HOW TO DESIGN ANALOG CIRCUITS

Junction Diodes — Rectifiers & Zeners

Made up of two different types of semiconductor material, the diode is found in a wide variety of applications. This month we'll learn how it is used.

MANNIE HOROWITZ

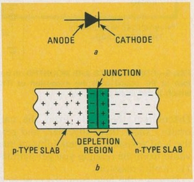

FIG. 1—A JUNCTION DIODE consists of a slab of n-type- and a slab of p-type-semiconductor material in close contact with each other, as shown in *b*. The schematic symbol for that device is shown in *a*.

THIS MONTH, WE ARE GOING TO LOOK AT another semiconductor device—the diode. Unlike the devices we discussed in the last part of this series, it is made up of both p-type and n-type semiconductor material. Diodes pass current readily in only one direction. It is that property that makes the device useful in a wide variety of circuits. Let's take a closer look at one of the most popular types of diode—the junction diode.

Junction diodes

The schematic symbol for a junction diode is shown in Fig. 1-a. Junction diodes are formed by placing a slab of n-type semiconductor material in contact with a slab of p-type semiconductor material. Now you would think that because you have an n-type slab touching a p-type slab, some of the excess electronics from the former will be attracted to and flow to the latter. That is exactly what happens.

Let's take a look at what happens at the junction of the p- and n-type semiconductor material, as shown in Fig. 1-b. Near the junction, there is a small area of electrons on the p-type slab and a small area of *holes* on the n-type slab. (A hole is an area with a shortage of electrons. In Fig. 1-b, the holes are indicated by the + signs.) That area near the junction is called the *depletion region*.

Normally, the number of electrons that cross over the junction is limited because the first electrons to do that repel the rest. The same holds true for the holes. To have more electrons and holes cross the junction, a negative voltage (relative to the voltage on the p-type slab) must be applied to the n-type slab. When that is done, the diode is said to be *forward biased*. If a positive voltage (again relative to the voