designed to operate with a complex impedance at, or near, $600 + j0$.

Second, you can sometimes talk about the length of the impedance vector, calling this a certain number of ohms. If you talk about "impedance" this way, you are being ambiguous, because you can have an infinite number of different vectors of a given length in the RX plane.

Sometimes, the capital letter Z is used in place of the word "impedance" in general discussions. This is what engineers mean when they say things like "$Z = 50\ \Omega$" or "$Z = 300\ \Omega$ nonreactive."

"$Z = 8\ \Omega$" in this context, if no specific complex impedance is given, can refer to the complex value $8 + j0$, or $0 + j8$, or $0 = j8$, or any value on a half circle of points in the RX plane that are at distance 8 units away from $0 + j0$. This is shown in Fig. 15-7. There exist an infinite number of different complex impedances with $Z = 8\ \Omega$.

Problems 15-1, 15-2, and 15-3 can be considered as problems in finding absolute-value impedance from complex impedance numbers.

Problem 15-4

Name seven different complex impedances having an absolute value of $Z = 10$.

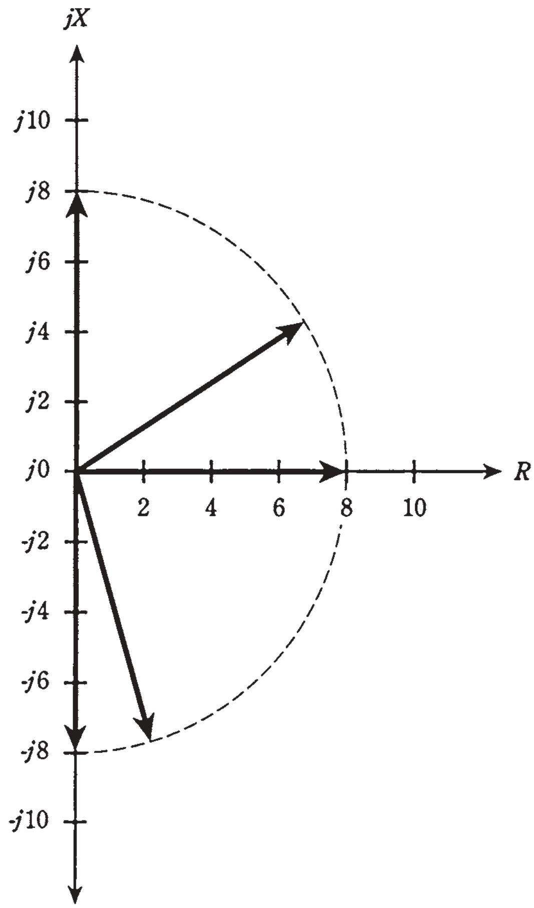
It's easy name three: $0 + j10$, $10 + j0$, and $0 - j10$. These are pure inductance, pure resistance, and pure capacitance, respectively.

A right triangle can exist having sides in a ratio of 6:8:10 units. This is true because $6^2 + 8^2 = 10^2$. (Check it and see!) Therefore, you might have $6 + j8$, $6 - j8$, $8 + j6$ and $8 - j6$, all complex impedances whose absolute value is 10 ohms. Obviously, the value $Z = 10$ was chosen for this problem because such a whole-number right-triangle exists. It becomes quite a lot messier to do this problem (but by no means impossible) if $Z = 11$ instead.

If you're not specifically told what complex impedance is meant when a single-number ohmic figure is quoted, it's best to assume that the engineers are talking about *nonreactive* impedances. That means they are pure resistances, and that the imaginary, or reactive, factor is zero. Engineers will often speak of nonreactive impedances, or of complex impedance vectors, as "low-Z or high-Z." For instance, a speaker might be called "low-Z" and a microphone "high-Z."

Characteristic impedance

There is one property of electronic components that you'll sometimes hear called *impedance,* that really isn't "impedance" at all. This is *characteristic impedance* or *surge*



15-7 Vectors representing an absolute value impedance of 8 Ω.

impedance. It is abbreviated Z_o, and is a specification of *transmission lines*. It can always be expressed as a positive real number.

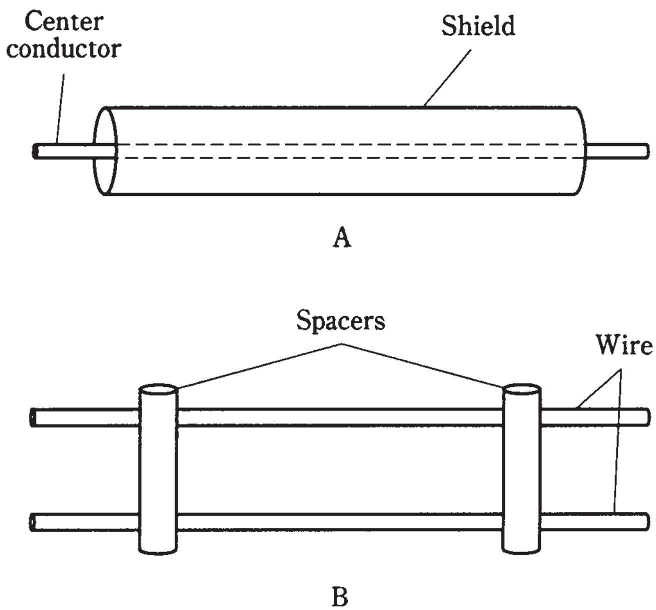
Transmission lines

Any time that it is necessary to get energy or signals from one place to another, a transmission line is required. These almost always take either of two forms, *coaxial* or *two wire*. These are illustrated qualitatively in Fig. 15-8.

Examples of transmission lines include the ribbon that goes from a television antenna to the receiver, the cable running from a hi-fi amplifier to the speakers, and the set of wires that carries electricity over the countryside.

Factors affecting Z_o

The Z_{oh} of a parallel-wire transmission line depends on the diameter of the wires, on the spacing between the wires, and on the nature of the insulating material separating the wires.



15-8 At A, coaxial transmission line. At B, parallel-wire transmission line.

In general, Z_0 increases as the wire diameter gets smaller, and decreases as the wire diameter gets larger, all other things being equal.

In a coaxial line, the thicker the center conductor, the lower the Z_0 if the shield stays the same size. If the center conductor stays the same size and the shield tubing increases in diameter, the Z_0 will increase.

Also in general, Z_0 increases as the spacing between wires, or between the center conductor and the *shield* or *braid,* gets greater, and decreases as the spacing is made less.

Solid dielectrics such as polyethylene reduce the Z_0 of a transmission line, compared with air or a vacuum between the conductors.

Z_0 in practice

In rigorous terms, the characteristic impedance of a line is determined according to the nature of the *load* with which the line works at highest efficiency.

Suppose that you have an 8-Ω hi-fi speaker, and you want to get audio energy to that speaker with the greatest possible efficiency, so that the least possible power is dissipated in the line. You would use large-diameter wires, of course, but for true optimization, you would want the spacing between the wires to be just right. Adjusting this spacing for optimum power transfer would result in a line Z_0 of 8 Ω. Then, the greatest possible efficiency would be had with a speaker of impedance $8 + j0$.

If you can't get the wires to have the right size and spacing for a good match to Z = 8 Ω, you might need to use an *impedance transformer.* This makes the speaker's impedance look like something different, such as 50 Ω or 600 Ω.

Imagine that you have a 300-Ω television antenna, and you want the best possible reception. You purchase "300-Ω" ribbon line, with a value of Z_o that has been optimized by the manufacturer for use with antennas whose impedances are close to $300 + j0$.

For a system having an "impedance" of "R Ω," the best line Z_o is also R Ω. If R is much different from Z_o, an unnecessary amount of power will be wasted in heating up the transmission line. This might not be a significant amount of power, but it often is.

Impedance matching will be discussed in more detail in the next chapter.

Conductance

In an ac circuit, electrical *conductance* works the same way as it does in a dc circuit. Conductance is symbolized by the capital letter G. It was introduced back in chapter 2.

The relationship between conductance and resistance is simple: $G = 1/R$. The unit is the siemens. The larger the value of conductance, the smaller the resistance, and the more current will flow. Conversely, the smaller the value of G, the greater the value of R, and the less current will flow.

Susceptance

Sometimes, you'll come across the term *susceptance* in reference to an ac circuit containing a capacitive reactance or an inductive reactance. Susceptance is symbolized by the capital letter B. It is the reciprocal of reactance. That is, $B = 1/X$. Susceptance can be either capacitive or inductive. These are symbolized as B_C and B_L respectively. Therefore, $B_C = 1/X_C$, and $B_L = 1/X_L$.

There is a trick to determining susceptances in terms of reactances. Or, perhaps better stated, a trickiness. Susceptance is imaginary, just as is reactance. That is, all values of B require the use of the j operator, just as do all values of X. But $1/j = -j$. This reverses the sign when you find susceptance in terms of reactance.

If you have an inductive reactance of, say, 2 ohms, then this is expressed as $j2$ in the imaginary sense. What is $1/(j2)$? You can break this apart and say that $1/(j2) = (1/j)(1/2) = (1/j)0.5$. But what is $1/j$? Without making this into a mathematical treatise, suffice it to say that $1/j = -j$. Therefore, the reciprocal of $j2$ is $-j0.5$. *Inductive susceptance* is *negative imaginary*.

If you have a capacitive reactance $X_C = 10$ ohms, then this is expressed as $X_C = -j10$. The reciprocal of this is $B_C = 1/(-j10) = (1/-j)(1/10) = (1/-j)0.1$. What is $1/-j$? Again, without going into deep theoretical math, it is equal to j. Therefore, the reciprocal of $-j10$ is $j0.1$. *Capacitive susceptance* is *positive imaginary*.

This is exactly reversed from the situation with reactances.

Problem 15-5

Suppose you have a capacitor of 100 pF at a frequency of 3.00 MHz. What is B_C?

First, find the reactance X_C by the formula

$$X_C = -1/(6.28fC)$$

Remembering that 100 pF = 0. 000100 µF, you can substitute in this formula for f = 3.00 and C = 0.000100, getting

$$\begin{aligned} X_C &= -1/(6.28 \times 3.00 \times 0.000100) \\ &= -1/0.001884 = -531\ \Omega = -j531 \end{aligned}$$

The susceptance, B_C, is equal to $1/X_C$. Thus, $B_C = 1/(-j531) = j0.00188$. Remember that capacitive susceptance is positive. This can "short-circuit" any frustration you might have in manipulating the minus signs in these calculations.

Note that above, you found a reciprocal of a reciprocal. You did something and then immediately turned around and undid it, slipping a minus sign in because of the idiosyncrasies of that little j operator. In the future, you can save work by remembering that the formula for capacitive susceptance simplified, is

$$B_C = 6.28fC \text{ siemens} = j(6.28fC)$$

This resembles the formula for inductive reactance.

Problem 15-6

An inductor has L = 163 µH at a frequency of 887 kHz. What is B_L?

First, calculate X_L, the inductive reactance

$$\begin{aligned} X_L &= 6.28fL = 6.28 \times 0.887 \times 163 \\ &= 908\ \Omega = j908 \end{aligned}$$

The susceptance, B_L is equal to $1/X_L$ Therefore, $B_L = -1/j908 = -j0.\ 00110$. Remember that inductive susceptance is negative.

The formula for inductive susceptance is similar to that for capacitive reactance:

$$B_L = -1/(6.28fl) \text{ siemens} = -j(1/(6.28fL)$$

Admittance

Conductance and susceptance combine to form *admittance,* symbolized by the capital letter Y.

Admittance, in an ac circuit, is analogous to conductance in a dc circuit.

Complex admittance

Admittance is a complex quantity and represents the ease with which current can flow in an ac circuit. As the absolute value of impedance gets larger, the absolute value of admittance becomes smaller, in general. Huge impedances correspond to tiny admittances, and vice-versa.

Admittances are written in complex form just like impedances. But you need to keep track of which quantity you're talking about! This will be obvious if you use the symbol, such as $Y = 3 - j0.5$ or $Y = 7 + j3$. When you see Y instead of Z, you know that negative j factors (such as $-j0.5$) mean that there is a net inductance in the circuit, and positive j factors (such as $+j3$) mean there is net capacitance.

Admittance is the complex composite of conductance and susceptance. Thus, admittance takes the form

$$Y = G + jB$$

The j factor might be negative, of course, so there are times you'll write $Y = G - jB$.

Parallel circuits

Recall how resistances combine with reactances in series to form complex impedances? In chapters 13 and 14, you saw series RL and RC circuits. Perhaps you wondered why parallel circuits were ignored in those discussions. The reason is that admittance, rather than impedance, is best for working with parallel ac circuits. Therefore, the subject of parallel circuits was deferred.

Resistance and reactance combine in rather messy fashion in parallel circuits, and it can be hard to envision what's happening. But conductance (G) and susceptance (B) just add together in parallel circuits, yielding admittance (Y). This greatly simplifies the analysis of parallel ac circuits.

The situation is similar to the behavior of resistances in parallel when you work with dc. While the formula is a bit cumbersome if you need to find the value of a bunch of resistances in parallel, it's simple to just add the conductances.

Now, with ac, you're working in two dimensions instead of one. That's the only difference.

Parallel circuit analysis is covered in detail in the next chapter.

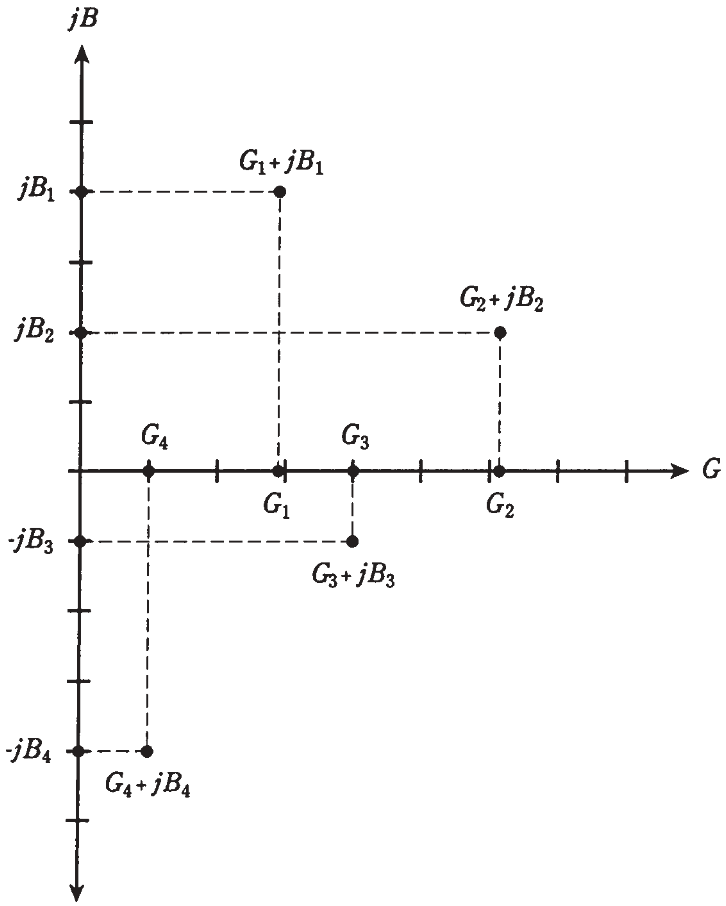
The GB plane

Admittance can be depicted on a plane that looks just like the complex impedance (RX) plane. Actually, it's a half plane, because there is ordinarily no such thing as negative conductance. (You can't have something that conducts worse than not at all.) Conductance is plotted along the horizontal, or G, axis on this coordinate half plane, and susceptance is plotted along the B axis. The plane is shown in Fig. 15-9 with several points plotted.

Although the *GB plane* looks superficially identical to the RX plane, the difference is great indeed! The GB plane is literally blown inside-out from the RX plane, as if you had jumped into a black hole and undergone a spatial transmutation, inwards out and outwards in, turning zero into infinity and vice-versa. Mathematicians love this kind of stuff.

The center, or origin, of the GB plane represents that point at which there is no conduction of any kind whatsoever, either for direct current or for alternating current. In the RX plane, the origin represents a perfect *short circuit;* in the GB plane it corresponds to a perfect *open circuit.*

The open circuit in the RX plane is way out beyond sight, infinitely far away from the origin. In the GB plane, it is the short circuit that is out of view.



15-9 Some points in the GB plane, and their components on the *G* and *B* axes.

Formula for conductance

As you move out towards the right ("east") along the *G*, or conductance, axis of the GB plane, the conductance improves, and the current gets greater, but only for dc. The formula for *G* is simply

$$G = 1/R$$

where *R* is the resistance in ohms and *G* is the conductance in siemens, also sometimes called mhos.

Formula for capacitive susceptance

It won't hurt to review the formulas for susceptance again. They can get a little bit confusing, especially after having worked with reactance.

When you move upwards ("north") along the jB axis from the origin, you have ever-increasing capacitive susceptance. The formula for this quantity, B_C, is

$$B_C = 6.28fL \text{ siemens}$$

where f is in Hertz and C is in farads. The value of B is in siemens. Alternatively, you can use frequency values in megahertz and capacitances in microfarads. The complex value is $jB = j(6.28fC)$.

Moving upwards along the jB axis indicates increasing capacitance values.

Formula for inductive susceptance

When you go down ("south") along the jB axis from the origin, you encounter increasingly negative susceptance. This is inductive susceptance; the formula for it is

$$B_L = -1/(6.28fL) \text{ siemens}$$

where f is in Hertz and L is in henrys. Alternatively, f can be expressed in megahertz, and L can be given in microhenrys. The complex value is $jB = -j(1/(6.28fL)$.

Moving downwards along the jB axis indicates decreasing values of inductance.

Vector representation of admittance

Complex admittances can be shown as vectors, just as can complex impedances. In Fig. 15-10, the points from Fig. 15-9 are rendered as vectors.

Generally, longer vectors indicate greater flow of current, and shorter ones indicate less current.

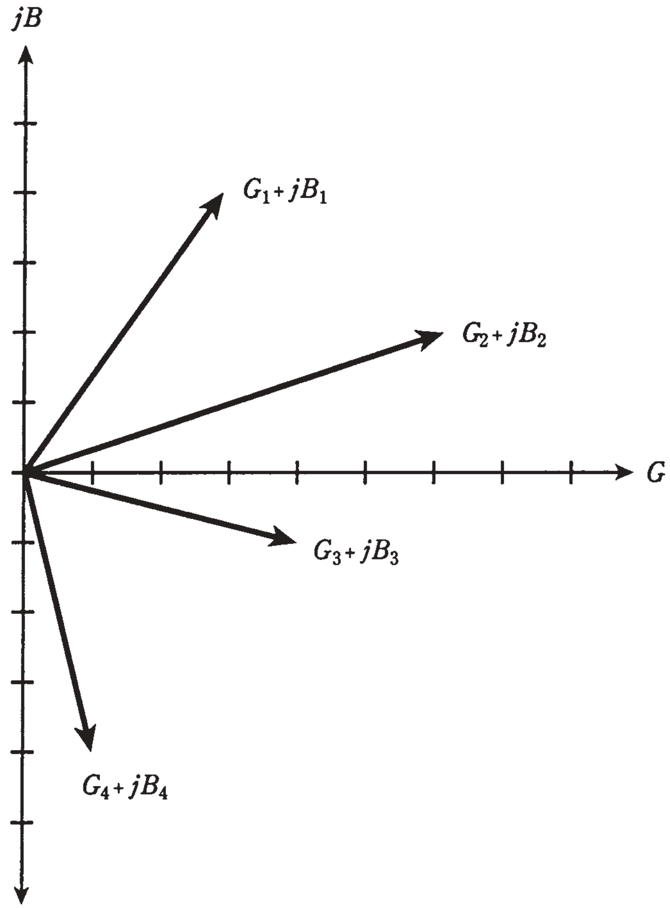
Imagine a point moving around on the GB plane, and think of the vector getting longer and shorter, and changing direction. Vectors pointing generally "northeast," or upwards and to the right, correspond to conductances and capacitances in parallel. Vectors pointing in a more or less "southeasterly" direction, or downwards and to the right, are conductances and inductances in parallel.

Why all these different expressions?

Do you think that the foregoing discussions are an elaborate mental gymnastics routine? Why do you need all these different quantities: resistance, capacitance, capacitive reactance, inductance, inductive reactance, impedance, conductance, capacitive susceptance, inductive susceptance, admittance?

Well, gymnastics are sometimes necessary to develop skill. Sometimes you need to "break a mental sweat." Each of these expressions is important.

The quantities that were dealt with before this chapter, and also early in this chapter, are of use mainly with series RLC (resistance-inductance-capacitance) circuits. The ones introduced in the second half of this chapter are important when you need to analyze parallel RLC circuits. Practice them and play with them, especially if they intimidate you. After awhile they'll become familiar.



15-10 Vectors representing the points of Fig. 15-9.

Think in two dimensions. Draw your own RX and GB planes. (Be thankful there are only two dimensions, and not three! Some scientists need to deal in dozens of dimensions.)

If you want to be an engineer, you'll need to know how to handle these expressions. If you plan to manage engineers, you'll want to know what these quantities are, at least, when the engineers talk about them.

If the math seems a bit thick right now, hang in there. Impedance and admittance are the most mathematical subjects you'll have to deal with.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 or more correct. Answers are in the back of the book.

1. The square of an imaginary number:
   A. Can never be negative.

B. Can never be positive.

C. Might be either positive or negative.

D. Is equal to j.

2. A complex number:

A. Is the same thing as an imaginary number.

B. Has a real part and an imaginary part.

C. Is one-dimensional.

D. Is a concept reserved for elite imaginations.

3. What is the sum of $3 + j7$ and $-3 - j7$?

A. $0 + j0$

B. $6 + j14$.

C. $-6 - j14$.

D. $0 - j14$.

4. What is $(-5 + j7) - (4 - j5)$?

A. $-1 + j2$.

B. $-9 - j2$.

C. $-1 - j2$.

D. $-9 + j12$.

5. What is the product $(-4 - j7)(6 - j2)$?

A. $24 - j14$.

B. $-38 - j34$.

C. $-24 - j14$.

D. $-24 + j14$.

6. What is the magnitude of the vector $18 - j24$?

A. 42.

B. -42.

C. 30.

D. -30.

7. The impedance vector $5 + j0$ represents:

A. A pure resistance.

B. A pure inductance.

C. A pure capacitance.

D. An inductance combined with a capacitance.

8. The impedance vector $0 - j22$ represents:

A. A pure resistance.

B. A pure inductance.

C. A pure capacitance.

D. An inductance combined with a resistance.

9. What is the absolute-value impedance of $3.0 - j6.0$?

A. $Z = 9.0\ \Omega$.

B. $Z = 3.0\ \Omega$.

C. $Z = 45\ \Omega$.

D. $Z = 6.7\ \Omega$.

10. What is the absolute-value impedance of $50 - j235$?

A. $Z = 240\ \Omega$.

B. $Z = 58{,}000\ \Omega$.

C. $Z = 285\ \Omega$.

D. $Z = -185\ \Omega$.

11. If the center conductor of a coaxial cable is made to have smaller diameter, all other things being equal, what will happen to the Z_0 of the transmission line?

A. It will increase.

B. It will decrease.

C. It will stay the same.

D. There is no way to know.

12. If a device is said to have an impedance of $Z = 100\ \Omega$, this would most often mean that:

A. $R + j\mathrm{X} = 100 + j0$.

B. $R + j\mathrm{X} = 0 + j100$.

C. $R + j\mathrm{X} = 0 - j100$.

D. You need to know more specific information.

13. A capacitor has a value of 0.050 µF at 665 kHz. What is the capacitive susceptance?

A. $j4.79$.

B. $-j4.79$.

C. $j0.\ 209$.

D. $-j0.\ 209$.

14. An inductor has a value of 44 mH at 60 Hz. What is the inductive susceptance?

A.. $-j0.060$.

B. $j0.060$.

C. $-j17$.

D. $j17$.

15. Susceptance and conductance add to form:
    A. Impedance.
    B. Inductance.
    C. Reactance.
    D. Admittance.
16. Absolute-value impedance is equal to the square root of:
    A. $G^2 + B^2$
    B. $R^2 + X^2$.
    C. Z_o.
    D. Y.
17. Inductive susceptance is measured in:
    A. Ohms.
    B. Henrys.
    C. Farads.
    D. Siemens.
18. Capacitive susceptance is:
    A. Positive and real valued.
    B. Negative and real valued.
    C. Positive and imaginary.
    D. Negative and imaginary.
19. Which of the following is false?
    A. $B_C = 1/X_C$.
    B. Complex impedance can be depicted as a vector.
    C. Characteristic impedance is complex.
    D. $G = 1/R$.
20. In general, the greater the absolute value of the impedance in a circuit:
    A. The greater the flow of alternating current.
    B. The less the flow of alternating current.
    C. The larger the reactance.
    D. The larger the resistance.

16
CHAPTER
RLC circuit analysis

WHENEVER YOU SEE AC CIRCUITS WITH INDUCTANCE AND/OR CAPACITANCE AS well as resistance, you should switch your mind into "2D" mode. You must be ready to deal with two-dimensional quantities.

While you can sometimes talk and think about impedances as simple ohmic values, there are times you can't. If you're sure that there is no reactance in an ac circuit, then it's all right to say "Z = 600 ohms, " or "This speaker is 8 ohms," or "The input impedance to this amplifier is 1,000 ohms."

As soon as you see coils and/or capacitors, you should envision the complex-number plane, either RX (resistance-reactance) or GB (conductance-admittance). The RX plane applies to series-circuit analysis. The GB plane applies to parallel-circuit analysis.

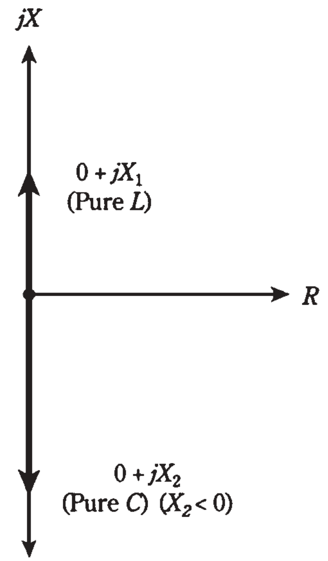
Complex impedances in series

When you see resistors, coils, and capacitors in series, you should envision the RX plane.

Each component, whether it is a resistor, an inductor, or a capacitor, has an impedance that can be represented as a vector in the RX plane. The vectors for resistors are constant regardless of the frequency. But the vectors for coils and capacitors vary with frequency, as you have learned.

Pure reactances

Pure inductive reactances (X_L) and capacitive reactances (X_C) simply add together when coils and capacitors are in series. Thus, $X = X_L + X_C$. In the RX plane, their vectors add, but because these vectors point in exactly opposite directions— inductive reactance upwards and capacitive reactance downwards—the resultant sum vector will also inevitably point either straight up or down (Fig. 16-1).



16-1 Pure inductance and pure capacitance are represented by reactance vectors that point straight up and down.

Problem 16-1

A coil and capacitor are connected in series, with $jX_L = j200$ and $jX_C = -j150$. What is the net reactance vector jX?

Just add the values $jX = jX_L + jX_C = j200 + (-j150) = j(200 - 150) = j50$. This is an *inductive* reactance, because it is *positive imaginary*.

Problem 16-2

A coil and capacitor are connected in series, with $jX_L = j30$ and $jX_C = -j110$. What is the net reactance vector jX?

Again, add $jX = j30 + (-j110) = j(30 - 110) = -j80$. This is a *capacitive* reactance, because it is *negative imaginary*.

Problem 16-3

A coil of $L = 5.00$ µH and a capacitor of $C = 200$ pF are in series. The frequency is $f = 4.00$ MHz. What is the net reactance vector jX?

First calculate

$$\begin{aligned} jX_L &= j6.28fL \\ &= j(6.28 \times 4.00 \times 5.00) = j126 \end{aligned}$$

Then calculate

$$\begin{aligned} jX_C &= -j(1/(6.28fC)) \\ &= -j(1/(6.28 \times 4.00 \times 0.000200) = -j199 \end{aligned}$$

Finally, add

$$\begin{aligned} jX &= jX_L + jX_C \\ &= j126 + (-j199) = -j73 \end{aligned}$$

This is a net capacitive reactance. There is no resistance in this circuit, so the impedance vector is $0 - j73$.

Problem 16-4

What is the net reactance vector jX for the above combination at a frequency of $f = 10.0$ MHz?

First calculate

$$\begin{aligned} jX_L &= j6.28fL \\ &= j(6.28 \times 10.0 \times 5.00) = j314 \end{aligned}$$

Then calculate

$$\begin{aligned} jX_C &= -j(1/(6.28fC)) \\ &= -j(1/(6.28 \times 10.0 \times 0.000200) = -j79.6 \end{aligned}$$

Finally, add

$$\begin{aligned} jX &= jX_L + jX_C \\ &= j314 + (-j79.6) = j234 \end{aligned}$$

This is a net inductive reactance. Again, there is no resistance, and therefore the impedance vector is pure imaginary, $0 + j234$.

Notice that the change in frequency, between Problems 16-3 and 16-4, caused the circuit to change over from a net capacitance to a net inductance. You might think that there must be some frequency, between 4.00 MHz and 10.0 MHz, at which jX_L and jX_C add up to $j0$—that is, at which they exactly cancel each other out, yielding $0 + j0$ as the complex impedance. Then the circuit, at that frequency, would appear as a short circuit. If, you suspect this, you're right.

Any series combination of coil and capacitor offers theoretically zero opposition to ac at one special frequency. This is called *series resonance*, and is dealt with in the next chapter.

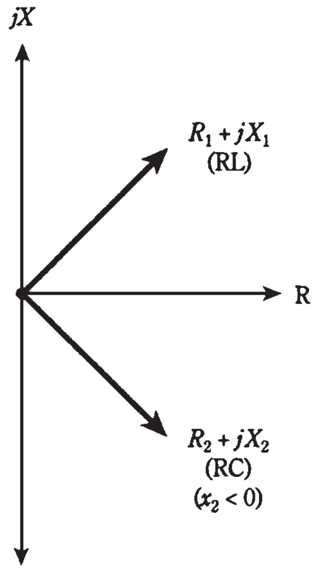
Adding impedance vectors

Often, there is resistance, as well as reactance, in an ac series circuit containing a coil and capacitor. This occurs when the coil wire has significant resistance (it's never a perfect conductor). It might also be the case because a resistor is deliberately connected into the circuit.

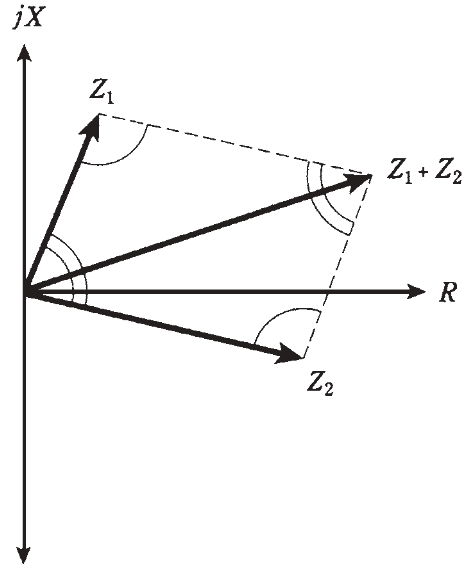
Whenever the resistance in a series circuit is significant, the impedance vectors no longer point straight up and straight down. Instead, they run off towards the "northeast" (for the inductive part of the circuit) and "southeast" (for the capacitive part). This is illustrated in Fig. 16-2.

When vectors don't lie along a single line, you need to use *vector addition* to be sure that you get the correct resultant. Fortunately, this isn't hard.

In Fig. 16-3, the geometry of vector addition is shown. Construct a *parallelogram*, using the two vectors $Z_1 = R_1 + jX_1$ and $Z_2 = R_2 + jX_2$ as two of the sides. The diagonal is the resultant. In a parallelogram, opposite angles have equal measure. These equalities are indicated by single and double arcs in the figure.



16-2 When resistance is present along with reactance, impedance vectors point "northeast" or "southeast."



16-3 Parallelogram method of vector addition.

Formula for complex impedances in series

Given two impedances, $Z_1 = R_1 + jX_1$ and $Z_2 = R_2 + jX_2$, the net impedance Z of these in series is their vector sum, given by

$$Z = (R_1 + R_2) + j(X_1 + X_2)$$

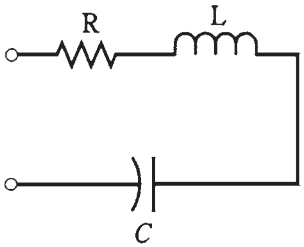
The reactances X_1 and X_2 might both be inductive; they might both be capacitive; or one might be inductive and the other capacitive.

Calculating a vector sum using the formula is easier than doing it geometrically with a parallelogram. The arithmetic method is also more nearly exact. The resistance and reactance components add separately. That's all there is to it.

Series RLC circuits

When a coil, capacitor, and resistor are connected in series (Fig. 16-4), the resistance R can be thought of as all belonging to the coil, when you use the above formulas. (Thinking of it all as belonging to the capacitor will also work.) Then you have two vectors to add, when finding the impedance of a series RLC circuit:

$$\begin{aligned} Z &= (R + jX_L) + (0 + jX_C) \\ &= R + j(X_L + X_C) \end{aligned}$$



16-4 A series RLC circuit.

Problem 16-5

A resistor, coil, and capacitor are connected in series with $R = 50\ \Omega$, $X_L = 22\ \Omega$, and X_C –33 Ω. What is the net impedance, Z?

Consider the resistor to be part of the coil, obtaining two complex vectors, $50 + j22$ and $0 - j33$. Adding these gives the resistance component of $50 + 0 = 50$, and the reactive component of $j22 - j33 = -j11$. Therefore, $Z = 50 - j11$.

Problem 16-6

A resistor, coil, and capacitor are connected in series with $R = 600\ \Omega$, $X_L = 444\ \Omega$, and $X_C = -444\ \Omega$. What is the net impedance, Z?

Again, consider the resistor to be part of the inductor. Then the vectors are $600 + j444$ and $0 - j444$. Adding these, the resistance component is $600 + 0 = 600$, and the reactive component is $j444 - j444 = j0$. Thus, $Z = 600 + j0$. This is a purely resistive impedance, and you can rightly call it "600 Ω."

Problem 16-7

A resistor, coil, and capacitor are connected in series. The resistor has a value of 330 Ω, the capacitance is 220 pF, and the inductance is 100 µH. The frequency is 7.15 MHz. What is the net complex impedance?

First, you need to calculate the inductive and capacitive reactances. Remembering the formula $X_L = 6.28fL$, multiply to obtain

$$jX_L = j(6.28 \times 7.15 \times 100) = j4490$$

*Mega*hertz and *micro*henrys go together in the formula. As for X_C, recall the formula $X_C = -1/(6.28fC)$. Convert 220 pF to microfarads to go with megahertz in the formula $C = 0.000220$ µF. Then

$$jX_C = -j(1/(6.28 \times 7.15 \times 0.000220)) = -j101$$

Now, you can consider the resistance and the inductive reactance to go together, so one of the impedance vectors is $330 + j4490$. The other is $0 - j101$. Adding these gives $330 + j4389$; this rounds off to $Z = 330 + j4390$.

Problem 16-8

A resistor, coil, and capacitor are in series. The resistance is 50.0 **Ω**, the inductance is 10.0 µH, and the capacitance is 1000 pF. The frequency is 1592 kHz. What is the complex impedance of this series RLC circuit at this frequency?

First, calculate $X_L = 6.28fL$. Convert the frequency to megahertz; 1592 kHz = 1.592 MHz. Then

$$jX_L = j(6.28 \times 1.592 \times 10.0) = j100$$

Then calculate $X_C = 1/(6.28fC)$. Convert picofarads to microfarads, and use megahertz for the frequency. Therefore,

$$jX_C = -j(1/(6.28 \times 1.592 \times 0.001000)) = -j100$$

Let the resistance and inductive reactance go together as one vector, $50.0 + j100$. Let the capacitance alone be the other vector, $0 - j100$. The sum is $50.0 + j100 - j100 = 50.0 + j0$. This is a pure resistance of 50.0 **Ω**. You can correctly say that the impedance is "50.0 **Ω**" in this case.

This concludes the analysis of series RLC circuit impedances. What about parallel circuits? To deal with these, you must calculate using conductance, susceptance, and admittance, converting to impedance only at the very end.

Complex admittances in parallel

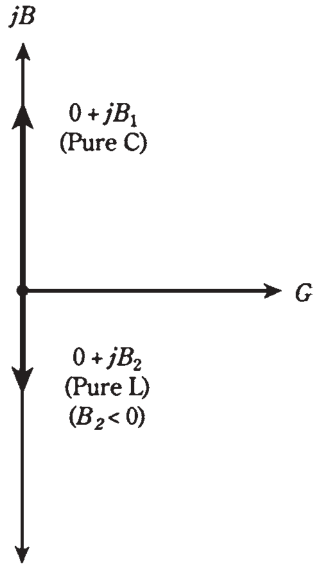
When you see resistors, coils, and capacitors in parallel, you should envision the GB (conductance-susceptance) plane.

Each component, whether it is a resistor, an inductor, or a capacitor, has an admittance that can be represented as a vector in the GB plane. The vectors for pure conductances are constant, even as the frequency changes. But the vectors for the coils and capacitors vary with frequency, in a manner similar to the way they vary in the RX plane.

Pure susceptances

Pure inductive susceptances (B_L) and capacitive susceptances (B_C) add together when coils and capacitors are in parallel. Thus, $B = B_L + B_C$. Remember that B_L is *negative* and B_C is *positive*, just the opposite from reactances.

In the GB plane, the jB_L and jB_C vectors add, but because these vectors point in exactly opposite directions—inductive susceptance down and capacitive susceptance up—the sum, jB, will also inevitably point straight down or up (Fig. 16-5).



16-5 Pure capacitance and pure inductance are represented by susceptance vectors that point straight up and down.

Problem 16-9

A coil and capacitor are connected in parallel, with $jB_{\mathrm{L}} = -j0.05$ and $j\mathrm{B}_{\mathrm{C}} = j0.08$. What is the net admittance vector?

Just add the values $jB = jB_{\mathrm{L}} + jB_{\mathrm{C}} = -j0.05 + j0.08 = j0.03$. This is a *capacitive* susceptance, because it is *positive imaginary*. The admittance vector is $0 + j0.03$.

Problem 16-10

A coil and capacitor are connected in parallel, with $jB_{\mathrm{L}} = -j0.60$ and $jB_{\mathrm{C}} = j\,0.25$. What is the net admittance vector?

Again, add $jB = -j0.\,60 + j0.25 = -j0.35$. This is an *inductive* susceptance, because it is *negative imaginary*. The admittance vector is $0 - j0.\,35$.

Problem 16-11

A coil of $L = 6.00$ µH and a capacitor of $C = 150$ pF are in parallel. The frequency is $f = 4.00$ MHz. What is the net admittance vector?

First calculate

$$\begin{aligned} jB_{\mathrm{L}} &= -j(\mathrm{l}/(6.28fL) \\ &= -j(\mathrm{l}/(6.28 \times 4.00 \times 6.00))\ -j0.00663 \end{aligned}$$

Then calculate

$$\begin{aligned} j\mathrm{B}_{\mathrm{C}} &= j(6.28fC) \\ &= j(6.28 \times 4.00 \times 0.000150) = j0.00377 \end{aligned}$$

Finally, add

$$\begin{aligned} jB &= jB_{\mathrm{L}} + jB_{\mathrm{C}} \\ &= -j0.00663 + j0.00377 = -j0.00286 \end{aligned}$$

This is a net inductive susceptance. There is no conductance in this circuit, so the admittance vector is $0 - j0.00286$.

Problem 16-12

What is the net admittance vector for the above combination at a frequency of f 5.31 MHz?

First calculate

$$\begin{aligned} jB_{\mathrm{L}} &= -j(1/(6.28fL) \\ &= -j(1/(6.28 \times 5.31 \times 6.00)) = -j0.00500 \end{aligned}$$

Then calculate

$$\begin{aligned} jB_{\mathrm{C}} &= j(6.28fC) \\ &= j(6.28 \times 5.31 \times 0.000150) = j0.00500 \end{aligned}$$

Finally, add

$$\begin{aligned} jB &= jB_{\mathrm{L}} + jB_{\mathrm{C}} \\ &= -j0.00500 + j0.00500 = j0 \end{aligned}$$

There is no susceptance. Because the conductance is also zero (there is nothing else in parallel with the coil and capacitor that might conduct), the admittance vector is $0 + j0$.

This situation, in which there is no conductance and no susceptance, seems to imply that this combination of coil and capacitor in parallel is an open circuit at 5.31 MHz. In theory this is true; zero admittance means no current can get through the circuit. In practice it's not quite the case. There is always a small leakage. This condition is known as *parallel resonance*. It's discussed in the next chapter.

Adding admittance vectors

In real life, there is a small amount of conductance, as well as susceptance, in an ac parallel circuit containing a coil and capacitor. This occurs when the capacitor lets a little bit of current leak through. More often, though, it is the case because a *load* is connected in parallel with the coil and capacitor. This load might be an antenna, or the input to an amplifier circuit, or some test instrument, or a transducer.

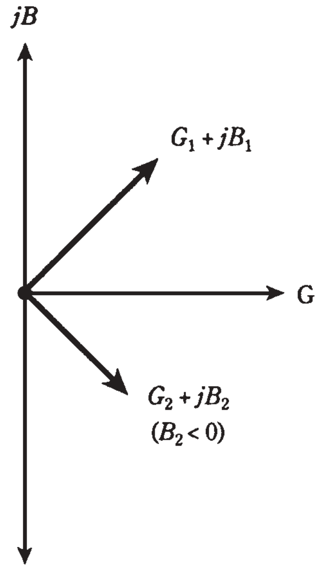
Whenever the conductance in a parallel circuit is significant, the admittance vectors no longer point straight up and down. Instead, they run off towards the "northeast" (for the capacitive part of the circuit) and "southeast" (for the inductive part). This is illustrated in Fig. 16-6.

In the problems above, you added numbers, but in fact you were adding vectors that just happened to fall along a single line, the imaginary (j) axis of the GB plane. In practical circuits, the vectors often do not lie along a single line. You've already seen how to deal with these in the RX plane. In the GB plane, the principle is the same.

Formula for complex admittances in parallel

Given two admittances, $Y_1 = G_1 + jB_1$ and $Y_2 = G_2 + jB_2$, the net admittance Y of these in parallel is their vector sum, given by

$$Y = (G_1 + G_2) + j(B_1 + B_2)$$



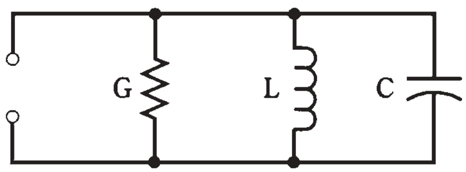
16-6 When conductance is present along with susceptance, admittance vectors point "northeast" or "southeast."

The susceptances B_1 and B_2 might both be inductive; they might both be capacitive; or one might be inductive and the other capacitive.

Parallel GLC circuits

When a coil, capacitor, and resistor are connected in parallel (Fig. 16-7), the resistor should be thought of as a *conductor,* whose value in siemens is equal to the reciprocal of the value in ohms. Think of the conductance as all belonging to the inductor. (Thinking of it all as belonging to the capacitor will also work.) Then you have two vectors to add, when finding the admittance of a parallel GLC (conductance-inductance-capacitance) circuit:

$$\begin{aligned} Y &= (G + jB_{\text{L}}) + (0 + jB_{\text{C}}) \\ &= G + j(B_{\text{L}} + B_{\text{C}}) \end{aligned}$$



16-7 A parallel GLC circuit.

Problem 16-13

A resistor, coil, and capacitor are connected in parallel with $G = 0.1$ siemens, $jB_L = -j0.010$, and $jB_{\text{C}} = j0.020$. What is the net admittance vector?

Consider the resistor to be part of the coil, obtaining two complex vectors, $0.1 - j0.010$ and $0 + j0.020$. Adding these gives the conductance component of $0.1 + 0$

= 0.1, and the susceptance component of $-j0.010 + j0.020 = j0.010$. Therefore, the admittance vector is $0.1 + j0.010$.

Problem 16-14

A resistor, coil, and capacitor are connected in parallel with $G = 0.0010$ siemens, $jB_L = -j0.0022$ and $jB_C = j0.0022$. What is the net admittance vector?

Again, consider the resistor to be part of the coil. Then the complex vectors are $0.0010 - j0.0022$ and $0 + j0.0022$. Adding these, the conductance component is $0.0010 + 0 = 0.0010$, and the susceptance component is $-j0.0022 + j0.0022 = j0$. Thus, the admittance vector is $0.0010 + j0$. This is a purely conductive admittance. There is no susceptance.

Problem 16-15

A resistor, coil, and capacitor are connected in parallel. The resistor has a value of 100 Ω, the capacitance is 200 pF, and the inductance is 100 µH. The frequency is 1.00 MHz. What is the net complex admittance?

First, you need to calculate the inductive and capacitive susceptances. Recall jB_L $-j(1/(6.28fL)$, and "plug in" the values, getting

$$jB_L = -j(1/(6.28 \times 1.00 \times 100) = -j0.00159$$

*Mega*hertz and *micro*henrys go together in the formula. As for jB_C, recall the formula $jB_C = j(6.28fC)$. Convert 200 pF to microfarads to go with megahertz in the formula; $C = 0.000200$ µF. Then

$$jB_C = j(6.28 \times 1.00 \times 0.000200) = j0.00126$$

Now, you can consider the conductance, which is $1/100 = 0.0100$ siemens, and the inductive susceptance to go together. So one of the vectors is $0.0100 - j0.00159$. The other is $0 + j0.00126$. Adding these gives $0.0100 - j0.00033$.

Problem 16-16

A resistor, coil, and capacitor are in parallel. The resistance is 10.0 Ω, the inductance is 10.0 µH, and the capacitance is 1000 pF. The frequency is 1592 kHz. What is the complex admittance of this circuit at this frequency?

First, calculate $jB_L = -j(1/(6.28fL))$. Convert the frequency to megahertz; 1592 kHz = 1.592 MHz. Then

$$jB_L = -j/(1/(6.28 \times 1.592 \times 10.0)) = -j0.0100$$

Then calculate $jB_C = j(6.28fC)$. Convert picofarads to microfarads, and use megahertz for the frequency. Therefore

$$jB_C = j(6.28 \times 1.592 \times 0.001000) = j0.0100$$

Let the conductance and inductive susceptance go together as one vector, $0.100 - j0.0100$. (Remember that conductance is the reciprocal of resistance; here $G =$

$1/R = 1/10.0 = 0.100$.) Let the capacitance alone be the other vector, $0 + j0.0100$. Then the sum is $0.100 - j0.0100 + j0.0100 = 0.100 + j0$. This is a pure conductance of 0.100 siemens.

Converting from admittance to impedance

The GB plane is, as you have seen, similar in appearance to the RX plane, although mathematically the two are worlds apart. Once you've found a complex admittance for a parallel RLC circuit, how do you transform this back to a complex impedance? Generally, it is the impedance, not the admittance, that technicians and engineers work with.

The transformation from complex admittance, or a vector $G + jB$, to a complex impedance, or a vector $R + jX$, requires the use of the following formulas:

$$R = G/(G^2 + B^2)$$
$$X = -B/(G^2 + B^2)$$

If you know the complex admittance, first find the resistance and reactance components individually. Then assemble them into the impedance vector, $R + jX$.

Problem 16-17

The admittance vector for a certain parallel circuit is $0.010 - j0.0050$. What is the impedance vector?

In this case, $G = 0.010$ and $B = -0.0050$. Find $G^2 + B^2$ first, because you'll need to use it twice as a denominator; it is $0.010^2 + (-0.0050)^2 = 0.000100 + 0.000025$ 0.000125. Then

$$R = G/0.000125 = 0.010/0.000125 = 80$$
$$X = -B/0.000125 = 0.0050/0.000125 = 40$$

The impedance vector is therefore $R + jX = 80 + j40$.

Putting it all together

When you're confronted with a parallel RLC circuit, and you want to know the complex impedance $R + jX$, take these steps:

1. Find the conductance $G = 1/R$ for the resistor. (It will be positive or zero.)
2. Find the susceptance B_L of the inductor using the appropriate formula. (It will be negative or zero.)
3. Find the susceptance B_C of the capacitor using the appropriate formula. (It will be positive or zero.)
4. Find the net susceptance $B = B_L + B_C$. (It might be positive, negative, or zero.)
5. Compute R and X in terms of G and B using the appropriate formulas.
6. Assemble the vector $R + jX$.

Problem 16-18

A resistor of 10.0 Ω, a capacitor of 820 pF, and a coil of 10.0 μH are in parallel. The frequency is 1.00 MHz. What is the impedance $R + jX$?

Proceed by the steps as numbered above.

1. $G = 1/R = 1/10.0 = 0.100$.
2. $B_L = -1/(6.28fL) = -1/(6.28 \times 1.00 \times 10.0) = -0.0159$.
3. $B_C = 6.28fC = 6.28 \times 1.00 \times 0.000820 = 0.00515$. (Remember to convert the capacitance to microfarads, to go with megahertz.)
4. $B = B_L + B_C = -0.0159 + 0.00515 = -0.0108$.
5. First find $G^2 + B^2 = 0.100^2 + (-0.0108)^2 = 0.010117$. (Go to a couple of extra places to be on the safe side.) Then $R = G/0.010117 = 0.100/0.010117 = 9.88$, and $X = -B/0.010117 = 0.0108/0.010117 = 1.07$.
6. The vector $R + jX$ is therefore $9.88 + j1.07$. This is the complex impedance of this parallel RLC circuit.

Problem 16-19

A resistor of 47.0 **Ω**, a capacitor of 500 pF, and a coil of 10.0 µH are in parallel. What is their complex impedance at a frequency of 2.252 MHz?

Proceed by the steps as numbered above.

1. $G = 1/R = 1/47.0 = 0.0213$.
2. $B_L = -1/(6.28fL) = -1/(6.28 \times 2.252 \times 10.0) = -0.00707$.
3. $B_C = 6.28fC = 6.28 \times 2.252 \times 0.000500 = 0.00707$.
4. $B = B_L + B_C = -0.00707 + 0.00707 = 0$.
5. Find $G^2 + B^2 = 0.0213^2 + 0.000^2 = 0.00045369$. (Again, go to a couple of extra places.) Then $R = G/0.00045369 = 0.0213/0.00045369 = 46.9$, and $X = -B/0.00045369 = 0$.
6. The vector $R + jX$ is therefore $46.9 + j0$. This is a pure resistance, almost exactly the value of the resistor in the circuit.

Reducing complicated RLC circuits

Sometimes you'll see circuits in which there are several resistors, capacitors, and/or coils in series and parallel combinations. It is not the intent here to analyze all kinds of bizarre circuit situations. That would fill up hundreds of pages with formulas, diagrams, and calculations, and no one would ever read it (assuming any author could stand to write it).

A general rule applies to "complicated" RLC circuits: Such a circuit can usually be reduced to an equivalent circuit that contains one resistor, one capacitor, and one inductor.

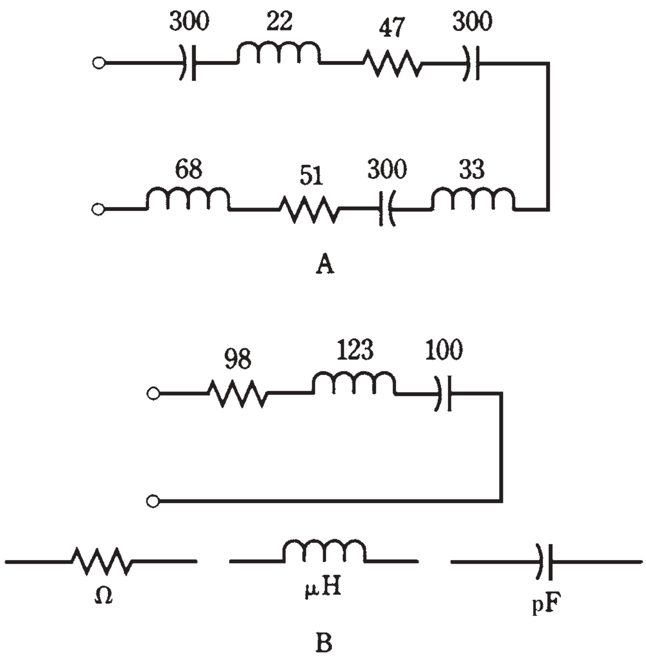
Series combinations

Resistances in series simply add. Inductances in series also add. Capacitances in series combine in a somewhat more complicated way. If you don't remember the formula, it is

$$1/C = 1/C_1 + 1/C_2 + \ldots + 1/C_n$$

where $C_1, C_2, \ldots, C_n$ are the individual capacitances and C is the total capacitance. Once you've found $1/C$, take its reciprocal to obtain C.

An example of a "complicated" series RLC circuit is shown in Fig. 16-8A. The equivalent circuit, with just one resistor, one capacitor, and one coil, is shown in Fig. 16-813.



16-8 At A, a "complicated" series RLC circuit; at B, the same circuit simplified.

Parallel combinations

In parallel, resistances and inductances combine the way capacitances do in series. Capacitances just add up.

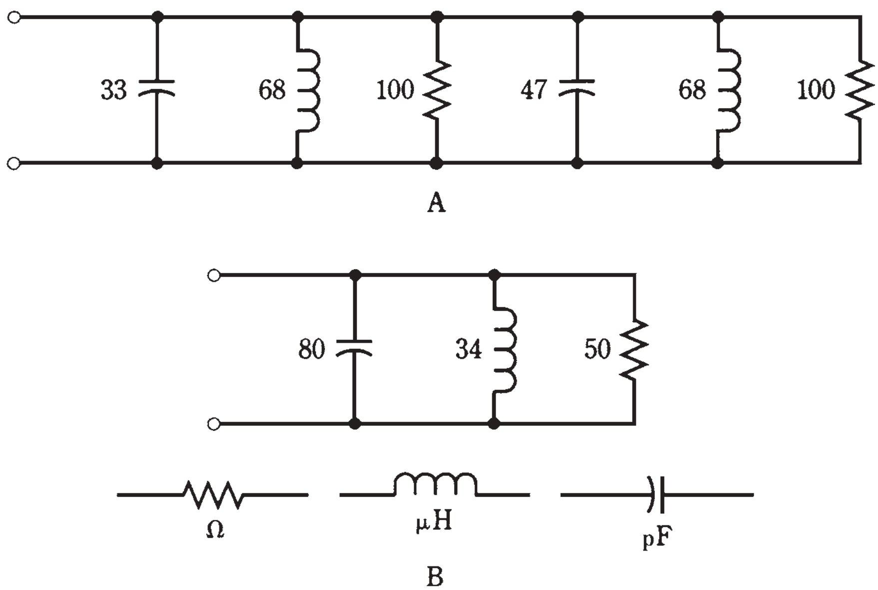
An example of a "complicated" parallel RLC circuit is shown in Fig. 16-9A. The equivalent circuit, with just one resistor, one capacitor, and one coil, is shown in Fig. 16-9B.

Complicated, messy nightmares

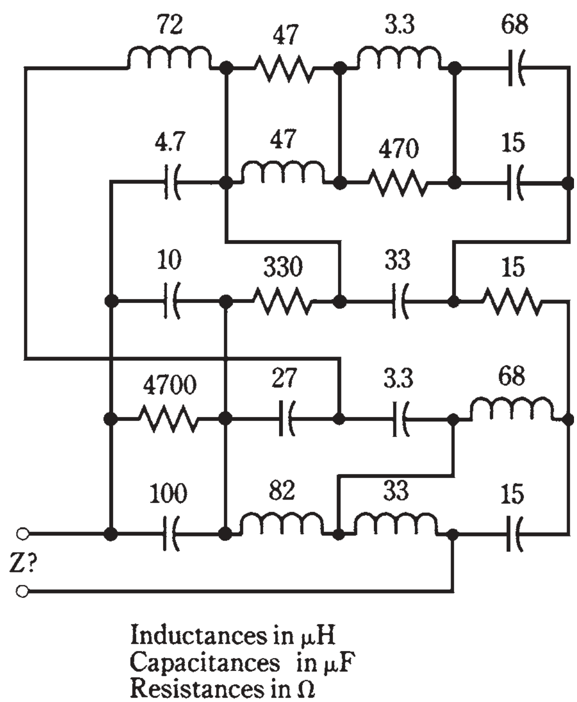
Some RLC circuits don't fall neatly into either of the above categories. An example of such a circuit is shown in Fig. 16-10. "Complicated" really isn't the word to use here! How would you find the complex impedance at some frequency, such as 8.54 MHz?

You needn't waste much time worrying about circuits like this. But be assured, given a frequency, a complex impedance does exist.

In real life, an engineer would use a computer to solve this problem. If a program didn't already exist, the engineer would either write one, or else hire it done by a professional programmer.



16-9 At A, a "complicated" parallel RLC circuit; at B, the same circuit simplified.



16-10 A series-parallel RLC nightmare.

Another way to find the complex impedance here would be to actually build the circuit, connect a signal generator to it, and measure R and X directly with an *impedance bridge*. Because "the proof of the pudding is in the eating," a performance test must eventually be done anyway, no matter how sophisticated the design theory. Engineers have to build things that work!

Ohm's law for ac circuits

Ohm's Law for a dc circuit is a simple relationship among three variables: current (I), voltage (E), and resistance (R). The formulas, again, are

$$I = E/R$$
$$E = IR$$
$$R = E/I$$

In ac circuits containing negligible or zero reactance, these same formulas apply, as long as you are sure that you use the *effective* current and voltage.

Effective amplitudes

The effective value for an ac sine wave is the root-mean-square, or rms, value. You learned about this in chapter 9. The rms current or voltage is 0.707 times the peak amplitude. Conversely, the peak value is 1.414 times the rms value.

If you're told that an ac voltage is 35 V, or that an ac current is 570 mA, it is generally understood that this refers to a sine-wave rms level, unless otherwise specified.

Purely resistive impedances

When the impedance in an ac circuit is such that the reactance X has a negligible effect, and that practically all of the current and voltage exists through and across a resistance R, Ohm's Law for an ac circuit is expressed as

$$I = E/Z$$
$$E = IZ$$
$$Z = E/I$$

where Z is essentially equal to R, and the values I and E are rms current and voltage.

Complex impedances

When determining the relationship among current, voltage and resistance in an ac circuit with resistance and reactance that are both significant, things get interesting.

Recall the formula for absolute-value impedance in a series RLC circuit,

$$Z^2 = R^2 + X^2$$

so Z is equal to the square root of $R^2 + X^2$. This is the length of the vector $R + jX$ in the complex impedance plane. You learned this in chapter 15. This formula applies only for series RLC circuits.

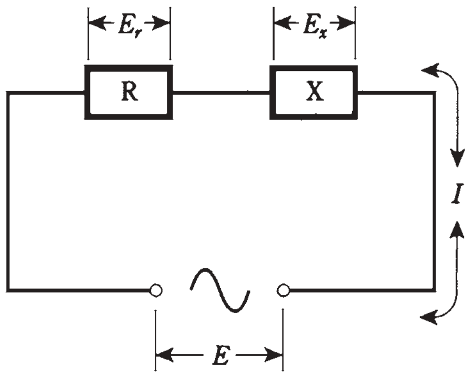
The absolute-value impedance for a parallel RLC circuit, in which the resistance is R and the reactance is X, is defined by the formula:

$$Z^2 = (RX)^2/(R^2 + X^2)$$

Thus, Z is equal to RX divided by the square root of $R^2 + X^2$.

Problem 16-20

A series RX circuit (Fig. 16-11) has $R = 50.0\ \Omega$ of resistance and $X = -50.0\ \Omega$ of reactance, and 100 Vac is applied. What is the current?



16-11 A series RX circuit. Notation is discussed in the text.

First, calculate $Z^2 = R^2 + X^2 = 50.0^2 + (-50.0)^2 = 2500 + 2500 = 5000$; Z is the square root of 5000, or 70.7 Ω. Then $I = E/Z = 100/70.7 = 1.41$ A.

Problem 16-21

What are the voltage drops across the resistance and the reactance, respectively, in the above problem?

The Ohm's Law formulas for dc will work here. Because the current is $I = 1.41$ A, the voltage drop across the resistance is equal to $E_R = IR = 1.41 \times 50.0 = 70.5$ V. The voltage drop across the reactance is the product of the current and the reactance: $E_X = IX = 1.41 \times (-50.0) = -70.5$. This is an ac voltage of equal magnitude to that across the resistance. But the phase is different.

The voltages across the resistance and the reactance—a capacitive reactance in this case, because it's negative—don't add up to 100. The meaning of the minus sign for the voltage across the capacitor is unclear, but there is no way, whether you consider this sign or not, that the voltages across the resistor and capacitor will arithmetically add up to 100. Shouldn't they? In a dc circuit, yes; in an ac circuit, generally, no.

In a resistance-reactance ac circuit, there is always a difference in phase between the voltage across the resistive part and the voltage across the reactive part. They always add up to the applied voltage *vectorially*, but not always *arithmetically*. You don't need to be concerned with the geometry of the vectors in this situation. It's enough to understand that the vectors don't fall along a single line, and this is why the voltages don't add arithmetically.

Problem 16-22

A series RX circuit (Fig. 16-11) has $R = 10.0\ \Omega$ and a net reactance $X = 40.0\ \Omega$. The applied voltage is 100. What is the current?

Calculate $Z^2 = R^2 + X^2 = 100 + 1600 = 1700$; thus, $Z = 41.2$. Therefore, $I = E/Z$ $100/41.2 = 2.43$ A. Note that the reactance in this circuit is inductive, because it is positive.

Problem 16-23

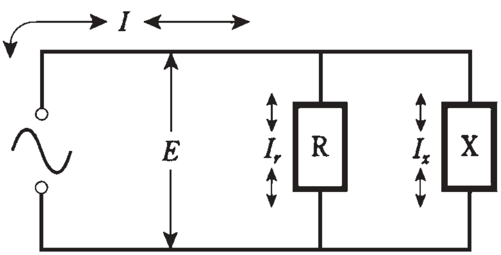
What is the voltage across R in the preceding problem? Across X?

Knowing the current, calculate $E_R = IR = 2.43 \times 10.0 = 24.3$ V. Also, $E_X = IX = 2.43 \times 40.0 = 97.2$ V. If you add $E_R + E_X$ arithmetically, you get $24.3\text{ V} + 97.2\text{ V} = 121.5$ V as the total across R and X. Again, the simple dc rule does not work here. The reason is the same as before.

Problem 16-24

A parallel RX circuit (Fig. 16-12) has resistance $R = 30$ ohms and a net reactance $X = -20\ \Omega$. The supply voltage is 50 V. What is the total current drawn from the supply?

Find the absolute-value impedance, remembering the formula for parallel circuits: $Z^2 = (RX)^2 / (R^2 + X^2) = 360{,}000/1300 = 277$. The impedance Z is the square root of 277, or 16.6 Ω. The total current is therefore $I = E/Z = 50/16.6 = 3.01$ A.



6-12 A parallel RX circuit. Notation is discussed in the text.

Problem 16-25

What is the current through R above? Through X?

The Ohm's Law formulas for dc will work here. For the resistance, $I_R = E/R = 50/30 = 1.67$ A. For the reactance, $I_X = E/X = 50/(-20) = -2.5$ A.

These currents don't add up to 3.01 A, the total current, whether the minus sign is taken into account, or not. It's not really clear what the minus sign means, anyhow. The reason that the constituent currents, I_R and I_X, don't add up to the total current, I, is the same as the reason the voltages don't add up in a series RX circuit. These currents are actually 2D vectors; you're seeing them through 1D glasses.

If you want to study the geometrical details of the voltage and current vectors in series and parallel RX circuits, a good circuit theory text is recommended.

One of the most important practical aspects of ac circuit theory involves the ways that reactances, and complex impedances, behave when you try to feed power to them. That subject will start off the next chapter.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. A coil and capacitor are connected in series. The inductive reactance is 250 Ω, and the capacitive reactance is −300 Ω. What is the net impedance vector, $R + jX$?

   A. $0 + j550$.

   B. $0 - j50$.

   C. $250 - j300$

   D. $-300 + j250$.

2. A coil of 25.0 µH and capacitor of 100 pF are connected in series. The frequency is 5.00 MHz. What is the impedance vector, $R + jX$?

   A $0 + j467$.

   B. $25 + j100$.

   C. $0 - j467$.

   D. $25 - j100$.

3. When $R = 0$ in a series RLC circuit, but the net reactance is not zero, the impedance vector:

   A. Always points straight up.

   B. Always points straight down.

   C. Always points straight towards the right.

   D. None of the above.

4. A resistor of 150 Ω, a coil with reactance 100 Ω and a capacitor with reactance –200 Ω are connected in series. What is the complex impedance $R + jX$?

   A. $150 + j100$.

   B. $150 - j200$.

   C. $100 - j200$.

   D. $150 - j100$.

5. A resistor of 330 Ω, a coil of 1.00 µH and a capacitor of 200 pF are in series. What is $R + jX$ at 10.0 MHz?

   A. $330 - j199$.

   B. $300 + j201$.

C. $300 + j142$.

D. $330 - j16.8$.

6. A coil has an inductance of 3.00 μH and a resistance of 10.0 Ω in its winding. A capacitor of 100 pF is in series with this coil. What is $R + jX$ at 10.0 MHz?

A. $10 + j3.00$.

B. $10 + j29.2$.

C. $10 - j97$.

D. $10 + j348$.

7. A coil has a reactance of 4.00 Ω. What is the admittance vector, $G + jB$, assuming nothing else is in the circuit?

A. $0 + j0.25$.

B. $0 + j4.00$.

C. $0 - j0.25$.

D. $0 - j4.00$.

8. What will happen to the susceptance of a capacitor if the frequency is doubled, all other things being equal?

A. It will decrease to half its former value.

B. It will not change.

C. It will double.

D. It will quadruple.

9. A coil and capacitor are in parallel, with $jB_L = -j0.05$ and $jB_C = j0.03$. What is the admittance vector, assuming that nothing is in series or parallel with these components?

A. $0 - j0.02$.

B. $0 - j0.07$.

C. $0 + j0.02$.

D. $-0.05 + j0.03$.

10. A coil, resistor, and capacitor are in parallel. The resistance is 1 Ω ; the capacitive susceptance is 1.0 siemens; the inductive susceptance is −1.0 siemens. Then the frequency is cut to half its former value. What will be the admittance vector, $G + jB$, at the new frequency?

A. $1 + j0$.

B. $1 + j1.5$.

C. $1 - j1.5$.

D. $1 - j2$.

11. A coil of 3.50 µH and a capacitor of 47.0 pF are in parallel. The frequency is 9.55 MHz. There is nothing else in series or parallel with these components. What is the admittance vector?
   A. $0 + j0.00282$.
   B. $0 - j0.00194$.
   C. $0 + j0.00194$.
   D. $0 - j0.00758$.
12. A vector pointing "southeast" in the GB plane would indicate the following:
   A. Pure conductance, zero susceptance.
   B. Conductance and inductive susceptance.
   C. Conductance and capacitive susceptance.
   D. Pure susceptance, zero conductance.
13. A resistor of 0.0044 siemens, a capacitor whose susceptance is 0.035 siemens, and a coil whose susceptance is −0.011 siemens are all connected in parallel. The admittance vector is:
   A. $0.0044 + j0.024$.
   B. $0.035 - j0.011$.
   C. $-0.011 + j0.035$.
   D. $0.0044 + j0.046$.
14. A resistor of 100 **Ω**, a coil of 4.50 µH, and a capacitor of 220 pF are in parallel. What is the admittance vector at 6.50 MHz?
   A. $100 + j0.00354$.
   B. $0.010 + j0.00354$.
   C. $100 - j0.0144$.
   D. $0.010 + j0.0144$.
15. The admittance for a circuit, $G + jB$, is $0.02 + j0.20$. What is the impedance, R $+jX$?
   A. $50 + j5.0$.
   B. $0.495 - j4.95$.
   C. $50 - j5.0$.
   D. $0.495 + j4.95$.
16. A resistor of 51.0 **Ω**, an inductor of 22.0 µH and a capacitor of 150 pF are in parallel. The frequency is 1.00 MHz. What is the complex impedance, $R + jX$?
   A. $51.0 - j14.9$.
   B. $51.0 + j14.9$.

C. $46.2 - j14.9$.

D. $46.2 + j14.9$.

17. A series circuit has 99.0 Ω of resistance and 88.0 Ω of inductive reactance. An ac rms voltage of 117 V is applied to this series network. What is the current?

A. 1.18 A.

B. 1.13 A.

C. 0.886 A.

D. 0.846 A.

18. What is the voltage across the reactance in the above example?

A. 78.0 V.

B. 55.1 V.

C. 99.4 V.

D. 74.4 V.

19. A parallel circuit has 10 ohms of resistance and 15 Ω of reactance. An ac rms voltage of 20 V is applied across it. What is the total current?

A. 2.00 A.

B. 2.40 A.

C. 1.33 A.

D. 0.800 A.

20. What is the current through the resistance in the above example?

A. 2.00 A.

B. 2.40 A.

C. 1.33 A.

D. 0.800 A.

17
CHAPTER

Power and resonance in ac circuits

YOU HAVE LEARNED HOW CURRENT, VOLTAGE, AND RESISTANCE BEHAVE IN ac circuits. How can all this theoretical knowledge be put to practical use?

One of the engineer's biggest challenges is the problem of efficient energy transfer. This is a major concern at radio frequencies. But audio design engineers, and even the utility companies, need to be concerned with ac circuit efficiency because it translates into energy conservation. The first two-thirds of this chapter is devoted to this subject.

Another important phenomenon, especially for the radio-frequency engineer, is *resonance.* This is an electrical analog of the reverberation you're familiar with if you've ever played a musical instrument. The last third of this chapter discusses resonance in series and parallel circuits.

What is power?

There are several different ways to define *power.* The applicable definition depends on the kind of circuit or device in use.

Energy per unit time

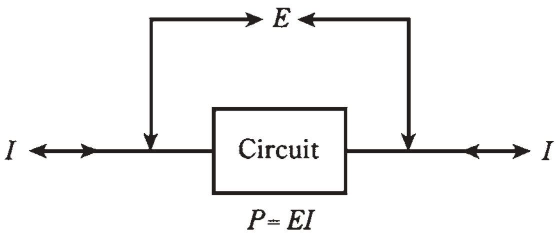
The most all-encompassing definition of power, and the one commonly used by physicists, is this: *Power is the rate at which energy is expended.* The standard unit of power is the watt, abbreviated W; it is equivalent to one joule per second.

This definition can be applied to motion, chemical effects, electricity, radio waves, sound, heat, light, ultraviolet, and X rays. In all cases, the energy is "used up" somehow, converted from one form into another form at a certain rate. This expression of power refers to an event that takes place at some definite place or places.

Sometimes power is given as kilowatts (kW or thousands of watts), megawatts (MW or millions of watts) or gigawatts (GW or billions of watts). It might be given as milliwatts (mW or thousandths of watts), microwatts (µW or millionths of watts), or nanowatts (nW or billionths of watts).

Volt-amperes

In dc circuits, and also in ac circuits having no reactance, power can be defined this way: *Power is the product of the voltage across a circuit or component, times the current through it*. Mathematically this is written $P = EI$. If E is in volts and I is in amperes, then P is in *volt-amperes (VA)*. This translates into watts when there is no reactance in the circuit (Fig. 17-1). The root-mean-square (rms) values for voltage and current are always used to derive the effective, or average, power.



17-1 When there is no reactance in a circuit, the power, P, is the product of the voltage E and the current I.

Like joules per second, volt-amperes, also called *VA power* or *apparent power*, can take various forms. A resistor converts electrical energy into heat energy, at a rate that depends on the value of the resistance and the current through it. A light bulb converts electricity into light and heat. A radio antenna converts high-frequency ac into radio waves. A speaker converts low-frequency ac into sound waves. The power in these forms is a measure of the intensity of the heat, light, radio waves, or sound waves.

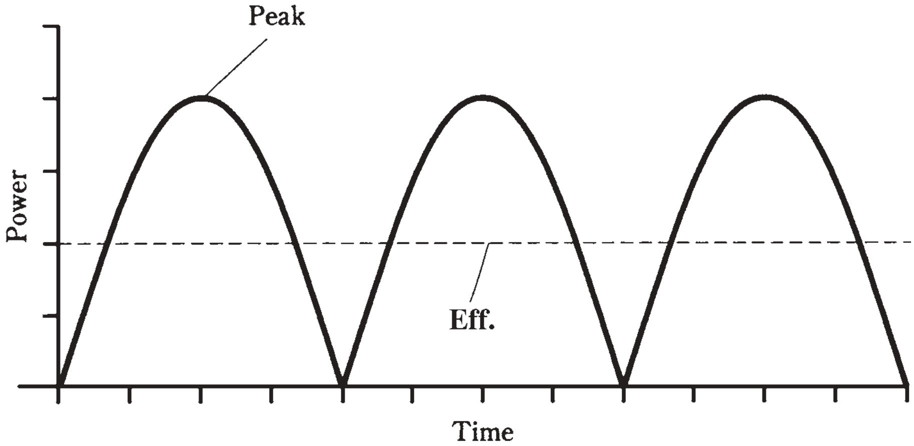
The VA power can have a meaning that the rate-of-energy-expenditure definition does not encompass. This is *reactive* or *imaginary* power, discussed shortly.

Instantaneous power

Usually, but not always, engineers think of power based on the rms, or effective, ac value. But for VA power, peak values are sometimes used instead. If the ac is a sine wave, the peak current is 1.414 times the rms current, and the peak voltage is 1.414 times the rms voltage. If the current and the voltage are exactly in phase, the product of their peak values is twice the product of their rms values.

There are instants in time when the VA power in a reactance-free, sine-wave ac circuit is twice the effective power. There are other instants in time when the VA power is zero; at still other moments, the VA power is somewhere between zero and twice the effective power level (Fig. 17-2). This constantly changing power is called *instantaneous power*.

In some situations, such as with a voice-modulated radio signal or a fast-scan television signal, the instantaneous power varies in an extremely complicated fashion. Perhaps you have seen the *modulation envelope* of such a signal displayed on an oscilloscope.



17-2 Peak versus effective power for a sine wave.

Imaginary power

If an ac circuit contains reactance, things get interesting. The rate of energy expenditure is the same as the VA power in a pure resistance. But when inductance and/or capacitance exists in an ac circuit, these two definitions of power part ways. The VA power becomes greater than the power actually manifested as heat, light, radio waves, or whatever. The extra "power" is called *imaginary power,* because it exists in the reactance, and reactance can be, as you have learned, rendered in mathematically imaginary numerical form. It is also known as *reactive power.*

Inductors and capacitors store energy and then release it a fraction of a cycle later. This phenomenon, like *true power,* is expressible as the rate at which energy is changed from one form to another. But rather than any immediately usable form of power, such as radio or sound waves, imaginary power is "stashed" as a magnetic or electric field and then "dumped" back into the circuit, over and over again.

You might think of the relationship between imaginary and true power in the same way as you think of *potential* versus *kinetic* energy. A brick held out of a seventh-story window has potential energy, just as a charged-up capacitor or inductor has imaginary power.

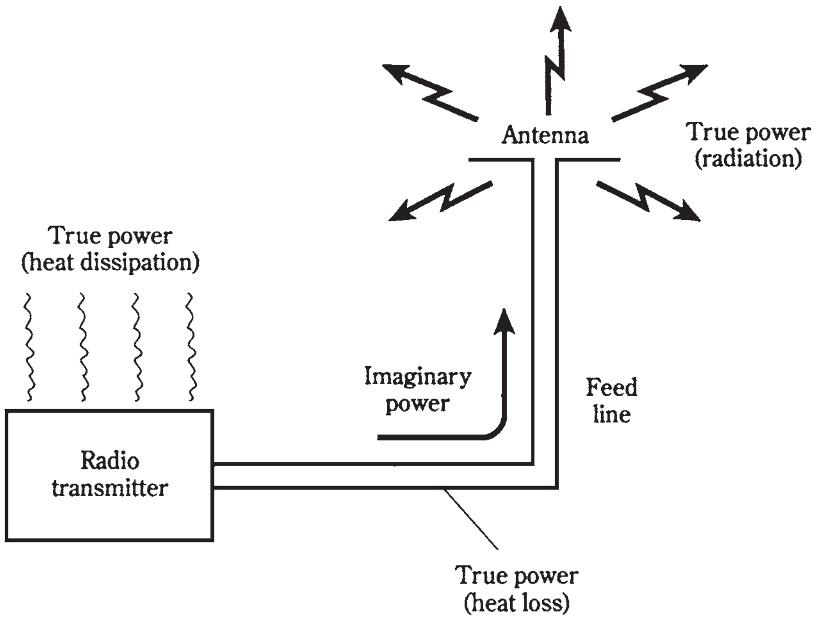
Although the label "imaginary power" carries a connotation that it's not real or important, it's significant indeed. Imaginary power is responsible for many aspects of ac circuit behavior.

True power doesn't travel

An important semantical point should be brought up concerning true power, not only in ac circuits, but in any kind of circuit or device.

A common and usually harmless misconception about true power is that it "travels." For example, if you connect a radio transmitter to a cable that runs outdoors to an antenna, you might say you're "feeding power" through the cable to the antenna. Everybody says this, even engineers and technicians. What's moving along the cable is *imaginary power,* not *true power.* True power always involves a change in form, such as from electrical current and voltage into radio waves.

Some true power is dissipated as heat in the transmitter amplifiers and in the feed line (Fig. 17-3). The useful dissipation of true power occurs when the imaginary power, in the form of high-frequency current and voltage, gets to the antenna, where it is changed into electromagnetic waves.



17-3 True and imaginary power in a radio antenna system.

You will often hear expressions such as "forward power" and "reflected power," or "power is fed from this amplifier to these speakers." It is all right to talk like this, but it can sometimes lead to wrong conclusions, especially concerning impedance and standing waves. Then, you need to be keenly aware of the distinction among true, imaginary and apparent power.

Reactance does not consume power

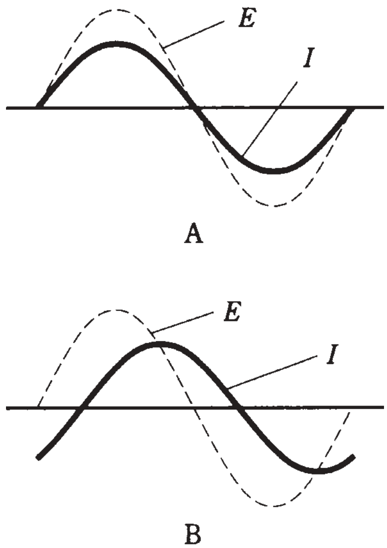
A coil or capacitor cannot dissipate power. The only thing that such a component can do is store energy and then give it back to the circuit a fraction of a cycle later. In real life, the dielectrics or wires in coils or capacitors dissipate some power as heat, but ideal components would not do this.

A capacitor, as you have learned, stores energy as an electric field. An inductor stores energy as a magnetic field.

A reactance causes ac current to shift in phase, so that it is no longer exactly in step with the voltage. In a circuit with inductive reactance, the current lags the voltage by up to 90 degrees, or one-quarter cycle. In a circuit with capacitive reactance, the current leads the voltage by up to 90 degrees.

In a resistance-reactance circuit, true power is dissipated only in the resistive components. The reactive components cause the VA power to be exaggerated compared with the true power.

Why does reactance cause this discrepancy between apparent (VA) power and true power? In a circuit that is purely resistive, the voltage and current march right along in step with each other, and therefore, they combine in the most efficient possible way (Fig. 17-4A). But in a circuit containing reactance, the voltage and current don't work together as well (Fig. 17-4B) because of their phase difference. Therefore, the actual energy expenditure, or true power, is not as great as the product of the voltage and the current.



17-4 At A, current (I) and voltage (E) are in phase in a nonreactive ac circuit. At B, I and E are not in phase when reactance is present.

True power, VA power, and reactive power

In a circuit containing both resistance and reactance, the relationships among true power P_T, apparent or VA power P_{VA}, and imaginary or reactive power P_X are

$$P_{VA}^2 = P_T^2 - P_X^2$$
$$P_T < P_{VA}$$
$$P_X < P_{VA}$$

If there is no reactance in the circuit, then

$$P_{VA} = P_T$$
$$P_X = 0$$

Engineers often strive to eliminate, or at least minimize, the reactance in a circuit. This is particularly true for radio antenna systems, or when signals must be sent over long spans of cable. It is also important in the design of radio-frequency amplifiers. To a lesser extent, minimizing the reactance is important in audio work and in utility power transmission.

Power factor

The ratio of the true power to the VA power, P_T/P_{VA}, is called the *power factor* in an ac circuit. If there is no reactance, the ideal case, then $P_T = P_{VA}$, and the power factor (PF) is equal to 1. If the circuit contains all reactance and no resistance of any significance (that is, zero or infinite resistance), then $P_T = 0$, and $PF = 0$.

If you try to get a pure reactance to dissipate power, it's a little like throwing a foam-rubber ball into a gale-force wind. The ball will come right back in your face. A pure reactance cannot, and will not, dissipate power.

When a *load*, or a circuit in which you want power to be dissipated, contains some resistance and some reactance, then PF will be between 0 and 1. That is, $0 < PF < 1$. PF might be expressed as a percentage between 0 and 100, written $PF_{\%}$. Mathematically,

$$PF = P_T/P_{VA}$$
$$PF_{\%} = 100P_T/P_{VA}$$

When a load has some resistance and some reactance, a portion (but not all) of the power is dissipated as true power, and some is "rejected" by the load and sent back to the source as VA power.

Calculation of power factor

There are two ways to determine the power factor in an ac circuit that contains reactance and resistance. Either method can be used in any situation, although sometimes one scheme is more convenient than the other.

Cosine of phase angle

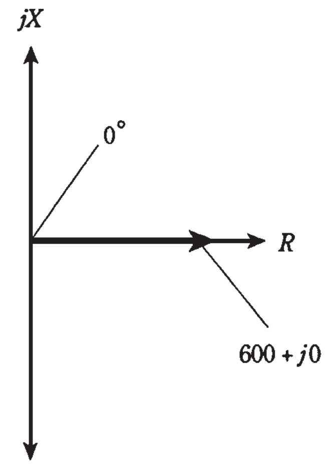
Recall that in a circuit having reactance and resistance, the current and the voltage are not in phase. The extent to which they differ in phase is the *phase angle.* If there is no reactance, then the phase angle is 0 degrees. If there is a pure reactance, then the phase angle is either +90 degrees or –90 degrees.

The power factor is equal to the cosine of the phase angle. You can use a calculator to find this easily.

Problem 17-1

A circuit contains no reactance, but a pure resistance of 600 Ω. What is the power factor?

Without doing any calculations, it is evident that $PF = 1$, because $P_{VA} = P_T$ in a pure resistance. So $P_T/P_{VA} = 1$. But you can also look at this by noting that the phase angle is 0 degrees, because the current is in phase with the voltage. Using your calculator, find cos 0 = 1. Therefore, $PF = 1 = 100$ percent. The vector for this case is shown in Fig. 17-5.

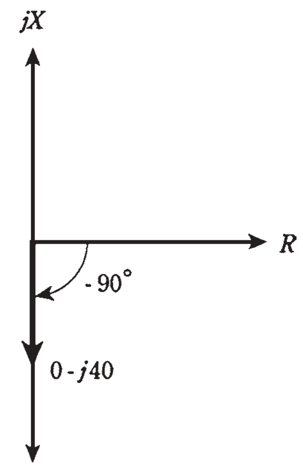


17-5 Phase angle for a pure resistance 600 + j0.

Problem 17-2

A circuit contains a pure capacitive reactance of –40 Ω, but no resistance. What is the power factor?

Here, the phase angle is –90 degrees (Fig. 17-6). A calculator will tell you that cos –90 = 0. Therefore, $PF = 0$. This means that $P_T/P_{VA} = 0 = 0$ percent. That is, none of the power is true power, and all of it is imaginary or reactive power.

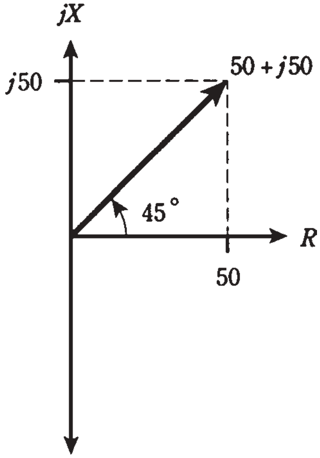


17-6 Phase angle for a pure capacitive impedance 0 – j40.

Problem 17-3

A circuit contains a resistance of 50 Ω and an inductive reactance of 50 Ω in series. What is the power factor?

The phase angle in this case is 45 degrees (Fig. 17-7).The resistance and reactance components are equal, and form two sides of a right triangle, with the complex impedance vector forming the hypotenuse. Find cos 45 = 0.707. This means that P_T/P_{VA} = 0.707 = 70.7 percent.



17-7 Phase angle for impedance 50 + j50.

Ratio R/Z

Another way to calculate the power factor is to find the ratio of the resistance *R* to the absolute-value impedance *Z*. In Fig. 17-7, this is visually apparent. A right triangle is formed by the resistance vector *R* (the base), the reactance vector *jX* (the height) and the absolute-value impedance *Z* (the hypotenuse). The cosine of the phase angle is equal to the ratio of the base length to the hypotenuse length; this represents *R*/*Z*.

Problem 17-4

A circuit has an absolute-value impedance *Z* of 100 Ω, with a resistance *R* = 80 Ω. What is the power factor?

Simply find the ratio: *PF* = *R*/*Z* = 80/100 = 0.8 = 80 percent. Note that it doesn't matter whether the reactance in this circuit is capacitive or inductive. In fact, you don't even have to worry about the value of the reactance here.

Problem 17-5

A circuit has an absolute-value impedance of 50 Ω, purely resistive. What is the power factor?

Here, *R* = *Z* = 50. Therefore, *PF* = *R*/*Z* = 50/50 = 1 = 100 percent.

Resistance and reactance

Sometimes you'll get data that tells you the resistance and reactance components in a circuit. To calculate the power factor from this, you can either find the phase angle and take its cosine, or find the absolute-value impedance and take the ratio R/Z.

Problem 17-6

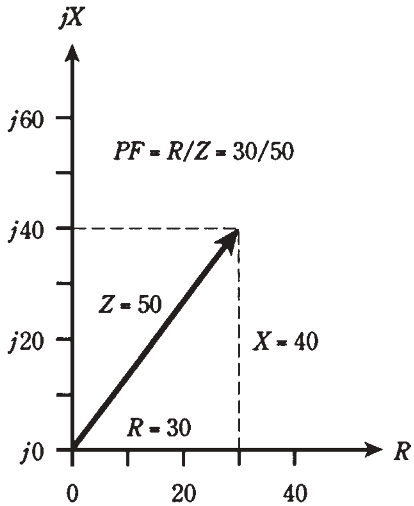
A circuit has a resistance of 50 Ω and a capacitive reactance of –30 Ω. What is the power factor? Use the cosine method.

The tangent of the phase angle is equal to X/R. Therefore, the phase angle is arctan $(X/R) = \arctan(-30/50) = \arctan(-0.60) = -31$ degrees. The power factor is the cosine of this angle; $PF = \cos(-31) = 0.86 = 86$ percent.

Problem 17-7

A circuit has a resistance of 30 Ω and an inductive reactance of 40 Ω. What is the power factor? Use the R/Z method.

Find the absolute-value impedance: $Z^2 = R^2 + X^2 = 30^2 + 40^2 = 900 + 1600 = 2500$; therefore $Z = 50$. The power factor is therefore $PF = R/Z = 30/50 = 0.60 = 60$ percent. This problem is represented very nicely by a 3:4:5 right triangle (Fig. 17-8).



17-8 Illustration for Problem 17-7.

How much of the power is true?

The above simple formulas allow you to figure out, given the resistance, reactance, and VA power, how many watts are true or real power, and how many watts are imaginary or reactive power. This is important in radio-frequency (RF) equipment, because RF wattmeters will usually display VA power, and this reading is exaggerated when there is reactance in a circuit.

Problem 17-8

A circuit has 50 Ω of resistance and 30 Ω of inductive reactance in series. A wattmeter shows 100 watts, representing the VA power. What is the true power?

First, calculate the power factor. You might use either the phase-angle method or the *R*/*Z* method. Suppose you use the phase-angle method, then,

$$\text{Phase angle} = \arctan (X/R)$$
$$= \arctan (30/50) = 31 \text{ degrees}$$

The power factor is the cosine of the phase angle. Thus,

$$PF = \cos 31 = 0.86 = 86 \text{ percent}$$

Remember that $PF = P_T/P_{VA}$. This means that the true power, P_T, is equal to 86 watts.

Problem 17-9

A circuit has a resistance of 1,000 Ω in parallel with a capacitance of 1000 pF. The frequency is 100 kHz. If a wattmeter reads a VA power of 88.0 watts, what is the true power?

This problem requires several steps in calculation. First, note that the components are in *parallel.* This means that you have to find the conductance and the capacitive susceptance, and then combine these to get the admittance. Convert the frequency to megahertz: f = 100 kHz = 0.100 MHz. Convert capacitance to microfarads: C = 1000 pF = 0.001000 µF. From the previous chapter, use the equation for capacitive susceptance:

$$B_C = 6.28fC = 6.28 \times 0.100 \times 0.001000$$
$$= 0.000628 \text{ siemens}$$

The conductance of the resistor, *G*, is found by taking the reciprocal of the resistance, *R*:

$$G = 1/R = 1/1000 = 0.001000 \text{ siemens}$$

Although you don't need to know the actual complex admittance vector to solve this problem, note in passing that it is

$$G + jB = 0.001000 + j0.000628$$

Now, use the formula for calculating the resistance and reactance of this circuit, in terms of the conductance and susceptance. First, find the resistance:

$$R = G/(G^2 + B^2)$$
$$0.001000/(0.001000^2 + 0.000628^2)$$
$$= 0.001000/0.000001394$$
$$=717\ \Omega$$

Then, find the reactance:

$$X = -B/(G^2 + B^2)$$
$$= -0.000628/0.000001394$$
$$= -451\ \Omega$$

Therefore, $R = 717$ and $X = -451$.

Using the phase-angle method to solve this (the numbers are more manageable that way than they are with the R/Z method), calculate

$$\text{Phase angle} = \arctan (X/R)$$
$$= \arctan (-451/717) = \arctan (-0.629)$$
$$= -32.2 \text{ degrees}$$

Then the power factor is

$$PF = \cos -32.2 = 0.846 = 84.6 \text{ percent}$$

The VA power, P_{VA}, is given as 88.0 watts, and $PF = P_{\mathrm{T}}/P_{\mathrm{VA}}$. Therefore, the true power is found this way:

$$P_{\mathrm{T}}/P_{\mathrm{VA}} = 0.846$$
$$P_{\mathrm{T}}/88.0 = 0.846$$
$$P_{\mathrm{T}} = 0.846 \times 88.0 = 74.4 \text{ watts}$$

This is a good example of a practical problem. Although there are several steps, each requiring careful calculation, none of the steps individually is very hard. It's just a matter of using the right equations in the right order, and plugging the numbers in. You do have to be somewhat careful in manipulating plus/minus signs, and also in placing decimal points.

Power transmission

One of the most multifaceted, and important, problems facing engineers is *power: transmission*.

Generators produce large voltages and currents at a power plant, say from turbines driven by falling water. The problem: getting the electricity from the plant to the homes, businesses, and other facilities that need it. This process involves the use of long wire *transmission lines*. Also needed are *transformers* to change the voltages to higher or lower values.

A radio transmitter produces a high-frequency alternating current. The problem: getting the power to be radiated by the antenna, located some distance from the transmitter. This involves the use of a radio-frequency transmission line. The most common type is coaxial cable. Two-wire line is also sometimes used. At ultra-high and microwave frequencies, another kind of transmission line, known as a *waveguide*, is often employed.

The overriding concern in any power-transmission system is minimizing the loss. Power wastage occurs almost entirely as heat in the line conductors and dielectric, and in objects near the line. Some loss can also take the form of unwanted electromagnetic radiation from a transmission line.

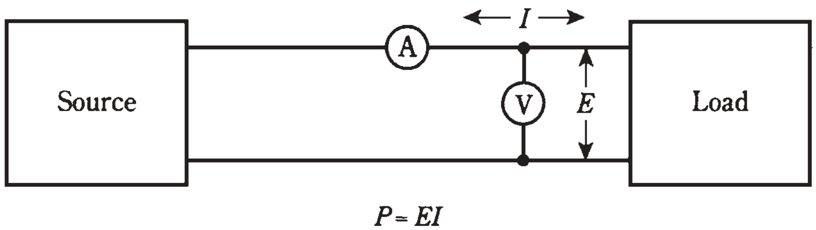
In an ideal transmission line, all of the power is VA power; that is, it is in the form of an alternating current in the conductors and an alternating voltage between them.

It is undesirable to have power in a transmission line exist in the form of true power. This translates either into heat loss in the line, radiation loss, or both. The place for true power dissipation is in the load, such as electrical appliances or radio antennas. Any true power in a transmission line represents power that can't be used by the load, because it doesn't show up there.

The rest of this chapter deals mainly with radio transmitting systems.

Power measurement in a transmission line

In a transmission line, power is measured by means of a voltmeter between the conductors, and an ammeter in series with one of the conductors (Fig. 17-9). Then the power, P (in watts) is equal to the product of the voltage E (in volts) and the current I (in amperes). This technique can be used in any transmission line, be it for 60-Hz utility service, or in a radio transmitting station.



17-9 Power measurement in a transmission line.

But is this indication of power the same as the power actually dissipated by the load at the end of the line? Not necessarily.

Recall, from the discussion of impedance, that any transmission line has a *characteristic impedance.* This value, Z_o, depends on the size of the line conductors, the spacing between the conductors, and the type of dielectric material that separates the conductors. For a coaxial cable, Z_o can be anywhere from about 50 to 150 Ω. For a parallel-wire line, it can range from about 75 Ω to 600 Ω.

If the load is a pure resistance R containing no reactance, and if $R = Z_o$, then the power indicated by the voltmeter/ammeter scheme will be the same as the true power dissipated by the load. The voltmeter and ammeter must be placed at the load end of the transmission line.

If the load is a pure resistance R, and $R < Z_o$ or $R > Z_o$, then the voltmeter and ammeter will not give an indication of the true power. Also, if there is any reactance in the load, the voltmeter/ammeter method will not be accurate.

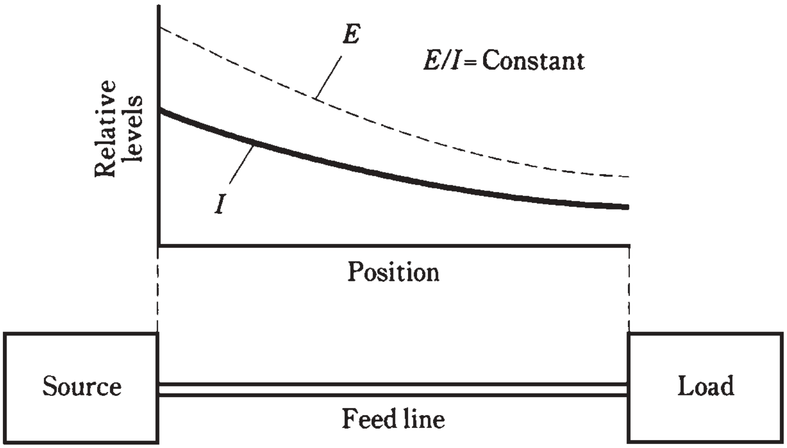
The physics of this is rather sophisticated, and a thorough treatment of it is beyond the scope of this course. But you should remember that it is always desirable to have the load impedance be a pure resistance, a complex value of $R + j0$, where $R = Z_o$. Small discrepancies, in the form of a slightly larger or smaller resistance, or a small reactance, can sometimes be tolerated. But in very-high-frequency (VHF), ultra-high-frequency (UHF) and microwave radio transmitting systems, even a small *impedance mismatch* between the load and the line can cause excessive power losses in the line.

An impedance mismatch can usually be corrected by means of *matching transformers* and/or reactances that cancel out any load reactance. This is discussed in the next chapter.

Loss in a mismatched line

When a transmission line is terminated in a resistance $R = Z_0$, then the current and the voltage are constant all along the line, if the line is perfectly lossless. The ratio of the voltage to the current, E/I, is equal to R and also equal to Z_0.

Actually, this is an idealized case; no line is completely without loss. Therefore, as a signal goes along the line from a source to a load, the current and voltage gradually decrease. But they are always in same ratio (Fig. 17-10).



17-10 In a matched line, E/I is constant, although both E and I decrease with increasing distance from the source.

Standing waves

If the load is not matched to the line, the current and voltage vary in a complicated way along the length of the line. In some places, the current is high; in other places it is low. The maxima and minima are called *loops* and *nodes* respectively. At a current loop, the voltage is minimum (a voltage node), and at a current node, the voltage is maximum (a voltage loop). Loops and nodes make it impossible to measure power by the voltmeter/ammeter method, because the current and voltage are not in constant proportion.

The loops and nodes, if graphed, form wavelike patterns along the length of the line. These patterns remain fixed over time. They are therefore known as *standing waves*—they just "stand there."

Standing-wave loss

At current loops, the loss in the line conductors is exaggerated. At voltage loops, the loss in the line dielectric is increased. At minima, the losses are decreased. But overall, in a mismatched line, the losses are greater than they are in a perfectly matched line.

This loss occurs at heat dissipation. It is true power. Any true power that goes into heating up a transmission line is wasted, because it cannot be dissipated in the load. The additional loss caused by standing waves, over and above the perfectly-matched line loss, is called *standing-wave loss.*

The greater the mismatch, the more severe the standing-wave loss becomes. The more loss a line has to begin with (that is, when it is perfectly matched), the more loss is caused by a given amount of mismatch. Standing-wave loss increases with frequency. It tends to be worst in long lengths of line at VHF, UHF, and microwaves.

Line overheating

A severe mismatch between the load and the transmission line can cause another problem: physical destruction of the line!

A feed line might be able to handle a kilowatt (1 kW) of power when it is perfectly matched. But if a severe mismatch exists and you try to feed 1 kW into the line, the extra current at the current loops can heat the conductors to the point where the dielectric material melts and the line shorts out.

It is also possible for the voltage at the voltage loops to cause arcing between the line conductors. This perforates and/or burns the dielectric, ruining the line.

When a line must be used with a mismatch, *derating functions* are required to determine how much power the line can safely handle. Manufacturers of prefabricated lines can supply you with this information.

Series resonance

One of the most important phenomena in ac circuits, especially in radio-frequency engineering, is the property of *resonance.* You've already learned that resonance is a condition that occurs when capacitive and inductive reactance cancel each other out. Resonant circuits and devices have a great many different applications in electricity and electronics.

Recall that capacitive reactance, X_C, and inductive reactance, X_L, can sometimes be equal in magnitude. They are always opposite in effect. In any circuit containing an inductance and capacitance, there will be a frequency at which $X_L = -X_C$. This is resonance. Sometimes $X_L = -X_C$ at just one frequency; in some special devices it can occur at many frequencies. Generally, if a circuit contains one coil and one capacitor, there will be one resonant frequency.

Refer to the schematic diagram of Fig. 17-11. You might recognize this as a series RLC circuit. At some particular frequency, $X_L = -X_C$. This is inevitable, if L and C are finite and nonzero. This is the *resonant frequency* of the circuit. It is abbreviated f_o.

17-11 A series RLC circuit.

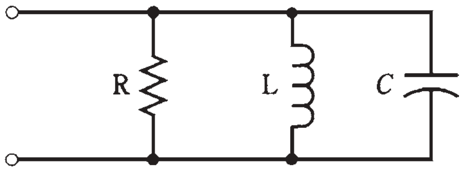
At f_o, the effects of capacitive reactance and inductive reactance cancel out. The result is that the circuit appears as a pure resistance, with a value very close to R. If

$R = 0$, that is, the resistor is a short circuit, then the circuit is called a *series LC circuit,* and the impedance at resonance will be extremely low. The circuit will offer practically no opposition to the flow of alternating current at the frequency f_o. This condition is *series resonance*.

Parallel resonance

Refer to the circuit diagram of Fig. 17-12. This is a parallel RLC circuit. You remember that, in this case, the resistance R is thought of as a conductance G, with $G = 1/R$. Then the circuit can be called a GLC circuit.

17-12 A parallel RLC circuit.



At some particular frequency f_o, the inductive susceptance B_L will exactly cancel the capacitive susceptance B_C; that is, $B_L = -B_C$. This is inevitable for some frequency f_o, as long as the circuit contains finite, nonzero inductance and finite, nonzero capacitance.

At the frequency f_O, the susceptances cancel each other out, leaving zero susceptance. The admittance through the circuit is then very nearly equal to the conductance, G, of the resistor. If the circuit contains no resistor, but only a coil and capacitor, it is called a *parallel LC circuit,* and the admittance at resonance will be extremely low. The circuit will offer great opposition to alternating current at f_o. Engineers think more often in terms of impedance than in terms of admittance; low admittance translates into high impedance. This condition is *parallel resonance*.

Calculating resonant frequency

The formula for calculating resonant frequency f_o, in terms of the inductance L in henrys and the capacitance C in farads, is

$$f_o = 0.159/(LC)^{1/2}$$

The $^1/_2$ power is the square root.

If you know L and C in henrys and farads, and you want to find f_o, do these calculations in this order: First, find the product LC, then take the square root, then divide 0.159 by this value. The result is f_o in hertz.

The formula will also work to find f_o in megahertz (MHz), when L is given in *micro*henrys (µH) and C is in *micro*farads (µF). These values are far more common than hertz, henrys, and farads in electronic circuits. Just remember that millions of hertz go with million*th*s of henrys and million*th*s of farads.

This formula works for both series-resonant and parallel-resonant RLC circuits.

Problem 17-10

Find the resonant frequency of a series circuit with an inductance of 100 µH and a capacitance of 100 pF.

First, convert the capacitance to microfarads: 100 pF = 0.000100 µF. Then find the product $LC = 100 \times 0.000100 = 0.0100$. Take the square root of this, getting 0.100. Finally, divide 0.159 by 0.100, getting $f_o = 1.59$ MHz.

Problem 17-11

Find the resonant frequency of a parallel circuit consisting of a 33-µH coil and a 47-pF capacitor.

Again, convert the capacitance to microfarads: 47 pF = 0.000047 µF. Then find the product $LC = 33 \times 0.000047 = 0.00155$. Take the square root of this, getting 0.0394. Finally, divide 0.159 by 0.0394, getting $f_o = 4.04$ MHz.

There are times when you might know the resonant frequency f_o that you want, and you need to find a particular inductance or capacitance instead. The next two problems illustrate this type of situation.

Problem 17-12

A circuit must be designed to have $f_o = 9.00$ MHz. You have a 33-pF fixed capacitor available. What size coil will be needed to get the desired resonant frequency?

Use the formula $f_o = 0.159/(LC)^{1/2}$, and plug in the values. Convert the capacitance to microfarads: 33 pF = 0.000033 µF. Then just manipulate the numbers, using familiar rules of arithmetic, until the value of L is "ferreted out":

$$9.00 = 0.159/(L \times 0.000033)^{1/2}$$
$$9.00^2 = 0.159^2/(0.000033 \times L)$$
$$81.0 = 0.0253/(0.000033 \times L)$$
$$81.0 \times 0.000033 \times L = 0.0253$$
$$0.00267 \times L = 0.0253$$
$$L = 0.0253/0.00267 = 9.48\ \mu H$$

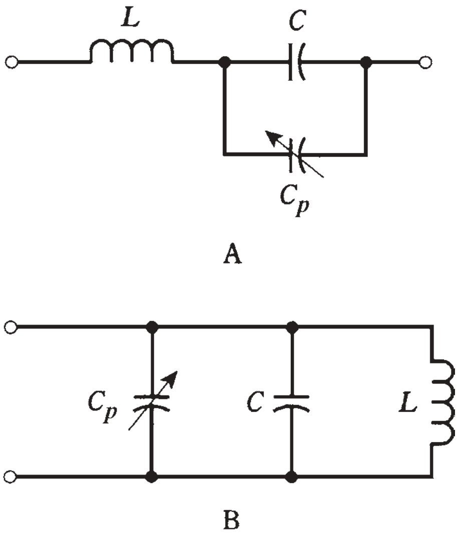
Problem 17-13

A circuit must be designed to have $f_o = 455$ kHz. A coil of 100 µH is available. What size capacitor is needed?

Convert kHz to MHz: 455 kHz = 0.455 MHz. Then the calculation proceeds in the same way as with the preceding problem:

$$f_o = 0.159/(LC)^{1/2}$$
$$0.455 = 0.159/(100 \times C)^{1/2}$$
$$0.455^2 = 0.159^2/(100 \times C)$$
$$0.207 = 0.0253/(100 \times C)$$
$$0.207 \times 100 \times C = 0.0253$$
$$20.7 \times C = 0.0253$$
$$C = 0.0253/20.7 = 0.00122\ \mu F = 1220\ pF$$

In practical circuits, variable inductors and/or variable capacitors are often placed in tuned circuits, so that small errors in the frequency can be compensated for. The most common approach is to design the circuit for a frequency slightly higher than f_o, and to use a *padder capacitor* in parallel with the main capacitor (Fig. 17-13).



17-13 Padding capacitors (C_p) allow adjustment of resonant frequency in a series LC circuit (A) or a parallel LC circuit (B).

Resonant devices

While resonant circuits often consist of coils and capacitors in series or parallel, there are other kinds of hardware that exhibit resonance. Some of these are as follows.

Crystals

Pieces of quartz, when cut into thin wafers and subjected to voltages, will vibrate at high frequencies. Because of the physical dimensions of such a *crystal,* these vibrations occur at a precise frequency f_o, and also at whole-number multiples of f_o. These multiples, $2f_o$, $3f_o$, $4f_o$, and so on, are called *harmonics.* The frequency f_o is called the *fundamental frequency* of the crystal.

Quartz crystals can be made to act like LC circuits in electronic devices. A crystal exhibits an impedance that varies with frequency. The reactance is zero at f_o and the harmonic frequencies.

Cavities

Lengths of metal tubing, cut to specific dimensions, exhibit resonance at very-high, ultra-high, and microwave frequencies. They work in much the same way as musical instruments resonate with sound waves. But the waves are electromagnetic, rather than acoustic.

Cavities, also called cavity *resonators,* have reasonable lengths at frequencies above about 150 MHz. Below this frequency, a cavity can be made to work, but it will be long and unwieldy. Like crystals, cavities resonate at a fundamental frequency f_o, and also at harmonic frequencies.

Sections of transmission line

When a transmission line is cut to $^1/_4$ wavelength, or to any whole-number multiple of this, it behaves as a resonant circuit. The most common length for a transmission-line resonantor is a quarter wavelength. Such a piece of transmission line is called a *quarter-wave section*.

When a quarter-wave section is short-circuited at the far end, it acts like a parallel-resonant LC circuit, and has a high impedance at the resonant frequency f_o. When it is open at the far end, it acts as a series-resonant LC circuit, and has a low impedance at f_o. Thus, a quarter-wave section turns a short circuit into an open circuit and vice-versa, at a specific frequency f_o.

The length of a quarter-wave section depends on f_o. It also depends on how fast the electromagnetic energy travels along the line. This speed is specified in terms of a *velocity factor*, abbreviated v. The value of v is given as a fraction of the speed of light. Typical transmission lines have velocity factors ranging from about 0.66 to 0.95. This factor is provided by the manufacturers of prefabricated lines such as coaxial cable.

If the frequency in megahertz is f_o and the velocity factor of a line is v, then the length L_{ft} of a quarter-wave section of transmission line, in feet, is

$$L_{ft} = 246v/f_o$$

The length in meters, L_m, is

$$L_m = 75.0v/f_o$$

Problem 17-14

How many feet long is a quarter-wave section of transmission line at 7.05 MHz, if the velocity factor is 0.800?

Just use the formula

$$\begin{aligned} L_{ft} &= 246v/f_o \\ &= (246 \times 0.800)/7.05 \\ &= 197/7.05 = 27.9 \text{ feet} \end{aligned}$$

Antennas

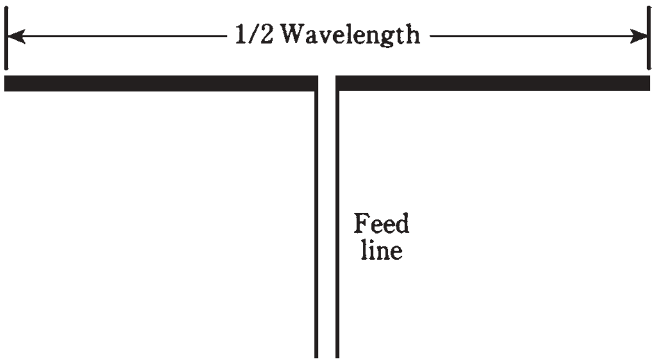
Many types of antennas exhibit resonant properties. The simplest type of resonant antenna, and the only kind that will be mentioned here, is the center-fed, half-wavelength *dipole antenna* (Fig. 17-14).

The length L_{ft}, in feet, for a 1/2-wave dipole at a frequency of f_o MHz is given by the following formula:

$$L_{ft} = 468/f_o$$

This takes into account the fact that electromagnetic fields travel along a wire at about 95 percent of the speed of light. If L_m, is specified in meters, then

$$L_m = 143/f_o$$



17-14 A half-wave dipole antenna.

A half-wave dipole has a purely resistive impedance of about 70 Ω at resonance. This is like a series-resonant RLC circuit with $R = 70\ \Omega$.

A half-wave dipole is resonant at all harmonics of its fundamental frequency f_0. The dipole is a full wavelength long at $2f_0$; it is 3/2 wavelength long at $3f_0$; it is two full wavelengths long at $4f_0$ and so on.

At f_0 and odd harmonics, that is, at f_0, $3f_0$, $5f_0$ and so on, the antenna behaves like a series-resonant RLC circuit with a fairly low resistance. At even harmonics, that is, at $2f_0$, $4f_0$, $6f_0$, and so on, the antenna acts like a parallel-resonant RLC circuit with a high resistance.

But, you say, there's no resistor in the diagram of Fig. 17-14! Where does the resistance come from? This is an interesting phenomenon that all antennas have. It is called *radiation resistance,* and is a crucial factor in the design of any antenna system.

When electromagnetic energy is fed into an antenna, power flies away into space as radio waves. This is a form of true power. True power is always dissipated in a resistance. Although you don't see any resistor in Fig. 17-14, the radiation of radio waves is, in effect, power dissipation in a resistance.

For details concerning the behavior of antennas, a text on antenna engineering is recommended. This subject is vast and many faceted. Some engineers devote their careers exclusively to antenna design and manufacture.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 or more correct. Answers are in the back of the book.

1. The power in a reactance is:
   - A. Radiated power.
   - B. True power.
   - C. Imaginary power.
   - D. Apparent power.

2. Which of the following is *not* an example of true power?
   A. Power that heats a resistor.
   B. Power radiated from an antenna.
   C. Power in a capacitor.
   D. Heat loss in a feed line.

3. The apparent power in a circuit is 100 watts, and the imaginary power is 40 watts. The true power is:
   A. 92 watts.
   B. 100 watts.
   C. 140 watts.
   D. Not determinable from this information.

4. Power factor is equal to:
   A. Apparent power divided by true power.
   B. Imaginary power divided by apparent power.
   C. Imaginary power divided by true power.
   D. True power divided by apparent power.

5. A circuit has a resistance of 300 W and an inductance of 13.5 µH in series at 10.0 MHz. What is the power factor?
   A. 0.334.
   B. 0.999.
   C. 0.595.
   D. It can't be found from the data given.

6. A series circuit has $Z = 88.4\ \Omega$, with $R = 50.0\ \Omega$. What is PF?
   A. 99.9 percent.
   B. 56.6 percent.
   C. 60.5 percent.
   D. 29.5 percent.

7. A series circuit has $R = 53.5\ \Omega$ and $X = 75.5\ \Omega$. What is PF?
   A. 70.9 percent.
   B. 81.6 percent.
   C. 57.8 percent.
   D. 63.2 percent.

8. Phase angle is equal to:
   A. Arctan Z/R.
   B. Arctan R/Z.

C. Arctan R/X.

D. Arctan X/R.

9. A wattmeter shows 220 watts of VA power in a circuit. There is a resistance of 50 Ω in series with a capacitive reactance of –20 Ω. What is the true power?

A. 237 watts.

B. 204 watts.

C. 88.0 watts.

D. 81.6 watts.

10. A wattmeter shows 57 watts of VA power in a circuit. The resistance is known to be 50 Ω, and the true power is known to be 40 watts. What is the absolute-value impedance?

A. 50 Ω.

B. 57 Ω.

C. 71 Ω.

D. It can't be calculated from this data.

11. Which of the following is the most important consideration in a transmission line?

A. The characteristic impedance.

B. The resistance.

C. Minimizing the loss.

D. The VA power.

12. Which of the following does *not* increase the loss in a transmission line?

A. Reducing the power output of the source.

B. Increasing the degree of mismatch between the line and the load.

C. Reducing the diameter of the line conductors.

D. Raising the frequency.

13. A problem that standing waves can cause is:

A. Feed line overheating.

B. Excessive power loss.

C. Inaccuracy in power measurement.

D. All of the above.

14. A coil and capacitor are in series. The inductance is 88 mH and the capacitance is 1000 pF. What is the resonant frequency?

A. 17 kHz.

B. 540 Hz.

C. 17 MHz.

D. 540 kHz.

15. A coil and capacitor are in parallel, with $L = 10.0$ µH and $C = 10$ pF. What is f_0?

A. 15.9 kHz.

B. 5.04 MHz.

C. 15.9 MHz.

D. 50.4 MHz.

16. A series-resonant circuit is to be made for 14.1 MHz. A coil of 13.5 µH is available. What size capacitor is needed?

A. 0.945 µF.

B. 9.45 pF.

C. 94.5 pF.

D. 945 pF.

17. A parallel-resonant circuit is to be made for 21.3 MHz. A capacitor of 22.0 pF is available. What size coil is needed?

A. 2.54 mH.

B. 254 µH.

C. 25.4 µH.

D. 2.54 µH.

18. A 1/4-wave line section is made for 21.1 MHz, using cable with a velocity factor of 0.800. How many meters long is it?

A. 11.1 m.

B. 3.55 m.

C. 8.87 m.

D. 2.84 m.

19. The fourth harmonic of 800 kHz is:

A. 200 kHz.

B. 400 kHz.

C. 3.20 MHz.

D. 4.00 MHz.

20. How long is a 1/2-wave dipole for 3.60 MHz?

A. 130 feet.

B. 1680 feet.

C. 39.7 feet.

D. 515 feet.

18
CHAPTER

Transformers and impedance matching

IN ELECTRICITY AND ELECTRONICS, TRANSFORMERS ARE EMPLOYED IN VARIOUS ways . Transformers are used to obtain the right voltage for the operation of a circuit or system. Transformers can match impedances between a circuit and a load, or between two different circuits. Transformers can be used to provide dc isolation between electronic circuits while letting ac pass. Another application is to mate balanced and unbalanced circuits, feed systems, and loads.

Principle of the transformer

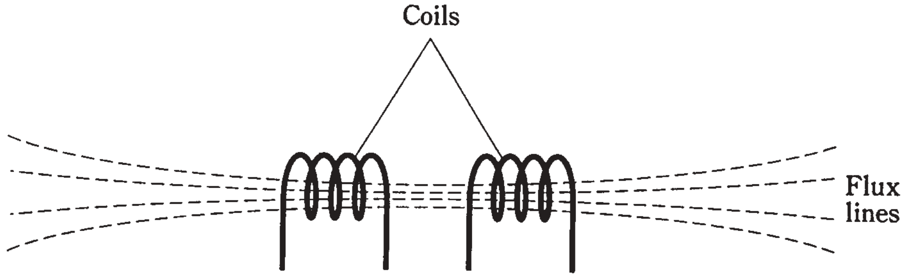
When two wires are near each other, and one of them carries a fluctuating current, a current will be induced in the other wire. This effect is known as *electromagnetic induction*. All ac transformers work according to the principle of electromagnetic induction. If the first wire carries sine-wave ac of a certain frequency, then the *induced current* will be sine-wave ac of the same frequency in the second wire.

The closer the two wires are to each other, the greater the induced current will be, for a given current in the first wire. If the wires are wound into coils and placed along a common axis (Fig. 18-1), the induced current will be greater than if the wires are straight and parallel. Even more *coupling,* or efficiency of induced-current transfer, is obtained if the two coils are wound one atop the other.

The first coil is called the *primary winding,* and the second coil is known as the *secondary winding.* These are often spoken of as simply the primary and secondary.

The induced current creates a voltage across the secondary. In a *step-down* transformer, the secondary voltage is less than the primary voltage. In a *step-up* transformer, the secondary voltage is greater than the primary voltage. The primary voltage is abbreviated E_{pri}, and the secondary voltage is abbreviated E_{sec}. Unless otherwise stated, effective (rms) voltages are always specified.

The windings of a transformer have inductance because they are coils. The required inductances of the primary and secondary depend on the frequency of operation, and also on the resistive part of the impedance in the circuit. As the frequency increases, the needed inductance decreases. At high-resistive impedances, more inductance is generally needed than at low-resistive impedances



18-1 Magnetic flux between two coils of wire.

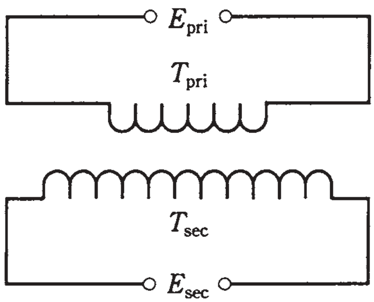
Turns ratio

The *turns ratio* in a transformer is the ratio of the number of turns in the primary, T_{pri}, to the number of turns in the secondary, T_{sec}. This ratio is written T_{pri}:T_{sec} or T_{pri}/T_{sec}.

In a transformer with excellent primary-to-secondary coupling, the following relationship always holds:

$$E_{pri}/E_{sec} = T_{pri}/T_{sec}$$

That is, the primary-to-secondary voltage ratio is always equal to the primary-to-secondary turns ratio (Fig. 18-2).



18-2 Primary and secondary turns and voltages in a transformer. See text for discussion.

Problem 18-1

A transformer has a primary-to-secondary turns ratio of exactly 9:1. The voltage at the primary is 117 V. What is the voltage at the secondary?

This is a step-down transformer. Simply plug in the numbers in the above equation and solve for E_{sec}:

$$E_{\text{pri}}/E_{\text{sec}} = T_{\text{pri}}/T_{\text{sec}}$$
$$117/E_{\text{sec}} = 9/1 = 9$$
$$1/E_{\text{sec}} = 9/117$$
$$E_{\text{sec}} = 117/9 = 13 \text{ V}$$

Problem 18-2

A transformer has a primary-to-secondary turns ratio of exactly 1:9. The voltage at the primary is 117 V. What is the voltage at the secondary?

This is a step-up transformer. Plug in numbers again:

$$117/E_{\text{sec}} = 1/9$$
$$E_{\text{sec}}/117 = 9/1 = 9$$
$$E_{\text{sec}} = 9 \times 117 = 1053 \text{ V}$$

This can be rounded off to 1050 V.

A step-down transformer always has a primary-to-secondary turns ratio greater than 1, and a step-up transformer has a primary-to-secondary turns ratio less than 1.

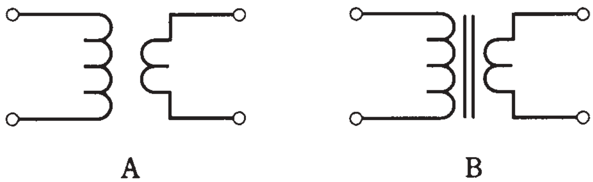
Sometimes the secondary-to-primary turns ratio is given. This is the reciprocal of the primary-to-secondary turns ratio, written $T_{\text{sec}}/T_{\text{pri}}$. In a step-down unit, $T_{\text{sec}}/T_{\text{pri}} < 1$; in a step-up unit, $T_{\text{sec}}/T_{\text{pri}} > 1$.

When you hear someone say that such-and-such a transformer has a certain turns ratio, say 10:1, you need to be sure of which ratio is meant, $T_{\text{pri}}/T_{\text{sec}}$ or $T_{\text{sec}}/T_{\text{pri}}$! If you get it wrong, you'll have the secondary voltage off by a factor of the *square* of the turns ratio. You might be thinking of 12 V when the engineer is talking about 1200 V. One way to get rid of doubt is to ask, "Step-up or step-down?"

Transformer cores

If a ferromagnetic substance such as iron, powdered iron, or ferrite is placed within the pair of coils, the extent of coupling is increased far above that possible with an air core. But this improvement in coupling takes place with a price; some energy is invariably lost as heat in the core. Also, ferromagnetic transformer cores limit the frequency at which the transformer will work well.

The schematic symbol for an air-core transformer consists of two inductor symbols back-to-back (Fig. 18-3A). If a ferromagnetic core is used, two parallel lines are added to the schematic symbol (Fig. 18-3B).



18-3 Schematic symbols for air-core (A) and ferromagnetic-core (B) transformers.

Laminated iron

In transformers for 60-Hz utility ac, and also at low audio frequencies, sheets of silicon steel, glued together in layers, are often employed as transformer cores. The silicon steel is sometimes called *transformer iron,* or simply iron.

The reason layering is used, rather than making the core from a single mass of metal, is that the magnetic fields from the coils cause currents to flow in a solid core. These *eddy currents* go in circles, heating up the core and wasting energy that would otherwise be transferred from the primary to the secondary. Eddy currents are choked off by breaking up the core into layers, so that currents cannot flow very well in circles.

Another type of loss, called *hysteresis loss,* occurs in any ferromagnetic transformer core. Hysteresis is the tendency for a core material to be "sluggish" in accepting a fluctuating magnetic field. Laminated cores exhibit high hysteresis loss above audio frequencies, and are therefore not good above a few kilohertz.

Ferrite and powdered iron

At frequencies up to several megahertz, *ferrite* works well for radio-frequency (RF) transformers. This material has high permeability and concentrates the flux efficiently. High permeability reduces the number of turns needed in the coils. But at frequencies higher than a few megahertz, ferrite begins to show loss, and is no longer effective.

For work well into the very-high-frequency (VHF) range, or up to 100 MHz or more, *powdered iron* cores work well. The permeability of powdered iron is less than that of ferrite, but at high frequencies, it is not necessary to have high magnetic permeability. In fact, at radio frequencies above a few megahertz, air core coils are often preferred, especially in transmitting amplifiers. At frequencies above several hundred megahertz, ferromagnetic cores can be dispensed with entirely.

Transformer geometry

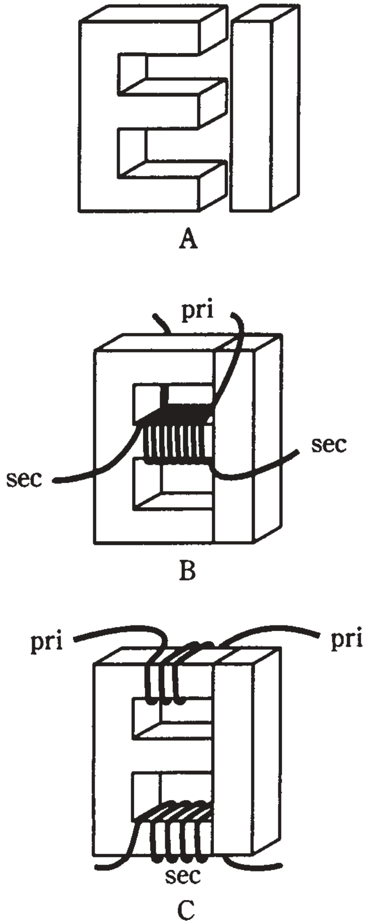
The shape of a transformer depends on the shape of its core. There are several different core geometries commonly used with transformers.

Utility transformers

A common core for a power transformer is the *E core*, so named because it is shaped like the capital letter E. A bar, placed at the open end of the E, completes the core once the coils have been wound on the E-shaped section (Fig. 18-4A).

The primary and secondary windings can be placed on an E core in either of two ways.

The simpler winding method is to put both the primary and the secondary around the middle bar of the E (Fig. 18-4B). This is called the *shell method* of transformer winding. It provides maximum coupling between the windings. However, the shell-winding scheme results in a considerable capacitance between the primary and the secondary. This capacitance can sometimes be tolerated; sometimes it cannot. Another disadvantage of the shell geometry is that, when windings are placed one atop the other, the transformer cannot handle very much voltage.



18-4 Utility transformer E core (A), shell winding method (B), and core winding method (C).

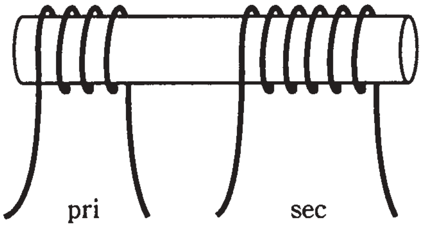
Another winding method is the *core method.* In this scheme, the primary is placed at the bottom of the E section, and the secondary is placed at the top (Fig. 18-4C). The coupling occurs via magnetic flux in the core. The capacitance between the primary and secondary is much lower with this type of winding. Also, a core-wound transformer can handle higher voltages than the shell-wound transformer. Sometimes the center part of the E is left out of the core when this winding scheme is used.

Shell-wound and core-wound transformers are almost universally employed at 60 Hz. These configurations are also common at audio frequencies.

Solenoidal core

A pair of cylindrical coils, wound around a rod-shaped piece of powdered iron or ferrite, was once a common configuration for transformers at radio frequencies. Sometimes this type of transformer is still seen, although it is most often used as a *loopstick antenna* in portable radio receivers and in radio direction-finding equipment.

The coil windings might be placed one atop the other, or they might be separated (Fig. 18-5) to reduce the capacitance between the primary and secondary.

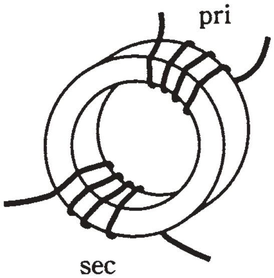


18-5 Solenoidal-core transformer.

In a loopstick antenna, the primary serves to pick up the radio signals. The secondary winding provides the best *impedance match* to the first amplifier stage, or *front end,* of the radio. The use of transformers for impedance matching is discussed later in this chapter.

Toroidal core

In recent years, the *toroidal core* has become the norm for winding radio-frequency transformers. The toroid is a donut-shaped ring of powdered iron or ferrite. The coils are wound around the donut. The primary and secondary might be wound one over the other, or they might be wound over different parts of the core (Fig. 18-6). As with other transformers, when the windings are one atop the other, there is more inter-winding capacitance than when they are separate.



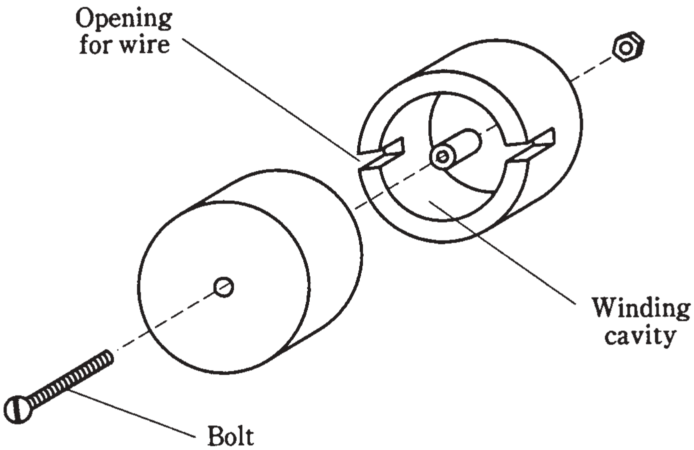
18-6 Toroidal-core transformer.

Toroids confine practically all the magnetic flux within the core material. This allows toroidal coils and transformers to be placed near other components without inductive interaction. Also, a toroidal coil or transformer can be mounted directly on a metal chassis, and the operation will not be affected (assuming the wire is insulated).

A toroidal core provides considerably more inductance per turn, for the same kind of ferromagnetic material, than a solenoidal core. It is not uncommon to see toroidal coils or transformers that have inductances of 10 mH or even 100 mH.

Pot core

Still more inductance per turn can be obtained with a *pot core*. This is a shell of ferromagnetic material that wraps around a loop-shaped coil. The core comes in two halves (Fig. 18-7). You wind the coil inside one of the halves, and then bolt the two together. The final core completely surrounds the loop, and the magnetic flux is confined to the core material.



18-7 Exploded view of pot core (windings not shown).

Like the toroid, the pot core is self-shielding. There is essentially no coupling to external components. A pot core can be used to wind a single, high-inductance coil; sometimes the value can be upwards of 1 H.

In a pot-core transformer, the primary and secondary must always be wound on top of, or right next to, each other; this is unavoidable because of the geometry of the shell. Therefore, the interwinding capacitance of a pot-core transformer is always rather high.

Pot cores are useful at the lower frequencies. They are generally not employed at higher frequencies because it isn't necessary to get that much inductance.

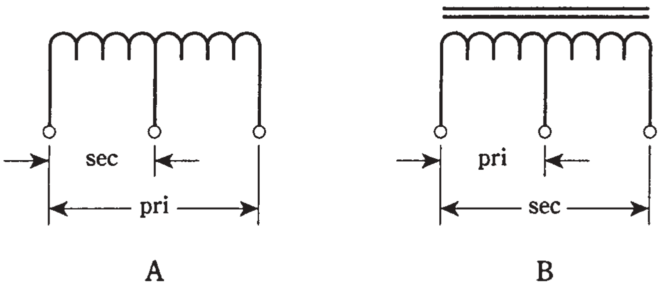
The autotransformer

Sometimes, it's not necessary to provide dc isolation between the primary and secondary windings of a transformer. Then an *autotransformer* can be used. This has a single, tapped winding. Its schematic symbol is shown in Fig. 18-8A for an air core, and Fig. 18-8B for a ferromagnetic core.

An autotransformer can be either a step-down or a step-up device. In Fig. 18-8, the autotransformer at A is step-down, and the one at B is step-up.

An autotransformer can have an air core, or it can be wound on any of the aforementioned types of ferromagnetic cores. You'll sometimes see this type of transformer in a radio-frequency receiver or transmitter. It works quite well in impedance-matching applications, and also in solenoidal loopsticks.

Autotransformers are occasionally, but not often, used at audio frequencies and in 60-Hz utility wiring. In utility circuits, autotransformers can step down by a large factor, but they aren't used to step up by more than a few percent.



18-8 Schematic symbols for autotransformers. At A, air core and at B, ferromagnetic core. The unit at A steps down and the one at B steps up.

Power transformers

Any transformer used in the 60-Hz utility line, intended to provide a certain rms ac voltage for the operation of electrical circuits, is a *power transformer.* Power transformers exist in sizes ranging from smaller than a tennis ball to as big as a room.

At the generating plant

The largest transformers are employed right at the place where electricity is generated. Not surprisingly, high-energy power plants have bigger transformers that develop higher voltages than low-energy, local power plants. These transformers must be able to handle not only huge voltages, but large currents as well. Their primaries and secondaries must withstand a product EI of *volt-amperes* that is equal to the power P ultimately delivered by the transmission line.

When electrical energy must be sent over long distances, extremely high voltages are used. This is because, for a given amount of power ultimately dissipated by the loads, the current is lower when the voltage is higher. Lower current translates into reduced loss in the transmission line.

Recall the formula $P = EI$, where P is the power in watts, E is the voltage in volts, and I is the current in amperes. If you can make the voltage 10 times larger, for a given power level, then the current is reduced to 1/10 as much. The *ohmic losses* in the wires are proportional to the *square* of the current; remember that $P = I^2R$, where P is the power in watts, I is the current in amperes, and R is the resistance in ohms.

Engineers can't do much about the wire resistance or the power consumed by the loads, but they can adjust the voltage, and thereby the current. Increasing the voltage 10 times will cut the current to 0.1 its previous value. This will render the I^2R loss $(0.1)^2 = 0.01$ (1 percent!) as much as before.

For this reason, regional power plants have massive transformers capable of generating hundreds of thousands of volts. A few can produce 1,000,000 V rms. A transmission line that carries this much voltage requires gigantic insulators, sometimes several meters long, and tall, sturdy towers.

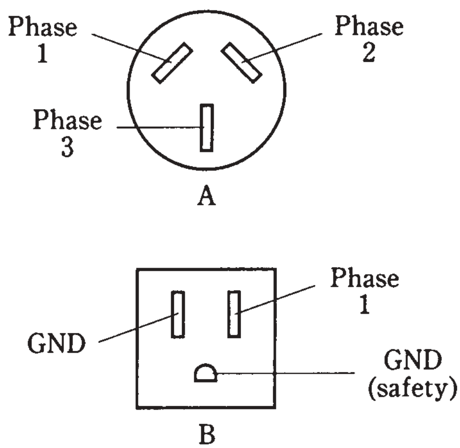
Along the line

Extreme voltage is good for *high-tension* power transmission, but it's certainly of no use to an average consumer. The wiring in a high-tension system must be done using precautions to prevent arcing (sparking) and short circuits. Personnel must be kept at least several feet, or even tens of feet, away from the wires. Can you imagine trying to use an appliance, say a home computer, by plugging it into a 500-kV electrical outlet? A bolt of artificial lightning would dispatch you before you even got near the receptacle.

Medium-voltage power lines branch out from the major lines, and step-down transformers are used at the branch points. These lines fan out to still lower-voltage lines, and step-down transformers are employed at these points, too. Each transformer must have windings heavy enough to withstand the product $P = EI$, the amount of power delivered to all the subscribers served by that transformer, at periods of peak demand.

Sometimes, such as during a heat wave, the demand for electricity rises above the normal peak level. This loads down the circuit to the point that the voltage drops several percent. This is a *brownout.* If consumption rises further still, a dangerous current load is placed on one or more intermediate power transformers. Circuit breakers in the transformers protect them from destruction by opening the circuit. Then there is a temporary *blackout.*

Finally, at individual homes and buildings, transformers step the voltage down to either 234 V or 117 V. Usually, 234-V electricity is in three phases, each separated by 120 degrees, and each appearing at one of the three prongs in the outlet (Fig. 18-9A). This voltage is commonly employed with heavy appliances, such as the kitchen oven/stove (if they are electric), heating (if it is electric), and the laundry washer and dryer. A 117-V outlet supplies just one phase, appearing between two of the three prongs in the outlet. The third prong is for ground (Fig. 18-9B).



18-9 At A, three-phase, 234-Vac outlet. At B, single-phase, 117-Vac outlet.

In electronic devices

The lowest level of power transformer is found in electronic equipment such as television sets, ham radios, and home computers.

Most solid-state devices use low voltages, ranging from about 5 V up to perhaps 50 V. This equipment needs step-down power transformers in its power supplies. Solid-state equipment usually (but not always) consumes relatively little power, so the transformers are usually not very bulky. The exception is high-powered audio-frequency or radio-frequency amplifiers, whose transistors can demand more than 1000 watts (1 kW) in some cases. At 12 V, this translates to 90 A or more.

Television sets have cathode-ray tubes that need several hundred volts. This is derived by using a step-up transformer in the power supply. Such transformers don't have to supply a lot of current, though, so they are not very big or heavy. Another type of device that needs rather high voltage is a ham-radio amplifier with vacuum tubes. Such an amplifier requires from 2 kV to 5 kV.

Any voltage higher than about 12 V should be treated with respect. *The voltages in televisions and ham radios are extremely dangerous, even after the equipment has been switched off. Do not try to service such equipment unless you are trained to do so!*

Audio-frequency transformers

Transformers for use at audio frequency (AF) are similar to those employed for 60-Hz electricity. The differences are that the frequency is somewhat higher (up to 20 kHz), and that audio signals exist in a band of frequencies (20 Hz to 20 kHz) rather than at just one frequency.

Most audio transformers look like, and are constructed like miniature utility transformers. They have laminated E cores with primary and secondary windings wound around the cross bars, as shown in Fig. 18-4.

Audio transformers can be either the step-up or step-down type. However, rather than being made to produce a specific voltage, audio transformers are designed to match impedances.

Audio circuits, and in fact all electronic circuits that handle sine-wave or complex-wave signals, exhibit impedance at the input and output. The load has a certain impedance; a source has another impedance. Good audio design strives to minimize the reactances in the circuitry, so that the absolute-value impedance, Z, is close to the resistance R in the complex vector $R + jX$. This means that X must be zero or nearly zero.

In the following discussion of impedance-matching transformers, both at audio and at radio frequencies, assume that the reactance is zero, so that the impedance is purely resistive, that is, $Z = R$.

Isolation transformers

One useful function of a transformer is that it can provide isolation between electronic circuits. While there is *inductive coupling* in a transformer, there is comparatively little *capacitive coupling*. The amount of capacitive coupling can be reduced by using cores that minimize the number of wire turns needed in the windings, and by keeping the windings separate from each other (rather than overlapping).

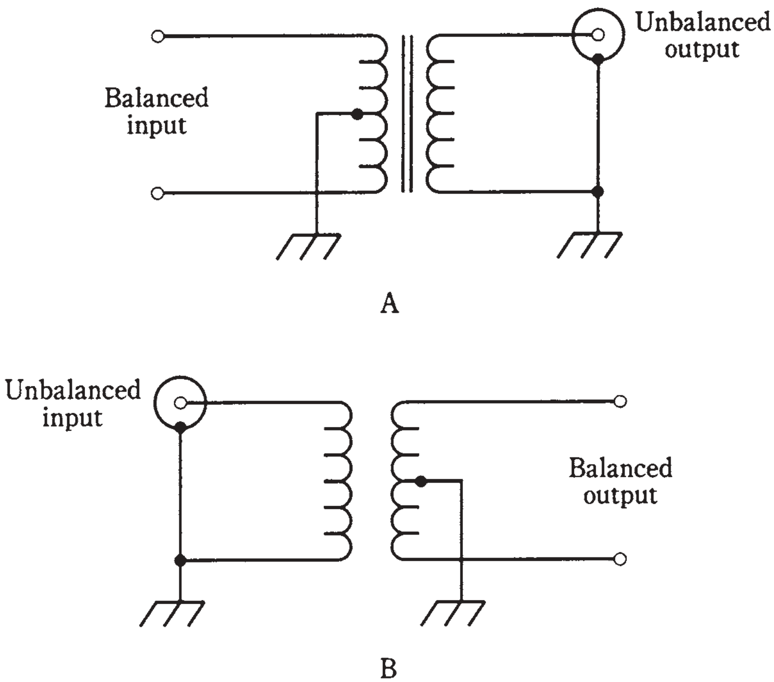
Balanced and unbalanced loads

A *balanced load* is one whose terminals can be reversed without significantly affecting circuit behavior. A plain resistor is a good example. The two-wire antenna input in a television receiver is another example of a balanced load. A *balanced transmission line* is usually a two-wire line, such as television antenna ribbon.

An *unbalanced load* is a load that must be connected a certain way; switching its leads will result in improper circuit operation. Some radio antennas are of this type. Usually, unbalanced sources and loads have one side connected to ground. The coaxial input of a television receiver is unbalanced; the shield (braid) of the cable is grounded. An *unbalanced transmission line* is usually a coaxial line, such as you find in a cable television system.

Normally, you cannot connect an unbalanced line to a balanced load, or a balanced line to an unbalanced load, and expect good performance from an electrical or electronic system. But a transformer can allow for mating between these two types of systems.

In Fig. 18-10A, a balanced-to-unbalanced transformer is shown. Note that the balanced side is center-tapped, and the tap is grounded. In Fig. 18-10B, an unbalanced-to-balanced transformer is illustrated. Again, the balanced side has a grounded center tap.



18-10 At A, a balanced-to-unbalanced transformer. At B, an unbalanced-to-balanced transformer.

The turns ratio of a balanced-to-unbalanced ("balun") or unbalanced-to-balanced ("unbal") transformer might be 1:1, but it doesn't have to be. If the impedances of the balanced and unbalanced parts of the systems are the same, then a 1:1 turns ratio is ideal. But if the impedances differ, the turns ratio should be such that the impedances are matched. This is discussed in the section on impedance transfer ratio that follows.

Transformer coupling

Transformers are sometimes used between amplifier stages in electronic equipment where a large *amplification factor* is needed. There are other methods of *coupling* from one amplifier stage to another, but transformers offer some advantages, especially in radio-frequency receivers and transmitters.

Part of the problem in getting a radio to work is that the amplifiers must operate in a stable manner. If there is too much feedback, a series of amplifiers will oscillate, and this will severely degrade the performance of the radio. Transformers that minimize the capacitance between the amplifier stages, while still transferring the desired signals, can help to prevent this oscillation.

Impedance-transfer ratio

One of the most important applications of transformers is in audio-frequency (AF) and radio-frequency (RF) electronic circuits. In these applications, transformers are generally employed to match impedances. Thus, you might hear of an *impedance step-up* transformer, or an *impedance step-down* device.

The *impedance-transfer ratio* of a transformer varies according to the square of the turns ratio, and also according to the square of the voltage-transfer ratio. Recall the formula for voltage-transfer ratio:

$$E_{pri}/E_{sec} = T_{pri}/T_{sec}$$

If the input and output, or source and load, impedances are purely resistive, and are denoted by Z_{pri} (at the primary winding) and Z_{sec} (at the secondary), then

$$Z_{pri}/Z_{sec} = (T_{pri}/T_{sec})^2$$

and

$$Z_{pri}/Z_{sec} = (E_{pri}/E_{sec})^2$$

The inverses of these formulas, in which the turns ratio or voltage-transfer ratio are expressed in terms of the impedance-transfer ratio, are

$$T_{pri}/T_{sec} = (Z_{pri}/Z_{sec})^{1/2}$$

and

$$E_{pri}/E_{sec} = (Z_{pri}/Z_{sec})^{1/2}$$

The $^1/_2$ power is the same thing as the square root.

Problem 18-3

A transformer is needed to match an input impedance of 50.0 **Ω**, purely resistive, to an output impedance of 300 **Ω**, also purely resistive. What should the ratio T_{pri}/T_{sec} be?

The required transformer will have a step-up impedance ratio of Z_{pri}/Z_{sec} = 50.0/300 = 1:6.00. From the above formulas,

$$\begin{aligned} T_{pri}/T_{sec} &= (Z_{pri}/Z_{sec})^{1/2} \\ &= (1/6.00)^{1/2} = 0.16667^{1/2} = 0.40829 \end{aligned}$$

A couple of extra digits are included (as they show up on the calculator) to prevent the sort of error introduction you recall from earlier chapters. The decimal value 0.40829 can be changed into ratio notation by taking its reciprocal, and then writing "1:" followed by that reciprocal value

$$0.40829 = 1{:}(1/0.40829) = 1{:}2.4492$$

This can be rounded to three significant figures, or 1:2.45. This is the primary-to-secondary turns ratio for the transformer. The secondary winding has 2.45 times as many turns as the primary winding.

Problem 18-4

A transformer has a primary-to-secondary turns ratio of 4.00:1. The load, connected to the transformer output, is a pure resistance of 37.5 Ω. What is the impedance at the primary?

The impedance-transfer ratio is equal to the square of the turns ratio. Therefore,

$$\begin{aligned} Z_{pri}/Z_{sec} &= (T_{pri}/T_{sec})^2 \\ &= (4.00/1)^2 = 4.00^2 = 16.0 \end{aligned}$$

This can be written 16.0:1. The input (primary) impedance is 16.0 times the secondary impedance. We know that the secondary impedance, Z_{sec} is 37.5 Ω. Therefore,

$$\begin{aligned} Z_{pri} &= 16.0 \times Z_{sec} \\ &= 16.0 \times 37.5 = 600 \end{aligned}$$

Anything connected to the transformer primary will "see" a purely resistive impedance of 600 Ω.

Radio-frequency transformers

In radio receivers and transmitters, transformers can be categorized generally by the method of construction used. Some have primary and secondary windings, just like utility and audio units. Others employ transmission-line sections. These are the two most common types of transformer found at radio frequencies.

Coil types

In the wound radio-frequency (RF) transformer, powdered-iron cores can be used up to quite high frequencies. Toroidal cores are most common, because they are self-shielding (all of the magnetic flux is confined within the core material). The number of turns depends on the frequency, and also on the permeability of the core.

In high-power applications, air-core coils are sometimes used, because air, although it has a low permeability, also has extremely low hysteresis loss. The disadvantage of air-core coils is that some of the magnetic flux extends outside of the coil. This affects the performance of the transformer when it must be placed in a cramped space, such as in a transmitter final-amplifier compartment.

A major advantage of coil type transformers, especially when they are wound on toroidal cores, is that they can be made to work over a wide band of frequencies, such as from 3.5 MHz to 30 MHz. These are called *broadband transformers.*

Transmission-line types

As you recall, any transmission line has a characteristic impedance, or Z_o, that depends on the line construction. This property is sometimes used to make impedance transformers out of coaxial or parallel-wire line.

Transmission-line transformers are always made from quarter-wave sections. From the previous chapter, remember the formula for the length of a quarter-wave section, that is,

$$L_{ft} = 246v/f_o$$

where L_{ft} is the length of the section in feet, v is the velocity factor expressed as a fraction, and f_o is the frequency of operation in megahertz. If the length L_m is in meters, then:

$$L_m = 75v/f_o$$

In the last chapter, you saw how a short circuit is changed into an open circuit, and vice versa, by a quarter-wave section of line. What happens to a pure resistive impedance at one end of such a line? What will be "seen" at the opposite end?

Let a quarter-wave section of line, with characteristic impedance Z_o, be terminated in a purely resistive impedance R_{out}. Then the input impedance is also a pure resistance R_{in}, and the following relationship holds:

$$Z_o^2 = R_{in}R_{out}$$

This is illustrated in Fig. 18-11. This formula can be broken down to solve for R_{in} in terms of R_{out}, or vice versa:

$$R_{in} = Z_o^2/R_{out}$$

and

$$R_{out} = Z_o^2/R_{in}.$$

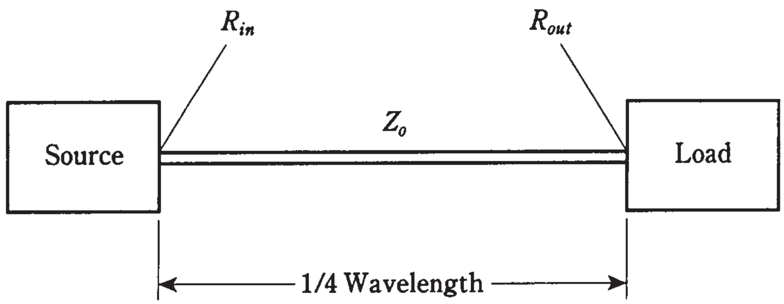
These relationships hold at the frequency, f_o, for which the line is 1/4 wavelength long. Neglecting line losses, these relationships will also hold at the *odd harmonics* of f_o, that is, at $3f_o$, $5f_o$, $7f_o$, and so on. At other frequencies, the line will not act as a transformer, but instead, will behave in complicated ways that are beyond the scope of this discussion.

Quarter-wave transformers are most often used in antenna systems, especially at the higher frequencies, where their dimensions become practical.

Problem 18-5

An antenna has a purely resistive impedance of 100 Ω. It is connected to a 1/4-wave section of 75-Ω coaxial cable. What will be the impedance at the input end of the section?

Use the formula from above:



18-11 A quarter-wave matching section of transmission line. Abbreviations are discussed in the text.

$$
\begin{aligned}
R_{\text{in}} &= Z_{\text{o}}^2/R_{\text{out}} \\
&= 75^2/100 = 5625/100 \\
&= 56\ \Omega
\end{aligned}
$$

Problem 18-6

An antenna is known to have a pure resistance of 600 Ω. You want to match it to 50.0 Ω pure resistance. What is the characteristic impedance needed for a quarter-wave matching section?

Use this formula:

$$
\begin{aligned}
Z_{\text{o}}^2 &= R_{\text{in}}R_{\text{out}} \\
Z_{\text{o}}^2 &= 600 \times 50 = 30{,}000 \\
Z_{\text{o}}^2 &= (30{,}000)^{1/2} = 173\ \Omega
\end{aligned}
$$

The challenge is to find a line that has this particular Z_{o}. Commercially manufactured lines come in standard Z_{o} values, and a perfect match might not be obtainable. In that case, the closest obtainable Z_{o} should be used. In this case, it would probably be 150 Ω.

If nothing is available anywhere near the characteristic impedance needed for a quarter-wave matching section, then a coil-type transformer will probably have to be used instead. A quarter-wave matching section should be made using unbalanced line if the load is unbalanced and balanced line if the load is balanced.

The major disadvantage of quarter-wave sections is that they work only at specific frequencies. But this is often offset by the ease with which they are constructed, if radio equipment is to be used at only one frequency, or at odd-harmonic frequencies.

What about reactance?

Things are simple when there is no reactance in an ac circuit using transformers. But often, especially in radio-frequency circuits, pure resistance doesn't occur naturally. It has to be obtained by using inductors and/or capacitors to cancel the reactance out.

Reactance makes a perfect match impossible, no matter what the turns ratio or Z_0 of the transformer. A small amount of reactance can be tolerated at lower radio frequencies (below about 30 MHz). A near-perfect match becomes more important at higher frequencies.

The behavior of reactance, as it is coupled through transformer windings, is too complicated for a thorough analysis here. But if you're interested in delving into it, there are plenty of good engineering texts that deal with it in all its mathematical glory.

Recall that inductive and capacitive reactances are opposite in effect, and that their magnitudes can vary. If a load presents a complex impedance $R + jX$, with X not equal to zero, it is always possible to cancel out the reactance X by adding an equal and opposite reactance ($-X$) in the circuit. This can be done by connecting an inductor or capacitor in series with the load.

For radio communications over a wide band, adjustable impedance-matching and reactance-canceling networks can be placed between a transmitter and an antenna system. Such a device is called a *transmatch* and is popular among radio hams, who use frequencies ranging from 1.8 MHz to the microwave spectrum.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 or more correct. Answers are in the back of the book.

1. In a step-up transformer:
   - A. The primary impedance is greater than the secondary impedance.
   - B. The secondary winding is right on top of the primary.
   - C. The primary voltage is less than the secondary voltage.
   - D. All of the above.

2. The capacitance between the primary and the secondary windings of a transformer can be minimized by:
   - A. Placing the windings on opposite sides of a toroidal core.
   - B. Winding the secondary right on top of the primary.
   - C. Using the highest possible frequency.
   - D. Using a center tap on the balanced winding.

3. A transformer steps a voltage down from 117 V to 6.00 V. What is its primary-to-secondary turns ratio?
   - A. 1:380.
   - B. 380:1.
   - C. 1:19.5.
   - D. 19.5:1.

4. A step-up transformer has a primary-to-secondary turns ratio of 1:5.00. If 117 V rms appears at the primary, what is the rms voltage across the secondary?
   - A. 23.4 V.

B. 585 V.
C. 117 V.
D. 2.93 kV.

5. A transformer has a secondary-to-primary turns ratio of 0.167. This transformer is:
   A. A step-up unit.
   B. A step-down unit.
   C. Neither step-up nor step-down.
   D. A reversible unit.
6. Which of the following is *false*, concerning air cores versus ferromagnetic cores?
   A. Air concentrates the magnetic lines of flux.
   B. Air works at higher frequencies than ferromagnetics.
   C. Ferromagnetics are lossier than air.
   D. A ferromagnetic-core unit needs fewer turns of wire than an equivalent air-core unit.
7. Eddy currents cause:
   A. An increase in efficiency.
   B. An increase in coupling between windings.
   C. An increase in core loss.
   D. An increase in usable frequency range.
8. A transformer has 117 V rms across its primary and 234 V rms across its secondary. If this unit is reversed, assuming it can be done without damaging the windings, what will be the voltage at the output?
   A. 234 V.
   B. 468 V.
   C. 117 V.
   D. 58.5 V.
9. The shell method of transformer winding:
   A. Provides maximum coupling.
   B. Minimizes capacitance between windings.
   C. Withstands more voltage than other winding methods.
   D. Has windings far apart but along a common axis.
10. Which of these core types, in general, is best if you need a winding inductance of 1.5 H?
   A. Air core.
   B. Ferromagnetic solenoid core.
   C. Ferromagnetic toroid core.
   D. Ferromagnetic pot core.

11. An advantage of a toroid core over a solenoid core is:
    - A. The toroid works at higher frequencies.
    - B. The toroid confines the magnetic flux.
    - C. The toroid can work for dc as well as for ac.
    - D. It's easier to wind the turns on a toroid.

12. High voltage is used in long-distance power transmission because:
    - A. It is easier to regulate than low voltage.
    - B. The I^2R losses are lower.
    - C. The electromagnetic fields are stronger.
    - D. Smaller transformers can be used.

13. In a household circuit, the 234-V power has:
    - A. One phase.
    - B. Two phases.
    - C. Three phases.
    - D. Four phases.

14. In a transformer, a center tap would probably be found in:
    - A. The primary winding.
    - B. The secondary winding.
    - C. The unbalanced winding.
    - D. The balanced winding.

15. An autotransformer:
    - A. Works automatically.
    - B. Has a center-tapped secondary.
    - C. Has one tapped winding.
    - D. Is useful only for impedance matching.

16. A transformer has a primary-to-secondary turns ratio of 2.00:1. The input impedance is 300 Ω resistive. What is the output impedance?
    - A. 75 Ω.
    - B. 150 Ω.
    - C. 600 Ω.
    - D. 1200 Ω.

17. A resistive input impedance of 50 Ω must be matched to a resistive output impedance of 450 Ω. The primary-to-secondary turns ratio of the transformer must be:
    - A. 9.00:1.
    - B. 3.00:1.

C. 1:3.00.

D. 1:9.00.

18. A quarter-wave matching section has a characteristic impedance of 75.0 Ω. The input impedance is 50.0 Ω resistive. What is the resistive output impedance?

A. 150 Ω.

B. 125 Ω.

C. 100 Ω.

D. 113 Ω.

19. A resistive impedance of 75 Ω must be matched to a resistive impedance of 300 Ω. A quarter-wave section would need:

A. $Z_o = 188\ \Omega$.

B. $Z_o = 150\ \Omega$.

C. $Z_o = 225\ \Omega$.

D. $Z_o = 375\ \Omega$.

20. If there is reactance at the output of an impedance transformer:

A. The circuit will not work.

B. There will be an impedance mismatch, no matter what the turns ratio of the transformer.

C. A center tap must be used at the secondary.

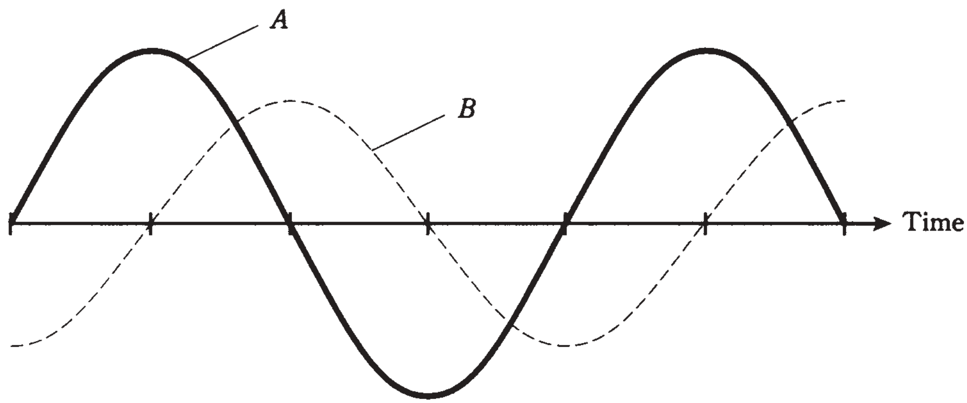
D. The turns ratio must be changed to obtain a match.

Test: Part two

DO NOT REFER TO THE TEXT WHEN TAKING THIS TEST. A GOOD SCORE IS AT least 37 correct. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. A series circuit has a resistance of 100 Ω and a capacitive reactance of -200 Ω. The complex impedance is:
   - A. −200 + j100.
   - B. 100 + j200.
   - C. 200 − j100.
   - D. 200 + j100.
   - E. 100 − j200.
2. Mutual inductance causes the net value of a set of coils to:
   - A. Cancel out, resulting in zero inductance.
   - B. Be greater than what it would be with no mutual coupling.
   - C. Be less than what it would be with no mutual coupling.
   - D. Double.
   - E. Vary, depending on the extent and phase of mutual coupling.
3. Refer to Fig. TEST 2-1. Wave A is:
   - A. Leading wave B by 90 degrees.
   - B. Lagging wave B by 90 degrees.
   - C. Leading wave B by 180 degrees.
   - D. Lagging wave B by 135 degrees.

E. Lagging wave B by 45 degrees.



TEST 2-1 Illustration for PART TWO test question 3.

4. A sine wave has a peak value of 30.0 V. Its rms value is:
   A. 21.2 V.
   B. 30.0 V.
   C. 42.4 V.
   D. 60.0 V.
   E. 90.0 V.

5. Four capacitors are connected in parallel. Their values are 100 pF each. The net capacitance is:
   A. 25 pF.
   B. 50 pF.
   C. 100 pF.
   D. 200 pF.
   E. 400 pF.

6. A transformer has a primary-to-secondary turns ratio of exactly 8.88:1. The input voltage is 234 V rms. The output voltage is:
   A. 2.08 kV rms.
   B. 18.5 kV rms.
   C. 2.97 V rms.
   D. 26.4 V rms.
   E. 20.8 V rms.

7. In a series RL circuit, as the resistance becomes small compared with the reactance, the angle of lag approaches:
   A. 0 degrees.

B. 45 degrees.

C. 90 degrees.

D. 180 degrees.

E. 360 degrees.

8. A transmission line carries 3.50 A of ac current and 150 V ac. The *true power* in the line is:

A. 525 W.

B. 42.9 W.

C. 1.84 W.

D. Meaningless; true power is dissipated, not transmitted.

E. Variable, depending on standing wave effects.

9. In a parallel configuration, susceptances:

A. Simply add up.

B. Add like capacitances in series.

C. Add like inductances in parallel.

D. Must be changed to reactances before you can work with them.

E. Cancel out.

10. A wave has a frequency of 200 kHz. How many degrees of phase change occur in a microsecond (a millionth of a second)?

A. 180 degrees.

B. 144 degrees.

C. 120 degrees.

D. 90 degrees.

E. 72 degrees.

11. At a frequency of 2.55 MHz, a 330-pF capacitor has a reactance of:

A. $-5.28\ \Omega$.

B. $-0.00528\ \Omega$.

C. $-189\ \Omega$.

D. $-18.9\text{k}\ \Omega$.

E. $-0.000189\ \Omega$.

12. A transformer has a step-up turns ratio of 1:3.16. The output impedance is 499 Ω purely resistive. The input impedance is:

A. $50.0\ \Omega$.

B. $158\ \Omega$.

C. $1.58\text{k}\ \Omega$.

D. $4.98\text{k}\ \Omega$.

E. Not determinable from the data given.

13. A complex impedance is represented by $34 - j23$. The absolute-value impedance is:

    A. 34 Ω.

    B. 11 Ω.

    C. −23 Ω.

    D. 41 Ω.

    E. 57 Ω.

14. A coil has an inductance of 750 µH. The inductive reactance at 100 kHz is:

    A. 75.0 Ω.

    B. 75.0 kΩ.

    C. 471 Ω.

    D. 47.1 kΩ.

    E. 212 Ω.

15. Two waves are 180 degrees out of phase. This is a difference of:

    A. 1/8 cycle.

    B. 1/4 cycle.

    C. 1/2 cycle.

    D. A full cycle.

    E. Two full cycles.

16. If R denotes resistance and Z denotes absolute-value impedance, then R/Z is the:

    A. True power.

    B. Imaginary power.

    C. Apparent power.

    D. Absolute-value power.

    E. Power factor.

17. Two complex impedances are in series. One is $30 + j50$ and the other is $50 - j30$. The net impedance is:

    A. $80 + j80$.

    B. $20 + j20$.

    C. $20 - j20$.

    D. $-20 + j20$.

    E. $80 + j20$.

18. Two inductors, having values of 140 µH and 1.50 mH, are connected in series. The net inductance is:

    A. 141.5 µH.

    B. 1.64 µH.

C. 0.1415 mH.

D. 1.64 mH.

E. 0.164 mH.

19. Which of the following types of capacitor is polarized?

A. Mica.

B. Paper.

C. Electrolytic.

D. Air variable.

E. Ceramic.

20. A toroidal-core coil:

A. Has lower inductance than an air-core coil with the same number of turns.

B. Is essentially self-shielding.

C. Works well as a loopstick antenna.

D. Is ideal as a transmission-line transformer.

E. Cannot be used at frequencies below about 10 MHz.

21. The efficiency of a generator:

A. Depends on the driving power source.

B. Is equal to output power divided by driving power.

C. Depends on the nature of the load.

D. Is equal to driving voltage divided by output voltage.

E. Is equal to driving current divided by output current.

22. Admittance is:

A. The reciprocal of reactance.

B. The reciprocal of resistance.

C. A measure of the opposition a circuit offers to ac.

D. A measure of the ease with which a circuit passes ac.

E. Another expression for absolute-value impedance.

23. The absolute-value impedance Z of a parallel RLC circuit, where R is the resistance and X is the net reactance, is found according to the formula:

A. $Z = R + X$.

B. $Z^2 = R^2 + X^2$.

C. $Z^2 = RX/(R^2 + X^2)$.

D. $\mathrm{Z} = 1/(R^2 + X^2)$.

E. $Z = R^2X^2/(R + X)$.

24. Complex numbers are used to represent impedance because:

A. Reactance cannot store power.

B. Reactance isn't a real physical thing.
C. They provide a way to represent what happens in resistance-reactance circuits.
D. Engineers like to work with sophisticated mathematics.
E. No! Complex numbers aren't used to represent impedance.

25. Which of the following does *not* affect the capacitance of a capacitor?
A. The mutual surface area of the plates.
B. The dielectric constant of the material between the plates (within reason).
C. The spacing between the plates (within reason).
D. The amount of overlap between plates.
E. The frequency (within reason).

26. The zero-degree phase point in an ac sine wave is usually considered to be the instant at which the amplitude is:
A. Zero and negative-going.
B. At its negative peak.
C. Zero and positive-going.
D. At its positive peak.
E. Any value; it doesn't matter.

27. The inductance of a coil can be continuously varied by:
A. Varying the frequency.
B. Varying the net core permeability.
C. Varying the current in the coil.
D. Varying the wavelength.
E. Varying the voltage across the coil.

28. Power factor is defined as the ratio of:
A. True power to VA power.
B. True power to imaginary power.
C. Imaginary power to VA power.
D. Imaginary power to true power.
E. VA power to true power.

29. A 50 Ω feed line needs to be matched to an antenna with a purely resistive impedance of 200 Ω. A quarter-wave matching section should have:
A. $Z_o = 150\ \Omega$.
B. $Z_o = 250\ \Omega$.
C. $Z_o = 125\ \Omega$.
D. $Z_o = 133\ \Omega$.
E. $Z_o = 100\ \Omega$.

30. The vector $40 + j30$ represents:
    A. 40 Ω resistance and 30 µH inductance.
    B. 40 uH inductance and 30 Ω resistance.
    C. 40 Ω resistance and 30 Ω inductive reactance.
    D. 40 Ω inductive reactance and 30 Ω resistance.
    E. 40 uH inductive reactance and 30 Ω resistance.

31. In a series RC circuit, where, R = 300 Ω and X_C = –30 Ω:
    A. The current leads the voltage by a few degrees.
    B. The current leads the voltage by almost 90 degrees.
    C. The voltage leads the current by a few degrees.
    D. The voltage leads the current by almost 90 degrees.
    E. The voltage leads the current by 90 degrees.

32. In a step-down transformer:
    A. The primary voltage is greater than the secondary voltage.
    B. The primary impedance is less than the secondary impedance.
    C. The secondary voltage is greater than the primary voltage.
    D. The output frequency is higher than the input frequency.
    E. The output frequency is lower than the input frequency.

33. A capacitor of 470 pF is in parallel with an inductor of 4.44 µH. What is the resonant frequency?
    A. 3.49 MHz.
    B. 3.49 kHz.
    C. 13.0 MHz.
    D. 13.0 GHz.
    E. Not determinable from the data given.

34. A sine wave contains energy at:
    A. Just one frequency.
    B. A frequency and its even harmonics.
    C. A frequency and its odd harmonics.
    D. A frequency and all its harmonics.
    E. A frequency and its second harmonic only.

35. Inductive susceptance is:
    A. The reciprocal of inductance.
    B. Negative imaginary.
    C. Equal to capacitive reactance.
    D. The reciprocal of capacitive susceptance.
    E. A measure of the opposition a coil offers to ac.

36. The rate of change (derivative) of a sine wave is itself a wave that:
    A. Is in phase with the original wave.
    B. Is 180 degrees out of phase with the original wave.
    C. Leads the original wave by 45 degrees of phase.
    D. Lags the original wave by 90 degrees of phase.
    E. Leads the original wave by 90 degrees of phase.

37. True power is equal to:
    A. VA power plus imaginary power.
    B. Imaginary power minus VA power.
    C. Vector difference of VA and reactive power.
    D. VA power; the two are the same thing.
    E. 0.707 times the VA power.

38. Three capacitors are connected in series. Their values are 47 μF, 68 μF, and 100 μF. The total capacitance is:
    A. 215 μF.
    B. Between 68 μF and 100 μF.
    C. Between 47 μF and 68 μF.
    D. 22 μF.
    E. Not determinable from the data given.

39. The reactance of a section of transmission line depends on all of the following *except*:
    A. The velocity factor of the line.
    B. The length of the section.
    C. The current in the line.
    D. The frequency.
    E. The wavelength.

40. When confronted with a parallel RLC circuit and you need to find the complex impedance:
    A. Just add the resistance and reactance to get $R + jX$.
    B. Find the net conductance and susceptance, then convert to resistance and reactance, and add these to get $R + jX$.
    C. Find the net conductance and susceptance, and just add these together to get $R + jX$.
    D. Rearrange the components so they're in series, and find the complex impedance of that circuit.
    E. Subtract reactance from resistance to get $R - jX$.

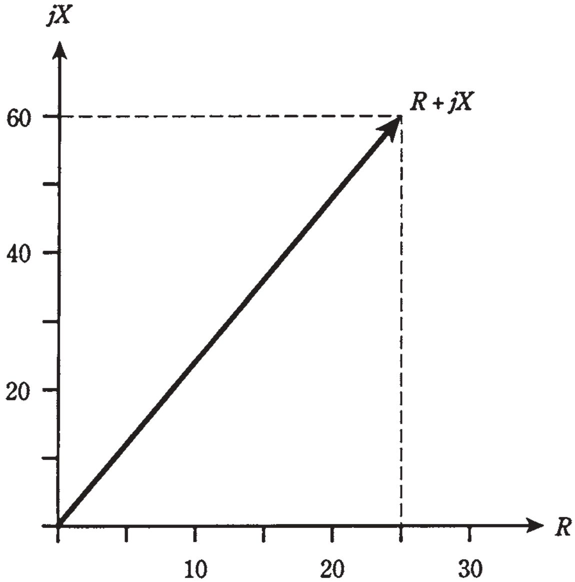
41. The illustration in Fig. Test 2-2 shows a vector $R + jX$ representing:
    A. $X_C = 60\ \Omega$ and $R = 25\ \Omega$.

B. $X_L = 60\ \Omega$ and $R = 25\ \Omega$.

C. $X_L = 60\ \mu H$ and $R = 25\ \Omega$.

D. $C = 60\ \mu F$ and $R = 25\ \Omega$.

E. $L = 60\ \mu H$ and $R = 25\ \Omega$.



TEST 2-2 Illustration for PART TWO test question 41.

42. If two sine waves have the same frequency and the same amplitude, but they cancel out, the phase difference is:

A. 45 degrees.

B. 90 degrees.

C. 180 degrees.

D. 270 degrees.

E. 360 degrees.

43. A series circuit has a resistance of 50 Ω and a capacitive reactance of –37 Ω. The phase angle is:

A. 37 degrees.

B. 53 degrees.

C. –37 degrees.

D. –53 degrees.

E. Not determinable from the data given.

44. A 200-Ω resistor is in series with a coil and capacitor; $X_L = 200\ \Omega$ and $X_C = -100\ \Omega$. The complex impedance is:

A. $200 - j100$.

B. $200 - j200$.
C. $200 + j100$.
D. $200 + j200$.
E. Not determinable from the data given.

45. The characteristic impedance of a transmission line:
A. Is negative imaginary.
B. Is positive imaginary.
C. Depends on the frequency.
D. Depends on the construction of the line.
E. Depends on the length of the line.

46. The period of a wave is 2×10^{-8} second. The frequency is:
A. 2×10^{8} Hz.
B. 20 MHz.
C. 50 kHz.
D. 50 MHz.
E. 500 MHz.

47. A series circuit has a resistance of 600 Ω and a capacitance of 220 pF. The phase angle is:
A. –20 degrees.
B. 20 degrees.
C. –70 degrees.
D. 70 degrees.
E. Not determinable from the data given.

48. A capacitor with a negative temperature coefficient:
A. Works less well as the temperature increases.
B. Works better as the temperature increases.
C. Heats up as its value is made larger.
D. Cools down as its value is made larger.
E. Has increasing capacitance as temperature goes down.

49. Three coils are connected in parallel. Each has an inductance of 300µH. There is no mutual inductance. The net inductance is:
A. 100 µH.
B. 300 µH.
C. 900 µH.
D. 17.3 µH.
E. 173 µH.

50. An inductor shows 100 Ω of reactance at 30.0 MHz. What is its inductance?
    A. 0.531 μH.
    B. 18.8 mH.
    C. 531 μH.
    D. 18.8 μH.
    E. It can't be found from the data given.

3
PART
Basic electronics

This page intentionally left blank

19

CHAPTER

Introduction to semiconductors

SINCE THE SIXTIES, WHEN THE TRANSISTOR BECAME COMMON IN CONSUMER devices, *semiconductors* have acquired a dominating role in electronics. This chapter explains what semiconducting materials actually are.

You've learned about electrical conductors, which pass current easily, and about insulators, which block the current flow. A semiconductor can sometimes act like a conductor, and at other times like an insulator in the same circuit.

The term *semiconductor* arises from the ability of these materials to conduct "part time." Their versatility lies in the fact that the conductivity can be controlled to produce effects such as amplification, rectification, oscillation, signal mixing, and switching.

The semiconductor revolution

It wasn't too long ago that vacuum tubes were the backbone of electronic equipment. Even in radio receivers and "portable" television sets, all of the amplifiers, oscillators, detectors, and other circuits required these devices. A typical vacuum tube ranged from the size of your thumb to the size of your fist.

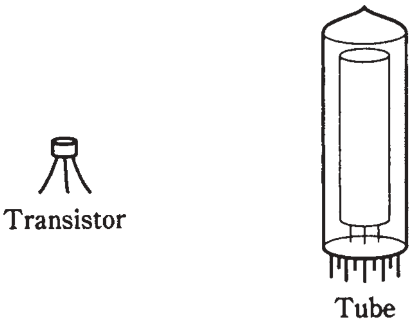
A radio might sit on a table in the living room; if you wanted to listen to it, you would turn it on and wait for the tube filaments to warm up. I can remember this. It makes me feel like an old man to think about it.

Vacuum tubes, sometimes called "tubes" or "valves" (in England), are still used in some power amplifiers, microwave oscillators, and video display units. There are a few places where tubes work better than semiconductor devices. Tubes tolerate momentary voltage and current surges better than semiconductors. They are discussed in chapter 29.

Tubes need rather high voltages to work. Even in radio receivers, turntables, and other consumer devices, 100 V to 200 V dc was required when tubes were employed. This mandated bulky power supplies and created an electrical shock hazard.

Nowadays, a transistor about the size of a pencil eraser can perform the functions of a tube in most situations. Often, the power supply can be a couple of AA cells or a 9-V "transistor battery."

Figure 19-1 is a size comparison drawing between a typical transistor and a typical vacuum tube.



19-1 Transistors are smaller than tubes.

Integrated circuits, hardly larger than individual transistors, can do the work of hundreds or even thousands of vacuum tubes. An excellent example of this technology is found in the personal computer, or PC. In 1950, a "PC" would have occupied a large building, required thousands of watts to operate, and probably cost well over a million dollars. Today you can buy one and carry it in a briefcase. Integrated-circuit technology is discussed in chapter 28.

Semiconductor materials

There are numerous different mixtures of elements that work as semiconductors. The two most common materials are *silicon* and a compound of gallium and arsenic known as *gallium arsenide* (often abbreviated GaAs).

In the early years of semiconductor technology, germanium formed the basis for many semiconductors; today it is seen occasionally, but not often. Other substances that work as semiconductors are selenium, cadmium compounds, indium compounds, and various metal oxides.

Many of the elements found in semiconductors can be mined from the earth. Others are "grown" as crystals under laboratory conditions.

Silicon

Silicon (chemical symbol Si) is widely used in diodes, transistors, and integrated circuits. Generally, other substances, or *impurities,* must be added to silicon to give it the desired properties. The best quality silicon is obtained by growing crystals in a laboratory. The silicon is then fabricated into *wafers* or *chips*.

Gallium arsenide

Another common semiconductor is the compound gallium arsenide. Engineers and technicians call this material by its acronym-like chemical symbol, GaAs, pronounced "gas." If you hear about "gasfets" and "gas ICs," you're hearing about gallium-arsenide technology.

Gallium arsenide works better than silicon in several ways. It needs less voltage, and will function at higher frequencies because the charge carriers move faster. GaAs devices are relatively immune to the effects of ionizing radiation such as X rays and gamma rays. GaAs is used in light-emitting diodes, infrared-emitting diodes, laser diodes, visible-light and infrared detectors, ultra-high-frequency amplifying devices, and a variety of integrated circuits.

The primary disadvantage of GaAs is that it is more expensive to produce than silicon.

Selenium

Selenium has resistance that varies depending on the intensity of light that falls on it. All semiconductor materials exhibit this property, known as *photoconductivity,* to a greater or lesser degree, but selenium is especially affected. For this reason, selenium is useful for making photocells.

Selenium is also used in certain types of *rectifiers.* This is a device that converts ac to dc; you'll learn about rectification in chapters 20 and 21. The main advantage of selenium over silicon is that selenium can withstand brief *transients,* or surges of abnormally high voltage.

Germanium

Pure germanium is a poor electrical conductor. It becomes a semiconductor when impurities are added. Germanium was used extensively in the early years of semiconductor technology. Some diodes and transistors still use it.

A germanium diode has a low voltage drop (0.3 V, compared with 0.6 V for silicon and 1 V for selenium) when it conducts, and this makes it useful in some situations. But germanium is easily destroyed by heat. Extreme care must be used when soldering the leads of a germanium component.

Metal oxides

Certain metal oxides have properties that make them useful in the manufacture of semiconductor devices. When you hear about MOS (pronounced "moss") or CMOS (pronounced "sea moss") technology, you are hearing about *metal-oxide semiconductor* and *complementary metal-oxide semiconductor* devices, respectively.

One advantage of MOS and CMOS devices is that they need almost no power to function. They draw so little current that a battery in a MOS or CMOS device lasts just about as long as it would on the shelf. Another advantage is high speed. This allows operation at high frequencies, and makes it possible to perform many calculations per second.

Certain types of transistors, and many kinds of integrated circuits, make use of this technology. In integrated circuits, MOS and CMOS allows for a large number of discrete diodes and transistors on a single chip. Engineers would say that MOS/CMOS has *high component density.*

The biggest problem with MOS and CMOS is that the devices are easily damaged by static electricity. Care must be used when handling components of this type.

Doping

For a semiconductor material to have the properties needed to work in electronic components, impurities are usually added. The impurities cause the material to conduct currents in certain ways. The addition of an impurity to a semiconductor is called *doping.* Sometimes the impurity is called a *dopant.*

Donor impurities

When an impurity contains an excess of electrons, the dopant is called a *donor impurity.* Adding such a substance causes conduction mainly by means of electron flow, as in a metal like copper. The excess electrons are passed from atom to atom when a voltage exists across the material. Elements that serve as donor impurities include antimony, arsenic, bismuth, and phosphorus.

A material with a donor impurity is called an *N type* semiconductor, because electrons have negative charge.

Acceptor impurities

If an impurity has a deficiency of electrons, the dopant is called an *acceptor impurity.* When a substance such as aluminum, boron, gallium, or indium is added to a semiconductor, the material conducts by means of *hole flow*. A *hole* is a missing electron; it is described in more detail shortly.

A material with an acceptor impurity is called a *P-type* semiconductor, because holes have positive charge.

Majority and minority charge carriers

Charge carriers in semiconductor materials are either electrons, which have a unit negative charge, or holes, having a unit positive charge. In any semiconductor material, some of the current is in the form of electrons passed from atom to atom in a negative-to-positive direction. Some current occurs as holes that move from atom to atom in a positive-to-negative direction.

Sometimes electrons dominate the current flow in a semiconductor; this is the case if the material has donor impurities. In substances having acceptor impurities, holes dominate. The dominating charge carriers (either electrons or holes) are the *majority carriers.* The less abundant ones are the *minority carriers.*

The ratio of majority to minority carriers can vary, depending on the nature of the semiconducting material.

Electron flow

In an N-type semiconductor, most of the current flows as electrons passed from atom to atom. But some of the current in a P-type material also takes this form. You learned about electron flow all the way back in chapter 1. It would be a good idea to turn back for a moment and review this material, because it will help you understand the concept of *hole flow.*

Hole flow

In a P-type semiconductor, most of the current flows in a way that some people find peculiar and esoteric. In a literal sense, in virtually all electronic devices, charge transfer is always the result of electron movement, no matter what the medium might be. The exceptions are particle accelerators and cloud chambers—apparatus of interest mainly to theoretical physicists.

The flow of current in a P-type material is better imagined as a flow of *electron absences,* not electrons. The behavior of P-type substances can be explained more easily this way. The absences, called “holes,” move in a direction opposite that of the electrons.

Imagine a sold-out baseball stadium. Suppose 19 of every 20 people are randomly issued candles. Imagine it’s nighttime, and the field lights are switched off. You stand at the center of the field, just behind second base.The candles are lit, and the people pass them around the stands. Each person having a candle passes it to the person on their right if, but only if, that person has no candle. You see moving dark spots: people without candles. The dark spots move against the candle movement. The physical image you see is produced by candle light, but the motion you notice is that of *candles absences.*

Figure 19-2 illustrates this phenomenon. Small dots represent candles or electrons. Imagine them moving from right to left in the figure as they are passed from person to person or from atom to atom. Circles represent candle absences or holes. They “move” from left to right, contrary to the flow of the candles or the electrons, because the candles or electrons are being *passed among stationary units* (people or atoms).

This is just the way holes flow in a semiconductor material.

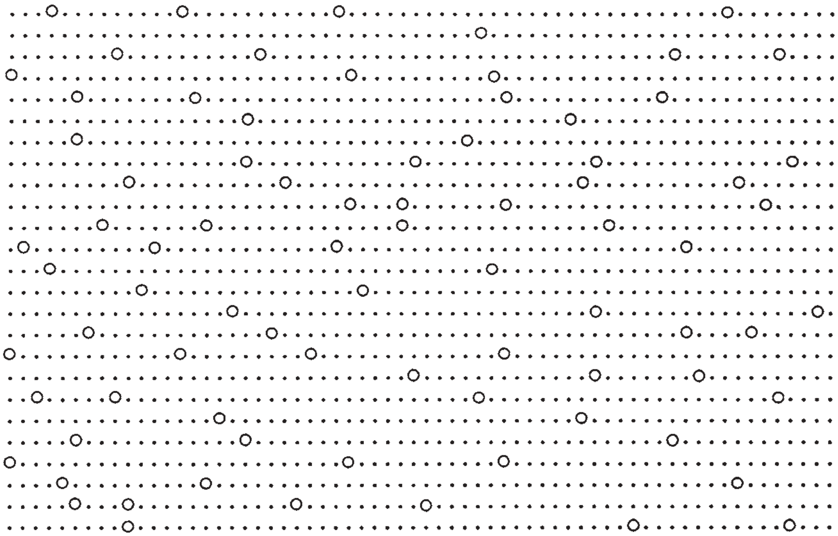
Behavior of a P-N junction

Simply having a semiconducting material, either P or N type, might be interesting, and a good object of science experiments. But when the two types of material are brought together, the *P-N junction* develops properties that make the semiconductor materials truly useful as electronic devices.

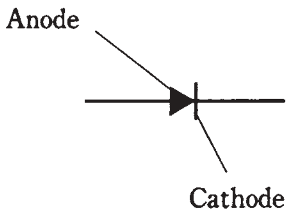
Figure 19-3 shows the schematic symbol for a *semiconductor diode,* formed by joining a piece of P-type material to a piece of N-type material. The N-type semiconductor is represented by the short, straight line in the symbol, and is called the *cathode.* The P-type semiconductor is represented by the arrow, and is called the *anode.*

In the diode as shown in the figure, electrons flow in the direction opposite the arrow. (Physicists consider current to flow from positive to negative, and this is in the same direction as the arrow points.) But current cannot, under most conditions, flow the other way. Electrons normally do not flow in the direction that the arrow points.

If you connect a battery and a resistor in series with the diode, you’ll get a current flow if the negative terminal of the battery is connected to the cathode and the positive



19-2 Pictorial representation of hole flow. Small dots represent electrons, moving one way; open circles represent holes, moving the other way.



19-3 Schematic symbol for a semiconductor diode.

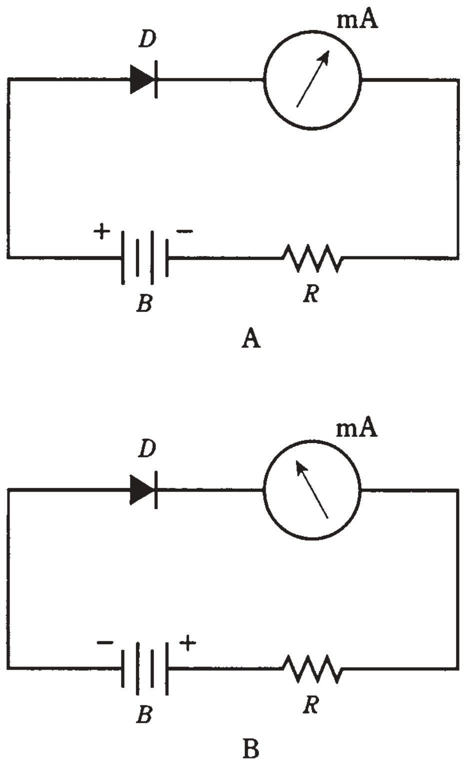
terminal is connected to the anode (Fig. 19-4A). No current will flow if the battery is reversed (Fig. 19-4B). The resistor is included in the circuit to prevent destruction of the diode by excessive current.

It takes a certain minimum voltage for conduction to occur. This is called the *forward breaker voltage* of the junction. Depending on the type of material, it varies from about 0.3 V to 1 V. If the voltage across the junction is not at least as great as the forward breaker value, the diode will not conduct. This effect can be of use in amplitude limiters, waveform clippers, and threshold detectors.

You'll learn about the various ways diodes are used in the next chapter.

How the junction works

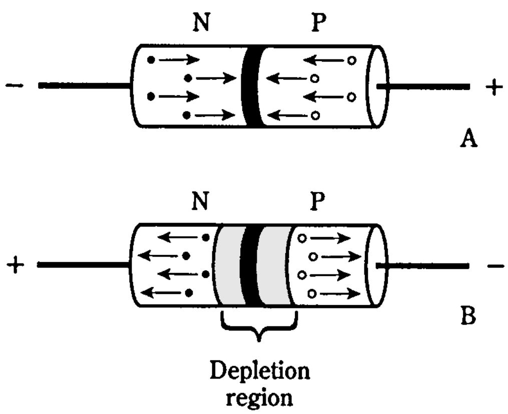
When the N-type material is negative with respect to the P-type, as in Fig. 19-4A, electrons flow easily from N to P. The N-type semiconductor, which already has an excess of electrons, gets even more; the P-type semiconductor, with a shortage of



19-4 Series connection of battery B, resistor R, milliammeter mA, and diode D. At A, forward bias results in current flow; at B, reverse bias results in no current.

electrons, is made even more deficient. The N-type material constantly feeds electrons to the P-type in an attempt to create an electron balance, and the battery or power supply keeps robbing electrons from the P-type material. This is shown in Fig. 19-5A and is known as *forward bias*.

When the polarity is switched so the N-type material is positive with respect to the P type, things get interesting. This is called *reverse bias.* Electrons in the N-type material



19-5 At A, forward bias of a P-N junction; at B, reverse bias. Electrons are shown as small dots, and holes are shown as open circles.

are pulled towards the positive charge, away from the junction. In the P-type material, holes are pulled toward the negative charge, also away from the junction. The electrons (in the N-type material) and holes (in the P type) are the majority charge carriers. They become depleted in the vicinity of the P-N junction (Fig. 19-5B). A shortage of majority carriers means that the semiconductor material cannot conduct well. Thus, the *depletion region* acts like an insulator.

Junction capacitance

Some P-N junctions can alternate between conduction (in forward bias) and nonconduction (in reverse bias) millions or billions of times per second. Other junctions are slower. The main limiting factor is the capacitance at the P-N junction during conditions of reverse bias. The amount of capacitance depends on several factors, including the operating voltage, the type of semiconductor material, and the cross-sectional area of the P-N junction.

By examining Fig. 19-5B, you should notice that the depletion region, sandwiched between two semiconducting sections, resembles the dielectric of a capacitor. In fact, the similarity is such that a reverse-biased P-N junction really is a capacitor. Some semiconductor components are made with this property specifically in mind.

The *junction capacitance* can be varied by changing the reverse-bias voltage, because this voltage affects the width of the depletion region. The greater the reverse voltage, the wider the depletion region gets, and the smaller the capacitance becomes.

In the next chapter, you'll learn how engineers take advantage of this effect.

Avalanche effect

The greater the reverse bias voltage, the "more determined an insulator" a P-N junction gets—to a point. If the reverse bias goes past this critical value, the voltage overcomes the ability of the junction to prevent the flow of current, and the junction conducts as if it were forward biased. This *avalanche effect* does not ruin the junction (unless the voltage is extreme); it's a temporary thing. When the voltage drops back below the critical value, the junction behaves normally again.

Some components are designed to take advantage of the avalanche effect. In other cases, avalanche effect limits the performance of a circuit.

In a device designed for voltage regulation, called a *Zener diode,* you'll hear about the *avalanche voltage* or *Zener voltage* specification. This might range from a couple of volts to well over 100 V. It's important in the design of voltage-regulating circuits in solid-state power supplies; this is discussed in the next chapter.

For *rectifier diodes* in power supplies, you'll hear about the *peak inverse voltage (PIV)* or *peak reverse voltage (PRV*) specification. It's important that rectifier diodes have PIV great enough so that avalanche effect will not occur (or even come close to happening) during any part of the ac cycle. Otherwise, the circuit efficiency will be compromised.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The term "semiconductor" arises from:
   A. Resistor-like properties of metal oxides.
   B. Variable conductive properties of some materials.
   C. The fact that there's nothing better to call silicon.
   D. Insulating properties of silicon and GaAs.

2. Which of the following is *not* an advantage of semiconductor devices over vacuum tubes?
   A. Smaller size.
   B. Lower working voltage.
   C. Lighter weight.
   D. Ability to withstand high voltages.

3. The most common semiconductor among the following substances is:
   A. Germanium.
   B. Galena.
   C. Silicon.
   D. Copper.

4. GaAs is a(n):
   A. Compound.
   B. Element.
   C. Conductor.
   D. Gas.

5. A disadvantage of gallium-arsenide devices is that:
   A. The charge carriers move fast.
   B. The material does not react to ionizing radiation.
   C. It is expensive to produce.
   D. It must be used at high frequencies.

6. Selenium works especially well in:
   A. Photocells.
   B. High-frequency detectors.
   C. Radio-frequency power amplifiers.
   D. Voltage regulators.

7. Of the following, which material allows the lowest forward voltage drop in a diode?
   - A. Selenium.
   - B. Silicon.
   - C. Copper.
   - D. Germanium.
8. A CMOS integrated circuit:
   - A. Can only work at low frequencies.
   - B. Is susceptible to damage by static.
   - C. Requires considerable power to function.
   - D. Needs very high voltage.
9. The purpose of doping is to:
   - A. Make the charge carriers move faster.
   - B. Cause holes to flow.
   - C. Give a semiconductor material certain properties.
   - D. Protect devices from damage in case of transients.
10. A semiconductor material is made into N type by:
    - A. Adding an acceptor impurity.
    - B. Adding a donor impurity.
    - C. Injecting electrons.
    - D. Taking electrons away.
11. Which of the following *does not* result from adding an acceptor impurity?
    - A. The material becomes P type.
    - B. Current flows mainly in the form of holes.
    - C. Most of the carriers have positive electric charge.
    - D. The substance has an electron surplus.
12. In a P-type material, electrons are:
    - A. Majority carriers.
    - B. Minority carriers.
    - C. Positively charged.
    - D. Entirely absent.
13. Holes flow from:
    - A. Minus to plus.
    - B. Plus to minus.
    - C. P-type to N-type material.
    - D. N-type to P-type material.

14. When a P-N junction does not conduct, it is:
    A. Reverse biased.
    B. Forward biased.
    C. Biased past the breaker voltage.
    D. In a state of avalanche effect.

15. Holes flow the opposite way from electrons because:
    A. Charge carriers flow continuously.
    B. Charge carriers are passed from atom to atom.
    C. They have the same polarity.
    D. No! Holes flow in the same direction as electrons.

16. If an electron has a charge of −1 unit, a hole has:
    A. A charge of −1 unit.
    B. No charge.
    C. A charge of +1 unit.
    D. A charge that depends on the semiconductor type.

17. When a P-N junction is reverse-biased, the capacitance depends on all of the following *except*:
    A. The frequency.
    B. The width of the depletion region.
    C. The cross-sectional area of the junction.
    D. The type of semiconductor material.

18. If the reverse bias exceeds the avalanche voltage in a P-N junction:
    A. The junction will be destroyed.
    B. The junction will insulate; no current will flow.
    C. The junction will conduct current.
    D. The capacitance will become extremely high.

19. Avalanche voltage is routinely exceeded when a P-N junction acts as a:
    A. Current rectifier.
    B. Variable resistor.
    C. Variable capacitor.
    D. Voltage regulator.

20. An *unimportant* factor concerning the frequency at which a P-N junction will work effectively is:
    A. The type of semiconductor material.
    B. The cross-sectional area of the junction.
    C. The reverse current.
    D. The capacitance with reverse bias.

20

CHAPTER

Some uses of diodes

THE TERM *DIODE* MEANS "TWO ELEMENTS." IN THE EARLY YEARS OF ELECTRONICS and radio, most diodes were vacuum tubes. The *cathode* element emitted electrons, and the *anode* picked up electrons. Thus, current would flow as electrons through the tube from the cathode to the anode, but not the other way.

Tubes had *filaments* to drive electrons from their cathodes. The filaments were heated via a low ac voltage, but the cathodes and anodes usually wielded hundreds or even thousands of dc volts.

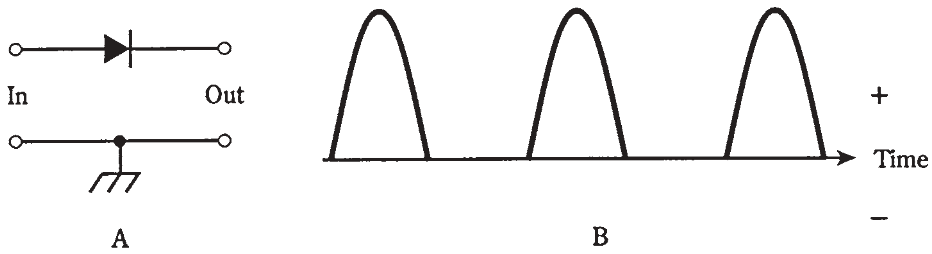
Today, you'll still hear about diodes, anodes, and cathodes. But rather than large, heavy, hot, high-voltage tubes, diodes are tiny things made from silicon or other semiconducting materials. Some diodes can handle voltages nearly as great as their tube counterparts. Semiconductor diodes can do just about everything that tube diodes could, plus a few things that people in the tube era probably never imagined.

Rectification

The hallmark of a *rectifier diode* is that it passes current in only one direction. This makes it useful for changing ac to dc. Generally speaking, when the cathode is negative with respect to the anode, current flows; when the cathode is positive relative to the anode, there is no current. The constraints on this behavior are the forward breakover and avalanche voltages, as you learned about in the last chapter.

Suppose a 60-Hz ac sine wave is applied to the input of the circuit in Fig. 20-1A. During half the cycle, the diode conducts, and during the other half, it doesn't. This cuts off half of every cycle. Depending on which way the diode is hooked up, either the positive half or the negative half of the ac cycle will be removed. Figure 20-1B shows the output of the circuit at A. Remember that electrons flow from negative to positive, against the arrow in the diode symbol.

The circuit and wave diagram of Fig. 20-1 show a *half-wave rectifier* circuit. This

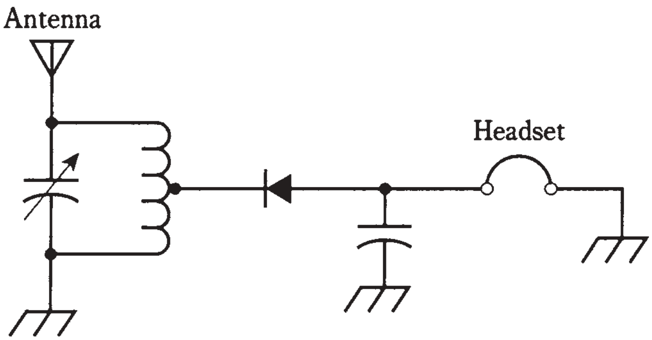


20-1 At A, half-wave rectifier. At B, output of the circuit of A with sine-wave ac input.

is the simplest possible rectifier. That's its chief advantage over other, more complicated rectifier circuits. You'll learn about the different types of rectifier diodes and circuits in the next chapter.

Detection

One of the earliest diodes, existing even before vacuum tubes, was a semiconductor. Known as a *cat whisker,* this semiconductor consisted of a fine piece of wire in contact with a small piece of the mineral *galena.* This bizarre-looking thing had the ability to act as a rectifier for small radio-frequency (RF) currents. When the cat whisker was connected in a circuit like that of Fig. 20-2, the result was a receiver capable of picking up amplitude-modulated (AM) radio signals.



20-2 Schematic diagram of a crystal set radio receiver.

A cat whisker was a finicky thing. Engineers had to adjust the position of the fine wire to find the best point of contact with the galena. A tweezers and magnifying glass were invaluable in this process. A steady hand was essential.

The galena, sometimes called a "crystal," gave rise to the nickname *crystal set* for this low-sensitivity radio. You can still build a crystal set today, using a simple RF diode, a coil, a tuning capacitor, a headset, and a long-wire antenna. Notice that there's no battery! The audio is provided by the received signal alone.

The diode in Fig. 20-2 acts to recover the audio from the radio signal. This is called *detection;* the circuit is a *detector.* If the detector is to be effective, the diode must be of the right type. It should have low capacitance, so that it works as a rectifier at radio frequencies, passing current in one direction but not in the other. Some modern RF diodes are actually microscopic versions of the old cat whisker, enclosed in a glass case with axial leads. You have probably seen these in electronics hobby stores.

Details about detector circuits are discussed in chapter 27. Some detectors use diodes; others do not. Modulation methods are examined in chapter 26.

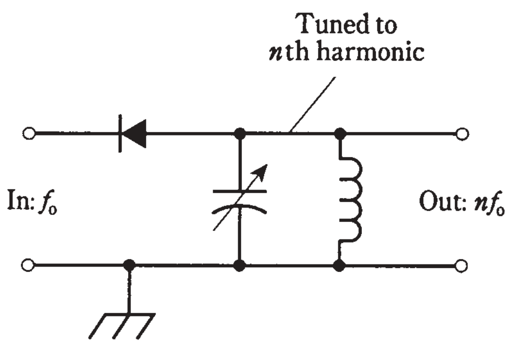
Frequency multiplication

When current passes through a diode, half of the cycle is cut off, as shown in Fig. 20-1. This occurs no matter what the frequency, from 60-Hz utility current through RF, as long as the diode capacitance is not too great.

The output wave from the diode looks much different than the input wave. This condition is known as *nonlinearity.* Whenever there is nonlinearity of any kind in a circuit—that is, whenever the output waveform is shaped differently from the input waveform—there will be harmonic frequencies in the output. These are waves at integer multiples of the input frequency. (If you've forgotten what harmonics are, refer back to chapter 9.)

Often, nonlinearity is undesirable. Then engineers strive to make the circuit *linear,* so that the output waveform has exactly the same shape as the input waveform. But sometimes a circuit is needed that will produce harmonics. Then nonlinearity is introduced deliberately. Diodes are ideal for this.

A simple frequency-multiplier circuit is shown in Fig. 20-3. The output LC circuit is tuned to the desired nth harmonic frequency, nf_0, rather than to the input or fundamental frequency, f_0.



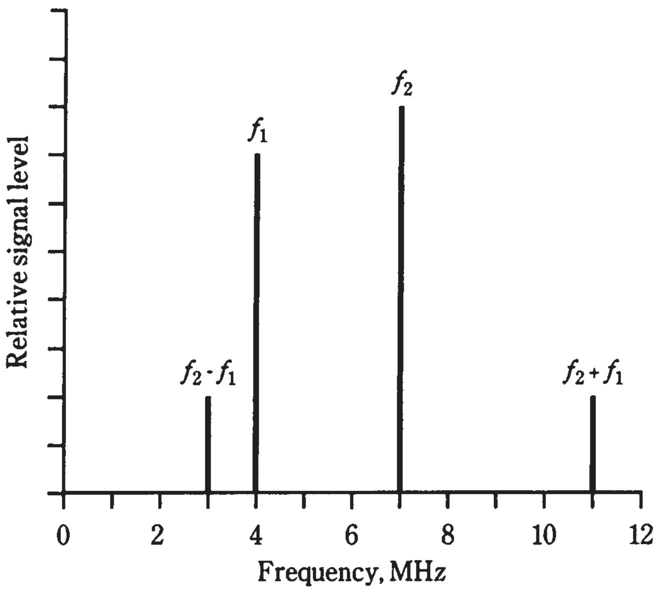
20-3 A frequency multiplier circuit.

For a diode to work as a frequency multiplier, it must be of a type that would also work well as a detector at the same frequencies. This means that the component should act like a rectifier, but not like a capacitor.

Mixing

When two waves having different frequencies are combined in a nonlinear circuit, new frequencies are produced. These new waves are at the sum and difference frequencies of the original waves. You've probably noticed this *mixing,* also called *heterodyning,* if you've ever heard two loud, sine wave tones at the same time.

Suppose there are two signals with frequencies f_1 and f_2. For mathematical convenience, assign f_2 to the wave with the higher frequency. If these signals are combined in a nonlinear circuit, new waves will result. One of them will have a frequency $f_2 - f_1$, and the other will be at $f_2 + f_1$. These are known as *beat frequencies.* The signals are called *mixing products* (Fig. 20-4).



20-4 Spectral (frequency-domain) illustration of mixing. Frequency designators are discussed in the text.

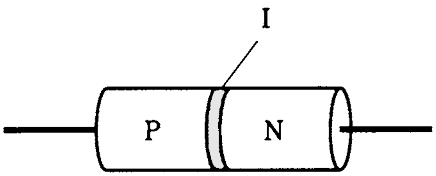
Figure 20-4 is a *frequency domain* graph. Amplitude (on the vertical scale) is shown as a function of frequency (on the horizontal scale). This kind of display is what engineers see when they look at the screen of a *spectrum analyzer.* Most of the graphs you've seen so far have been *time domain* graphs, in which things are shown as a function of time. The screen of an oscilloscope normally shows things in the time domain.

How do you get the nonlinearity necessary to obtain a *mixer* circuit? There are various different schemes, but one common way is—you guessed it—to use diodes. Mixer circuits are discussed in chapter 27.

Switching

The ability of diodes to conduct with forward bias, and to insulate with reverse bias, makes them useful for switching in some electronic applications. Diodes can switch at extremely high rates, much faster than any mechanical device.

One type of diode, made for use as an RF switch, has a special semiconductor layer sandwiched in between the P-type and N-type material. This layer, called an *intrinsic semiconductor,* reduces the capacitance of the diode, so that it can work at higher frequencies than an ordinary diode. The intrinsic material is sometimes called *I type.* A diode with I-type semiconductor is called a *PIN diode* (Fig. 20-5).



20-5 The PIN diode has a layer of intrinsic (I-type) semiconductor at the P-N junction.

Direct-current bias, applied to one or more PIN diodes, allows RF currents to be effectively channeled without using complicated relays and cables. A PIN diode also makes a good RF detector, especially at frequencies above 30 MHz.

Voltage regulation

Most diodes have avalanche breakdown voltages much higher than the reverse bias ever gets. The value of the avalanche voltage depends on how a diode is manufactured. *Zener diodes* are made to have well-defined, constant avalanche voltages.

Suppose a certain Zener diode has an avalanche voltage, also called the *Zener voltage,* of 50 V. If a reverse bias is applied to the P-N junction, the diode acts as an open circuit below 50 V. When the voltage reaches 50 V, the diode starts to conduct. The more the reverse bias tries to increase, the more current flows through the P-N junction. This effectively prevents the reverse voltage from exceeding 50 V.

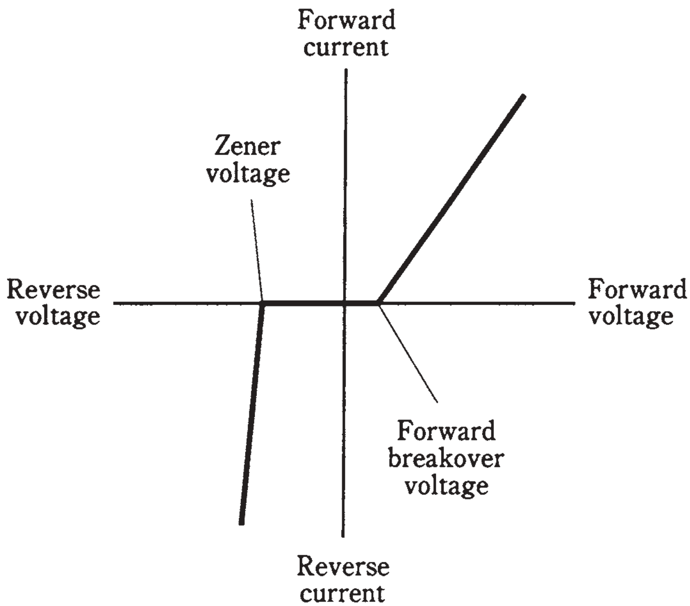
The current through a Zener diode, as a function of the voltage, is shown in Fig. 20-6. The Zener voltage is indicated by the abrupt rise in reverse current as the reverse bias increases. A typical Zener-diode voltage-limiting circuit is shown in Fig. 20-7.

There are other ways to get voltage regulation besides the use of Zener diodes, but Zener diodes often provide the simplest and least expensive alternative. Zener diodes are available with a wide variety of voltage and power-handling ratings. Power supplies for solid-state equipment commonly employ Zener diode regulators.

Details about power supply design are coming up in chapter 21.

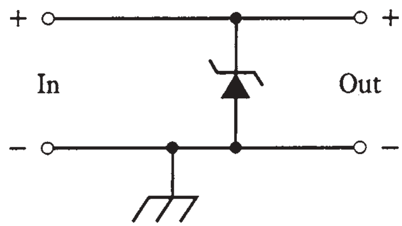
Amplitude limiting

The forward breakover voltage of a germanium diode is about 0.3 V; for a silicon diode it is about 0.6 V. In the last chapter, you learned that a diode will not conduct until the forward bias voltage is at least as great as the forward breakover voltage. The “flip side”



20-6 Current through a Zener diode, as a function of the bias voltage.

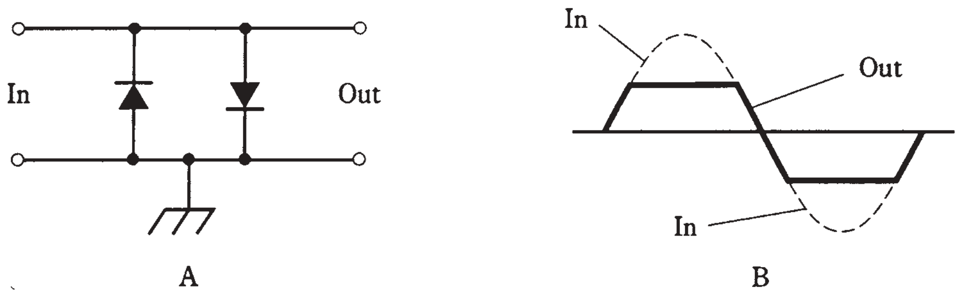
20-7 Connection of Zener diode for voltage regulation.



is that the diode will always conduct when the forward bias exceeds the breakover value. In this case, the voltage across the diode will be constant: 0.3 V for germanium and 0.6 V for silicon.

This property can be used to advantage when it is necessary to limit the amplitude of a signal, as shown in Fig. 20-8. By connecting two identical diodes back-to-back in parallel with the signal path (A), the maximum peak amplitude is limited, or *clipped,* to the forward breakover voltage of the diodes. The input and output waveforms of a clipped signal are illustrated at B. This scheme is sometimes used in radio receivers to prevent "blasting" when a strong signal comes in.

The downside of the *diode limiter* circuit is that it introduces distortion when limiting is taking place. This might not be a problem for reception of Morse code, or for signals that rarely reach the limiting voltage. But for voice signals with amplitude peaks that rise well past the limiting voltage, it can seriously degrade the audio quality, perhaps even rendering the words indecipherable.



20-8 At A, two diodes can work as a limiter. At B, the peaks are cut off by the action of the diodes.

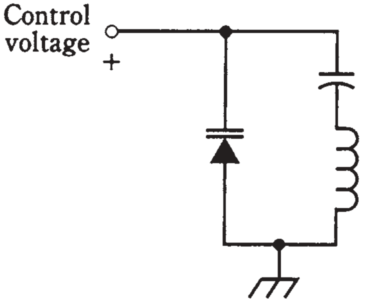
Frequency control

When a diode is reverse-biased, there is a region at the P-N junction with dielectric properties. As you know from the last chapter, this is called the depletion region, because it has a shortage of majority charge carriers. The width of this zone depends on several things, including the reverse voltage.

As long as the reverse bias is less than the avalanche voltage, varying the bias can change the width of the depletion region. This results in a change in the capacitance of the junction. The capacitance, which is always quite small (on the order of picofarads), varies inversely with the square root of the reverse bias.

Some diodes are manufactured especially for use as variable capacitors. These are *varactor diodes*. Sometimes you'll hear them called *varicaps*. They are made from silicon or gallium arsenide.

A common use for a varactor diode is in a circuit called a *voltage-controlled oscillator (VCO)*. A voltage-tuned circuit, using a coil and a varactor, is shown in Fig. 20-9. This is a parallel-tuned circuit. The fixed capacitor, whose value is large compared with that of the varactor, serves to keep the coil from short-circuiting the control voltage across the varactor. Notice that the symbol for the varactor has two lines on the cathode side. This is its "signature," so that you know that it's a varactor, and not just an ordinary diode.



20-9 Connection of a varactor diode in a tuned circuit.

Oscillation and amplification

Under certain conditions, diodes can be made to produce microwave radio signals. There are three types of diodes that do this: *Gunn diodes, IMPATT diodes,* and *tunnel diodes.*

Gunn diodes

A Gunn diode can produce up to 1 W of RF power output, but more commonly it works at levels of about 0.1 W. Gunn diodes are usually made from gallium arsenide.

A Gunn diode oscillates because of the *Gunn effect,* named after J. Gunn of International Business Machines (IBM) who observed it in the sixties. A Gunn diode doesn't work anything like a rectifier, detector, or mixer; instead, the oscillation takes place as a result of a quirk called *negative resistance.*

Gunn-diode oscillators are often tuned using varactor diodes. A Gunn-diode oscillator, connected directly to a microwave horn antenna, is known as a *Gunnplexer.* These devices are popular with amateur-radio experimenters at frequencies of 10 GHz and above.

IMPATT diodes

The acronym *IMPATT* comes from the words *imp*act *a*valanche *t*ransit *t*ime. This, like negative resistance, is a phenomenon the details of which are rather esoteric. An *IMPATT diode* is a microwave oscillating device like a Gunn diode, except that it uses silicon rather than gallium arsenide.

An IMPATT diode can be used as an amplifier for a microwave transmitter that employs a Gunn-diode oscillator. As an oscillator, an IMPATT diode produces about the same amount of output power, at comparable frequencies, as the Gunn diode.

Tunnel diodes

Another type of diode that will oscillate at microwave frequencies is the *tunnel diode,* also known as the *Esaki diode.* It produces only a very small amount of power, but it can be used as a local oscillator in a microwave radio receiver.

Tunnel diodes work well as amplifiers in microwave receivers, because they generate very little unwanted noise. This is especially true of gallium arsenide devices.

The behavior of Gunn, IMPATT, and tunnel diodes is a sophisticated topic and is beyond the scope of this book. College-level electrical-engineering texts are good sources of information on this subject. You will want to know about how these devices work if you plan to become a microwave engineer.

Energy emission

Some semiconductor diodes emit radiant energy when a current passes through the P-N junction in a forward direction. This phenomenon occurs as electrons fall from higher to lower energy states within atoms.

LEDs and IREDs

Depending on the exact mixture of semiconductors used in manufacture, visible light of almost any color can be produced. Infrared-emitting devices also exist. The most common color for a *light-emitting diode (LED)* is bright red. An *infrared-emitting diode (IRED)* produces wavelengths too long to see.

The intensity of the light or infrared from an LED or IRED depends to some extent on the forward current. As the current rises, the brightness increases up to a certain point. If the current continues to rise, no further increase in brilliance takes place. The LED or IRED is then said to be in a state of *saturation.*

Digital displays

Because LEDs can be made in various different shapes and sizes, they are ideal for use in digital displays. You've probably seen digital clock radios that use them. They are common in car radios. They make good indicators for "on/off," "a. m. /p. m.," "battery low," and other conditions.

In recent years, LED displays have been largely replaced by *liquid-crystal displays (LCDs).* This technology has advantages over LEDs, including much lower power consumption and better visibility in direct sunlight.

Communications

Both LEDs and IREDs are useful in communications because their intensity can be modulated to carry information. When the current through the device is sufficient to produce output, but not enough to cause saturation, the LED or IRED output will follow along with rapid current changes. Voices, music, and digital signals can be conveyed over light beams in this way. Some modern telephone systems make use of modulated light, transmitted through clear fibers. This is known as *fiberoptic* technology.

Special LEDs and IREDs produce *coherent* radiation; these are called *laser diodes.* The rays from these diodes aren't the intense, parallel beams that you probably imagine when you think about lasers. A laser LED or IRED generates a cone-shaped beam of low intensity. But it can be focused, and the resulting rays have some of the same advantages found in larger lasers.

Photosensitive diodes

Virtually all P-N junctions exhibit characteristics that change when electromagnetic rays strike them. The reason that conventional diodes are not affected by these rays is that most diodes are enclosed in opaque packages.

Some photosensitive diodes have variable resistance that depends on light intensity. Others actually generate dc voltages in the presence of electromagnetic radiation.

Silicon photodiodes

A silicon diode, housed in a transparent case and constructed in such a way that visible light can strike the barrier between the P-type and N-type materials, forms a *photodiode.*

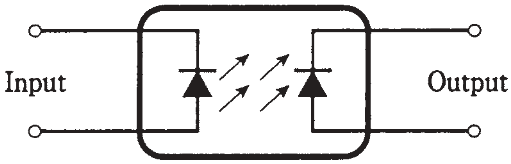
A reverse bias is applied to the device. When light falls on the junction, current flows. The current is proportional to the intensity of the light, within certain limits.

Silicon photodiodes are more sensitive at some wavelengths than at others. The greatest sensitivity is in the *near infrared* part of the spectrum, at wavelengths a little longer than visible red light.

When light of variable brightness falls on the P-N junction of a reverse-biased silicon photodiode, the output current follows the light-intensity variations. This makes silicon photodiodes useful for receiving modulated-light signals of the kind used in fiberoptic systems.

The optoisolator

An LED or IRED and a photodiode can be combined in a single package to get a component called an *optoisolator.* This device (Fig. 20-10) actually creates a modulated light signal and sends it over a small, clear gap to a receptor. The LED or IRED converts an electrical signal to visible light or infrared; the photodiode changes the visible light or infrared back into an electrical signal.



20-10 An optoisolator uses an LED or IRED (input) and a photodiode (output).

A major source of headache for engineers has always been the fact that, when a signal is electrically coupled from one circuit to another, the impedances of the two stages interact. This can lead to nonlinearity, unwanted oscillation, loss of efficiency, or other problems. Optoisolators overcome this effect, because the coupling is not done electrically. If the input impedance of the second circuit changes, the impedance that the first circuit "sees" will remain unaffected, being simply the impedance of the LED or IRED.

Photovoltaic cells

A silicon diode, with no bias voltage applied, will generate dc all by itself if enough electromagnetic radiation hits its P-N junction. This is known as the *photovoltaic effect.* It is the principle by which solar cells work.

Photovoltaic cells are made to have the greatest possible P-N junction surface area. This maximizes the amount of light that falls on the junction. A single silicon photovoltaic cell can produce about 0.6 V of dc electricity. The amount of current that it can deliver depends on the surface area of the junction. For every square inch of P-N surface area, a silicon photovoltaic cell can produce about 160 mA in direct sunlight.

Photovoltaic cells are often connected in series-parallel combinations to provide power for solid-state electronic devices like portable radios. A large assembly of solar cells is called a *solar panel.*

Solar-cell technology has advanced rapidly in the last several years. Solar power is expensive to produce, but the cost is going down—and solar cells do not pollute.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. When a diode is forward-biased, the anode:
   A. Is negative relative to the cathode.
   B. Is positive relative to the cathode.
   C. Is at the same voltage as the cathode.
   D. Alternates between positive and negative relative to the cathode.

2. If ac is applied to a diode, and the peak ac voltage never exceeds the avalanche voltage, then the output is:
   A. Ac with half the frequency of the input.
   B. Ac with the same frequency as the input.
   C. Ac with twice the frequency of the input.
   D. None of the above.

3. A crystal set:
   A. Can be used to transmit radio signals.
   B. Requires a battery with long life.
   C. Requires no battery.
   D. Is useful for rectifying 60-Hz ac.

4. A diode detector:
   A. Is used in power supplies.
   B. Is employed in some radio receivers.
   C. Is used commonly in high-power radio transmitters.
   D. Changes dc into ac.

5. If the output wave in a circuit has the same shape as the input wave, then:
   A. The circuit is linear.
   B. The circuit is said to be detecting.
   C. The circuit is a mixer.
   D. The circuit is a rectifier.

6. The two input frequencies of a mixer circuit are 3.522 MHz and 3.977 MHz. Which of the following frequencies might be used at the output?
   A. 455 kHz.
   B. 886 kHz.
   C. 14.00 MHz.
   D. 1.129 MHz.

7. A time-domain display might be found in:
   A. An ammeter.
   B. A spectrum analyzer.
   C. A digital voltmeter.
   D. An oscilloscope.

8. Zener voltage is also known as:
   A. Forward breakover voltage.
   B. Peak forward voltage.
   C. Avalanche voltage.
   D. Reverse bias.

9. The forward breakover voltage of a silicon diode is:
   A. About 0.3 V.
   B. About 0.6 V.
   C. About 1.0 V.
   D. Dependent on the method of manufacture.

10. A diode audio limiter circuit:
    A. Is useful for voltage regulation.
    B. Always uses Zener diodes.
    C. Rectifies the audio to reduce distortion.
    D. Can cause objectionable signal distortion.

11. The capacitance of a varactor varies with:
    A. Forward voltage.
    B. Reverse voltage.
    C. Avalanche voltage.
    D. Forward breakover voltage.

12. The purpose of the I layer in a PIN diode is to:
    A. Minimize the diode capacitance.
    B. Optimize the avalanche voltage.
    C. Reduce the forward breakover voltage.
    D. Increase the current through the diode.

13. Which of these diode types might be found in the oscillator circuit of a microwave radio transmitter?
    A. A rectifier diode.
    B. A cat whisker.
    C. An IMPATT diode.
    D. None of the above.

14. A Gunnplexer can be used as a:
    - A. Communications device.
    - B. Radio detector.
    - C. Rectifier.
    - D. Signal mixer.

15. The most likely place you would find an LED would be:
    - A. In a rectifier circuit.
    - B. In a mixer circuit.
    - C. In a digital frequency display.
    - D. In an oscillator circuit.

16. Coherent radiation is produced by a:
    - A. Gunn diode.
    - B. Varactor diode.
    - C. Rectifier diode.
    - D. Laser diode.

17. You want a circuit to be stable with a variety of amplifier impedance conditions. You might consider a coupler using:
    - A. A Gunn diode.
    - B. An optoisolator.
    - C. A photovoltaic cell.
    - D. A laser diode.

18. The power from a solar panel depends on all of the following except:
    - A. The operating frequency of the panel.
    - B. The total surface area of the panel.
    - C. The number of cells in the panel.
    - D. The intensity of the light.

19. Emission of energy in an IRED is caused by:
    - A. High-frequency radio waves.
    - B. Rectification.
    - C. Electron energy-level changes.
    - D. None of the above.

20. A photodiode, when not used as a photovoltaic cell, has:
    - A. Reverse bias.
    - B. No bias.
    - C. Forward bias.
    - D. Negative resistance.

21

CHAPTER

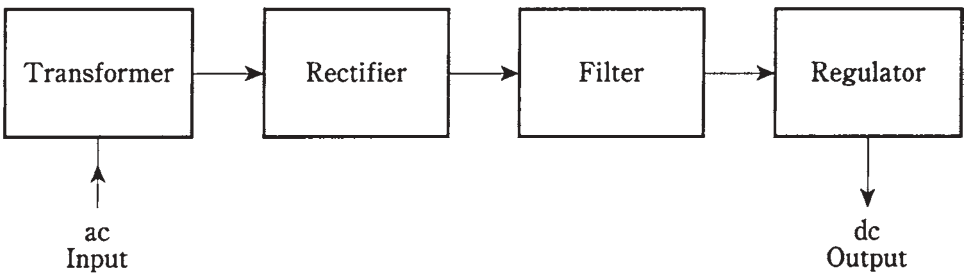
Power supplies

MOST ELECTRONIC EQUIPMENT NEEDS DIRECT CURRENT (DC) TO WORK. BATTERies produce dc, but there is a limit to how much energy and how much voltage a battery can provide. The same is true of solar panels.

The electricity from the utility company is alternating current (ac) with a frequency of 60 Hz. In your house, most wall outlets carry an effective voltage of 117 V; some have 234 V. The energy from a wall outlet is practically unlimited, but it must be converted from ac to dc, and tailored to just the right voltage, to be suitable for electronic equipment.

Parts of a power supply

A *power supply* provides the proper voltage and current for electronic apparatus. Most power supplies consist of several stages, always in the same order (Fig. 21-1).



21-1 Block diagram of a power supply. Sometimes a regulator is not needed.

First, the ac encounters a transformer that steps the voltage either down or up, depending on the exact needs of the electronic circuits.

Second, the ac is rectified, so that it becomes *pulsating dc* with a frequency of either 60 Hz or 120 Hz. This is almost always done by one or more semiconductor diodes.

Third, the pulsating dc is *filtered,* or smoothed out, so that it becomes a continuous voltage having either positive or negative polarity with respect to ground.

Finally, the dc voltage might need to be *regulated.* Some equipment is finicky, insisting on just the right amount of voltage all the time. Other devices can put up with some voltage changes.

Power supplies that provide more than a few volts must have features that protect the user (that's you!) from receiving a dangerous electrical shock. All power supplies need fuses and/or circuit breakers to minimize the fire hazard in case the equipment shorts out.

The power transformer

Power transformers can be categorized as step-down or step-up. As you remember, the output, or secondary, voltage of a step-down unit is lower than the input, or primary, voltage. The reverse is true for a step-up transformer.

Step-down

Most solid-state electronic devices, such as radios, need only a few volts. The power supplies for such equipment use step-down power transformers. The physical size of the transformer depends on the current.

Some devices need only a small current and a low voltage. The transformer in a radio receiver, for example, can be quite small physically. A ham radio transmitter or hi-fi amplifier needs much more current. This means that the secondary winding of the transformer must be of heavy-gauge wire, and the core must be bulky to contain the magnetic flux. Such a transformer is massive.

Step-up

Some circuits need high voltage. The picture tube in a TV set needs several hundred volts. Some ham radio power amplifiers use vacuum tubes working at kilovolts dc. The transformers in these appliances are step-up types. They are moderate to large in size because of the number of turns in the secondary, and also because high voltages can spark, or *arc,* between wire turns if the windings are too tight.

If a step-up transformer needs to supply only a small amount of current, it need not be big. But for ham radio transmitters and radio/TV broadcast amplifiers, the transformers are large and heavy—and expensive.

Transformer ratings

Transformers are rated according to output voltage and current. For a given unit, the *volt-ampere (VA)* capacity is often specified. This is the product of the voltage and current. A transformer with a 12-V output, capable of delivering 10 A, would have $12\ \text{V} \times 10\ \text{A} = 120\ \text{VA}$ of capacity.

The nature of power-supply filtering, to be discussed a bit later in this chapter, makes it necessary for the power-transformer VA rating to be greater than just the wattage needed by the load.

A high-quality, rugged power transformer, capable of providing the necessary currents and/or voltages, is crucial in any power supply. The transformer is usually the most expensive component to replace. When designing a power supply, it's wise to spend a little extra to get a reliable transformer. Engineers might call this "maintenance insurance."

The diode

Rectifier diodes are available in various sizes, intended for different purposes. Most rectifier diodes are made of silicon and are therefore known as *silicon rectifiers.* A few are fabricated from selenium, and are called *selenium rectifiers.*

Two important features of a power-supply diode are the *average forwvard current* (I_O) rating and the *peak inverse voltage (PIV)* rating. There are other specifications that engineers need to know when designing a specialized power supply, but in this course, you only need to be concerned about I_O and PIV.

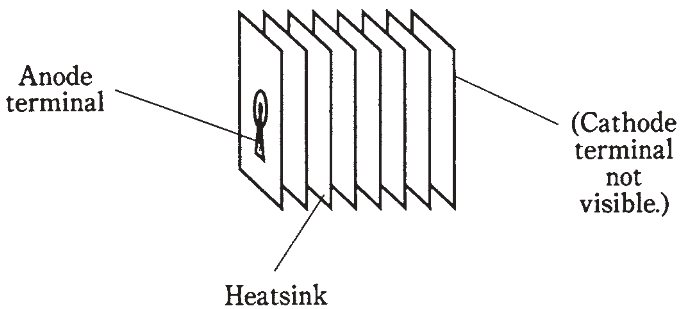
Average forward current

Electric current produces heat. If the current through a diode is too great, the heat will destroy the P-N junction.

Generally speaking, when designing a power supply, it's wise to use diodes with an I_O rating of at least 1.5 times the expected average dc forward current. If this current is 4.0 A, the rectifier diodes should be rated at I_O = 6.0 A or more. Of course, it would be wasteful of money to use a 100-A diode in a circuit where the average forward current is 4.0 A. While it would work, it would be a bit like shooting a sparrow with a cannon.

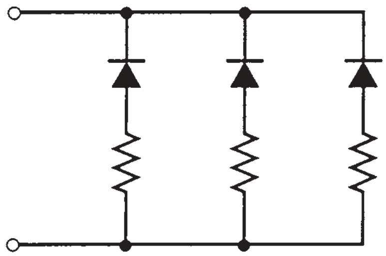
Note that I_O flows through the *diodes.* The current drawn by the load is often quite different from this. Also, note that I_O is an *average* figure. The *instantaneous* forward current is another thing, and can be 15 or even 20 times I_O, depending on the nature of the power-supply filtering circuitry.

Some diodes have *heatsinks* to help carry heat away from the P-N junction. A selenium diode can be recognized by the appearance of its heatsink (Fig. 21-2).



21-2 A selenium rectifier can be recognized by its heatsink.

Diodes can be connected in parallel to increase the current rating. When this is done, small-value resistors are placed in series with each diode in the set to equalize the current burden among the diodes (Fig. 21-3). Each resistor should have a voltage drop of about 1 V.



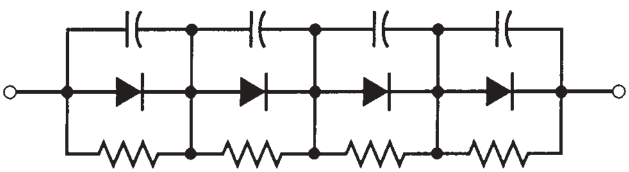
21-3 When diodes are connected in parallel, resistors help equalize the current load.

Peak inverse voltage

The PIV rating of a diode is the instantaneous inverse, or reverse-bias, voltage that it can withstand without avalanche taking place. A good power supply has diodes whose PIV ratings are significantly greater than the peak voltage of the ac at the input.

If the PIV rating is not great enough, the diode or diodes in a supply will conduct for part of the reverse cycle. This will degrade the efficiency of the supply; the reverse current will "buck" the forward current. It would be like having a team of rowers in a long boat, with one or two rowers trying to propel the boat backwards instead of forwards.

Diodes can be connected in series to get a higher PIV capacity than a single diode alone. This scheme is sometimes seen in high-voltage supplies, such as those needed for tube-type ham radio power amplifiers. High-value resistors, of about 500 Ω for each peak-inverse volt, are placed across each diode in the set to distribute the reverse bias equally among the diodes (Fig. 21-4). Also, each diode is shunted by a capacitor of 0.005 µF or 0.1 µF.



21-4 Diodes in series should be shunted by resistors and capacitors.

The half-wave rectifier

The simplest rectifier circuit uses just one diode (or a series or parallel combination) to "chop off" half of the ac input cycle. You saw this circuit in the previous chapter, diagrammed in Fig. 20-1.

In a half-wave circuit, the average output voltage is approximately 45 percent of the rms ac input voltage. But the PIV across the diode can be as much as 2.8 times the rms ac input voltage. It's a good idea to use diodes whose PIV ratings are at least 1.5 times the maximum expected PIV; therefore, with a half-wave supply, the diodes should be rated for at least 4.2 times the rms ac input voltage.

Half-wave rectification has some shortcomings. First, the output is hard to smooth out, because the waveform is so irregular. Second, the voltage output tends to drop when the supply is connected to a load. (This can be overcome to some extent by means of a good voltage regulator. Voltage regulation is discussed later in this chapter.) Third, half-wave rectification puts a disproportionate strain on the power transformer and the diodes.

Half-wave rectification is useful in supplies that don't have to deliver much current, or that don't need to be especially well regulated. The main advantage of using a half-wave circuit in these situations is that it costs a little less than full-wave or bridge circuits.

The full-wave, center-tap rectifier

A much better scheme for changing ac to dc is to use both halves of the ac cycle. Suppose you want to convert an ac wave to dc with positive polarity. Then you can allow the positive half of the ac cycle to pass unchanged, and flip the negative portion of the wave upside-down, making it positive instead. This is the principle behind *full-wave rectification*.

One common full-wave circuit uses a transformer with a center-tapped secondary, as shown in Fig. 21-5A. The center tap, a wire coming out of the exact middle of the secondary winding, is connected to common ground. This produces out-of-phase waves at the ends of the winding. These two waves can be individually half-wave rectified, cutting off the negative half of the cycle. Because the waves are 180 degrees (half a cycle) out of phase, the output of the circuit has positive pulses for both halves of the cycle (Fig. 21-5B).

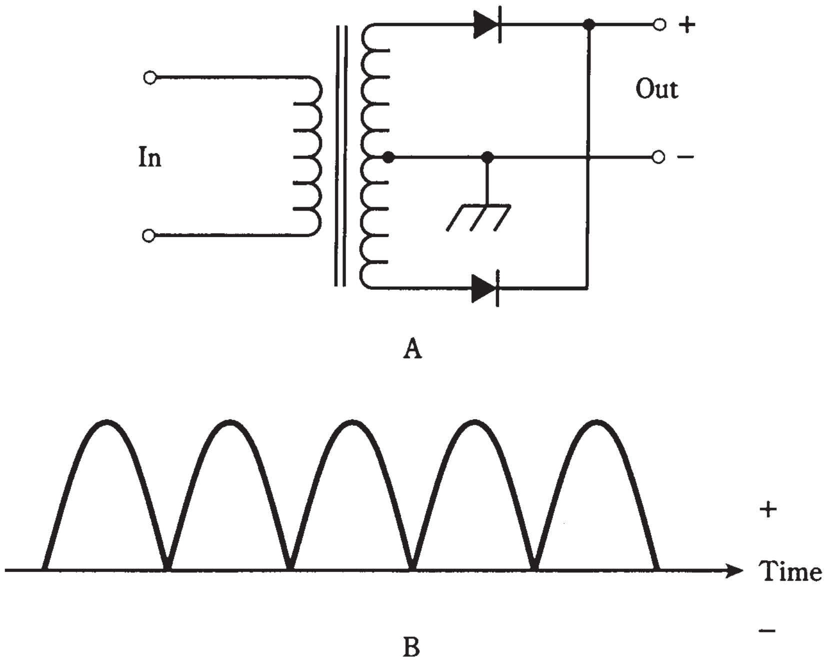
In this rectifier circuit, the average dc output voltage is about 90 percent of the rms ac input voltage. The PIV across the diodes can be as much as 2.8 times the rms input voltage. Therefore, the diodes should have a PIV rating of at least 4.2 times the rms ac input.

Compare Fig. 21-5B with Fig. 20-1B from the last chapter. Can you see that the waveform of the full-wave rectifier ought to be easier to smooth out? In addition to this advantage, the *full-wave, center-tap rectifier* is kinder to the transformer and diodes than a half-wave circuit. Furthermore, if a load is applied to the output of the full-wave circuit, the voltage will drop much less than it would with a half-wave supply, because the output has more "substance."

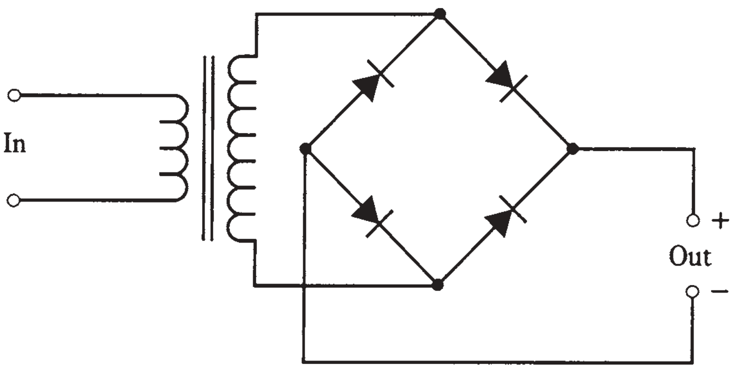
The bridge rectifier

Another way to get full-wave rectification is the *bridge rectifier*. It is diagrammed in Fig. 21-6. The output waveform is just like that of the full-wave, center-tap circuit.

The average dc output voltage in the bridge circuit is 90 percent of the rms ac input voltage, just as is the case with center-tap rectification. The PIV across the diodes is 1.4 times the rms ac input voltage. Therefore, each diode needs to have a PIV rating of at least 2.1 times the rms ac input voltage.



21-5 At A, schematic diagram of a full-wave, center-tap rectifier. At B, output waveform from this rectifier.



21-6 Schematic diagram of a full-wave bridge rectifier.

The bridge circuit does not need a center-tapped transformer secondary. This is its main practical advantage. Electrically, the bridge circuit uses the entire secondary on both halves of the wave cycle; the center-tap circuit uses one side of the secondary for one half of the cycle, and the other side for the other half of the cycle. For this reason, the bridge circuit makes more efficient use of the transformer.

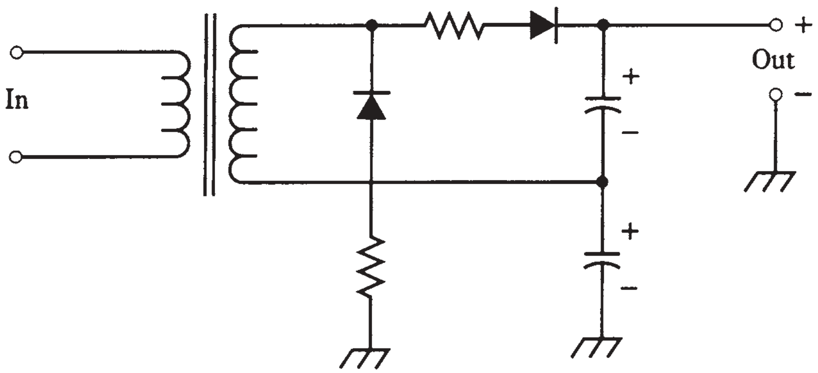
The main disadvantage of the bridge circuit is that it needs four diodes rather than two. This doesn't always amount to much in terms of cost, but it can be important when a power supply must deliver a high current. Then, the extra diodes—two for each half of the cycle, rather than one—dissipate more overall heat energy. When current is used up as heat, it can't go to the load. Therefore, center-tap circuits are preferable in high-current applications.

The voltage doubler

By using diodes and capacitors connected in certain ways, a power supply can be made to deliver a multiple of the peak ac input voltage. Theoretically, large whole-number multiples are possible. But you won't often see power supplies that make use of multiplication factors larger than 2.

In practice, *voltage multipliers* are practical only when the load draws low current. Otherwise, the regulation is poor; the output voltage changes considerably with changes in the load resistance. This bugaboo gets worse and worse as the multiplication factor increases. This is why engineers don't attempt to make, say, a factor-of-16 voltage multiplier. For a good high-voltage power supply, the best approach is to use a step-up transformer, not a voltage multiplier.

A *voltage-doubler* circuit is shown in Fig. 21-7. This circuit works on the whole ac input wave cycle, and is therefore called a *full-wave voltage doubler.* Its dc output voltage, when the current drawn is low, is about twice the *peak* ac input voltage, or about 2.8 times the rms ac input voltage.



21-7 A full-wave voltage doubler.

Notice the capacitors in this circuit. The operation of any voltage multiplier is dependent on the ability of these capacitors to hold a charge, even when a load is connected to the output of the supply. Thus, the capacitors must have large values. If the intent is to get a high dc voltage from the supply, massive capacitors will be necessary.

Also, notice the resistors in series with the diodes. These have low values, similar to those needed when diodes are connected in parallel. When the supply is switched on, the capacitors draw a huge initial charging current. Without the resistors, it would be necessary to use diodes with astronomical I_0 ratings. Otherwise the *surge current* would burn them out.

This circuit subjects the diodes to a PIV of 2.8 times the rms ac input voltage. Therefore, they should be rated for PIV of at least 4.2 times the rms ac input voltage.

In this circuit, each capacitor charges to the peak ac input voltage when there is no load (the output current is zero). As the load draws current, the capacitors will have trouble staying charged to the peak ac input voltage. This isn't much of a problem as long as the load is light, that is, if the current is low. But,for heavy loads, the output voltage will drop, and it will not be smooth dc.

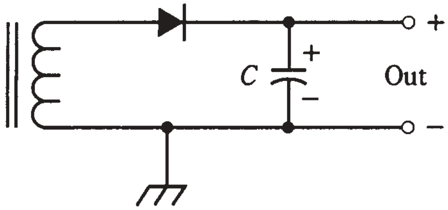
The major difference between the voltage doubler and the supplies discussed previously, besides the increased output voltage, is the fact that the dc output is filtered. The capacitors serve two purposes: to boost the voltage and to filter the output. Additional filtering might be wanted to smooth out the dc still more, but the circuit of Fig. 21-7 is a complete, if crude, power supply all by itself.

The filter

Electronic equipment doesn't like the pulsating dc that comes straight from a rectifier. The *ripple* in the waveform must be smoothed out, so that pure, battery-like dc is supplied. The *filter* does this.

Capacitors alone

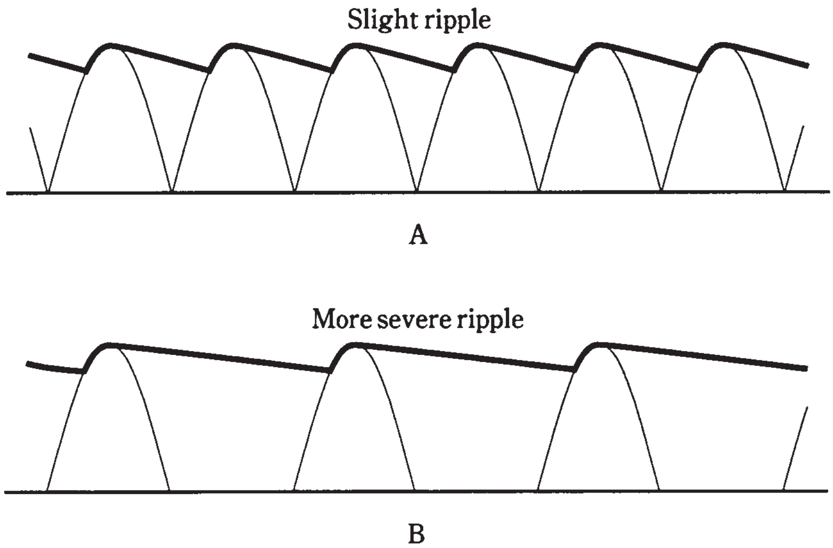
The simplest filter is one or more large-value capacitors, connected in parallel with the rectifier output (Fig. 21-8). Electrolytic capacitors are almost always used. They are *polarized*; they must be hooked up in the right direction. Typical values range in the hundreds or thousands of microfarads.



21-8 A simple filter. The capacitor, C, should have a large capacitance.

The more current drawn, the more capacitance is needed for good filtering. This is because the load resistance decreases as the current increases. The lower the load resistance, the faster the filter capacitors will discharge. Larger capacitances hold charge for a longer time with a given load.

Filter capacitors work by "trying" to keep the dc voltage at its peak level (Fig. 21-9). This is easier to do with the output of a full-wave rectifier (shown at A) as compared with a half-wave circuit (at B). The remaining waveform bumps are the ripple. With a half-wave rectifier, this ripple has the same frequency as the ac, or 60 Hz. With a full-wave supply, the ripple is 120 Hz. The capacitor gets recharged twice as often with a full-wave rectifier, as compared with a half-wave rectifier. This is why the ripple is less severe, for a given capacitance, with full-wave circuits.

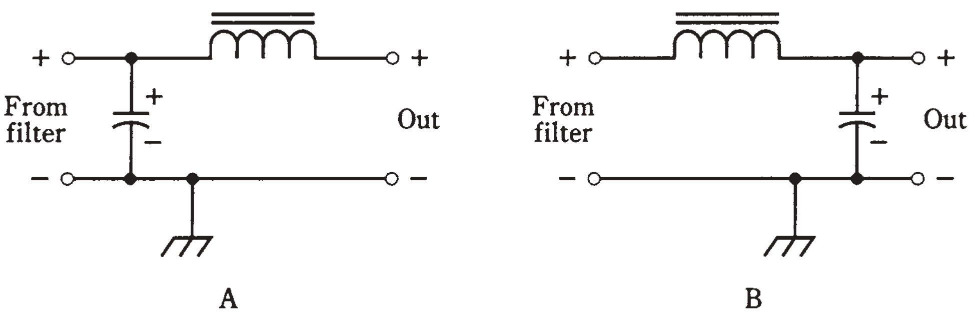


21-9 Filtered output for full-wave rectification (A) and half-wave rectification (B).

Capacitors and chokes

Another way to smooth out the dc from a rectifier is to use an extremely large inductance in series with the output. This is always done in conjunction with parallel capacitance. The inductance, called a *filter choke,* is on the order of several henrys. If the coil must carry a lot of current, it will be physically bulky.

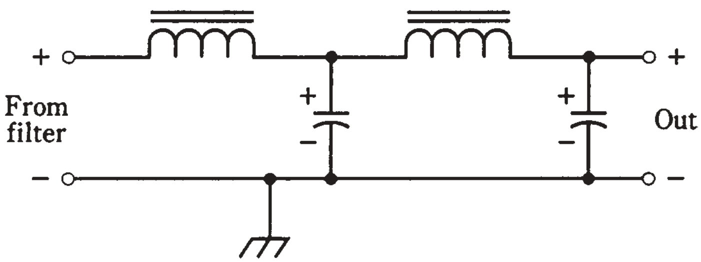
Sometimes the capacitor is placed ahead of the choke. This circuit is a *capacitor-input filter* (Fig. 21-10A). If the coil comes ahead of the capacitor, the circuit is a *choke-input filter* (Fig. 21-10B).



21-10 Capacitor-input (A) and choke-input (B) filtering.

Engineers might use capacitor-input filtering when the load is not expected to be very great. The output voltage is higher with a capacitor-input circuit than with a choke-input circuit. If the supply needs to deliver large or variable amounts of current, a choke-input filter is a better choice, because the output voltage is more stable.

If a supply must have a minimum of ripple, two or three capacitor/choke pairs might be *cascaded,* or connected one after the other (Fig. 21-11). Each pair is called a *section.* Multisection filters can consist of either capacitor-input or choke-input sections, but the two types are never mixed.



21-11 Two choke-input filter sections in cascade.

Voltage regulation

A full-wave rectifier, followed by a choke-input filter, offers fairly stable voltage under varying load conditions. But *voltage regulator* circuitry is needed for electronic devices that are finicky about the voltage they get.

Zener diodes

You learned about Zener diodes in the last chapter. If a reverse-biased Zener diode is connected across the output of a power supply, as shown back in Fig. 20-7, the diode will limit the output voltage of the supply by "brute force" as long as it has a high enough power rating.

Zener/transistor regulation

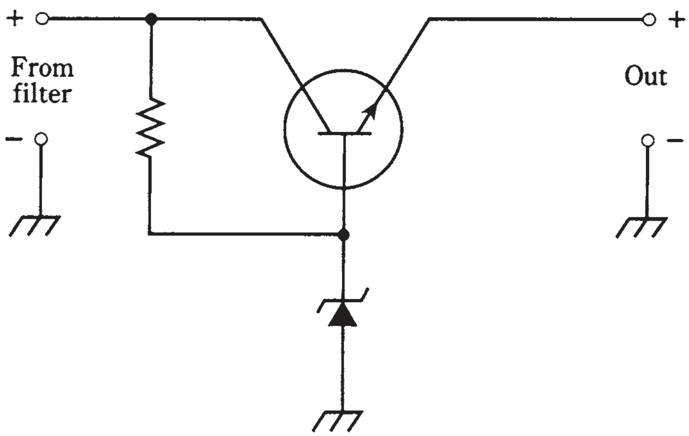
A Zener-diode voltage regulator is not very efficient if the load is heavy. When a supply must deliver high current, a power transistor is used along with the Zener diode to obtain regulation (Fig. 21-12). This greatly reduces the strain on the Zener diode, so that a lower-power (and therefore less costly) diode can be used.

Integrated circuits

In recent years, voltage regulators have become available in *integrated-circuit (IC)* form. You just connect the IC, perhaps along with some external components, at the output of the filter. This method provides the best possible regulation at low and moderate voltages. Even if the output current changes from zero to maximum, the output voltage stays exactly the same, for all practical purposes.

Regulator tubes

Occasionally, you'll find a power supply that uses a gas-filled *tube,* rather than solid-state components, to obtain regulation. The tube acts something like a very-high-power Zener



21-12 A voltage regulator circuit using a Zener diode and a power transistor.

diode. The voltage drop across a gaseous tube, designed for voltage regulation, is nearly constant. Tubes are available for regulation at moderately high voltages.

Surge current

At the instant a power supply is switched on, a sudden current surge occurs, even with no load at the output. This is because the filter capacitor(s) need an initial charge, and they draw a lot of current for a short time. The surge current is far greater than the operating current. This can destroy the rectifier diodes. The phenomenon is worst in high-voltage supplies and voltage-multiplier circuits. Diode failure can be prevented in at least four different ways.

The first method uses "brute force." You can simply use diodes with a current rating of many times the operating level. The main disadvantage is cost. High-voltage, high-current diodes can get expensive.

A second method involves connecting several units in parallel wherever a diode is called for in the circuit. This is actually a variation on the first method. The overall cost might be less. Current-equalizing resistors are necessary.

A third scheme for surge protection is to apply the input voltage little by little. A variable transformer, called a *Variac*, is useful for this. You start at zero input and turn a knob to get up to the full voltage. This can completely get rid of the current surge.

A fourth way to limit the current surge is to use an automatic switching circuit in the transformer primary. This applies a reduced ac voltage for a second or two, and then switches in the full input voltage.

Which of these methods is best? It depends on the overall cost, the operating convenience, and the whim of the design engineer.

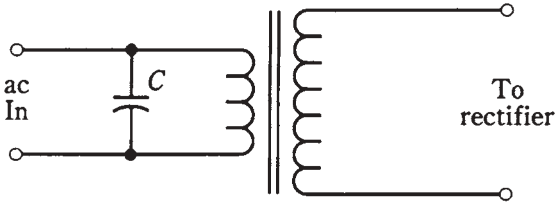
Transient suppression

The ac on the utility line is a sine wave with a constant rms voltage near 117 V. But there are "spikes," known as *transients*, lasting microseconds or milliseconds, that attain peak values of several hundred or even several thousand volts.

Transients are caused by sudden changes in the load in a utility circuit. Lightning can also produce them. Unless they are suppressed, they can destroy the diodes in a power supply. Transients can also befuddle the operation of sensitive equipment like personal computers.

The simplest way to get rid of most transients is to place a capacitor of about 0.01 µF, rated for 600 V or more, across the transformer primary (Fig. 21-13). Commercially made *transient suppressors* are also available.

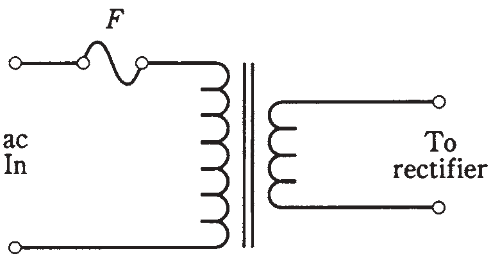
In the event of a thunderstorm locally, the best way to protect equipment is to unplug it from the wall outlet. This is inconvenient, of course. But if you have a personal computer, hi-fi set, or other electronic appliance that you like a lot, it's not a bad idea.



21-13 A capacitor, C, in parallel with the primary of the transformer, helps suppress transients.

Fuses and breakers

A *fuse* is a piece of soft wire that melts, breaking a circuit if the current exceeds a certain level. Fuses are placed in series with the transformer primary (Fig. 21-14). Any component failure, short circuit, or overload that might cause catastrophic damage (or fire!) will burn the fuse out. Fuses are easy to replace, although it's aggravating if a fuse blows and you don't have replacements on hand.



21-14 A fuse, F, in series with the ac input protects the transformer and diode, in case of overload.

If a fuse blows, it *must* be replaced with another of the same rating. If the replacement fuse is rated too low in current, it will probably blow out right away, or soon after it has been installed. If the replacement fuse is rated too high in current, it might not protect the equipment.

Fuses are available in two types: *quick-break* and *slow-blow.* You can usually recognize a slow-blow fuse by the spring inside. A quick-break fuse has only a wire or foil strip. When replacing a fuse, use the right kind. Quick-break fuses in slow-blow situations might burn out needlessly; slow-blow units in quick-break environments won't provide the proper protection.

Circuit breakers do the same thing as fuses, except that a breaker can be reset by turning off the power supply, waiting a moment, and then pressing a button or flipping a switch. Some breakers reset automatically when the equipment has been shut off for a certain length of time.

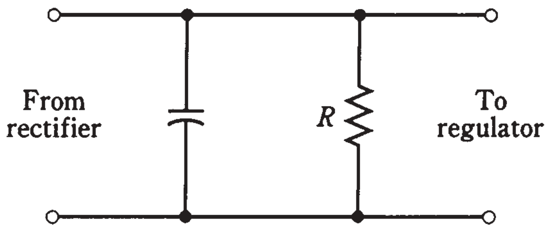
If a fuse or breaker keeps blowing out often, or if it blows immediately after you've replaced or reset it, then something is wrong with the supply or with the equipment connected to it.

Personal safety

Power supplies can be dangerous. This is especially true of high-voltage circuits, but anything over 12 V should be treated as potentially lethal.

A power supply is not necessarily safe after it has been switched off. Filter capacitors can hold the charge for a long time. In high-voltage supplies of good design, *bleeder resistors* of a high ohmic value (Fig. 21-15) are connected across each filter capacitor, so that the capacitors will discharge in a few minutes after the supply is turned off. But *don't* bet your life on components that might not be there, and that *can and do* sometimes fail.

21-15 A bleeder resistor, R, allows filter capacitors to discharge when a supply is shut off.



Most manufacturers supply safety instructions and precautions with equipment carrying hazardous voltages. But *don't* assume something is safe just because dangers aren't mentioned in the instructions.

Warning If you have any doubt about your ability to safely work with a power supply, then leave it to a professional.

In this chapter, you've had a look at power supplies from a general standpoint. Whole books have been written on the subject of power-supply engineering. If you want to design or build a power supply, you should refer to a college-level text or, better yet, a professional power-supply design manual.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The output of a rectifier is:
   A. 60-Hz ac.
   B. Smooth dc.
   C. Pulsating dc.
   D. 120-Hz ac.

2. Which of the following might not be needed in a power supply?
   A. The transformer.
   B. The filter.
   C. The rectifier.
   D. All of the above are generally needed.

3. Of the following appliances, which would need the biggest transformer?
   A. A clock radio.
   B. A TV broadcast transmitter.
   C. A shortwave radio receiver.
   D. A home TV set.

4. An advantage of full-wave bridge rectification is:
   A. It uses the whole transformer secondary for the entire ac input cycle.
   B. It costs less than other rectifier types.
   C. It cuts off half of the ac wave cycle.
   D. It never needs a regulator.

5. In. a supply designed to provide high power at low voltage, the best rectifier design would probably be:
   A. Half-wave.
   B. Full-wave, center-tap.
   C. Bridge.
   D. Voltage multiplier.

6. The part of a power supply immediately preceding the regulator is:
   A. The transformer.
   B. The rectifier.
   C. The filter.
   D. The ac input.

7. If a half-wave rectifier is used with 117-V rms ac (house mains), the average dc output voltage is about:
   - A. 52.7 V.
   - B. 105 V.
   - C. 117 V.
   - D. 328 V.
8. If a full-wave bridge circuit is used with a transformer whose secondary provides 50 V rms, the PIV across the diodes is about:
   - A. 50 V.
   - B. 70 V.
   - C. 100 V.
   - D. 140 V.
9. The principal disadvantage of a voltage multiplier is:
   - A. Excessive current.
   - B. Excessive voltage.
   - C. Insufficient rectification.
   - D. Poor regulation.
10. A transformer secondary provides 10 V rms to a voltage-doubler circuit. The dc output voltage is about:
    - A. 14 V.
    - B. 20 V.
    - C. 28 V.
    - D. 36 V.
11. The ripple frequency from a full-wave rectifier is:
    - A. Twice that from a half-wave circuit.
    - B. The same as that from a half-wave circuit.
    - C. Half that from a half-wave circuit.
    - D. One-fourth that from a half-wave circuit.
12. Which of the following would make the best filter for a power supply?
    - A. A capacitor in series.
    - B. A choke in series.
    - C. A capacitor in series and a choke in parallel.
    - D. A capacitor in parallel and a choke in series.
13. If you needed exceptionally good ripple filtering for a power supply, the best approach would be to:
    - A. Connect several capacitors in parallel.

B. Use a choke-input filter.
C. Connect several chokes in series.
D. Use two capacitor/choke sections one after the other.

14. Voltage regulation can be accomplished by a Zener diode connected in:
A. Parallel with the filter output, forward-biased.
B. Parallel with the filter output, reverse-biased.
C. Series with the filter output, forward-biased.
D. Series with the filter output, reverse-biased.

15. A current surge takes place when a power supply is first turned on because:
A. The transformer core is suddenly magnetized.
B. The diodes suddenly start to conduct.
C. The filter capacitor(s) must be initially charged.
D. Arcing takes place in the power switch.

16. Transient suppression minimizes the chance of:
A. Diode failure.
B. Transformer failure.
C. Filter capacitor failure.
D. Poor voltage regulation.

17. If a fuse blows, and it is replaced with one having a lower current rating, there's a good chance that:
A. The power supply will be severely damaged.
B. The diodes will not rectify.
C. The fuse will blow out right away.
D. Transient suppressors won't work.

18. A fuse with nothing but a straight wire inside is probably:
A. A slow-blow type.
B. A quick-break type.
C. Of a low current rating.
D. Of a high current rating.

19. Bleeder resistors are:
A. Connected in parallel with filter capacitors.
B. Of low ohmic value.
C. Effective for transient suppression.
D. Effective for surge suppression.

20. To service a power supply with which you are not completely familiar, you should:

    A. Install bleeder resistors.

    B. Use proper fusing.

    C. Leave it alone and have a professional work on it.

    D. Use a voltage regulator.

22

CHAPTER

The bipolar transistor

THE WORD *TRANSISTOR* IS A CONTRACTION OF "CURRENT-*TRANS*FERRING *RESIStor*. " This is an excellent description of what a bipolar transistor does.

Bipolar transistors have two P-N junctions connected together. This is done in either of two ways: a P-type layer sandwiched between two N-type layers, or an N type layer between two P-type layers.

Bipolar transistors, like diodes, can be made from various semiconductor substances. Silicon is probably the most common material used.

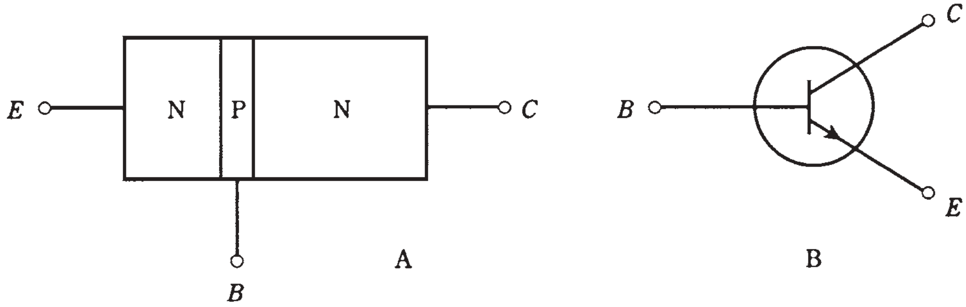
NPN versus PNP

A simplified drawing of an *NPN* transistor, and its schematic symbol, are shown in Fig. 22-1. The P-type, or center, layer is called the *base.* The thinner of the N-type semiconductors is the *emitter,* and the thicker is the *collector.* Sometimes these are labeled *B*, *E*, and *C* in schematic diagrams, although the transistor symbol alone is enough to tell you which is which.

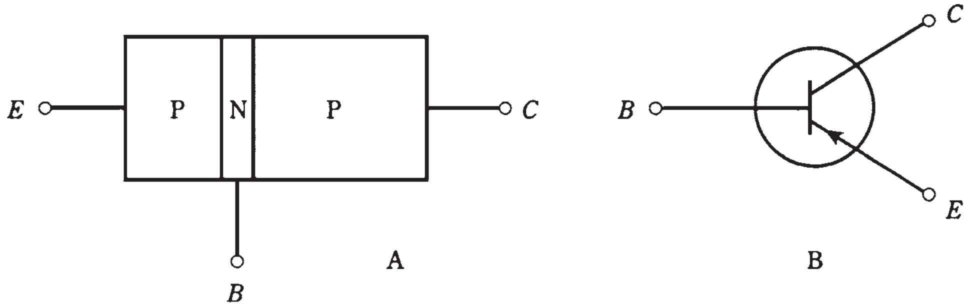
A *PNP* bipolar transistor is just the opposite of an NPN device, having two P-type layers, one on either side of a thin, N-type layer (Fig. 22-2). The emitter layer is thinner, in most units, than the collector layer.

You can always tell whether a bipolar transistor in a diagram is NPN or PNP. With the NPN, the arrow points outward; with the PNP it points inward. The arrow is always at the emitter.

Generally, PNP and NPN transistors can do the same things in electronic circuits. The only difference is the polarities of the voltages, and the directions of the currents. In most applications, an NPN device can be replaced with a PNP device or vice versa, and the power-supply polarity reversed, and the circuit will still work as long as the new device has the appropriate specifications.



22-1 At A, pictorial diagram of an NPN transistor. At B, the schematic symbol. Electrodes are E = emitter, B = base, C = collector.



22-2 At A, pictorial diagram of a PNP transistor. At B, schematic symbol; E = emitter, B = base, C = collector.

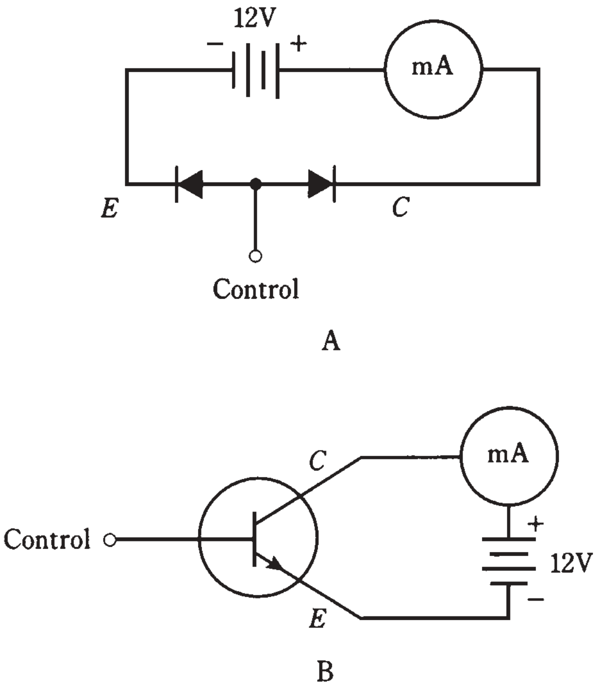
There are many different kinds of NPN or PNP bipolar transistors. Some are used for radio-frequency amplifiers and oscillators; others are intended for audio frequencies. Some can handle high power, and others cannot, being made for weak-signal work. Some bipolar transistors are manufactured for the purpose of switching, rather than signal processing. If you look through a catalog of semiconductor components, you'll find hundreds of different bipolar transistors, each with its own unique set of specifications.

Why, you might ask, need there be two different kinds of bipolar transistor (NPN and PNP), if they do exactly the same things? Sometimes engineers need to have both kinds in one circuit. Also, there are some subtle differences in behavior between the two types. These considerations are beyond the scope of this book. But you should know that the NPN/PNP duality is not just whimsy on the part of people who want to make things complicated.

NPN biasing

You can think of a bipolar transistor as two diodes in reverse series. You can't normally connect two diodes together this way and get a good transistor, but the analogy is good for *modeling* the behavior of bipolar transistors, so that their operation is easier to understand.

A dual-diode NPN transistor model is shown in Fig. 22-3. The base is formed by the connection of the two diode anodes. The emitter is one of the cathodes, and the collector is the other.



22-3 At A, simple NPN circuit using dual-diode modeling. At B, the actual transistor circuit.

The normal method of biasing an NPN transistor is to have the emitter negative and the collector positive. This is shown by the connection of the battery in Fig. 22-3. Typical voltages for this battery (although it might be, and often is, a dc power supply) range from 3 V to about 50 V. Most often, 6 V, 9 V, or 12 V supplies are used.

The base is labeled "control" in the figure. This is because the flow of current through the transistor depends critically on the *base bias* voltage, E_B, relative to the *emitter-collector bias* voltage, E_C.

Zero bias

Suppose that the base isn't connected to anything, or is at the same potential as the emitter. This is *zero base bias,* sometimes simply called *zero bias.* How much current will flow through the transistor? What will the milliammeter (mA) show?

The answer is that there will be no current. The meter will register zero.

Recall the discussion of diode behavior from the previous chapter. No current flows through a P-N junction unless the forward bias is at least equal to the forward breakover voltage. (For silicon, this is about 0.6 V.) But here, the forward bias is zero. Therefore, the emitter-base current, often called simply *base current* and denoted I_B, is zero, and the emitter-base junction does not conduct. This prevents any current from flowing between the emitter and collector, unless some signal is injected at the base to change the situation. This signal would have to be of positive polarity and would need to be at least equal to the forward breakover voltage of the junction.

Reverse bias

Now imagine that another battery is connected to the base at the point marked "control," so that E_B is negative with respect to the emitter. What will happen? Will current flow through the transistor?

The answer is no. The addition of this new battery will cause the emitter-base (E-B) junction to be *reverse-biased*. It is assumed that this new battery is not of such a high voltage that avalanche breakdown takes place at the junction.

A signal might be injected to overcome the reverse-bias battery and the forward breakover voltage of the E-B junction, but such a signal would have to be of a high, positive voltage.

Forward bias

Now suppose that E_B is made positive, starting at small voltages and gradually increasing.

If this *forward bias* is less than the forward breakover voltage, no current will flow. But as the base voltage E_B reaches breakover, the E-B junction will start to conduct.

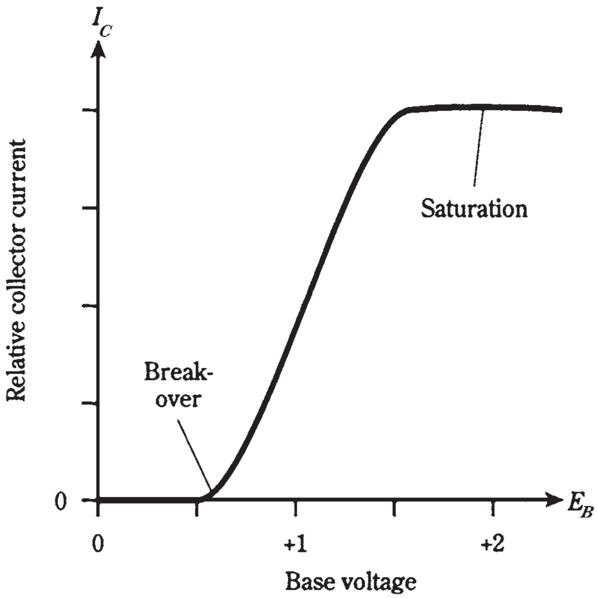
The base-collector (B-C) junction will remain reverse-biased as long as E_B is less than the supply voltage (in this case 12 V). In practical transistor circuits, it is common for E_B to be set at a fraction of the supply voltage.

Despite the reverse bias of the B-C junction, the emitter-collector current, called *collector current* and denoted I_C, will flow once the E-B junction conducts. In a real transistor (Fig. 22-3B), the meter reading will jump when the forward breakover voltage of the E-B junction is reached. Then even a small rise in E_B, attended by a rise in I_B, will cause a big increase in I_C. This is shown graphically in Fig. 22-4.

Saturation

If E_B continues to rise, a point will eventually be reached where I_C increases less rapidly. Ultimately, the I_C vs. E_B curve will level off. The transistor is then *saturated* or *in saturation.* It is conducting as much as it possibly can; it's "wide open."

This property of three-layer semiconductors, in which reverse-biased junctions can sometimes pass current, was first noticed in the late forties by the engineers Bardeen, Brattain, and Shockley at the Bell Laboratories. When they saw how current variations were magnified by a three-layer device of this kind, they knew they were on to something. They envisioned that the effect could be exploited to amplify weak signals, or to use small currents to switch much larger ones. They must have been excited, but they surely had no idea how much their discovery would affect the world.



22-4 Relative collector current (I_C) as a function of base voltage (E_B) for a hypothetical silicon transistor.

PNP biasing

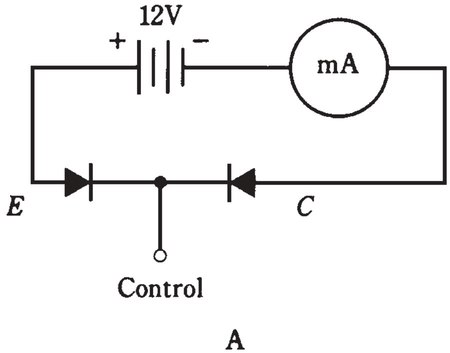
For a PNP transistor, the situation is just a "mirror image" of the case for an NPN device. The diodes are turned around the opposite way, the arrow points inward rather than outward in the transistor symbol, and all the polarities are reversed. The dual-diode PNP model, along with the actual bipolar transistor circuit, are shown in Fig. 22-5. In the discussion above, simply replace every occurrence of the word "positive" with the word "negative."

You need not be concerned with what actually goes on inside the semiconductor materials in NPN and PNP transistors. The important thing is the fact that either type of device can serve as a sort of "current valve. " Small changes in the base voltage, E_B, cause small changes in the base current, I_B. This induces large fluctuations in the current I_C through the transistor.

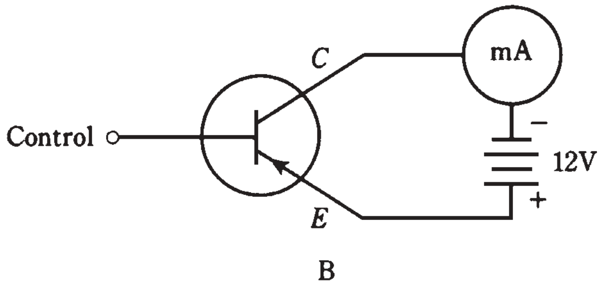
In the following discussion, and in most circuits that appear later in this book, you'll see NPN transistors used almost exclusively. This doesn't mean that NPN is better than PNP; in almost every case, you can replace each NPN transistor with a PNP, reverse the polarity, and get the same results. The motivation is to save space and avoid redundancy.

Biasing for current amplification

Because a small change in the base current, I_B, results in a large collector-current (I_C) variation when the bias is just right, a transistor can operate as a *current amplifier*. It might be more technically accurate to say that it is a "current-fluctuation amplifier," because it's the magnification of current variations, not the absolute current, that's important.



22-5 At A, simple PNP circuit using dual-diode modeling. At B, the actual transistor circuit.



If you look at Fig. 22-4 closely, you'll see that there are some bias values at which a transistor won't give current amplification. If the E-B junction is not conducting, or if the transistor is in saturation, the curve is horizontal. A small change (to the left and right) of the base voltage, E_B, in these portions of the curve, will cause little or no up-and-down variation of I_C.

But if the transistor is biased near the middle of the straight-line part of the curve in Fig. 22-4, the transistor will work as a current amplifier.

Static current amplification

Current amplification is often called *beta* by engineers. It can range from a factor of just a few times up to hundreds of times.

One method of expressing the beta of a transistor is as the *static forward current transfer ratio*, abbreviated H_{FE}. Mathematically, this is

$$H_{FE} = I_C/I_B$$

Thus, if a base current, I_B, of 1 mA results in a collector current, I_C, of 35 mA, $H_{FE} = 35/1 = 35$. If $I_B = 0.5$ mA yields $I_C = 35$ mA, then $H_{FE} = 35/0.5 = 70$.

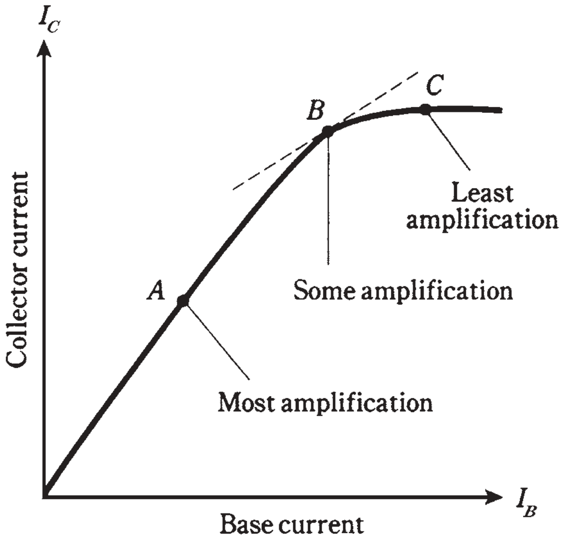
This definition represents the greatest current amplification possible with a given transistor.

Dynamic current amplification

Another way of specifying current amplification is as the ratio of *the difference in* I_C to *the difference in* I_B. Abbreviate the words "the difference in" by the letter *d.* Then, according to this second definition:

$$\text{Current amplification} = dI_C/dI_B$$

A graph of collector current versus base current (I_C vs I_B) for a hypothetical transistor is shown in Fig. 22-6. This graph resembles Fig. 22-4, except that current, rather than voltage, is on the horizontal scale. Three different points are shown, corresponding to various bias values.



22-6 Three different transistor bias points. See text for discussion.

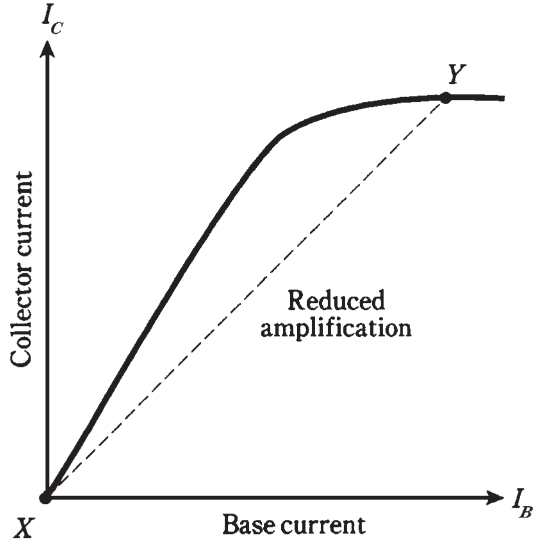
The ratio dI_C/dI_B is different for each of the points in this graph. Geometrically, dI_C/dI_B at a given point is the slope of a line tangent to the curve at that point. The tangent line for point *B* in Fig. 22-6 is a dotted, straight line; the tangent lines for points *A* and C lie right along the curve. The steeper the slope of the line, the greater is dI_C/dI_B.

Point A provides the highest dI_C/dI_B, as long as the input signal is small. This value is very close to H_{FE}. For small-signal amplification, point A represents a good bias level. Engineers would say that it's a good *operating point.*

At point B, dI_C/dI_B is smaller than at point A. (It might actually be less than 1.) At point C, dI_C/dI_B is practically zero. Transistors are rarely biased at these points.

Overdrive

Even when a transistor is biased for best operation (near point A in Fig. 22-6), a strong input signal can drive it to point B or beyond during part of the cycle. Then, dI_C/dI_B is reduced, as shown in Fig. 22-7. Points X and Y in the graph represent the instantaneous current extremes during the signal cycle.



22-7 Excessive input reduces amplification.

When conditions are like those in Fig. 22-7, there will be distortion in a transistor amplifier. The output waveform will not have the same shape as the input waveform. This *nonlinearity* can sometimes be tolerated; sometimes it cannot.

The more serious trouble with *overdrive* is the fact that the transistor is in or near saturation during part of the cycle. When this happens, you're getting "no bang for the buck." The transistor is doing futile work for a portion of every wave cycle. This reduces circuit efficiency, causes excessive collector current, and can overheat the base-collector (B-C) junction. Sometimes overdrive can actually destroy a transistor.

Gain versus frequency

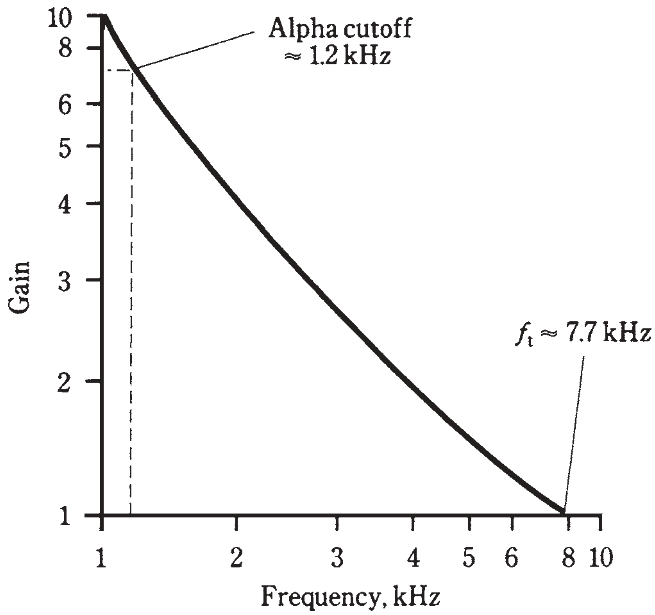
Another important specification for a transistor is the range of frequencies over which it can be used as an amplifier. All transistors have an amplification factor, or *gain,* that decreases as the signal frequency increases. Some devices will work well only up to a few megahertz; others can be used to several gigahertz.

Gain can be expressed in various different ways. In the above discussion, you learned a little about *current gain,* expressed as a ratio. You will also sometimes hear about *voltage gain* or *power gain* in amplifier circuits. These, too, can be expressed as ratios. For example, if the voltage gain of a circuit is 15, then the output signal voltage (rms, peak, or peak-to-peak) is 15 times the input signal voltage. If the power gain of a circuit is 25, then the output signal power is 25 times the input signal power.

There are two expressions commonly used for the gain-versus-frequency behavior of a bipolar transistor. The *gain bandwidth product,* abbreviated f_T, is the frequency at which the gain becomes equal to 1 with the emitter connected to ground. If you try to make an amplifier using a transistor at a frequency higher than its f_T, you'll fail! Thus f_T represents an absolute upper limit of sorts.

The *alpha cutoff frequency* of a transitor is the frequency at which the gain becomes 0.707 times its value at 1 kHz. A transistor might still have considerable gain at its alpha cutoff. By looking at the alpha cutoff frequency, you can get an idea of how rapidly the transistor loses gain as the frequency goes up. Some devices "die-off" faster than others.

Figure 22-8 shows the gain band width product and alpha cutoff frequency for a hypothetical transistor, on a graph of gain versus frequency. Note that the scales of this graph are nonlinear; they're "scrunched up" at the higher values. This type of graph is useful for showing some functions. It is called a *log-log* graph because both scales are *logarithmic* rather than linear.



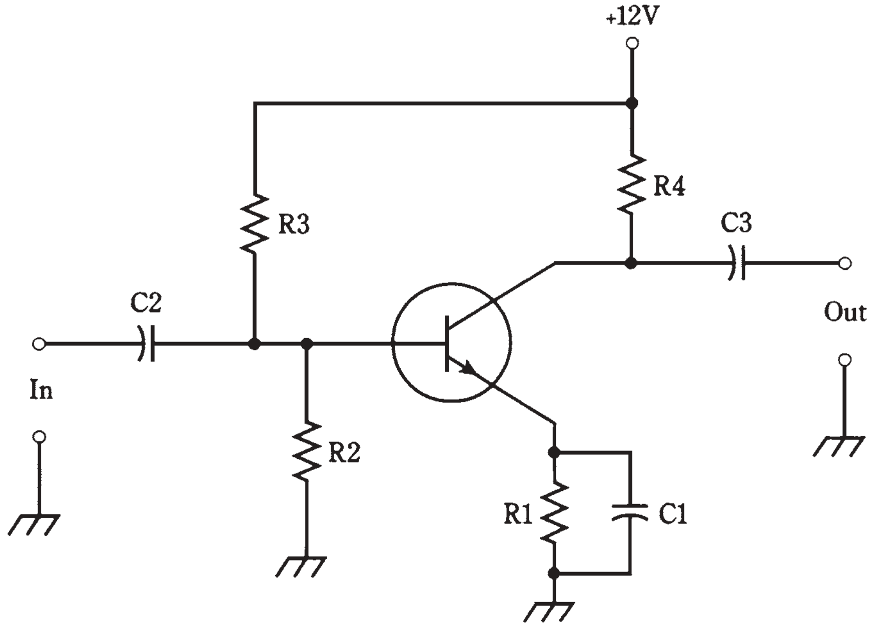
22-8 Alpha cutoff and gain bandwidth product for a hypothetical transistor.

Common emitter circuit

A transistor can be hooked up in three general ways. The emitter can be grounded for signal, the base can be grounded for signal, or the collector can be grounded for signal.

Probably the most often-used arrangement is the *common-emitter circuit.* "Common" means "grounded for the signal." The basic configuration is shown in Fig. 22-9.

A terminal can be at ground potential for the signal, and yet have a significant dc voltage. In the circuit shown, C1 looks like a dead short to the ac signal, so the emitter is at *signal ground.* But R1 causes the emitter to have a certain positive dc voltage with respect to ground (or a negative voltage, if a PNP transistor is used). The exact dc voltage at the emitter depends on the value of R1, and on the bias.



22-9 Common-emitter circuit configuration.

The bias is set by the ratio of resistances *R*2 and *R*3. It can be anything from zero, or ground potential, to + 12 V, the supply voltage. Normally it will be a couple of volts.

Capacitors C2 and C3 block dc to or from the input and output circuitry (whatever that might be) while letting the ac signal pass. Resistor R4 keeps the output signal from being shorted out through the power supply.

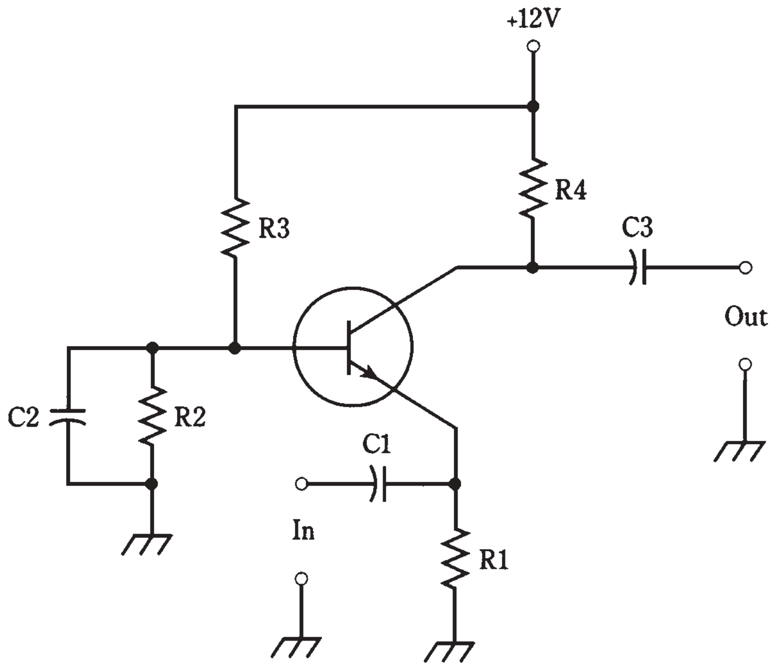
A signal voltage enters the common-emitter circuit through C2, where it causes the base current, I_B to vary. The small fluctuations in I_B cause large changes in the collector current, I_C. This current passes through R4, causing a fluctuating dc voltage to appear across this resistor. The ac part of this passes unhindered through C3 to the output.

The circuit of Fig. 22-9 is the basis for many amplifiers, from audio frequencies through ultra-high radio frequencies. The common-emitter configuration produces the largest gain of any arrangement. The output is 180 degrees out of phase with the input.

Common-base circuit

As its name implies, the *common-base circuit,* shown in general form by Fig. 22-10, has the base at signal ground.

The dc bias on the transistor is the same for this circuit as for the common-emitter circuit. The difference is that the input signal is applied at the emitter, instead of at the base. This causes fluctuations in the voltage across R1, causing variations in I_B. The result of these small current fluctuations is a large change in the dc current through R4. Therefore amplification occurs.



22-10 Common-base circuit configuration.

Instead of varying I_B by injecting the signal at the base, it's being done by injecting the signal at the emitter. Therefore, in the common-base arrangement, the output signal is in phase with the input, rather than out of phase.

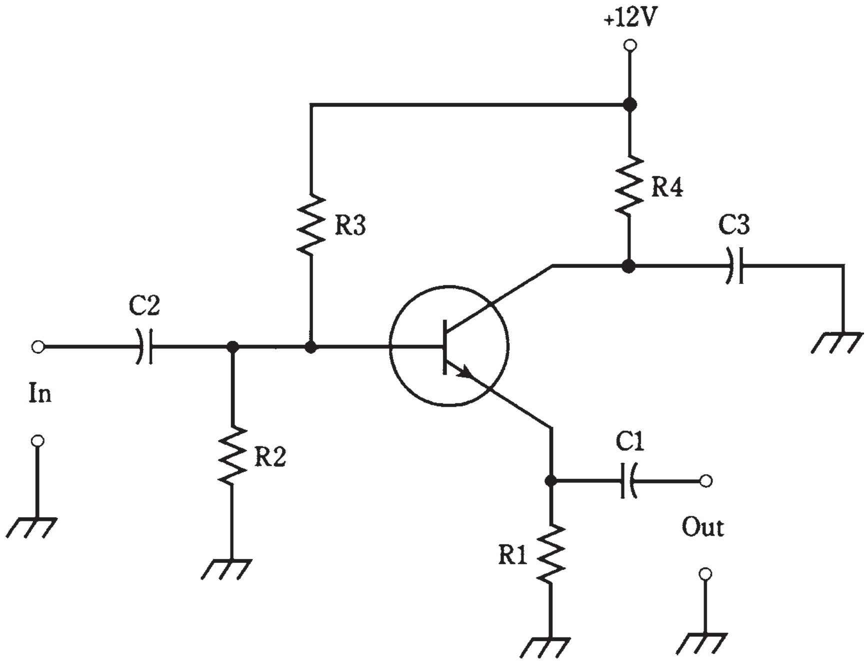
The signal enters through C1. Resistor R1 keeps the input signal from being shorted to ground. Bias is provided by R2 and R3. Capacitor C2 keeps the base at signal ground. Resistor R4 keeps the signal from being shorted out through the power supply. The output is through C3.

The common-base circuit provides somewhat less gain than a common-emitter circuit. But it is more stable than the common-emitter configuration in some applications, especially in radio-frequency power amplifiers.

Common-collector circuit

A *common-collector circuit* (Fig. 22-11) operates with the collector at signal ground. The input is applied at the base just as it is with the common-emitter circuit.

The signal passes through C2 onto the base of the transistor. Resistors R2 and R3 provide the correct bias for the base. Resistor R4 limits the current through the transistor. Capacitor C3 keeps the collector at signal ground. A fluctuating direct current flows through R1, and a fluctuating dc voltage therefore appears across it. The ac part of this voltage passes through C1 to the output. Because the output follows the emitter current, this circuit is sometimes called an *emitter follower circuit.*



22-11 Common-collector circuit configuration. This arrangement is also known as an emitter follower.

The output of this circuit is in phase with the input. The input impedance is high, and the output impedance is low. For this reason, the common-collector circuit can be used to match high impedances to low impedances. When well designed, an emitter follower works over a wide range of frequencies, and is a low-cost alternative to a broadband impedance-matching transformer.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. In a PNP circuit, the collector:
   - A. Has an arrow pointing inward.
   - B. Is positive with respect to the emitter.
   - C. Is biased at a small fraction of the base bias.
   - D. Is negative with respect to the emitter.

2. In many cases, a PNP transistor can be replaced with an NPN device and the circuit will do the same thing, provided that:
   - A. The supply polarity is reversed.

B. The collector and emitter leads are interchanged.
C. The arrow is pointing inward.
D. No! A PNP device cannot be replaced with an NPN.

3. A bipolar transistor has:
A. Three P-N junctions.
B. Three semiconductor layers.
C. Two N-type layers around a P-type layer.
D. A low avalanche voltage.

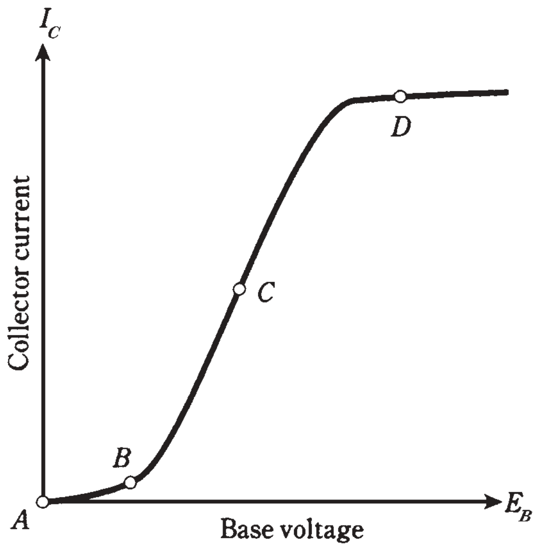
4. In the dual-diode model of an NPN transistor, the emitter corresponds to:
A. The point where the cathodes are connected together.
B. The point where the cathode of one diode is connected to the anode of the other.
C. The point where the anodes are connected together.
D. Either of the diode cathodes.

5. The current through a transistor depends on:
A. E_C.
B. E_B relative to E_C.
C. I_B.
D. More than one of the above.

6. With no signal input, a bipolar transistor would have the least I_C when:
A. The emitter is grounded.
B. The E-B junction is forward biased.
C. The E-B junction is reverse biased.
D. The E-B current is high.

7. When a transistor is conducting as much as it possibly can, it is said to be:
A. In cutoff.
B. In saturation.
C. Forward biased.
D. In avalanche.

8. Refer to Fig. 22-12. The best point at which to operate a transistor as a small-signal amplifier is:
A. A.
B. B.
C. C.
D. D.



22-12 Illustration for quiz questions 8, 9, 10, and 11.

9. In Fig. 22-12, the forward-breakover point for the E-B junction is nearest to:
   A. No point on this graph.
   B. B.
   C. C.
   D. D.
10. In Fig. 22-12, saturation is nearest to point:
    A. A.
    B. B.
    C. C.
    D. D.
11. In Fig. 22-12, the greatest gain occurs at point:
    A. A.
    B. B.
    C. C.
    D. D.
12. In a common-emitter circuit, the gain bandwidth product is:
    A. The frequency at which the gain is 1.
    B. The frequency at which the gain is 0.707 times its value at 1 MHz.
    C. The frequency at which the gain is greatest.
    D. The difference between the frequency at which the gain is greatest, and the frequency at which the gain is 1.

13. The configuration most often used for matching a high input impedance to a low output impedance puts signal ground at:
    A. The emitter.
    B. The base.
    C. The collector.
    D. Any point; it doesn't matter.

14. The output is in phase with the input in a:
    A. Common-emitter circuit.
    B. Common-base circuit.
    C. Common-collector circuit.
    D. More than one of the above.

15. The greatest possible amplification is obtained in:
    A. A common-emitter circuit.
    B. A common-base circuit.
    C. A common-collector circuit.
    D. More than one of the above.

16. The input is applied to the collector in:
    A. A common-emitter circuit.
    B. A common-base circuit.
    C. A common-collector circuit.
    D. None of the above.

17. The configuration noted for its stability in radio-frequency power amplifiers is the:
    A. Common-emitter circuit.
    B. Common-base circuit.
    C. Common-collector circuit.
    D. Emitter-follower circuit.

18. In a common-base circuit, the output is taken from the:
    A. Emitter.
    B. Base.
    C. Collector.
    D. More than one of the above.

19. The input signal to a transistor amplifier results in saturation during part of the cycle. This produces:
    A. The greatest possible amplification.
    B. Reduced efficiency.

C. Avalanche effect.

D. Nonlinear output impedance.

20. The gain of a transistor in a common-emitter circuit is 100 at a frequency of 1000 Hz. The gain is 70.7 at 335 kHz. The gain drops to 1 at 210 MHz. The alpha cutoff is:

A. 1 kHz.

B. 335 kHz.

C. 210 MHz.

D. None of the above.

23

CHAPTER

The field-effect transistor

BIPOLAR TRANSISTORS BEHAVE AS THEY DO BECAUSE CURRENT VARIATIONS AT one P-N junction produce larger current variations at another. You've seen a simplified picture of how this happens, and how the effect can be exploited to get current amplification.

The bipolar transistor isn't the only way that semiconductors can be combined to get amplification effects. The other major category of transistor, besides the bipolar device, is the *field-effect transistor* or *FET.* There are two main types of FET: the *junction FET (JFET)* and the *metal-oxide FET (MOSFET).*

Principle of the JFET

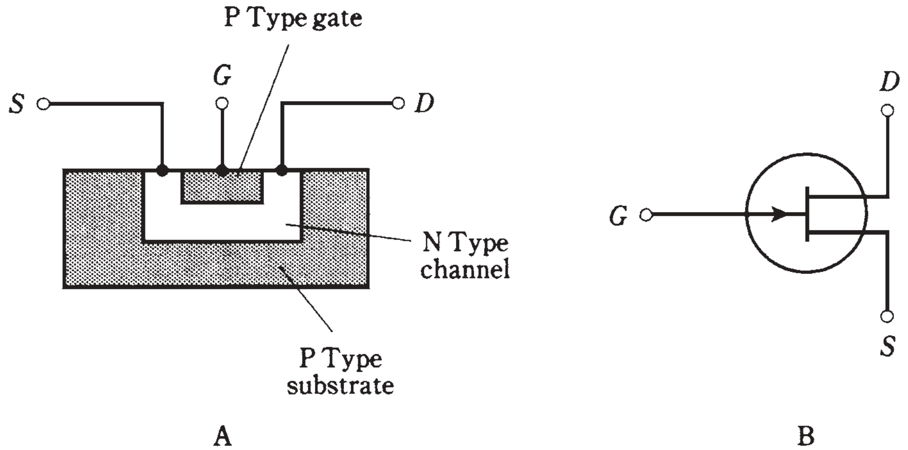
A JFET can have any of several different forms. They all work the same way: the current varies because of the effects of an electric field within the device.

The workings inside a JFET can be likened to the control of water flow through a garden hose. Electrons or holes pass from the *source* (S) electrode to the *drain* (D). This results in a *drain current,* I_D, that is generally the same as the *source current,* I_S. This is analogous to the fact that the water comes out of a garden hose at the same rate it goes in (assuming that there aren't any leaks in the hose).

The rate of flow of charge carriers—that is, the current—depends on the voltage at a regulating electrode called the *gate (G).* Fluctuations in *gate voltage,* E_G, cause changes in the current through the *channel,* I_S or I_D. Small fluctuations in the control voltage E_G can cause large variations in the flow of charge carriers through the JFET. This translates into voltage amplification in electronic circuits.

N-channel versus P-channel

A simplified drawing of an *N-channel* JFET, and its schematic symbol, are shown in Fig. 23-1. The N-type material forms the channel, or the path for charge carriers. In the N-channel device, the majority carriers are electrons. The source is at one end of the channel, and the drain is at the other. You can think of electrons as being "injected" into the source and "collected" from the drain as they pass through the channel. The drain is positive with respect to the source.



23-1 Simplified cross-sectional drawing of an N-channel JFET (at A) and its schematic symbol (at B).

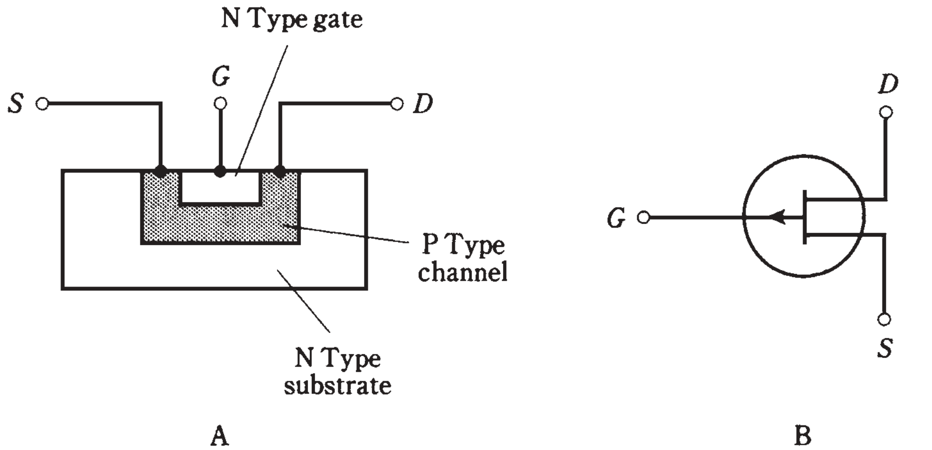
In an N-channel device, the gate consists of P-type material. Another, larger section P-type material, called the *substrate*, forms a boundary on the side of the channel opposite the gate. The JFET is formed in the substrate during manufacture by a process known as *diffusion*.

The voltage on the gate produces an electric field that interferes with the flow of charge carriers through the channel. The more negative E_G becomes, the more the electric field chokes off the current though the channel, and the smaller I_D becomes.

A *P-channel* JFET (Fig. 23-2) has a channel of P-type semiconductor. The majority charge carriers are holes. The drain is negative with respect to the source. In a sense, holes are "injected" into the source and are "collected" from the drain. The gate and the substrate are of N-type material.

In the P-channel JFET, the more positive E_G gets, the more the electric field chokes off the current through the channel, and the smaller I_D becomes.

You can recognize the N-channel device by the arrow pointing inward at the gate, and the P-channel JFET by the arrow pointing outward. Also, you can tell which is which (sometimes arrows are not included in schematic diagrams) by the power-supply polarity. A positive drain indicates an N-channel JFET, and a negative drain indicates a P-channel type.



23-2 Simplified cross-sectional drawing of a P-channel JFET (at A) and its schematic symbol (at B).

In electronic circuits, N-channel and P-channel devices can do the same kinds of things. The main difference is the polarity. An N-channel device can almost always be replaced with a P-channel JFET, and the power-supply polarity reversed, and the circuit will still work if the new device has the right specifications. Just as there are different kinds of bipolar transistors, there are various types of JFETs, each suited to a particular application. Some JFETs work well as weak-signal amplifiers and oscillators; others are made for power amplification.

Field-effect transistors have some advantages over bipolar devices. Perhaps the most important is that FETs are available that generate less internal noise than bipolar transistors. This makes them excellent for use in sensitive radio receivers at very high or ultra-high frequencies.

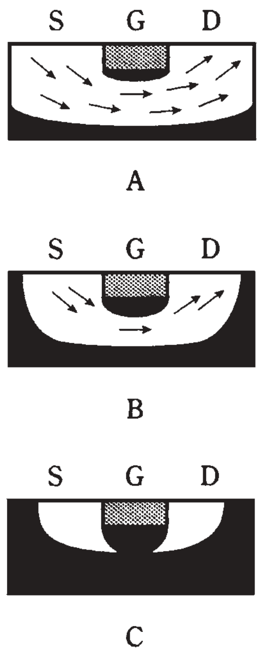
Field-effect transistors have high input impedances. The gate controls the flow of charge carriers by means of an electric *field*, rather than via an electric *current*.

Depletion and pinchoff

Either the N-channel or the P-channel JFET works because the voltage at the gate causes an electric field that interferes, more or less, with the flow of charge carriers along the channel. A simplified drawing of the situation for an N-channel device is shown in Fig. 23-3. For a P-channel device, just interchange polarity (minus/plus) and semiconductor types (N/P) in this discussion.

As the drain voltage E_D increases, so does the drain current I_D, up to a certain level-off value. This is true as long as the gate voltage E_G is constant, and is not too large negatively.

But as E_G becomes increasingly negative (Fig. 23-3A), a *depletion region* (solid black) begins to form in the channel. Charge carriers cannot flow in this region; they must pass through a narrowed channel. The more negative E_G becomes, the wider the depletion region gets, as shown at B. Ultimately, if the gate becomes negative enough,



23-3 At A, depletion region (solid area) is not wide, and many charge carriers (arrows) flow. At B, depletion region is wider, channel is narrower, and fewer carriers flow. At C, channel is completely obstructed, and no carriers flow.

the depletion region will completely obstruct the flow of charge carriers. This is called *pinchoff,* and is illustrated at C.

Again, think of the garden-hose analogy. More negative gate voltages, E_G, correspond to stepping harder and harder on the hose. When pinchoff takes place, you've cut off the water flow entirely, perhaps by bearing down with all your weight on one foot! Biasing beyond pinchoff is something like loading yourself up with heavy weights as you balance on the hose, thereby shutting off the water flow with extra force.

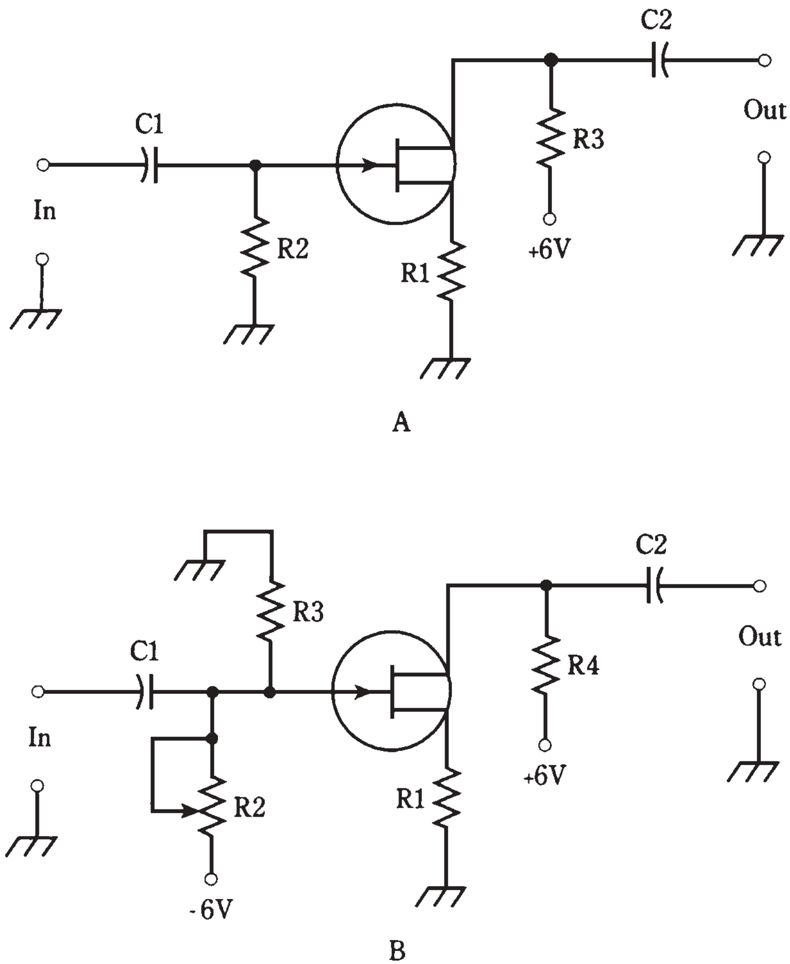
JFET biasing

Two biasing arrangements for an N-channel JFET are shown in Fig. 23-4. These hookups are similar to the way an NPN bipolar transistor is connected, except that the source-gate (SG) junction is not forward-biased.

At A, the gate is grounded through resistor R2. The source resistor, R1, limits the current through the JFET. The drain current, I_D, flows through R3, producing a voltage across this resistor. The ac output signal passes through C2.

At B, the gate is connected to a voltage that is negative with respect to ground through potentiometer R2. Adjusting this potentiometer results in a variable negative E_G between R2 and R3. Resistor R1 limits the current through the JFET. The drain current, I_D, flows through R4, producing a voltage across it; the ac output signal passes through C2.

In both of these circuits, the drain is positive relative to ground. For a P-channel JFET, reverse the polarities in Fig. 23-4. The connections are somewhat similar to the way a PNP bipolar transistor is used, except the SG junction isn't forward-biased.



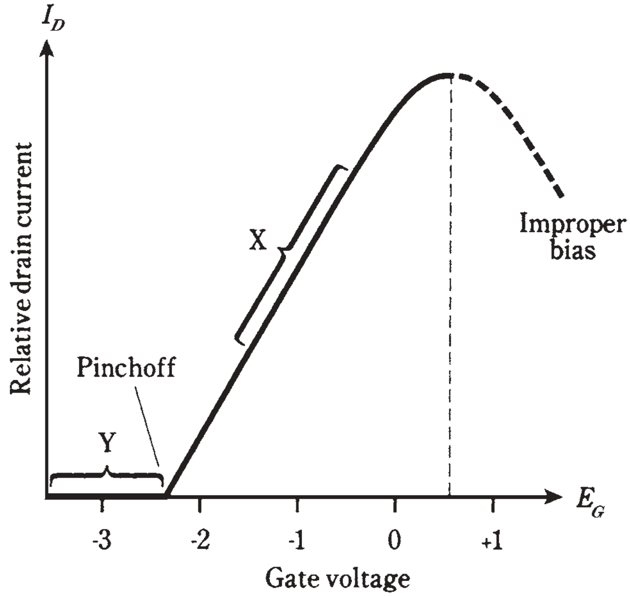
23-4 Two methods of biasing an N-channel JFET. At A, fixed gate bias; at B, variable gate bias.

Typical JFET power-supply voltages are comparable to those with bipolar transistors. The voltage between the source and drain, abbreviated E_D, can range from about 3 V to 150 V; most often it is 6 V to 12 V.

The biasing arrangement in Fig. 23-4A is commonly used for weak-signal amplifiers, low-level amplifiers and oscillators. The scheme at B is more often employed in power amplifiers having a substantial input signal.

Voltage amplification

The graph of Fig. 23-5 shows the drain (channel) current, I_D, as a function of the gate bias voltage, E_G, for a hypothetical N-channel JFET. The drain voltage, E_D, is assumed to be constant.



23-5 Relative drain current as a function of gate voltage for a hypothetical N-channel JFET.

When E_G is fairly large and negative, the JFET is pinched off, and no current flows through the channel. As E_G gets less negative, the channel opens up, and current begins flowing. As E_G gets still less negative, the channel gets wider and the current I_D increases. As E_G approaches the point where the SG junction is at forward breakover, the channel conducts as well as it possibly can.

If E_G becomes positive enough so that the SG junction conducts, the JFET will no longer work properly. Some of the current in the channel will then be shunted off through the gate, a situation that is never desired in a JFET. The hose will spring a leak!

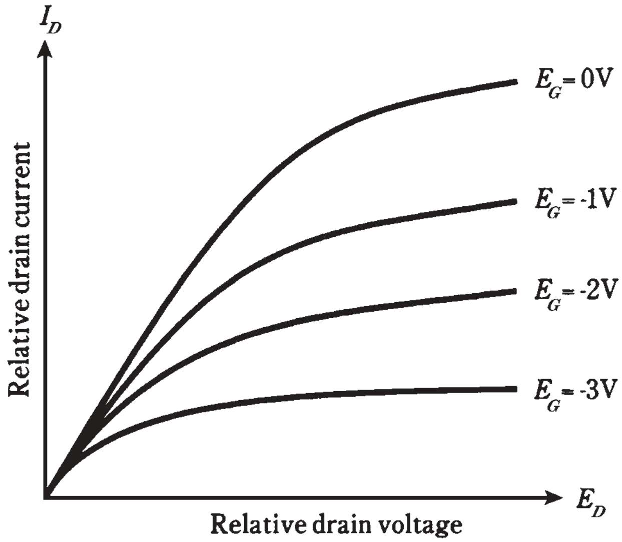
The best amplification for weak signals is obtained when the gate bias, E_G, is such that the slope of the curve in Fig. 23-5 is the greatest. This is shown roughly by the range marked X in the figure. For power amplification, however, results are often best when the JFET is biased at, or even beyond, pinchoff, in the range marked Y.

The current I_D passes through the drain resistor, as shown in either diagram of Fig. 23-4. Small fluctuations in E_G cause large changes in I_D, and these variations in turn produce wide swings in the dc voltage across R3 (at A) or R4 (at B). The ac part of this voltage goes through capacitor C2, and appears at the output as a signal of much greater ac voltage than that of the input signal at the gate. That's voltage amplification.

Drain current versus drain voltage

You might expect that the current I_D, passing through the channel of a JFET, would increase linearly with increasing drain voltage E_D. But this is not, in general, what happens. Instead, the current I_D rises for awhile, and then starts to level off.

The drain current I_D (which is the same as the channel current) is often plotted as a function of drain voltage, E_D, for various values of gate voltage, E_G. The resulting set of curves is called a *family of characteristic curves* for the device. The graph of Fig. 23-6 shows a family of characteristic curves for a hypothetical N-channel JFET. Engineers make use of these graphs when deciding on the best JFET type for an electronic circuit. Also of importance is the curve of I_D vs E_G, one example of which is shown in Fig. 23-5.



23-6 A family of characteristic curves for a hypothetical N-channel JFET.

Transconductance

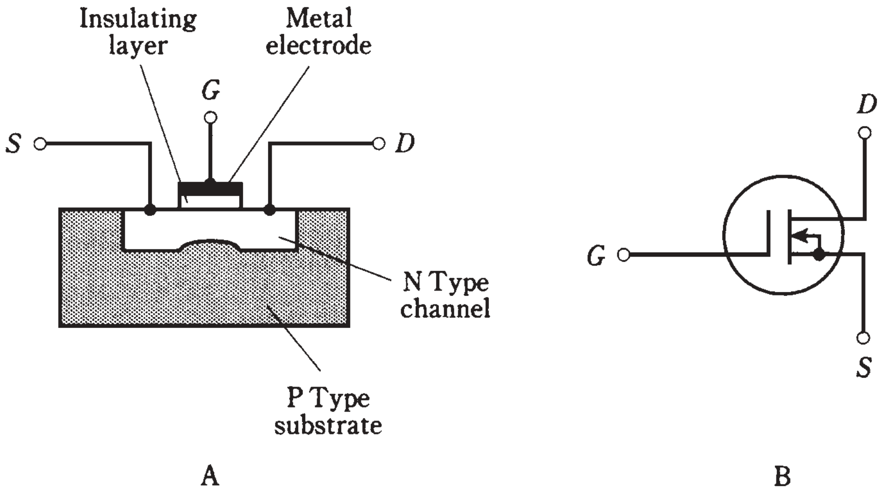
Recall the discussion of *dynamic current amplification* from the last chapter. This is a measure of how well a bipolar transistor amplifies a signal. The JFET analog of this is called *dynamic mutual conductance* or *transconductance*.

Refer again to Fig. 23-5. Suppose that E_G is a certain value, with a corresponding I_D resulting. If the gate voltage changes by a small amount dE_G then the drain current will also change by a certain increment dI_D. The transconductance is the ratio dI_D/dE_G. Geometrically, this translates to the slope of a line tangent to the curve of Fig. 23-5.

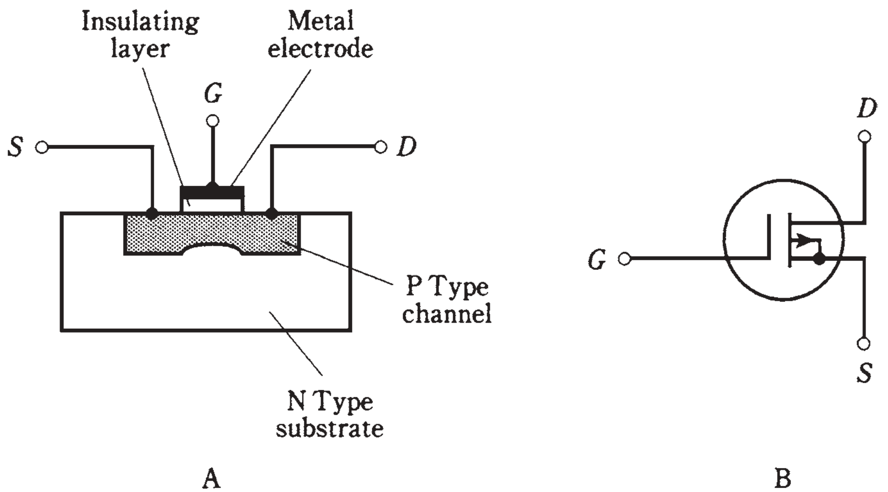
The value of dI_D/dE_G is obviously not the same everywhere along the curve. When the JFET is biased beyond pinchoff, in the region marked Y in the figure, the slope of the curve is zero. There is no drain current, even if the gate voltage changes. Only when the channel is conducting will there be a change in I_D when there is a change in E_G. The region where the transconductance, dI_D/dE_G, is the greatest is the region marked *X*, where the slope of the curve is steepest. This is where the most gain can be obtained from the JFET.

The MOSFET

The acronym *MOSFET* (pronounced "*moss*-fet") stands for *m*etal-*o*xide-*s*emiconductor *f*ield-*e*ffect *t*ransistor. A simplified cross-sectional drawing of an N-channel MOSFET, along with the schematic symbol, is shown in Fig. 23-7. The P-channel device is shown in the drawings of Fig. 23-8. The N-channel device is diffused into a substrate of P-type semiconductor material. The P-channel device is diffused into a substrate of N-type material.



23-7 At A, simplified cross-sectional drawing of an N-channel MOSFET. At B, the schematic symbol.



23-8 At A, simplified cross-sectional drawing of a P-channel MOSFET. At B, the schematic symbol.

Super-high input impedance

When the MOSFET was first developed, it was called an *insulated-gate FET* or *IGFET*. This is perhaps more descriptive of the device than the currently accepted name. The gate electrode is actually insulated, by a thin layer of dielectric, from the channel. As a result, the input impedance is even higher than that of a JFET; the gate-to-source resistance of a typical MOSFET is comparable to that of a capacitor! This means that a MOSFET draws essentially no current, and therefore no power, from the signal source. Some MOSFETs have input resistance exceeding a trillion (10^{12}) ohms.

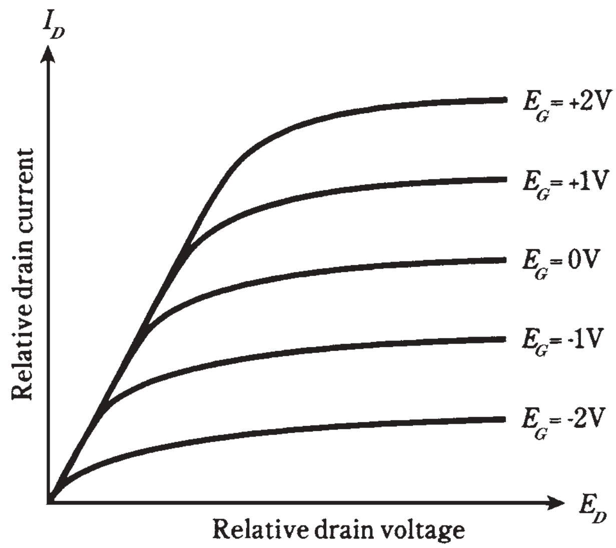
The main problem

The trouble with MOSFETs is that they can be easily damaged by static electric discharges. When building or servicing circuits containing MOS devices, technicians must use special equipment to ensure that their hands don't carry static charges that might ruin the components. If a static discharge occurs through the dielectric of a MOS device, the component will be destroyed permanently. Warm and humid climates do not offer protection against the hazard. (This author's touch has dispatched several MOSFETs in Miami during the summer.)

Flexibility

In actual circuits, an N-channel JFET can sometimes be replaced directly with an N-channel MOSFET; P-channel devices can be similarly interchanged. But the characteristic curves for MOSFETs are not the same as those for JFETs. The main difference is that the SG junction in a MOSFET is not a P-N junction. Therefore, forward breakover cannot occur. An E_G of more than $+\ 0.6$ V can be applied to an N-channel MOSFET, or an E_G more negative than -0.6 V to a P-channel device, without a current "leak" taking place.

A family of characteristic curves for a hypothetical N-channel MOSFET is shown in the graph of Fig. 23-9. The device will work with positive gate bias as well as with negative gate bias. A P-channel MOSFET behaves in a similar way, being usable with either positive or negative E_G.



23-9 A family of characteristic curves for a hypothetical N-channel MOSFET.

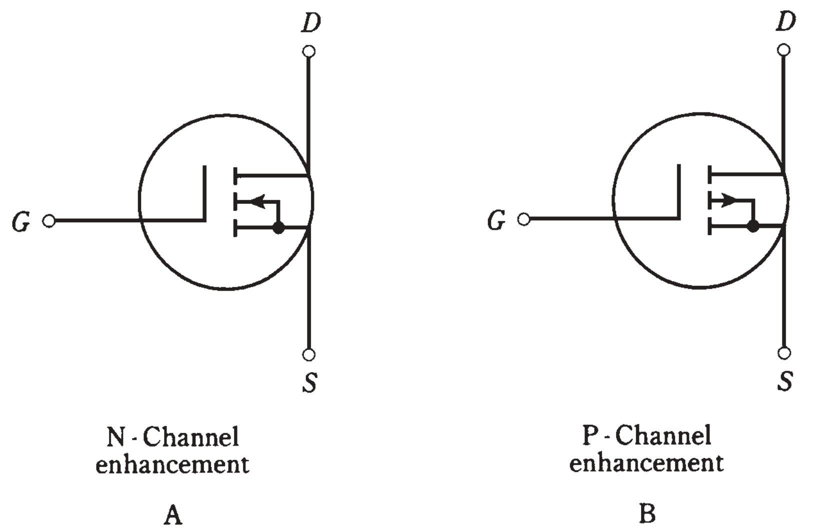
Depletion mode versus enhancement mode

The JFET works by varying the width of the channel. Normally the channel is wide open; as the depletion region gets wider and wider, choking off the channel, the charge carriers are forced to pass through a narrower and narrower path. This is known as the *depletion mode* of operation for a field-effect transistor.

A MOSFET can also be made to work in the depletion mode. The drawings and schematic symbols of Figs. 23-7 and 23-8 show depletion-mode MOSFETs.

However, MOS technology also allows an entirely different means of operation. An *enhancement-mode* MOSFET normally has a pinched-off channel. It is necessary to apply a bias voltage, E_G, to the gate so that a channel will form. If $E_G = 0$ in such a MOSFET, that is, if the device is at zero bias, the drain current I_D is zero when there is no signal input.

The schematic symbols for N-channel and P-channel enhancement-mode devices are shown in Fig. 23-10. The vertical line is broken. This is how you can recognize an enhancement-mode device in circuit diagrams.

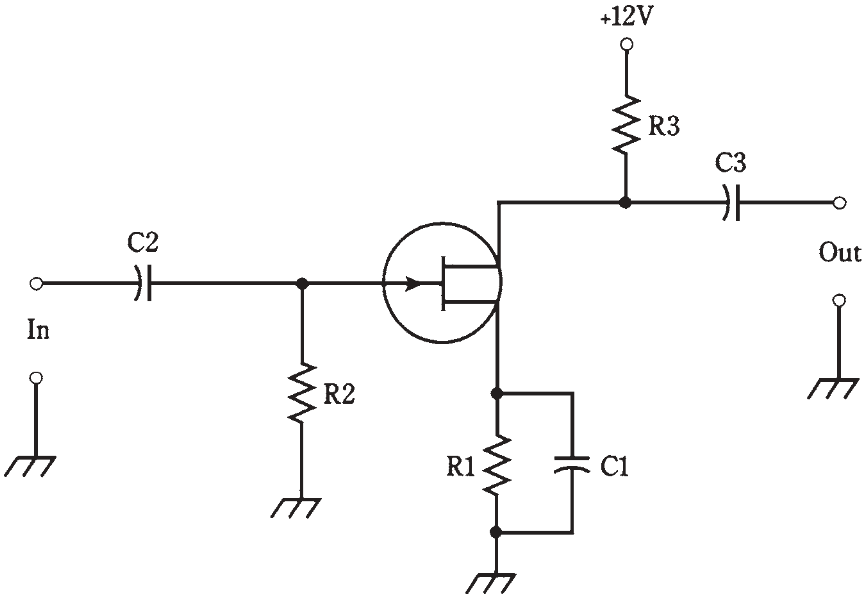


23-10 Schematic symbols for enhancement-mode MOSFETs. At A, N-channel; at B, P-channel.

Common-source circuit

There are three different circuit hookups for FETs, just as there are for bipolar transistors. These three arrangements have the source, the gate, or the drain at signal ground.

The *common-source circuit* places the source at signal ground. The input is at the base. The general configuration is shown in Fig. 23-11. An N-channel JFET is used here, but the device could be an N-channel, depletion-mode MOSFET and the circuit diagram would be the same. For an N-channel, enhancement-mode device, an extra resistor would be necessary, running from the gate to the positive power supply terminal. For P-channel devices, the supply would provide a negative, rather than a positive, voltage.



23-11 Common-source circuit configuration.

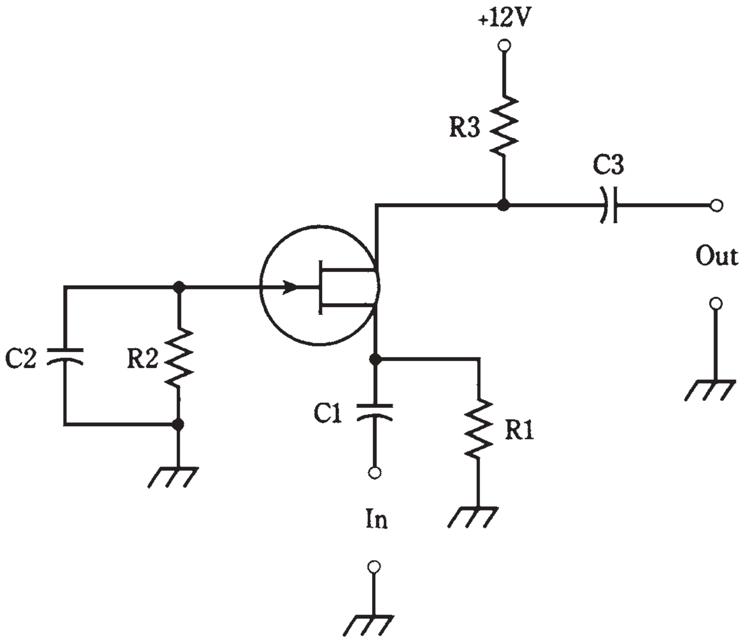
This circuit is an almost exact replica of the grounded-emitter bipolar arrangement. The only difference is the lack of a voltage-dividing network for bias on the control electrode.

Capacitor C1 and resistor R1 place the source at signal ground while elevating this electrode above ground for dc. The ac signal enters through C2; resistor R2 adjusts the input impedance and provides bias for the gate. The ac signal passes out of the circuit through C3. Resistor R3 keeps the output signal from being shorted out through the power supply.

The circuit of Fig. 23-11 is the basis for amplifiers and oscillators, especially at radio frequencies. The common-source arrangement provides the greatest gain of the three FET circuit configurations. The output is 180 degrees out of phase with the input.

Common-gate circuit

The *common-gate circuit* (Fig. 23-12) has the gate at signal ground. The input is applied to the source. The illustration shows an N-channel JFET. For other types of FETs, the same considerations apply as described above for the common-source circuit. Enhancement-mode devices would require a resistor between the gate and the positive supply terminal (or the negative terminal if the MOSFET is P-channel).



23-12 Common-gate circuit configuration.

The dc bias for the common-gate circuit is basically the same as that for the common-source arrangement. But the signal follows a different path. The ac input signal enters through C1. Resistor R1 keeps the input from being shorted to ground. Gate bias is provided by R1 and R2; capacitor C2 places the gate at signal ground. In some common-gate circuits, the gate electrode is directly grounded, and components R2 and C2 are not used. The output leaves the circuit through C3. Resistor R3 keeps the output signal from being shorted through the power supply.

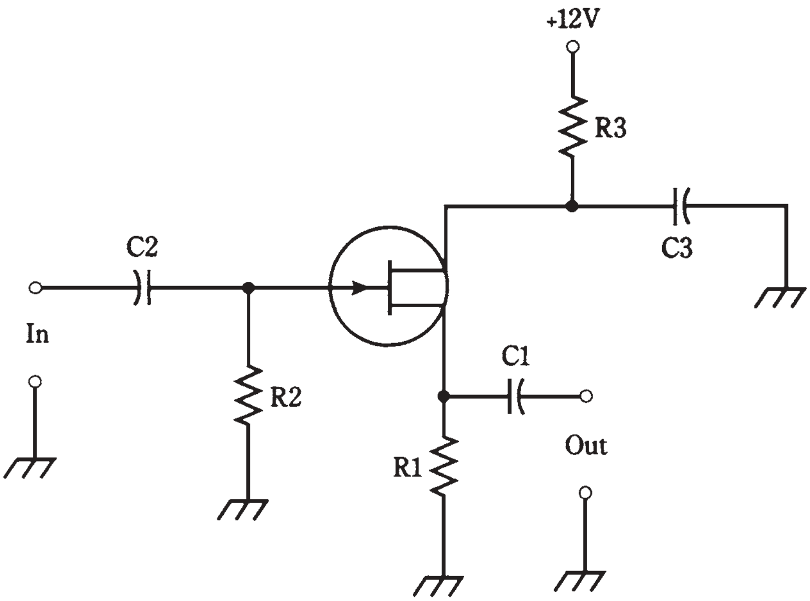
The common-gate arrangement produces less gain than its common-source counterpart. But this is not all bad; a common-gate amplifier is very stable, and is not likely to break into unwanted oscillation. The output is in phase with the input.

Common-drain circuit

A *common-drain circuit* is shown in Fig. 23-13. This circuit has the collector at signal ground. It is sometimes called a *source follower*.

The FET is biased in the same way as for the common-source and common-gate circuits. In the illustration, an N-channel JFET is shown, but any other kind of FET could be used, reversing the polarity for P-channel devices. Enhancement-mode MOSFETs would need a resistor between the gate and the positive supply terminal (or the negative terminal if the MOSFET is P-channel).

The input signal passes through C2 to the gate. Resistors R1 and R2 provide gate bias. Resistor R3 limits the current. Capacitor C3 keeps the drain at signal ground. Fluctuating dc (the channel current) flows through R1 as a result of the input signal; this



23-13 Common-drain circuit configuration.

causes a fluctuating dc voltage to appear across the resistor. The output is taken from the source, and its ac component passes through $C1$.

The output of the common-drain circuit is in phase with the input. This scheme is the FET analog of the bipolar common-collector arrangement. The output impedance is rather low, making this circuit a good choice for broadband impedance matching.

Table 23-1. Transistor circuit abbreviations.

Quantity	Abbreviations
Base-emitter voltage	E_{B}, V_{B}, E_{BE}, V_{BE}
Collector-emitter voltage	E_{C}, V_{C}, E_{CE}, V_{CE}
Collector-base voltage	E_{BC}, V_{BC}, E_{CB}, V_{CB}
Gate-source voltage	E_{G}, V_{G}, E_{GS}, V_{GS}
Drain-source voltage	E_{D}, V_{D}, E_{DS}, V_{DS}
Drain-gate voltage	E_{DG}, V_{DG}, E_{GD}, V_{GD}
Emitter current	I_{E}
Base current	I_{B}, I_{BE}, I_{EB}
Collector current	I_{C}, I_{CE}, I_{EC}
Source current	I_{S}
Gate current	I_{G}, I_{GS}, I_{SG}*
Drain current	I_{D}, I_{DS}, I_{SD}

*This is almost always insignificant.

A note about notation

In electronics, you'll encounter various different symbols that denote the same things. You might have already noticed that voltage is sometimes abbreviated by the letter E, and sometimes by the letter V. In bipolar and field-effect transistor circuits, you'll sometimes come across symbols like V_{CE} and V_{GS}; in this book they appear as E_C and E_G, respectively. Subscripts can be either uppercase or lowercase.

Remember that, although notations vary, the individual letters almost always stand for the same things. A variable might be denoted in different ways, depending on the author or engineer; but it's rare for one notation to acquire multiple meanings. The most common sets of abbreviations from this chapter and chapter 22 are shown in Table 23-1.

Wouldn't it be great if there were complete standardization in electronics? And it would be wonderful if everything were standardized in all other aspects of life, too, would it not?

Or would it?

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The current through the channel of a JFET is directly affected by all of the following *except*:
   - A. Drain voltage.
   - B. Transconductance.
   - C. Gate voltage.
   - D. Gate bias.
2. In an N-channel JFET, pinchoff occurs when the gate bias is:
   - A. Slightly positive.
   - B. Zero.
   - C. Slightly negative.
   - D. Very negative.
3. The current consists mainly of holes when a JFET:
   - A. Has a P-type channel.
   - B. Is forward-biased.
   - C. Is zero-biased.
   - D. Is reverse-biased.
4. A JFET might work better than a bipolar transistor in:
   - A. A rectifier.
   - B. A radio receiver.
   - C. A filter.
   - D. A transformer.

5. In a *P*-channel JFET:
   A. The drain is forward-biased.
   B. The gate-source junction is forward biased.
   C. The drain is negative relative to the source.
   D. The gate must be at dc ground.
6. A JFET is sometimes biased at or beyond pinchoff in:
   A. A power amplifier.
   B. A rectifier.
   C. An oscillator.
   D. A weak-signal amplifier.
7. The gate of a JFET has:
   A. Forward bias.
   B. High impedance.
   C. Low reverse resistance.
   D. Low avalanche voltage.
8. A JFET circuit essentially never has:
   A. A pinched-off channel.
   B. Holes as the majority carriers.
   C. A forward-biased P-N junction.
   D. A high-input impedance.
9. When a JFET is pinched off:
   A. dI_D/dE_G is very large with no signal.
   B. dI_D/dE_G might vary considerably with no signal.
   C. dI_D/dE_G is negative with no signal.
   D. dI_D/dE_G is zero with no signal.
10. Transconductance is the ratio of:
    A. A change in drain voltage to a change in source voltage.
    B. A change in drain current to a change in gate voltage.
    C. A change in gate current to a change in source voltage.
    D. A change in drain current to a change in drain voltage.
11. Characteristic curves for JFETs generally show:
    A Drain voltage as a function of source current.
    B. Drain current as a function of gate current.
    C. Drain current as a function of drain voltage.
    D. Drain voltage as a function of gate current.
12. A disadvantage of a MOS component is that:

A. It is easily damaged by static electricity.
B. It needs a high input voltage.
C. It draws a large amount of current.
D. It produces a great deal of electrical noise.

13. The input impedance of a MOSFET:
 A. Is lower than that of a JFET.
 B. Is lower than that of a bipolar transistor.
 C. Is between that of a bipolar transistor and a JFET.
 D. Is extremely high.

14. An advantage of MOSFETs over JFETs is that:
 A. MOSFETs can handle a wider range of gate voltages.
 B. MOSFETs deliver greater output power.
 C. MOSFETs are more rugged.
 D. MOSFETs last longer.

15. The channel in a zero-biased JFET is normally:
 A. Pinched off.
 B. Somewhat open.
 C. All the way open.
 D. Of P-type semiconductor material.

16. When an enhancement-mode MOSFET is at zero bias:
 A. The drain current is high with no signal.
 B. The drain current fluctuates with no signal.
 C. The drain current is low with no signal.
 D. The drain current is zero with no signal.

17. An enhancement-mode MOSFET can be recognized in schematic diagrams by:
 A. An arrow pointing inward.
 B. A broken vertical line inside the circle.
 C. An arrow pointing outward.
 D. A solid vertical line inside the circle.

18. In a source follower, which of the electrodes of the FET receives the input signal?
 A. None of them.
 B. The source.
 C. The gate.
 D. The drain.

19. Which of the following circuits has its output 180 degrees out of phase with its input?
    - A. Common source.
    - B. Common gate.
    - C. Common drain.
    - D. All of them.
20. Which of the following circuits generally has the greatest gain?
    - A. Common source.
    - B. Common gate.
    - C. Common drain.
    - D. It depends only on bias, not on which electrode is grounded.

24
CHAPTER
Amplifiers

IN THE PRECEDING TWO CHAPTERS, YOU SAW SCHEMATIC DIAGRAMS WITH BIPOLAR and field-effect transistors. The main intent was to acquaint you with biasing schemes. Some of the diagrams were of basic amplifier circuits. This chapter examines amplifiers more closely, but the subject is vast. For a thorough treatment, you should consult a book devoted to amplifiers and amplification.

The decibel

The extent to which a circuit amplifies is called the *amplification factor*. This can be given as a simple number, such as 100, meaning that the output signal is 100 times as strong as the input. More often, amplification factor is specified in units called *decibels,* abbreviated *dB.*

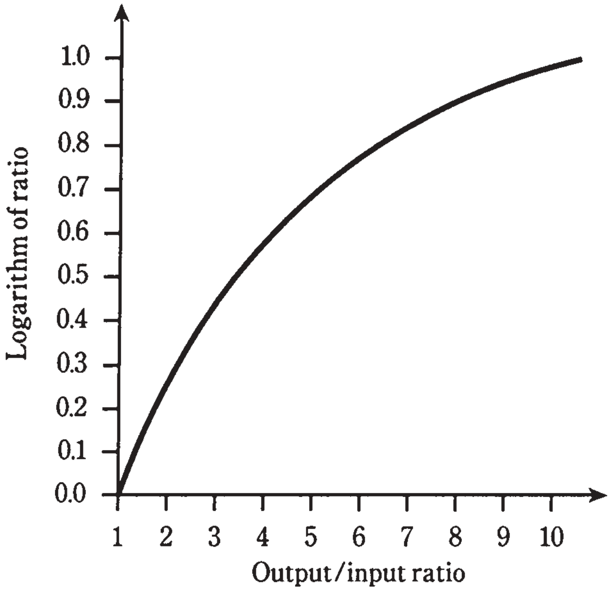
It's important to keep in mind what is being amplified: current, voltage, or power. Circuits are designed to amplify one of these aspects of a signal, but not necessarily the others. In a given circuit, the amplification factor is not the same for all three parameters.

Perception is logarithmic

You don't perceive loudness directly. Instead, you sense it in a nonlinear way. Physicists and engineers have devised the decibel system, in which amplitude changes are expressed according to the *logarithm* of the actual value (Fig. 24-1), to define relative signal strength.

Gain is assigned positive decibel values; loss is assigned negative values. Therefore, if signal A is at $+6$ dB relative to signal B, then A is stronger than B; if signal A is at -14 dB relative to B, then A is weaker than B.

An amplitude change of plus or minus 1 dB is about equal to the smallest change a listener can detect if the change is expected. If the change is not expected, then the smallest difference a listener can notice is about plus or minus 3 dB.



24-1 The base-10 logarithm function for output/input ratios of 1 to 10.

For voltage

Suppose there is a circuit with an rms ac input voltage of E_{in} and an rms ac output voltage of E_{out}. Then the *voltage gain* of the circuit, in decibels, is given by the formula:

$$\text{Gain (dB)} = 20 \log_{10}(E_{out}/E_{in})$$

The logarithm function is abbreviated *log*. The subscript 10 means that the *base* of the logarithm is 10. (Logarithms can have bases other than 10. This gets a little sophisticated, and it won't be discussed here.) You don't have to know the mathematical theory of logarithms to calculate them. All you need, and should buy right this minute if you don't have one, is a calculator that includes logarithm functions.

From now on, the base-10 logarithm will be called just the "logarithm," and the subscript 10 will be omitted.

Problem 24-1

A circuit has an rms ac input of 1.00 V and an rms ac output of 14.0 V. What is the gain in decibels?

First, find the ratio E_{out}/E_{in}. Because E_{out} = 14. 0 V and E_{in} = 1.00 V, this ratio is 14.0/1.00, or 14.0. Then, find the logarithm of 14.0. Your calculator will tell you that log 14.0 = 1.146128036 (it adds a lot of unnecessary digits). Finally, press the various buttons to multiply this number by 20, getting (with my calculator, anyway) 22.92256071. Round off to three significant figures, because that's all you're entitled to: Gain = 22.9 dB.

Problem 24-2

A circuit has an rms ac input voltage of 24.2 V and an rms ac output voltage of 19.9 V. What is the gain in decibels?

Find the ratio E_{out}/E_{in} = 19.9/24.2 = 0.822.... (The three dots indicate extra digits introduced by the calculator. You can leave them in untill the final roundoff.) Find the logarithm of this: log 0.822... = −0.0849.... Then multiply by 20: Gain = 1.699... dB, rounded off to − 1.70 dB.

Negative gain translates into *loss*. A gain of −1.70 dB is equivalent to a loss of 1.70 dB. The circuit of Problem 24-2 is not an amplifier—or if it is supposed to be, it isn't working!

If a circuit has the same output voltage as input voltage, that is, if $E_{out} = E_{in}$, then the gain is 0 dB. The ratio E_{out}/E_{in} is always equal to 1 in this case, and log 1 = 0.

It's important to remember, when doing gain calculations, always to use the same units for the input and the output signal levels. Don't use millivolts for E_{out} and microvolts for E_{in}, for example. This applies to current and power also.

For current

The *current gain* of a circuit is calculated just the same way as for voltage. If I_{in} is the rms ac input current and I_{out} is the rms ac output current, then

$$\text{Gain (dB)} = 20 \log (I_{out}/I_{in})$$

Often, a circuit that produces voltage gain will produce current loss, and vice versa. An excellent example is a simple ac transformer.

Some circuits have gain for both the voltage and the current, although not the same decibel figures. The reason is that the output impedance is different from the input impedance, altering the ratio of voltage to current.

For power

The *power gain* of a circuit, in decibels, is calculated according to the formula

$$\text{Gain (dB)} = 10 \log (P_{out}/P_{in})$$

where P_{out} is the output signal power and P_{in} is the input signal power.

Problem 24-3

A power amplifier has an input of 5.03 W and an output of 125 W. What is the gain in decibels?

First find the ratio P_{out}/P_{in} = 125/5.03 = 24.85.... Then find the logarithm: log 24.85... = 1.395.... Finally, multiply by 10 and round off: Gain = 10 × 1.395... = 14.0 dB.

Problem 24-4

An attenuator provides 10 dB power reduction. The input power is 94 W. What is the output power?

This problem requires you to "plug values into" the formula. An *attenuator* produces a power *loss*. When you hear that the attenuation is 10 dB, it is the same thing as a gain of −10 dB. You know P_{in} = 94 W, the unknown is P_{out}. Therefore,

$$-10 = 10 \log (P_{out}/94)$$

Solving this formula proceeds in several steps. First, divide each side by 10, getting

$$-1 = \log (P_{out}/94)$$

Then, take the *base-10 antilogarithm,* also known as the *antilog,* of each side. The antilog function is the *inverse* of the log function; that is, it "undoes" the log function. The function antilog (x) is sometimes written as 10^x. Thus

$$\text{antilog } (-1) = 10^{-1} = 0.1 = P_{out}/94$$

Now, multiply each side of the equation by 94, getting

$$94 \times 0.1 = 9.4 = P_{out}$$

Therefore, the output power is 9.4 W.

Don't confuse the voltage/current and power formulas. In general, for a given output/input ratio, the dB gain for voltage or current is twice the dB gain for power. Table 24-1 gives dB gain figures for various ratios of voltage, current, and power.

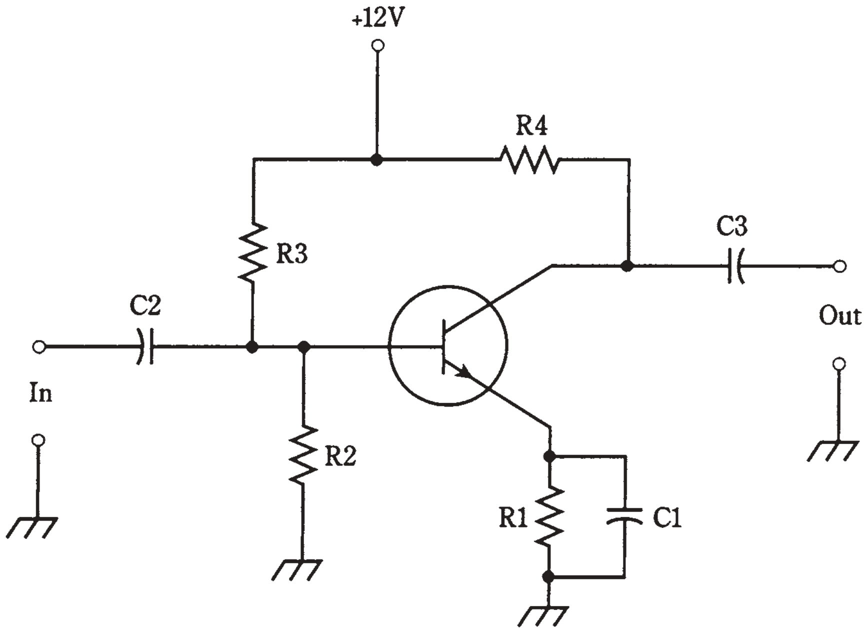
Table 24-1. Decibel gain figures for various ratios of voltage, current, and power.

Ratio	**Voltage or current gain**	**Power gain**
0.000 000 001 (10^{-9})	−180 dB	−90 dB
0.000 000 01 (10^{-8})	−160 dB	−80 dB
0.000 000 1 (10^{-7})	−140 dB	−70 dB
0.000 001 (10^{-6})	−120 dB	−60 dB
0.000 01 (10^{-5})	−100 dB	−50 dB
0.000 1 (10^{-4})	−80 dB	−40 dB
0.001	−60 dB	−30 dB
0.01	−40 dB	−20 dB
0.1	−20 dB	−10 dB
0.25	−12 dB	−6 dB
0.5	−6 dB	−3 dB
1	0 dB	0 dB
2	6 dB	3 dB
4	12 dB	6 dB
10	20 dB	10 dB
100	40 dB	20 dB
1000	60 dB	30 dB
10,000 (10^4)	80 dB	40 dB
100,000 (10^5)	100 dB	50 dB
1,000,000 (10^6)	120 dB	60 dB
10,000,000 (10^7)	140 dB	70 dB
100,000,000 (10^8)	160 dB	80 dB
1,000,000,000 (10^9)	180 dB	90 dB
10,000,000,000 (10^{10})	200 dB	100 dB

Basic bipolar amplifier circuit

In the previous chapters, you saw some circuits that will work as amplifiers. The principle is the same for all electronic amplification circuits. A signal is applied at some control point, causing a much greater signal to appear at the output.

In Fig. 24-2, an NPN bipolar transistor is connected as a common-emitter amplifier. The input signal passes through C2 to the base. Resistors R2 and R3 provide bias. Resistor R1 and capacitor C1 allow for the emitter to have a dc voltage relative to ground, while being grounded for signals. Resistor R1 also limits the current through the transistor. The ac output signal goes through capacitor C3. Resistor R4 keeps the ac output signal from being short-circuited through the power supply.



24-2 An amplifier using a bipolar transistor. Component designators and values are discussed in the text.

In this amplifier, the capacitors must have values large enough to allow the ac signal to pass with ease. But they shouldn't be much larger than the minimum necessary for this purpose. If an 0.1-µF capacitor will suffice, there's no point in using a 47-µF capacitor. That would introduce unwanted losses into the circuit, and would also make the circuit needlessly expensive to build.

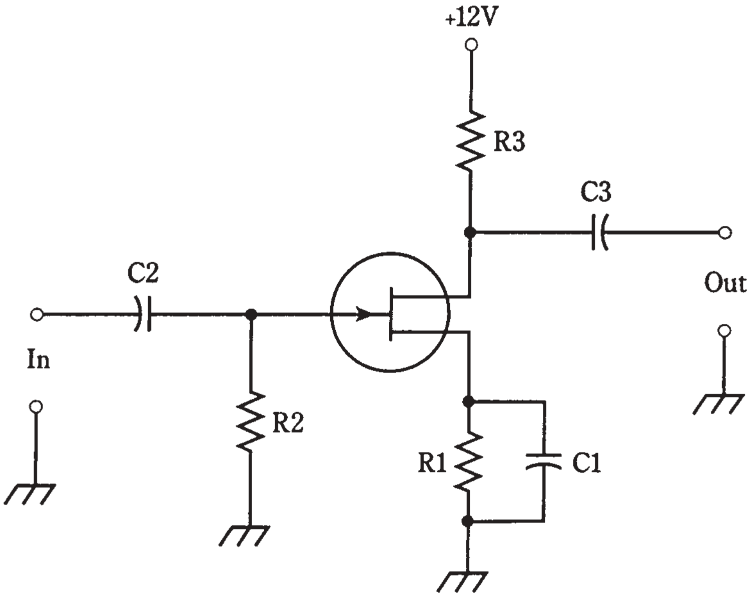
The ideal capacitance values depend on the design frequency of the amplifier, and also on the impedances at the input and output. In general, as the frequency and/or circuit impedance increase, less and less capacitance is needed. At audio frequencies, say 300 Hz to 20 kHz, and at low impedance, the capacitors might be as large as 100 µF. At radio frequencies, such as 1 MHz to 50 MHz, and with high impedances, values will be only a fraction of a microfarad, down to picofarads at the highest frequencies and impedances. The exact values are determined by the design engineers, working to optimize circuit performance in the lab.

The resistor values likewise depend on the application. Typical values are R1 = 470 Ω, R2 = 4.7 KΩ, R3 = 10KΩ, and R4 = 4.7 KΩ for a weak-signal, broadband amplifier.

If the circuit is used as a power amplifier, such as in a radio transmitter or a stereo hi-fi amplifier, the values of the resistors will be different. It might be necessary to bias the base negatively with respect to the emitter, using a second power supply with a voltage negative with respect to ground.

Basic FET amplifier circuit

In Fig. 24-3, an N-channel JFET is hooked up as a common-source amplifier. The input signal passes through C2 to the gate. Resistor R2 provides the bias. Resistor R1 and capacitor C1 give the source a dc voltage relative to ground, while grounding it for ac signals. The ac output signal goes through capacitor C3. Resistor R3 keeps the ac output signal from being short-circuited through the power supply.



24-3 An amplifier using an FET. Component designators and values are discussed in the text.

Concerning the values of the capacitors, the same considerations apply for this amplifier, as apply in the bipolar circuit. A JFET amplifier almost always has a high input impedance, and therefore the value of C2 will usually be small. If the device is a MOSFET, the input impedance is even higher, and C2 will be smaller yet, sometimes as little as 1 pF or less.

The resistor values depend on the application. In some instances, R1 and C1 are not used, and the source is grounded directly. If R1 is used, its value will depend on the input impedance and the bias needed for the FET. Nominal values might be R1 = 680 Ω, R2 = 10 KΩ, and R3 = 100 Ω for a weak-signal, wideband amplifier.

If the circuit is used as a power amplifier, the values of the resistors will be different. It might be necessary to bias the gate negatively with respect to the source, using a second power supply with a voltage negative relative to ground.

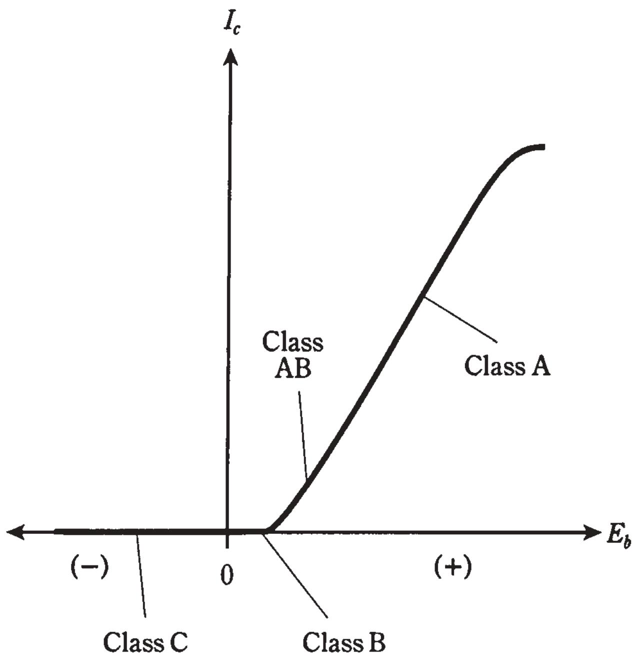
The class-A amplifier

With the previously mentioned component values, the amplifier circuits in Figs. 24-2 and 24-3 will operate in *class A*. Weak-signal amplifiers, such as the kind used in the first stage of a sensitive radio receiver, are always class-A. The term does not arise from inherent superiority of the design or technique (it's not like saying "grade-A eggs"). It's just a name chosen by engineers so that they know the operating conditions in the bipolar transistor or FET.

A class-A amplifier is always linear. That means that the output waveform has the same shape as (although a much greater amplitude than) the input waveform.

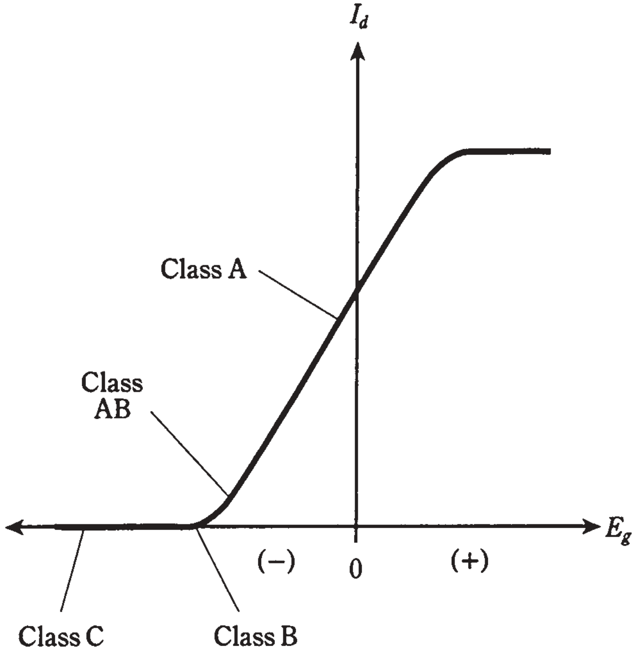
For class-A operation with a bipolar transistor, the bias must be such that, with no signal input, the device is near the middle of the straight-line portion of the I_C vs E_B (collector current versus base voltage) curve. This is shown for an NPN transistor in Fig. 24-4. For PNP, reverse the polarity signs.

With a JFET or MOSFET, the bias must be such that, with no signal input, the device is near the middle of the straight-line part of the I_D vs E_G (drain current versus gate



24-4 Various classes of amplifier operation for an NPN bipolar transistor.

voltage) curve. This is shown in Fig. 24-5 for an N-channel device. For P-channel, reverse the polarity signs.



24-5 Various classes of amplifier operation for an N-channel JFET.

It is important with class-A amplifiers that the input signal not be too strong. Otherwise, during part of the cycle, the base or gate voltage will be driven outside of the straight-line part of the curve. When this occurs, the output waveshape will not be a faithful reproduction of the input waveshape; the amplifier will be nonlinear. This will cause distortion of the signal. In an audio amplifier, the output might sound "raspy" or "scratchy." In a radio-frequency amplifier, the output signal will contain a large amount of energy at harmonic frequencies, The problem of harmonics, however, can be dealt with by means of resonant circuits in the output. These circuits attenuate harmonic energy, and allow amplifiers to be biased near, at, or even past cutoff or pinchoff.

The class-AB amplifier

When a class-A amplifier is working properly, it has low distortion. But class-A operation is inefficient. (Amplifier efficiency will be discussed later in this chapter.) This is mainly because the bipolar transistor or FET draws a large current, whether there is a signal input or not. Even with zero signal, the device is working hard.

For weak-signal work, efficiency is not very important; it's gain and sensitivity that matter. In power amplifiers, efficiency is a significant consideration, and gain and sensitivity are not so important. Any power not used toward generating a strong output signal will end up as heat in the bipolar transistor or FET. If an amplifier is designed to produce high power output, inefficiency translates to a lot of heat.

When a bipolar transistor is biased close to cutoff under no-signal conditions (Fig. 24-4), or when an FET is near pinchoff (Fig. 24-5), the input signal will drive the device into the nonlinear part of the operating curve. A small collector or drain current will flow when there is no input, but it will be less than the no-signal current that flows in a class-A amplifier. This is called *class-AB operation.*

With class-AB operation, the input signal might or might not cause the device to go into cutoff or pinchoff for a small part of the cycle. Whether or not this happens depends on the actual bias point, and also on the strength of the input signal. You can visualize this by imagining the *dynamic operating point* oscillating back and forth along the curve, in either direction from the *static* (no-signal) *operating point.*

If the bipolar transistor or FET is never driven into cutoff/pinchoff during any part of the signal cycle, the amplifier is working in *class-*AB_1. If the device goes into cutoff pinchoff for any part of the cycle (up to almost half), the amplifier is working in *class-*AB_2.

In a class-AB amplifier, the output waveshape is not identical with the input waveshape. But if the wave is *modulated,* such as in a voice radio transmitter, the waveform of the modulations will come out undistorted. Thus class-AB operation is useful in radio-frequency power amplifiers.

The class-B amplifier

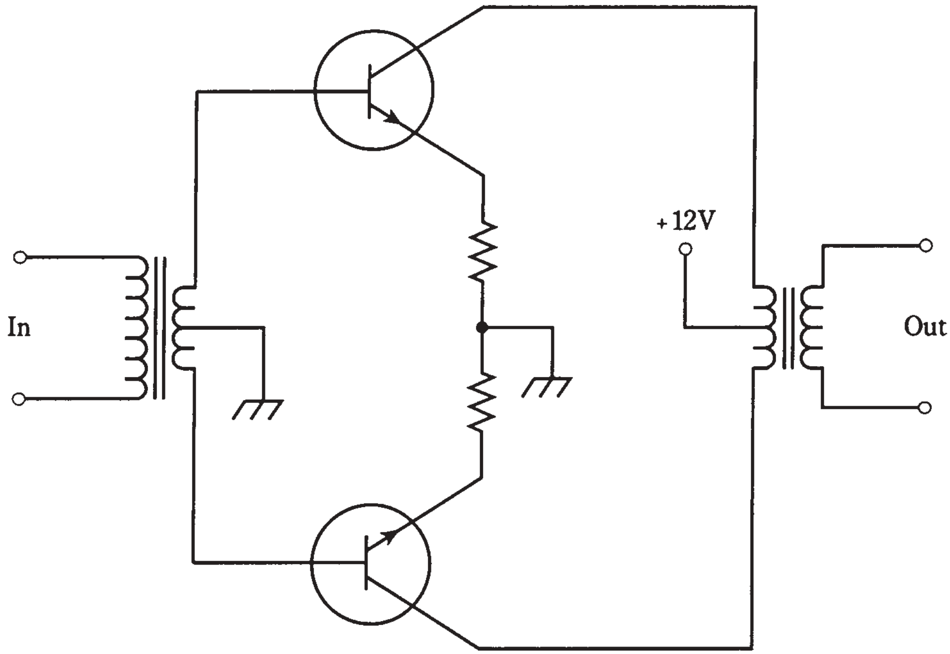
When a bipolar transistor is biased exactly at cutoff, or an FET at pinchoff, under zero-input-signal conditions, an amplifier is working in *class B*. These operating points are labeled on the curves in Figs. 24-4 and 24-5.

In class-B operation, there is no collector or drain current when there is no signal. This saves energy, because the circuit is not eating up any power unless there is a signal going into it. (Class-A and class-AB amplifiers draw current even when the input is zero.) When there is an input signal, current flows in the device during exactly half of the cycle. The output waveshape is greatly different from the input waveshape in a class-B amplifier; in fact, it is half-wave rectified.

Sometimes two bipolar transistors or FETs are used in a class-B circuit, one for the positive half of the cycle and the other for the negative half. In this way, distortion is eliminated. This is called a *class-B push-pull amplifier.* A class-B push-pull circuit using two NPN bipolar transistors is illustrated in Fig. 24-6. This configuration is popular for audio-frequency power amplification. It combines the efficiency of class B with the low distortion of class A. Its main disadvantage is that it needs two center-tapped transformers, one at the input and the other at the output. This translates into two things that engineers don't like: bulk and high cost. Nonetheless, the advantages often outweigh these problems.

The class-B scheme lends itself well to radio-frequency power amplification. Although the output waveshape is distorted, resulting in harmonic energy, this problem can be overcome by a resonant LC circuit in the output. If the signal is modulated, the modulation waveform will not be distorted.

You'll sometimes hear of class-AB or class-B "linear amplifiers," especially in ham radio. The term "linear" refers to the fact that the modulation waveform is not distorted by the amplifier. The *carrier wave* is*,* as you've seen, affected in a nonlinear fashion, because the amplifiers are not biased in the straight-line part of the operating curve.



24-6 A class-B push-pull amplifier.

Class-AB_2 and class-B amplifiers take some power from the input signal source. Engineers say that such amplifiers require a certain amount of *drive* or *driving power* to function. Class-A and class-AB_1 amplifiers theoretically need no driving power, although there must be an input voltage.

The class-C amplifier

A bipolar transistor or FET can be biased past cutoff or pinchoff, and it will still work as a power amplifier (PA), provided that the drive is sufficient to overcome the bias during part of the cycle. You might think, at first, that this bias scheme couldn't possibly result in amplification. Intuitively, it seems as if this could produce a marginal signal loss, at best. But in fact, if there is significant driving power, *class-C* operation can work very well. And, it is more efficient than any of the aforementioned methods. The operating points for class C are labeled in Figs. 24-4 and 24-5.

Class-C PAs are never linear, even for amplitude modulation on a signal. Because of this, a class-C circuit is useful only for signals that are either full-on or full-off. *Continuous-wave (CW),* also known as Morse code, and *radioteletype (RYTY)* are examples of such signals. Class-C PAs also work well with *frequency modulation (FM)* because the amplitude never changes.

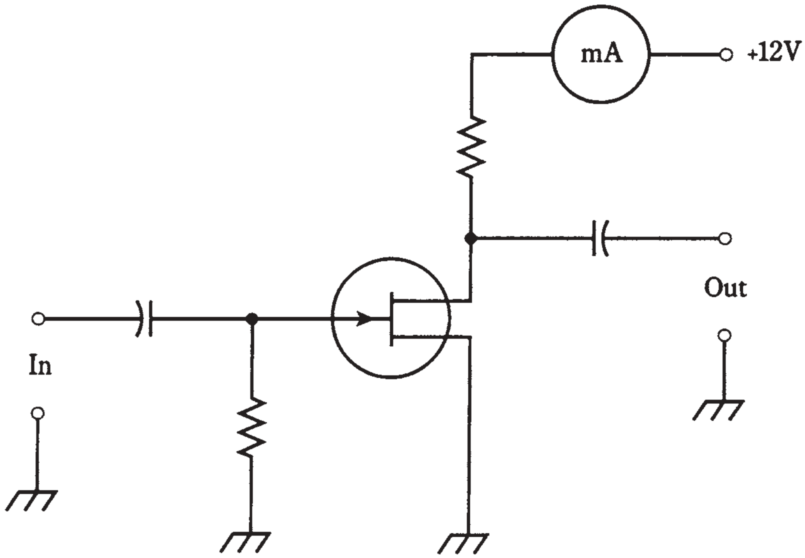
A class-C PA needs a lot of drive. The gain is fairly low. It might take 300 W of radio-frequency (RF) drive to get 1 kW of RF power output. That's a gain of only a little over 5 dB. Nonetheless, the efficiency is excellent, and class-C operation is common in CW, RTTY, or FM radio transmitters.

PA efficiency

Saving energy is a noble thing. But in electronic power amplifiers, as with many other kinds of hardware, energy conservation also translates into lower cost, smaller size and weight, and longer equipment life.

Power input

Suppose you connect an ammeter or milliammeter in series with the collector or drain of an amplifier and the power supply, as shown in Fig. 24-7. While the amplifier is in operation, this meter will have a certain reading. The reading might appear constant, or it might fluctuate with changes in the input signal level.



24-7 Connection of a current meter for dc power input measurement.

The dc *collector power input* to a bipolar-transistor amplifier circuit is the product of the collector current in amperes, and the collector voltage in volts. Similarly, for an FET, the *dc drain power input* is the product of the drain current and the drain voltage. These power figures can be further categorized as *average* or *peak* values. (This discussion involves only average power.)

The dc collector or drain power input can be high even when there is no signal applied to an amplifier. A class-A circuit operates this way. In fact, when a signal is applied to a class-A amplifier, the meter reading, and therefore the dc collector or drain power input, will not change compared to the value under no-signal conditions!

In class-AB_1 or class-AB_2, there is low current (and therefore low dc collector or drain power input) with zero signal, and a higher current (and therefore a higher dc power input) with signal.

In class-B and class-C, there is no current (and therefore zero dc collector or drain power input) when there is no input signal. The current, and therefore the dc power input, increases with increasing signal.

The dc collector or drain power input is usually measured in watts, the product of amperes and volts. It might be indicated in milliwatts for low-power amplifiers, or kilowatts for high-power amplifiers.

Power output

The *power output* of an amplifier must be measured by means of a specialized ac wattmeter. A dc ammeter/voltmeter combination won't work. The design of audio frequency and radio-frequency wattmeters is a sophisticated specialty in engineering.

When there is no signal input to an amplifier, there is no signal output, and therefore the power output is zero. This is true no matter what the class of amplification. The greater the signal input, in general, the greater the power output of a power amplifier, up to a certain point.

Power output, like dc collector or drain input, is measured in watts. For very low-power circuits, it might be in milliwatts; for high-power circuits it is often given in kilowatts.

Definition of efficiency

The *efficiency* of a power amplifier is the ratio of the ac power output to the dc collector or drain power input.

For a bipolar-transistor amplifier, let P_C be the dc collector power input, and P_{out} be the ac power output. For an FET amplifier, let P_D be the dc drain power input. Then the efficiency, eff, is given by

$$\text{eff} = P_{out}/P_C$$

for a bipolar-transistor circuit, and

$$\text{eff} = P_{out}/P_D$$

for an FET circuit. These are ratios, and they will always be between 0 and 1.

Efficiency is often expressed as a percentage, so that the formulas become

$$\text{eff}(\%) = 100\, P_{out}/P_C$$

and

$$\text{eff}(\%) = 100\, P_{out}/P_D$$

Problem 24-5

A bipolar-transistor amplifier has a dc collector input of 115 W and an ac power output of 65.0 W. What is the efficiency in percent?

Use the formula $\text{eff}(\%) = 100P_{out}/P_C = 100 \times 65/115 = 100 \times 0.565 = 56.5$ percent.

Problem 24-6

An FET amplifier is 60 percent efficient. If the power output is 3.5 W, what is the dc drain power input?

"Plug in" values to the formula eff(%) = 100 P_{out}/P_D. This gives

$$60 = 100 \times 3.5/P_D$$
$$60 = 350/P_D$$
$$60/350 = 1/P_D$$
$$P_D = 350/60 = 5.8 \text{ W}$$

Efficiency versus class

Class-A amplifiers are the least efficient, in general. The efficiency ranges from 25 to 40 percent, depending on the nature of the input signal and the type of bipolar or field-effect transistor used.

If the input signal is very weak, such as might be the case in a shortwave radio receiver, the efficiency of a class-A circuit is near zero. But in that application, the circuit is not working as a power amplifier, and efficiency is not of primary importance.

Class-AB amplifiers have better efficiency. A good class-AB_1 amplifier might be 35 to 45 percent efficient; a class-AB_2 amplifier will be a little better, approaching 60 percent with the best designs.

Class-B amplifiers are typically 50 to 60 percent efficient, although some radio frequency PA circuits work up to 65 percent or so.

Class-C amplifiers are the best of all. This author has seen well-designed class-C circuits that are 75 percent efficient.

These are not absolute figures, and you shouldn't memorize them as such. It's sufficient to know ballpark ranges, and that efficiency improves as the operating point moves towards the left on the curves shown in Figs. 24-4 and 24-5.

Drive and overdrive

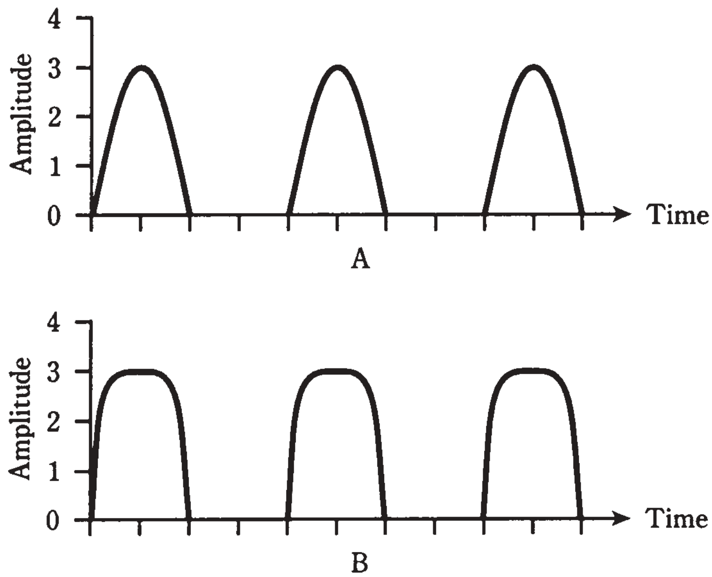
Class-A power amplifiers do not, in theory, take any power from the signal source in order to produce a significant amount of output power. This is one of the advantages of class-A operation. The same is true for class-AB_1 amplifiers. It is only necessary that a certain voltage be present at the control electrode (the base, gate, emitter, or source).

Class-AB_2 amplifiers need some driving power to produce ac power output. Class-B amplifiers require more drive than class-AB_2, and class-C amplifiers need still more drive.

Whatever kind of PA is used in a given situation, it is important that the driving signal not be too strong. If *overdrive* takes place, there will be distortion in the output signal.

An oscilloscope can be used to determine whether or not an amplifier is being overdriven. The scope is connected to the amplifier output terminals, and the waveshape of the output signal is examined. The output waveform for a particular class of amplifier always has a characteristic shape. Overdrive is indicated by a form of distortion known as *flat topping.*

In Fig. 24-8A, the output signal waveshape for a properly operating class-B amplifier is shown. It looks like the output of a half-wave rectifier, because the bipolar transistor or FET is drawing current for exactly half (180 degrees) of the cycle.



24-8 At A, waveshape at the output of a properly operating class-B amplifier. At B, distortion in the output waveshape caused by overdrive.

In Fig. 24-8B, the output of an overdriven class-B amplifier is shown. The wave is no longer a half sine wave, but instead, it shows evidence of flat topping. The peaks are blunted or truncated. The result of this is audio distortion in the modulation on a radio signal, and also an excessive amount of energy at harmonic frequencies.

The efficiency of a circuit can be degraded by overdrive. The "flat tops" of the distorted waves don't contribute anything to the strength of the signal at the desired frequency. But they do cause a higher-than-normal P_C or P_D value, which translates into a lower-than-normal efficiency P_{out}/P_C or P_{out}/P_D.

A thorough discussion of overdrive and distortion in various amplifier classes and applications would require an entire book. If you're interested in more detail, a good college or trade-school text on radio-frequency (RF) amplification is recommended.

Audio amplification

The circuits you've seen so far have been general, not application-specific. With capacitors of several microfarads, and when biased for class A, these circuits are representative of audio amplifiers. As with RF amplifiers, there isn't room enough to go into great depth about audio amplifiers in this book, but a couple of important characteristics deserve mention.

Frequency response

High-fidelity audio amplifiers, of the kind used in music systems, must have more or less constant gain from 20 Hz to 20 kHz. This is a frequency range of 1000: 1. Audio amplifiers for voice communications must work from 300 Hz to 3 kHz, a 10: 1 span of frequencies. In digital communications, audio amplifiers are designed to work over a narrow range of frequencies, sometimes less than 100 Hz wide.

Hi-fi amplifiers are usually equipped with resistor-capacitor (RC) networks that tailor the frequency response. These are *tone controls,* also called *bass* and *treble* controls. The simplest hi-fi amplifiers use a single knob to control the tone. More sophisticated “amps” have separate controls, one for bass and the other for treble. The most advanced hi-fi systems make use of *graphic equalizers,* having controls that affect the amplifier gain over several different frequency spans.

Gain-versus-frequency curves for three hypothetical audio amplifiers are shown in Fig. 24-9. At A, a wideband, flat curve is illustrated. This is typical of hi-fi system amplifiers. At B, a voice communications response is shown. At C, a narrowband response curve, typical of audio amplifiers in Morse code or low-speed digital-signal receivers, is illustrated.

Volume control

Audio amplifier systems usually consist of two or more *stages.* A stage is one bipolar transistor or FET (or a push-pull combination), plus peripheral resistors and capacitors. Stages are cascaded one after the other to get high gain.

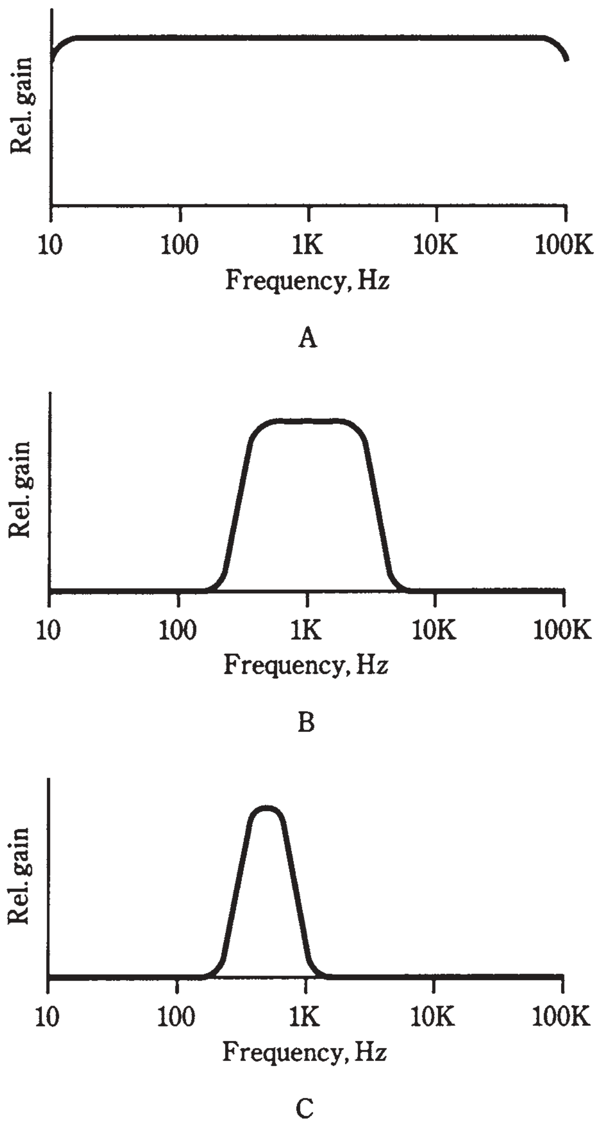
In one of the stages in an audio system, a *volume control* is used. This control is usually a potentiometer that allows the gain of a stage to be adjusted without affecting its linearity. An example of a simple volume control is shown in Fig. 24- 10. In this amplifier, the gain through the transistor itself is constant. The ac output signal passes through C1 and appears across R1, a potentiometer. The *wiper* (indicated by the arrow) of the potentiometer “picks off” more or less of the ac output signal, depending on the position of the control shaft. When the shaft is fully counterclockwise, the arrow is at the bottom of the zig-zaggy line, and none of the signal passes to the output. When the shaft is fully clockwise, the arrow is at the top of the zig-zaggy line, and all of the signal passes to the output. At intermediate positions of the control shaft, various proportions of the full output signal will appear at the output. Capacitor C2 isolates the potentiometer from the dc bias of the following stage.

Volume control is usually done in a stage where the audio power level is quite low. This allows the use of a small potentiometer, rated for perhaps 1 W. If volume control were done at high audio power levels, the potentiometer would need to be able to dissipate large amounts of power, and would be needlessly expensive.

Coupling methods

In all of the amplifiers you’ve seen so far, with the exception of the push-pull circuit (Fig. 24-6), capacitors have been used to allow ac to pass while blocking dc. But there is another way to do this, and in some amplifier systems, it is preferred. This is the use of a transformer to couple signals from one stage to the next.

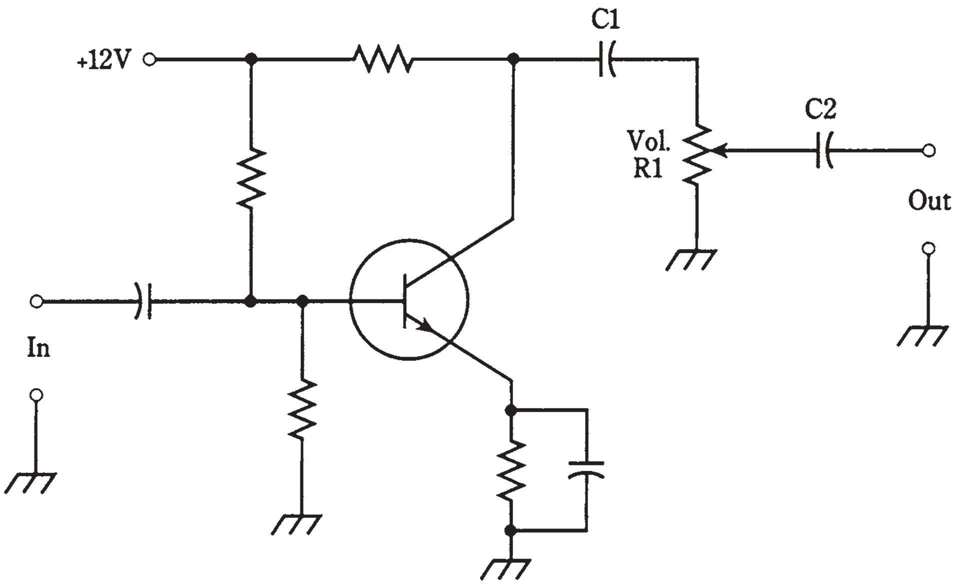
An example of *transformer coupling* is shown in Fig. 24-11. Capacitors C1 and C2 keep one end of the transformer primary and secondary at signal ground. Resistor R1



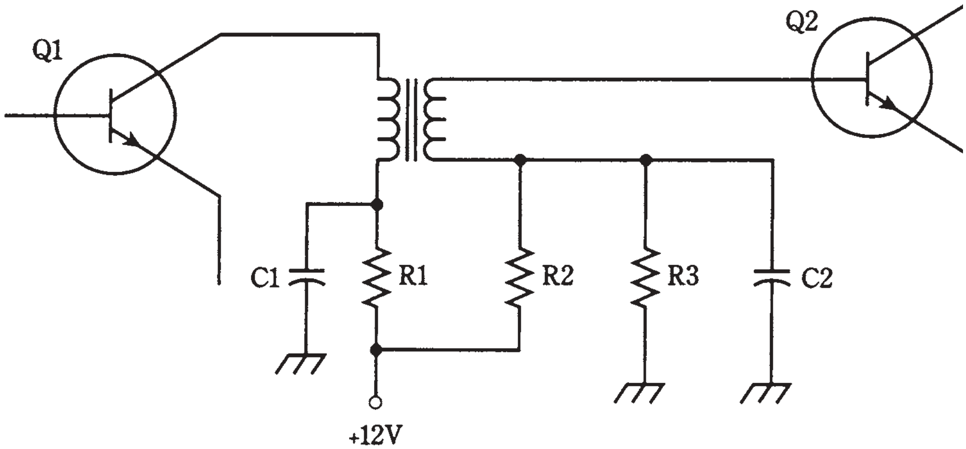
24-9 At A, frequency response for music; at B, for voice signals; at C, for narrowband digital signals.

limits the current through the first transistor, Q1. (In some cases, R1 might be eliminated.) Resistors R2 and R3 provide the proper base bias for transistor Q2.

The main disadvantage of this scheme is that it costs more than *capacitive coupling*. But transformer coupling can provide an optimum signal transfer between amplifier stages with a minimum of loss. This is because of the impedance-matching ability of transformers. Remember that the turns ratio of a transformer affects not only the input and output voltage, but the ratio of impedances. By selecting the right transformer, the output impedance of Q1 can be perfectly matched to the input impedance of Q2.



24-10 A simple volume control. Component designators and functions are discussed in the text.



24-11 Transformer coupling. Component designators and functions are discussed in the text.

In some amplifier systems, capacitors are added across the primary and/or secondary of the transformer. This results in resonance at a frequency determined by the capacitance and the transformer winding inductance. If the set of amplifiers is intended for just one frequency (and this is often the case in RF systems), this method of coupling, called *tuned-circuit coupling,* enhances the system efficiency. But care must be taken to be sure that the amplifier chain doesn't get so efficient that it *oscillates* at the resonant frequency of the tuned circuits! You'll learn about oscillation in the next chapter.

Radio-frequency amplification

The RF spectrum begins at about 9 kHz and extends upward in frequency to well over 300 GHz, or 300,000,000,000 Hz. A complete discussion of RF amplifier design would occupy a book. Therefore, again, only a sketch of the most important characteristics can be given here.

Weak-signal versus power amplifiers

Some RF amplifiers are designed for weak-signal work. The general circuits, shown earlier in this chapter, are representative of such amplifiers, when the capacitors have values of about 1 µF or less. The higher the frequency, the smaller the values of the capacitors.

The *front end,* or first amplifying stage, of a radio receiver requires the most sensitive possible amplifier. Sensitivity is determined by two factors: *gain* and *noise figure.*

The noise figure of an amplifier is a measure of how well it can amplify the desired signal, without injecting unwanted noise. All bipolar transistors or FETs; create some *white noise* because of the movement of charge carriers. In general, JFETs produce less noise than bipolar transistors. Gallium arsenide FETs, also called *GaAsFETs* (pronounced “gasfets”), are the least noisy of all.

The higher the frequency at which a weak-signal amplifier is designed, the more important the noise figure gets. This is because there is less atmospheric noise at the higher radio frequencies, as compared with the lower frequencies. At 1.8 MHz, for example, the airwaves contain much atmospheric noise, and it doesn’t make a significant difference if the receiver introduces a little noise itself. But at 1.8 GHz the atmospheric noise is almost nonexistent, and receiver performance depends much more critically on the amount of internally generated noise.

Weak-signal amplifiers almost always use resonant circuits. This optimizes the amplification at the desired frequency, while helping to cut out noise on unwanted frequencies. A typical tuned GaAsFET weak-signal RF amplifier is diagrammed in Fig. 24-12. It is designed for about 10 MHz.

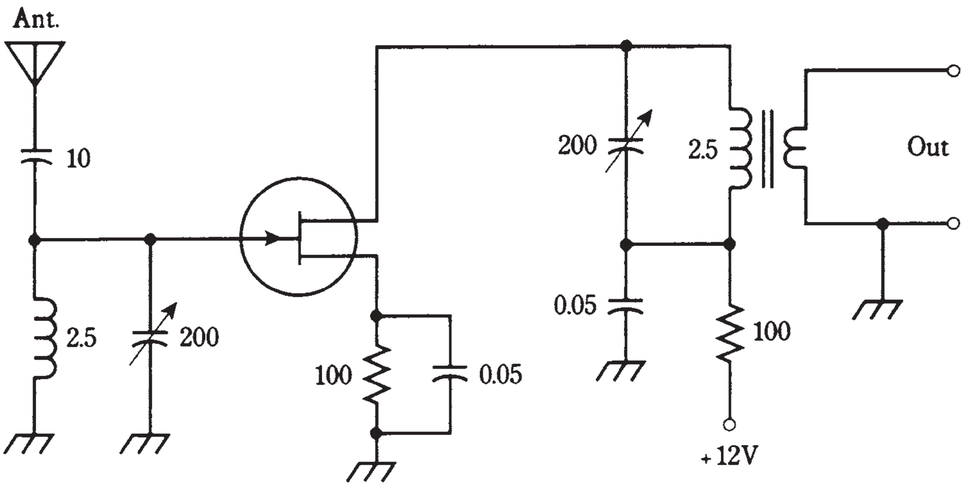
Broadband PAs

At RF, a PA might be either *broadband* or *tuned.*

The main advantage of a broadband PA is ease of operation, because it does not need tuning. A broadbanded amplifier is not “particular” with respect to the frequency within its design range, such as 1.5 MHz through 15 MHz. The operator need not worry about critical adjustments, nor bother to change them when changing the frequency.

One disadvantage of broadband PAs is that they are slightly less efficient than tuned PAs. This usually isn’t too hard to put up with, though, considering the convenience of not having to fiddle with the tuning.

The more serious problem with broadband PAs is that they’ll amplify *anything* in the design range, whether or not you want it to go over the air. If some earlier stage in a transmitter is oscillating at a frequency nowhere near the intended signal frequency, and



24-12 A tuned RF amplifier for use at about 10 MHz. Resistances are in ohms. Capacitances are in μF if less than 1, and in pF if more than 1. Inductances are in μH.

if this undesired "signal" falls within the design frequency range of the broadband PA, it will be amplified. The result will be unintended (and illegal!) RF emission from the radio transmitter. Such unwanted signals are called *spurious emissions,* and they occur more often than you might think.

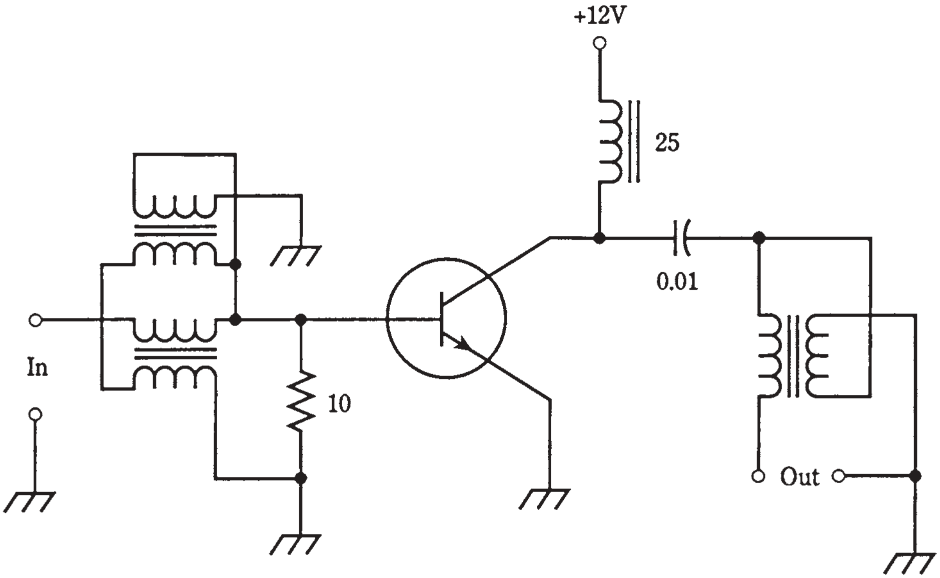
A typical broadband PA circuit is diagrammed schematically in Fig. 24-13. The NPN bipolar transistor is a power transistor. It will reliably provide about 3 W of continuous RF output from 1.5 MHz through 15 MHz. The transformers are a critical part of this circuit; they must be designed to work well over a 10:1 range of frequencies. This circuit is suitable for use on the ham radio bands at 160, 80, 75, 40, 30, and 20 meters.

Tuned PAs

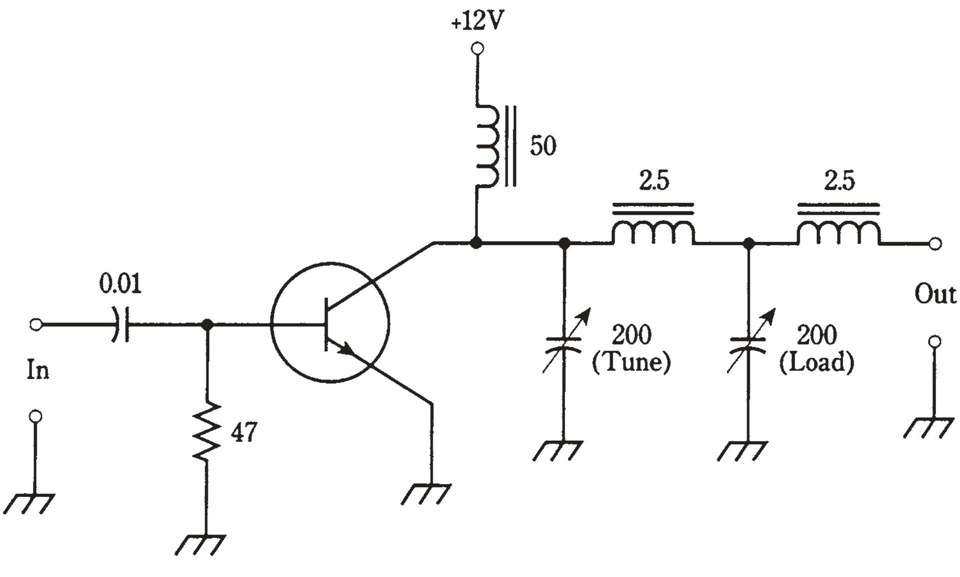
A tuned RF power amplifier offers improved efficiency compared with broadband designs. Also, the tuning helps to reduce the chances of spurious signals being amplified and transmitted over the air.

Another advantage of tuned PAs is that they can work into a wide range of load impedances. In addition to a *tuning control,* or resonant circuit that adjusts the output of the amplifier to the operating frequency, there is a *loading control* that optimizes the signal transfer between the amplifier and the load (usually an antenna).

The main drawback of a tuned PA is that the adjustment takes time, and improper adjustment can result in damage to the amplifying device (bipolar transistor or FET). If the tuning and/or loading controls are out of kilter, the efficiency of the amplifier will be extremely low—sometimes practically zero—while the dc collector or drain power input is unnaturally high. Solid-state devices overheat quickly under these conditions.



24-13 A broadband RF power amplifier, capable of producing a few watts output.



24-14 A tuned RF power amplifier, capable of producing a few watts output.

A tuned RF PA, providing 3 W output at 10 MHz or so, is shown in Fig. 24-14. The transistor is the same as for the broadband amplifier discussed above. The tuning and loading controls should be adjusted for maximum RF power output as indicated on a wattmeter in the feed line going to the load.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The decibel is a unit of:
   - A. Relative signal strength.
   - B. Voltage.
   - C. Power.
   - D. Current.
2. If a circuit has a voltage-amplification factor of 20, then the voltage gain is:
   - A. 13 dB.
   - B. 20 dB.
   - C. 26 dB.
   - D. 40 dB.
3. A gain of −15 dB in a circuit means that:
   - A. The output signal is stronger than the input.
   - B. The input signal is stronger than the output.
   - C. The input signal is 15 times as strong as the output.
   - D. The output signal is 15 times as strong as the input.
4. A device has a voltage gain of 23 dB. The input voltage is 3.3 V. The output voltage is:
   - A. 76 V.
   - B. 47 V.
   - C. 660 V.
   - D. Not determinable from the data given.
5. A power gain of 44 dB is equivalent to an output/input power ratio of:
   - A. 44.
   - B. 160.
   - C. 440.
   - D. 25,000.
6. A resistor between the base of an NPN bipolar transistor and the positive supply voltage is used to:
   - A. Provide proper bias.
   - B. Provide a path for the input signal.
   - C. Provide a path for the output signal.
   - D. Limit the collector current.

7. The capacitance values in an amplifier circuit depend on:
   A. The supply voltage.
   B. The polarity.
   C. The signal strength.
   D. The signal frequency.

8. A class-A circuit would not work well as:
   A. A stereo hi-fi amplifier.
   B. A television transmitter PA.
   C. A low-level microphone preamplifier.
   D. The first stage in a radio receiver.

9. In which of the following FET amplifier types does drain current flow for 50 percent of the signal cycle?
   A. Class A.
   B. Class AB_1.
   C. Class AB_2.
   D. Class B.

10. Which of the following amplifier types produces the least distortion of the signal waveform?
   A. Class A.
   B. Class AB_1.
   C. Class AB_2.
   D. Class B.

11. Which bipolar amplifier type has some distortion in the signal wave, with collector current during most, but not all, of the cycle?
   A. Class A.
   B. Class AB_1.
   C. Class AB_2.
   D. Class B.

12. How can a class-B amplifier be made suitable for hi-fi audio applications?
   A. By increasing the bias.
   B. By using two transistors in push-pull.
   C. By using tuned circuits in the output.
   D. A class-B amplifier cannot work well for hi-fi audio.

13. How can a class-C amplifier be made linear?
   A. By reducing the bias.

B. By increasing the drive.

C. By using two transistors in push-pull.

D. A class-C amplifier cannot be made linear.

14. Which of the following amplifier classes generally needs the most driving power?

A. Class A.

B. Class AB_1.

C. Class AB_2.

D. Class B.

15. A graphic equalizer is a form of:

A. Bias control.

B. Gain control.

C. Tone control.

D. Frequency control.

16. A disadvantage of transfer coupling, as opposed to capacitive coupling, is that:

A. Transformers can't match impedances.

B. Transformers can't work above audio frequencies.

C. Transformers cost more.

D. Transformers reduce the gain.

17. A certain bipolar-transistor PA is 66 percent efficient. The output power is 33 W. The dc collector power input is:

A. 22 W.

B. 50 W.

C. 2.2 W.

D. None of the above.

18. A broadband PA is:

A. Generally easy to use.

B. More efficient than a tuned PA.

C. Less likely than a tuned PA to amplify unwanted signals.

D. Usable only at audio frequencies.

19. A tuned PA must always be:

A. Set to work over a wide range of frequencies.

B. Adjusted for maximum power output.

C. Made as efficient as possible.

D. Operated in class C.

20. A loading control in a tuned PA:
    A. Provides an impedance match between the bipolar transistor or FET and the load.
    B. Allows broadband operation.
    C. Adjusts the resonant frequency.
    D. Controls the input impedance.

25
CHAPTER

Oscillators

SOMETIMES AMPLIFIERS WORK TOO WELL. YOU'VE PROBABLY HEARD THIS WHEN someone was getting ready to speak over a public-address system. The gain was set too high. The person began to speak; sound from the speakers got into the microphone, was amplified, went to the speakers again, and back to the microphone. A vicious cycle of *feedback* ensued. The result might have been a rumble, a howl, or a shriek. The system broke into *oscillation.* The amplifiers became temporarily useless until the gain was reduced.

Oscillation can be controlled, so that it takes place at a specific, stable, predictable frequency. An *oscillator* is a circuit that is deliberately designed to oscillate.

Uses of oscillators

Some oscillators work at audio frequencies, and others are intended to produce radio signals. Most generate sine waves, although some are built to emit square waves, sawtooth waves, or other waveshapes.

The subject of oscillators, once you understand amplifiers, is elementary. All oscillators are amplifiers with positive feedback. In this chapter, radio-frequency (RF) oscillators are discussed in some detail, and then audio oscillators are examined.

In radio communications, oscillators generate the "waves" or signals, that are ultimately sent over the air. For data to be sent, the signal from an oscillator must be *modulated.* Modulation is covered in chapter 26.

Oscillators are used in radio and TV receivers for frequency control and for *detection* and *mixing.* Detectors and mixers are discussed in chapter 27.

Audio-frequency oscillators find applications in such devices as music synthesizers, FAX modems, doorbells, beepers, sirens and alarms, and electronic toys.

Positive feedback

Feedback can be in phase or out of phase. For a circuit to oscillate, the feedback must be in phase, or *positive. Negative feedback* (out of phase) simply reduces the gain.

The output of a common-emitter or common-source amplifier is out of phase from the input. If you couple the collector to the base through a capacitor, you won't get oscillation.

The output of a common-base or common-gate amplifier is in phase with the input. But these circuits have limited gain. It's hard to make them oscillate, even with positive feedback.

Common-collector and common-drain circuits don't have enough gain to make oscillators.

But take heart: There are lots of ways to make circuits oscillate. Obtaining oscillation has never been a problem in electronics. Public-address systems do it willingly enough!

Concept of the oscillator

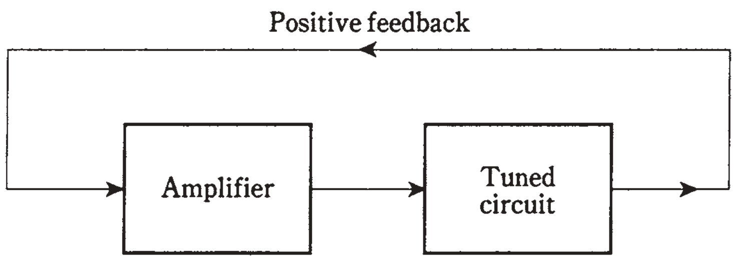
For a circuit to oscillate, the gain must be high, the feedback must be positive, and the *coupling* from output to input must be good. The feedback path must be easy for a signal to follow. The phase of a fed-back signal can be reversed without any trouble, so that common-emitter or common-source amplifiers can be made to oscillate.

Feedback at a single frequency

Recalling the public-address fiasco, some variation of which you've doubtless heard many times, could you know in advance whether the feedback would have a low pitch, a midrange pitch, or a high pitch? No. The oscillation was not intended, and might have started at any audio frequency.

The frequency of an oscillator is controlled by means of tuned, or resonant, circuits. These are usually inductance-capacitance (LC) or resistance-capacitance (RC) combinations. The LC scheme is common at RF; the RC method is more often used for audio oscillators.

The tuned circuit makes the feedback path easy for a signal to follow at one frequency, but hard to follow at all other frequencies (Fig. 25-1). The result is that the oscillation takes place at a predictable and stable frequency, determined by the inductance and capacitance or by the resistance and capacitance.

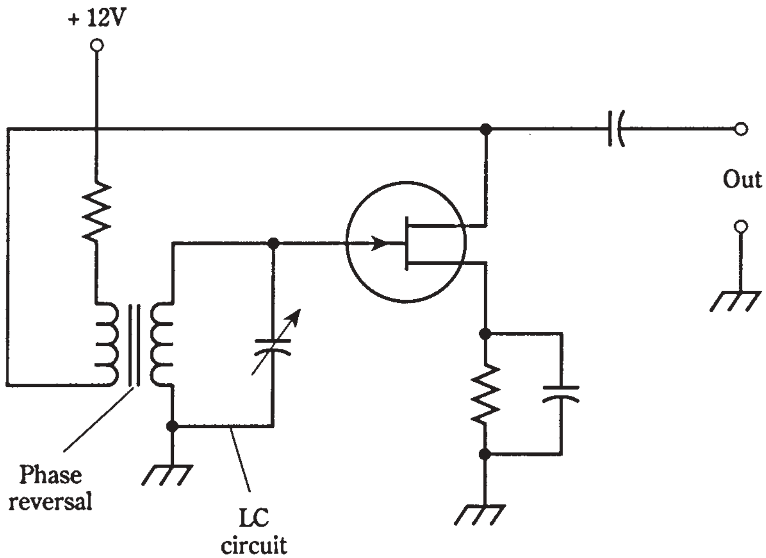


25-1 Basic concept of the oscillator.

The Armstrong oscillator

A common-emitter or common-source amplifier can be made to oscillate by coupling the output back to the input through a transformer that reverses the phase of the fed-back signal. The phase at a transformer output can be inverted by reversing the secondary terminals.

The schematic diagram of Fig. 25-2 shows a common-source amplifier whose drain circuit is coupled to the gate circuit via a transformer. In practice, getting oscillation is easy. If the circuit won't oscillate with the transformer secondary hooked up one way, you can just switch the wires.



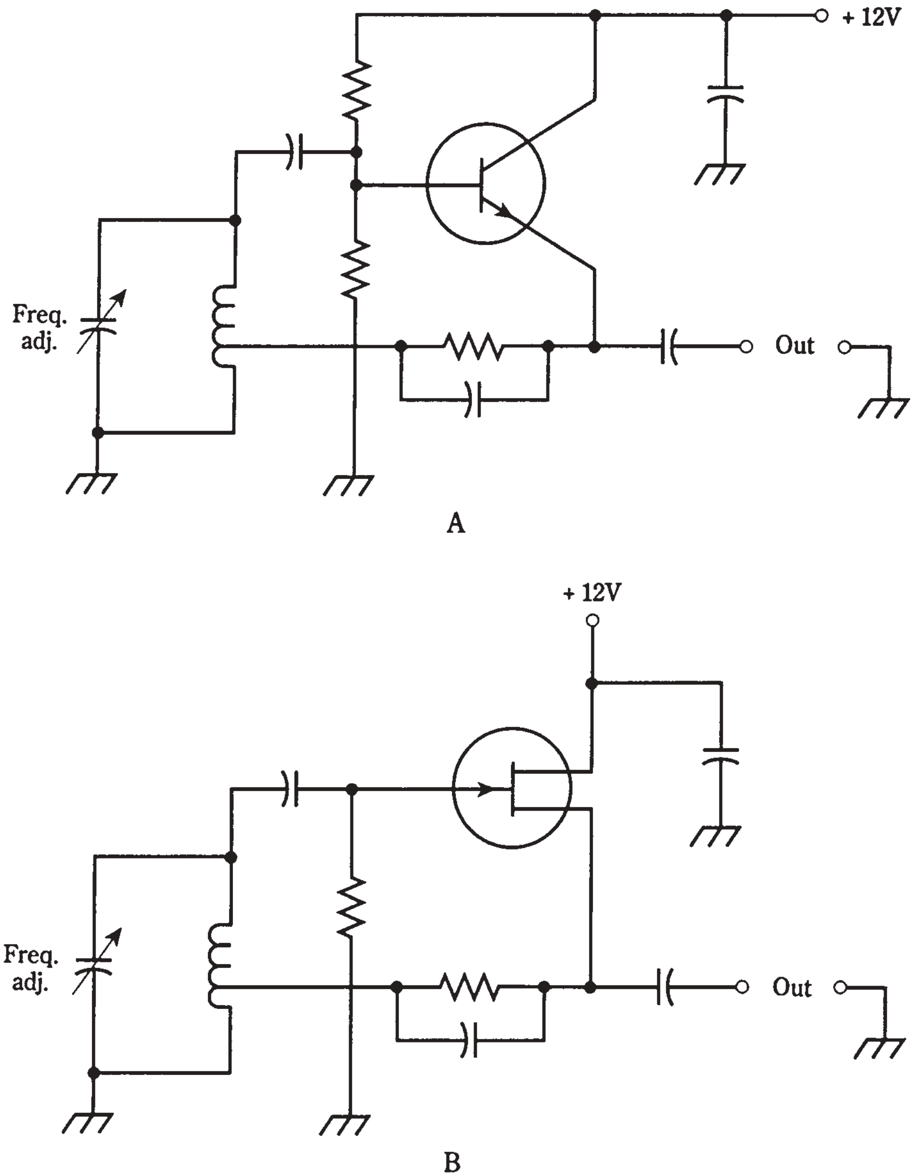
25-2 An Armstrong oscillator.

The frequency of this oscillator is controlled by means of a capacitor across either the primary or the secondary winding of the transformer. The inductance of the winding, along with the capacitance, forms a resonant circuit. The formula for determining the LC resonant frequency is in chapter 17. If you've forgotten it, now is a good time to review it.

The oscillator of Fig. 25-2 is known as an Armstrong oscillator. A bipolar transistor can be used in place of the JFET. It would need to be biased, using a resistive voltage-divider network, like a class-A amplifier.

The Hartley circuit

A method of obtaining controlled feedback at RF is shown in Fig. 25-3. At A, an NPN bipolar transistor is used; at B, an N-channel JFET is employed. The PNP and P-channel circuits are identical, but the power supply is negative instead of positive.



25.3 Hartley oscillators. At A, NPN bipolar transistor; at B, N-channel JFET.

The circuit uses a single coil with a tap on the windings to provide the feedback. A variable capacitor in parallel with the coil determines the oscillating frequency, and allows for frequency adjustment. This circuit is called a *Hartley oscillator.*

The Hartley oscillator uses about one-quarter of its amplifier power to produce feedback. (Remember, all oscillators are really specialized amplifiers.) The other three-quarters of the power can be used as output. Oscillators do not, in general, produce more than a fraction of a watt of power. If more power is needed, the signal can be boosted by one or more stages of amplification.

It's important to use only the minimum amount of feedback necessary to get oscillation. The amount of feedback is controlled by the position of the coil tap.

The Colpitts circuit

Another way to provide RIF feedback is to tap the capacitance instead of the inductance in the tuned circuit. In Fig. 25-4, NPN bipolar (at A) and N-channel JFET (at B) *Colpitt*s *oscillator* circuits are diagrammed.

The amount of feedback is controlled by the ratio of capacitances. The coil, rather than the capacitors, is variable in this circuit. This is a matter of convenience. It's almost impossible to find a dual variable capacitor with the right capacitance ratio between sections. Even if you find one, you cannot change the ratio of capacitances. It's easy to adjust the capacitance ratio using a pair of fixed capacitors.

Unfortunately, finding a good variable inductor might not be much easier than getting hold of a suitable dual-gang variable capacitor. A *permeability-tuned* coil can be used, but ferromagnetic cores impair the frequency stability of an RF oscillator. A *roller inductor* might be employed, but these are bulky and expensive. An inductor with several switch-selectable taps can be used, but this wouldn't allow for continuous frequency adjustment. The tradeoff is that the Colpitts circuit offers exceptional stability and reliability when properly designed.

As with the Hartley circuit, the feedback should be kept to the minimum necessary to sustain oscillation.

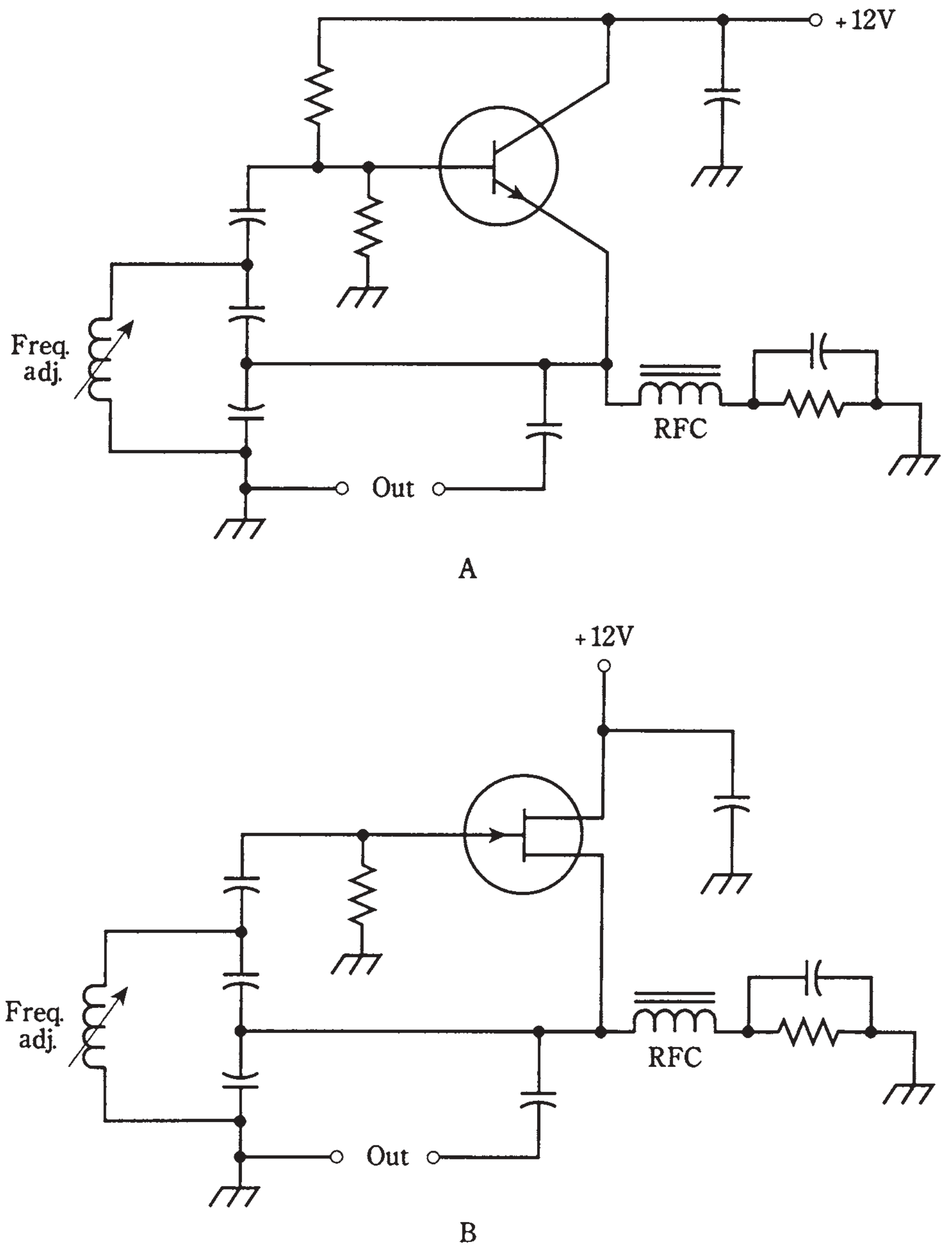
In these circuits, the outputs are taken from the emitter or source. Why? Shouldn't the output be taken from the collector or drain? The answer is that the output *can* be taken from the collector or drain circuit, and an oscillator will usually work just fine. But gain is not important in an oscillator; what matters is stability under varying load conditions. Stability is enhanced when the output of an oscillator is taken from the emitter or source portion of the circuit.

To prevent the output signal from being short-circuited to ground, an *RF choke (RFC)* is connected in series with the emitter or source lead in the Colpitts circuit. The choke lets dc pass while blocking ac (just the opposite of a blocking capacitor). Typical values for RF chokes range from about 100 µH at high frequencies, like 15 MHz, to 10 mH at low frequencies, such as 150 kHz.

The Clapp circuit

A variation of the Colpitts oscillator makes use of series resonance, instead of parallel resonance, in the tuned circuit. Otherwise, the circuit is basically the same as the parallel-tuned Colpitts oscillator. A schematic diagram of an N-channel JFET *Clapp oscillator* circuit is shown in Fig. 25-5. The P-channel circuit is identical, except for the power supply polarity, which is reversed.

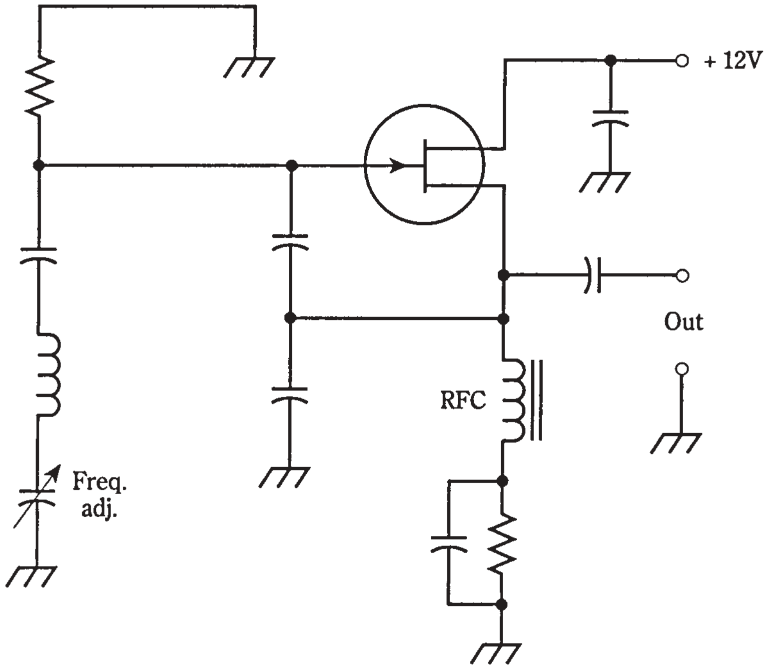
The bipolar-transistor Clapp circuit is almost exactly the same as the circuit of Fig. 25-5, with the emitter in place of the source, the base in place of the gate, and the collector



25-4 Colpitts oscillators. At A, NPN bipolar transistor; at B, N-channel JFET.

in place of the drain. The only difference, as you can probably guess by now, is the addition of a resistor between the base and the positive supply voltage (for NPN) or the negative supply voltage (for PNP).

The Clapp oscillator offers excellent stability at RF. Its frequency won't change much when high-quality components are used. The Clapp oscillator is a reliable circuit;



25-5 Series-tuned Colpitts oscillators using an N-channel JFET.

it isn't hard to get it to oscillate. Another advantage of the Clapp circuit is that it allows the use of a variable capacitor for frequency control, while still accomplishing feedback through a capacitive voltage divider.

Stability

The term *stability* is used often by engineers when they talk about oscillators. In an oscillator, stability has two meanings: constancy of frequency and reliability of performance. Obviously, both of these considerations are important in the design of a good oscillator circuit.

Constancy of frequency

The foregoing oscillator types allow for frequency adjustment using variable capacitors or variable inductors. The component values are affected by temperature, and sometimes by humidity. When designing a *variable-frequency oscillator (VFO),* it's crucial that the components maintain constant values, as much as possible, under all anticipated conditions.

Some types of capacitors maintain their values better than others, when the temperature goes up or down. Among the best are polystyrene capacitors. Silver-mica capacitors also work well when polystyrene units can't be found.

Inductors are most temperature-stable when they have air cores. They should be wound, when possible, from stiff wire with strips of plastic to keep the windings in place. Some air-core coils are wound on hollow cylindrical cores, made of ceramic or phenolic material. Ferromagnetic solenoidal or toroidal cores aren't very good for VFO coils, because these materials change their permeability as the temperature varies. This changes the inductance, in turn affecting the oscillator frequency.

Engineers spend much time and effort in finding components that will minimize *drift* (unwanted changes in frequency over time) in VFOs.

Reliability of performance

An oscillator should always start working as soon as power is supplied. It should keep oscillating under all normal conditions, not quitting if the load changes slightly or if the temperature rises or falls. A "finicky" oscillator is a great annoyance. The failure of a single oscillator can cause an entire receiver, transmitter, or transceiver to stop working. An oscillator is sometimes called *unstable* if it has to be "coaxed" into starting, or if it quits unpredictably.

Some oscillator circuits are more reliable than others. The circuits generalized in this chapter are those that engineers have found, through trial and error over the years, to work the best.

When an oscillator is built and put to use in a radio receiver, transmitter, or audio device, *debugging* is always necessary. This is a trial-and-error process of getting the flaws, or "bugs" out of the circuit. Rarely can an engineer build something straight from the drawing board and have it work just right the first time. In fact, if two oscillators are built from the same diagram, with the same component types and values in the same geometric arrangement, one circuit might work fine, and the other might be unstable. This usually happens because of differences in the quality of components that don't show up until the "acid test."

Oscillators are designed to work into a certain range of load impedances. It's important that the load impedance not be too low. (You need never be concerned that it might be too high. In general, the higher the load impedance, the better.) If the load impedance is too low, the load will try to draw power from an oscillator. Then, even a well-designed oscillator might be unstable. Oscillators aren't meant to produce powerful signals. High power can be obtained using amplification after the oscillator.

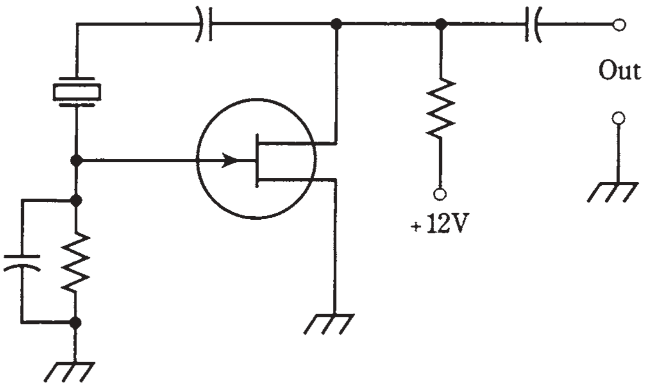
Crystal-controlled oscillators

Quartz crystals can be used in place of tuned LC circuits in RF oscillators, if it isn't necessary to change the frequency often. Crystal oscillators offer excellent frequency stability— far superior to that of LC-tuned VFOs.

There are several ways that crystals can be connected in bipolar or FET circuits to get oscillation. One common circuit is the *Pierce oscillator.* An N-channel JFET and quartz crystal are connected in a Pierce configuration as shown in the schematic diagram of Fig. 25-6.

The crystal frequency can be varied somewhat (by about 0.1 percent) by means of an inductor or capacitor in parallel with the crystal. But the frequency is determined mainly by the thickness of the crystal, and by the angle at which it is cut from the quartz rock.

Crystals change in frequency as the temperature changes. But they are far more stable than LC circuits, most of the time. Some crystal oscillators are housed in temperature-controlled chambers called *ovens.* They maintain their frequency so well that



25-6 A JFET Pierce oscillator.

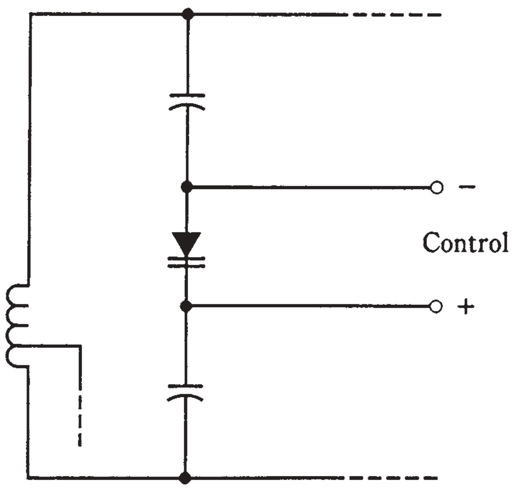
they are often used as *frequency standards,* against which other oscillators are calibrated. The accuracy can be within a few Hertz at working frequencies of several megahertz.

The voltage-controlled oscillator

The frequency of a VFO can be adjusted via a varactor diode in the tuned LC circuit. Recall that a varactor, also called a varicap, is a semiconductor diode that works as a variable capacitor when it is reverse-biased. The capacitance depends on the reverse-bias voltage. The greater this voltage, the lower the value of the capacitance.

The Hartley and Clapp oscillator circuits lend themselves well to varactor-diode frequency control. The varactor is placed in series or parallel with the tuning capacitor, and is isolated for dc by blocking capacitors. The schematic diagram of Fig. 25-7 shows an example of how a varactor can be connected in a tuned circuit. The resulting oscillator is called a *voltage-controlled oscillator (VCO).*

25-7 Connection of a varactor in a tuned LC circuit.



Why control the frequency of an oscillator in this way? It is commonly done in modern communications equipment; there must be a reason. In fact there are several good reasons why varactor control is better than the use of mechanically variable capacitors or inductors. But it all comes down to basically one thing: Varactors are cheaper. They're also less bulky than mechanically variable capacitors and inductors.

Nowadays, many frequency readouts are digital. You look at a numeric display instead of interpolating a dial scale. Digital control is often done by a *microcomputer.* You program the operating frequency by pressing a sequence of buttons, rather than by rotating a knob. The microcomputer might set the frequency via a synchro on the shaft of a variable capacitor or inductor. But that would be unwieldy. It would also be ridiculous, a "Rube Goldberg" contraption. A varactor can control the frequency without all that nonsense.

The PLL frequency synthesizer

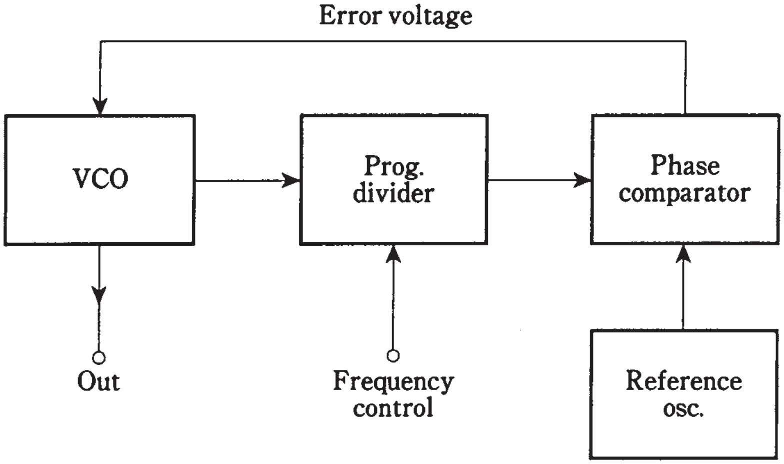
One type of oscillator that combines the flexibility of a VFO with the stability of a crystal oscillator is known as a *PLL frequency synthesizer.* This scheme is extensively used in modern digital radio transmitters and receivers.

The output of a VCO is passed through a *programmable divider,* a digital circuit that divides the VCO frequency by any of hundreds or even thousands of numerical values chosen by the operator. The output frequency of the programmable divider is locked, by means of a *phase comparator,* to the signal from a crystal-controlled *reference oscillator.*

As long as the output from the programmable divider is exactly on the reference oscillator frequency, the two signals are in phase, and the output of the phase comparator is zero volts dc. If the VCO frequency begins to drift, the output frequency of the programmable divider will drift, too (although at a different rate). But even the tiniest frequency change—a fraction of 1 Hz—causes the phase comparator to produce a dc *error voltage.* This error voltage is either positive or negative, depending on whether the VCO has drifted higher or lower in frequency. The error voltage is applied to a varactor in the VCO, causing the VCO frequency to change in a direction opposite to that of the drift. This forms a dc feedback circuit that maintains the VCO frequency at a precise multiple of the reference-oscillator frequency, that multiple having been chosen by the programmable divider. It is a *loop* circuit that *locks* the VCO onto a precise frequency, by means of *phase* sensing, hence the term *phase-locked loop* (PLL).

The key to the stability of the PLL frequency synthesizer lies in the fact that the reference oscillator is crystal-controlled. A block diagram of such a synthesizer is shown in Fig. 25-8. When you hear that a radio receiver, transmitter, or transceiver is "synthesized," it usually means that the frequency is determined by a PLL frequency synthesizer.

The stability of a synthesizer can be enhanced by using an amplified signal from the National Bureau of Standards, transmitted on shortwave by WWV at 5, 10, or 15 MHz, directly as the reference oscillator. These signals are frequency-exact to a minuscule fraction of 1 Hz, because they are controlled by atomic clocks. Most people don't need precision of this caliber, so you won't see consumer devices like ham radios and shortwave receivers with *primary-standard* PLL frequency synthesis. But it is employed by some corporations and government agencies, such as the military.



25-8 Block diagram of a PLL frequency synthesizer.

Diode oscillators

At ultra-high and microwave frequencies, certain types of diodes can be used as oscillators. These diodes, called *Gunn, IMPATT,* and *tunnel* diodes, were discussed in chapter 20.

Audio waveforms

The above described oscillators work above the human hearing range. At audio frequencies (AF), oscillators can use RC or LC combinations to determine frequency. If LC circuits are used, the inductances must be rather large, and ferromagnetic cores are usually necessary.

All RF oscillators produce a sine-wave output. A pure sine wave represents energy at one and only one frequency. Audio oscillators, by contrast, don't necessarily concentrate all their energy at a single frequency. A pure AF sine wave, especially if it is continuous and frequency-constant, causes ear/mind fatigue. Perhaps you've experienced it.

The various musical instruments in a band or orchestra all sound different from each other, even when they play the same note (such as middle C). The reason for this is that each instrument has its own unique waveform. A clarinet sounds different than a trumpet, which in turn sounds different than a cello or piano.

Suppose you were to use an oscilloscope to look at the waveforms of musical instruments. This can be done using a high-fidelity microphone, a low-distortion amplifier, and a scope. You'd see that each instrument has its own "signature." Each instrument's unique sound qualities can be reproduced using AF oscillators whose waveform outputs match those of the instrument.

The art of electronic music is a subject to which whole books have been devoted. All electronic music synthesizers use audio oscillators to generate the tones you hear.

Audio oscillators

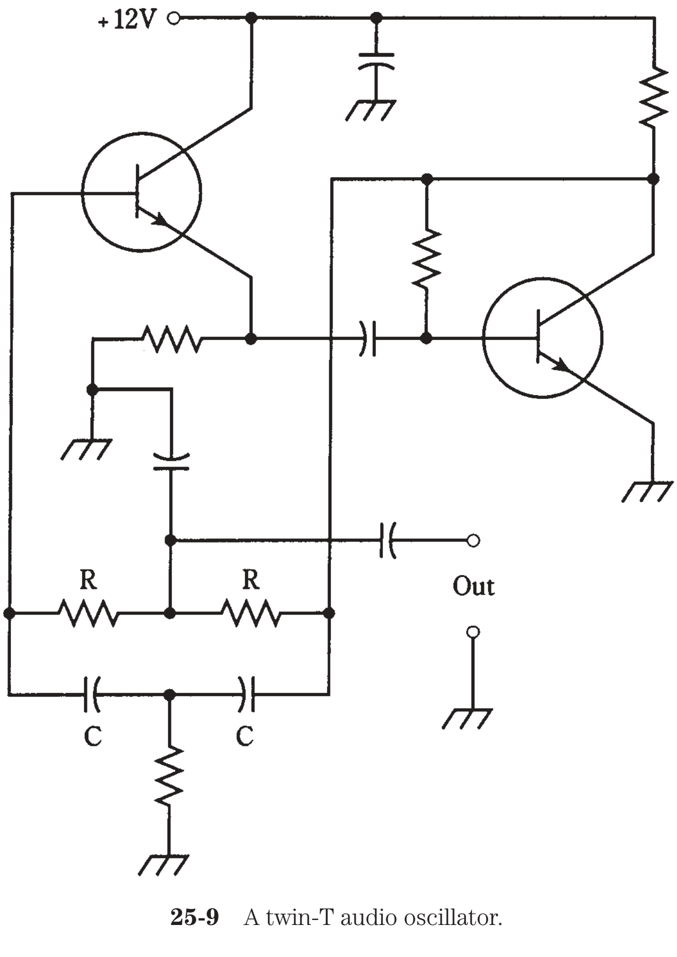
Audio oscillators find uses in doorbells, ambulance sirens, electronic games, and those little toys that play simple musical tunes. All AF oscillators work in the same way, consisting of amplifiers with positive feedback.

A simple audio oscillator

One form of AF oscillator that is popular for general-purpose use is the *twin-T oscillator* (Fig. 25-9). The frequency is determined by the values of the resistors R and capacitors C. The output is a near-perfect sine wave. The small amount of distortion helps to alleviate the irritation produced by an absolutely pure sinusoid. This circuit uses two NPN bipolar transistors. Two JFETs could also be used, biased for class-A amplifier operation.

The multivibrator

Another audio-oscillator circuit uses two identical common-emitter or common-source amplifier circuits, hooked up so that the signal goes around and around between them.

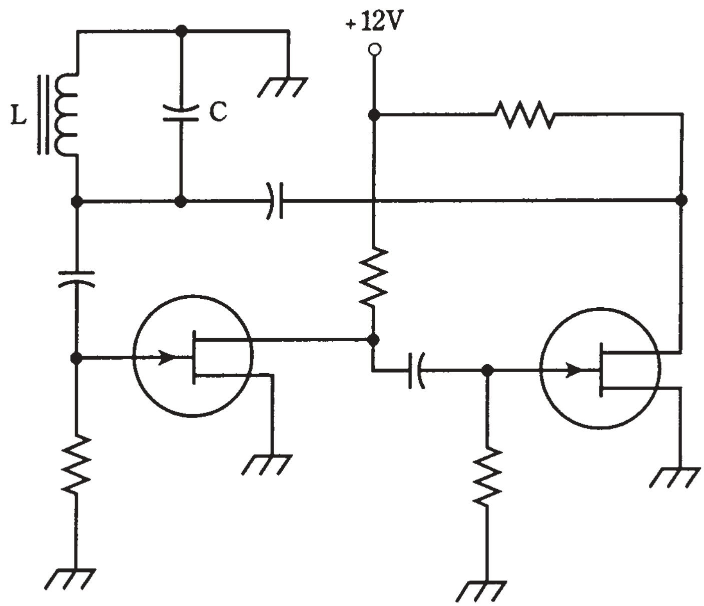


25-9 A twin-T audio oscillator.

This is sometimes called a *multivibrator* circuit, although that is technically a misnomer, the term being more appropriate to various digital signal-generating circuits.

Two N-channel JFETs are connected to form an oscillator as shown in Fig. 25-10. Each "stage" amplifies the signal in class-A, and reverses the phase by 180 degrees. Thus, the signal goes through a 360-degree shift each time it gets back to any particular point. A 360-degree shift results in positive feedback, being effectively equivalent to no phase shift.

The frequency is set by means of an LC circuit. The coil uses a ferromagnetic core, because stability is not of great concern and because such a core is necessary to obtain the large inductance needed for resonance at audio frequencies. The value of L is typically from 10 mH to as much as 1 H. The capacitance is chosen according to the formula for resonant circuits, to obtain an audio tone at the frequency desired.



25-10 A "multivibrator" type audio oscillator.

IC oscillators

In recent years, solid-state technology has advanced to the point that whole circuits can be etched onto silicon chips. Such devices are called *integrated circuits* (*ICs*). The *operational amplifier (op amp)* is one type of IC that is especially useful as an oscillator. Op-amp oscillators are most commonly employed as audio oscillators. Integrated circuits are discussed in chapter 28.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Negative feedback in an amplifier:
   A. Causes oscillation.
   B. Increases sensitivity.
   C. Reduces the gain.
   D. Is used in an Armstrong oscillator.
2. Oscillation requires:
   A. A common-drain or common-collector circuit.
   B. A stage with gain.
   C. A tapped coil.
   D. Negative feedback.
3. A Colpitts oscillator can be recognized by:
   A. A split capacitance in the tuned circuit.
   B. A tapped coil in the tuned circuit.
   C. A transformer for the feedback.
   D. A common-base or common-gate arrangement.
4. In an oscillator circuit, the feedback should be:
   A. As great as possible.
   B. Kept to a minimum.
   C. Just enough to sustain oscillation.
   D. Done through a transformer whose wires can be switched easily.
5. A tapped coil is used in a(n):
   A. Hartley oscillator.
   B. Colpitts oscillator.
   C. Armstrong oscillator.
   D. Clapp oscillator.
6. An RF choke:
   A. Passes RF but not dc.
   B. Passes both RF and dc.
   C. Passes dc but not RF.
   D. Blocks both dc and RF.
7. Ferromagnetic coil cores are not generally good for use in RF oscillators because:
   A. The inductances are too large.
   B. It’s hard to vary the inductance of such a coil.
   C. Such coils are too bulky.
   D. Air-core coils have better thermal stability.

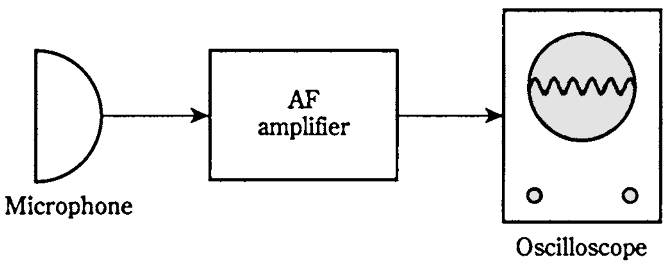
8. An oscillator might fail to start for any of the following reasons *except*:
   A. Low-power-supply voltage.
   B. Low stage gain.
   C. In-phase feedback.
   D. Very low output impedance.
9. An advantage of a crystal-controlled oscillator over a VFO is:
   A. Single-frequency operation.
   B. Ease of frequency adjustment.
   C. High output power.
   D. Low drift.
10. The frequency at which a crystal oscillator functions is determined mainly by:
    A. The values of the inductor and capacitor.
    B. The thickness of the crystal.
    C. The amount of capacitance across the crystal.
    D. The power-supply voltage.
11. The different sounds of musical instruments are primarily the result of:
    A. Differences in the waveshape.
    B. Differences in frequency.
    C. Differences in amplitude.
    D. Differences in phase.
12. A radio-frequency oscillator usually:
    A. Has an irregular waveshape.
    B. Has most or all of its energy at a single frequency.
    C. Produces a sound that depends on its waveform.
    D. Uses RC tuning.
13. A varactor diode:
    A. Is mechanically flexible.
    B. Has high power output.
    C. Can produce different waveforms.
    D. Is good for use in frequency synthesizers.
14. A frequency synthesizer has:
    A. High power output.
    B. High drift rate.
    C. Exceptional stability.
    D. Adjustable waveshape.

15. A ferromagnetic-core coil is preferred for use in the tuned circuit of an RF oscillator:
    - A. That must have the best possible stability.
    - B. That must have high power output.
    - C. That must work at microwave frequencies.
    - D. No! Air-core coils work better in RF oscillators.

16. If the load impedance for an oscillator is too high:
    - A. The frequency might drift.
    - B. The power output might be reduced.
    - C. The oscillator might fail to start.
    - D. It's not a cause for worry; it can't be too high.

17. The bipolar transistors or JFETs in a multivibrator are usually connected in:
    - A. Class B.
    - B. A common-emitter or common-source arrangement.
    - C. Class C.
    - D. A common-collector or common-drain arrangement.

18. The arrangement in the block diagram of Fig. 25-11 represents:
    - A. A waveform analyzer.
    - B. An audio oscillator.
    - C. An RF oscillator.
    - D. A sine-wave generator.



25-11 Illustration for quiz question 18.

19. Acoustic feedback in a public-address system:
    - A. Is useful for generating RF sine waves.
    - B. Is useful for waveform analysis.
    - C. Can be used to increase the amplifier gain.
    - D. Serves no useful purpose.

20. An IMPATT diode:
    - A. Makes a good audio oscillator.
    - B. Can be used for waveform analysis.
    - C. Is used as a microwave oscillator.
    - D. Allows for frequency adjustment of a VCO.

26

CHAPTER

Data transmission

TO CONVEY DATA, SOME ASPECT OF A SIGNAL MUST BE VARIED. THERE ARE several different characteristics of a signal that can be made to change in a controlled way, so that data is "imprinted" on it. *Modulation* is the process of imprinting data onto an electric current or radio wave.

Modulation can be accomplished by varying the amplitude, the frequency, or the phase of a wave. Another method is to transmit a series of pulses, whose duration, amplitude, or spacing is made to change in accordance with the data to be conveyed.

The carrier wave

The "heart" of most communications signals is a sine wave, usually of a frequency well above the range of human hearing. This is called a *carrier* or *carrier wave.* The lowest carrier frequency used for radio communications is 9 kHz. The highest frequency is in the hundreds of gigahertz.

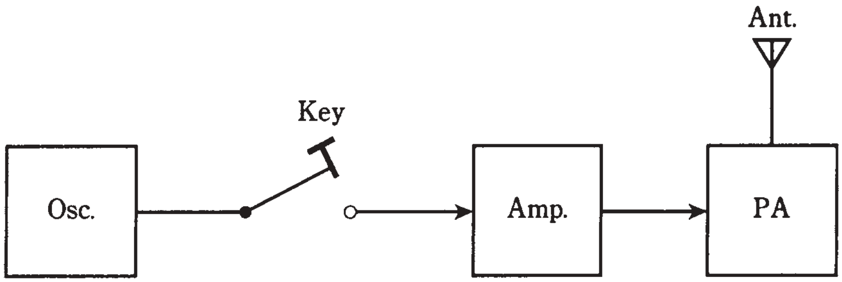
For modulation to work effectively, the carrier must have a frequency many times the highest frequency of the modulating signal. For example, if you want to modulate a radio wave with hi-fi music, which has a frequency range from a few hertz up to 20 kHz or so, the carrier wave must have a frequency well above 20 kHz. A good rule is that the carrier must have a frequency of at least 10 times the highest modulating frequency. So for good hi-fi music transmission, a radio carrier should be at 200 kHz or higher.

This rule holds for all kinds of modulation, whether it be of the amplitude, phase, or frequency. If the rule is violated, the efficiency of transmission will be degraded, resulting in less-than-optimum data transfer.

The Morse code

The simplest, and oldest, form of modulation is on-off *keying.* Early telegraph systems used direct currents that were keyed on and off, and were sent along wires. The first radio transmitters employed spark-generated "hash" signals that were keyed using the telegraph code. The noise from the sparks, like ignition noise from a car, could be heard in crystal-set receivers several miles away. At the time of their invention, this phenomenon was deemed miraculous: a wireless telegraph!

Keying is usually accomplished at the oscillator of a *continuous-wave (CW)* radio transmitter. A block diagram of a simple CW transmitter is shown in Fig. 26-1. This is the basis for a mode of communications that is, and always has been, popular among radio amateurs and experimenters.



26-1 A simple CW transmitter.

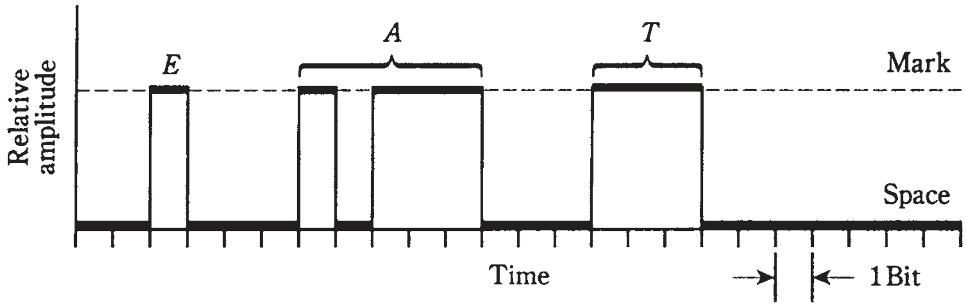
While the use of Morse code might seem old-fashioned, even archaic, a CW transmitter is extremely simple to build. A human operator, listening to Morse code and writing down the characters as they are sent, is one of the most efficient data receivers ever devised. Until computers are built to have intuition, there'll always be a place for Morse code radio communications. Besides being efficient, as any "CW fanatic" radio ham will tell you, it's just plain fun to send and receive signals in Morse code.

Morse code is a form of *digital* communications. It can be broken down into bits, each having a length of one *dot.* A *dash* is three bits long. The space between dots and dashes, within a single character, is one bit. The space between characters in a word is three bits. The space between words is seven bits. Punctuation marks are sent as characters attached to their respective words. An *amplitude-versus-time* rendition of the Morse word "eat" is shown in Fig. 26-2.

Morse code is a rather slow way to send and receive data. Human operators typically use speeds ranging from about 5 words per minute (wpm) to 40 or 50 wpm.

Frequency-shift keying

Morse code keying is the most primitive form of *amplitude modulation (AM).* The strength, or amplitude, of the signal is varied between two extreme conditions: full-on and full-off. There is another way to achieve two-state keying that works better with



26-2 The Morse code word "eat" as sent on CW.

teleprinter machines than on-off switching. That is to shift the frequency of the carrier wave back and forth. It is called *frequency-shift keying (FSK).*

Teleprinter codes

The Morse code is not the only digital code of its kind. There are two common teleprinter codes used to send and receive *radioteletype (RTTY)* signals. These codes are known as the *Baudot* (pronounced "Bo-*doe")* and ASCII (pronounced "*ask*-ee") codes. (You needn't worry about where these names come from.) A carrier wave can be keyed on and off using either of these codes, at speeds ranging from 60 wpm to over 1000 wpm. In recent years, ASCII has been replacing Baudot as the standard teleprinter code.

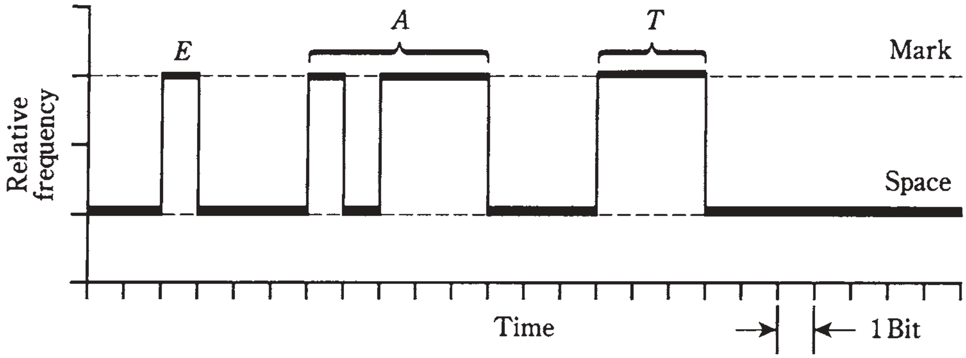
A special circuit, called a *terminal unit,* converts RTTY signals into electrical impulses to work a teleprinter or to display the characters on a monitor screen. The terminal unit also generates the signals necessary to send RTTY, as the operator types on the keyboard of a teleprinter *terminal.* A personal computer can be made to work as an RTTY terminal by means of *terminal emulation software.* This software is available in several different forms, and is popular among radio amateurs and electronics hobbyists.

Mark and space

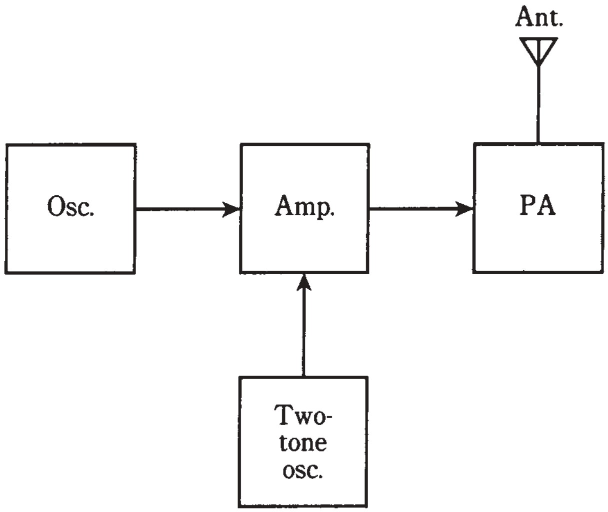
The trouble with using simple on-off keying for RTTY is that noise pulses, such as thunderstorm static crashes, can be interpreted by the terminal unit as signal pulses. This causes misprints. There is no problem if a crash takes place during the full-on, or *mark,* part of the signal; but if it happens during a pause or *space* interval, the terminal unit can be fooled into thinking it's a mark pulse instead.

This problem can be helped greatly by sending a signal during the space part of the signal, but at a different frequency from the mark pulse. Then the terminal unit knows for sure that a mark is not being sent. Instead of sending "Not mark," the transmitter sends "Not mark, but space instead." The easiest way to do this is to send the mark part of the signal at one carrier frequency, and the space part at another frequency a few hundred hertz higher or lower. This is FSK. The difference between the mark and space frequencies is called the *shift,* and is usually between 100 Hz and 1 kHz.

A *frequency-versus-time* graph of the Morse code word "eat," sent using FSK, is shown in Fig. 26-3. Normally, Baudot or ASCII, rather than Morse, is used for teleprinter operation. A block diagram of an FSK transmitter is shown in Fig. 26-4. The FSK mode, like on-off code keying, is a digital form of communications. But unlike on-off Morse keying, FSK is *frequency modulation (FM).*



26-3 The Morse code word "eat" as sent using FSK.



26-4 A simple FSK transmitter.

The telephone modem

Teleprinter data can be sent over the telephone lines using FSK entirely within the audio range. Two audio tones are generated, one for mark and the other for space. There are three sets of standard tone frequencies: 1200 Hz and 2200 Hz for general communications, 1070 Hz and 1270 Hz for message origination, and 2025 and 2225 Hz for answering. These represent shifts of 1000 Hz or 200 Hz.

Because this FSK takes place at audio, it is sometimes called *audio-frequency-shift keying (AFSK).* A device that sends and receives AFSK teleprinter over the phone lines is known as a *telephone modem.* If you've used a personal computer via the phone lines, you've used a telephone modem. Perhaps you've heard the "bleep-bleep" of the tones as data is sent or received.

Amplitude modulation for voice

A voice signal is a complex waveform with frequencies mostly in the range 300 Hz to 3 kHz. Direct currents can be varied, or *modulated,* by these waveforms, thereby transmitting voice information over wires. This is how early telephones worked.

Around 1920, when CW oscillators were developed to replace spark-gap transmitters, engineers wondered, "Can a radio wave be modulated with a voice, like dc in a telephone?" If so, voices could be sent by "wireless." Radio communications via Morse code was being done over thousands of miles, but CW was slow. The idea of sending voices was fascinating, and engineers set about to find a way to do it.

A simple amplitude modulator

An amplifier was built to have variable gain. The idea was to make the gain fluctuate at voice-frequency rates—up to 3 kHz or so. *Vacuum tubes* were used as amplifiers back then, because solid-state components hadn't been invented yet. But the principle of *amplitude modulation (AM)* is the same, whether the active devices are tubes, bipolar transistors, or FETs.

If bipolar transistors had been around in 1920, the first *amplitude modulator* would have resembled the circuit shown in Fig. 26-5. This circuit is simply a class-A RF amplifier, whose gain is varied in step with a voice signal coupled into the emitter circuit. The voice signal affects the instantaneous voltage between the emitter and base, varying the instantaneous bias. The result is that the instantaneous RF output increases and decreases, in a way that exactly duplicates the waveform of the voice signal.

The circuit of Fig. 26-5 will work quite well as an AM voice modulator, provided that the audio input isn't too great. If the AF is excessive, *overmodulation* will occur. This will result in a distorted signal.

The AM transmitter

Two complete AM transmitters are shown in block-diagram form in Fig. 26-6. At A, modulation is done at a low power level. This is *low-level AM.* All the amplification stages after the modulator must be linear. That means class AB or class B must be used. If a class-C PA is used, the signal will be distorted.

In some broadcast transmitters, AM is done in the final PA, as shown in Fig. 26-6B. This is *high-level AM.* The PA operates in class C; it is the modulator as well as the final amplifier. As long the PA is modulated correctly, the output will be a "clean" AM signal; RF linearity is of no concern.

The extent of modulation is expressed as a percentage, from 0 percent, representing an unmodulated carrier, to 100 percent, representing full modulation. Increasing the modulation past 100 percent will cause distortion of the signal, and will degrade, not enhance, the effectiveness of data transmission.

In an AM signal that is modulated 100 percent, only 1/3 of the power is actually used to convey the data; the other 2/3 is consumed by the carrier wave. For this reason, AM is rather inefficient. There are voice modulation techniques that make better use of available transmitter power. Perhaps the most widely used is *single sideband (SSB),* which you'll learn about shortly.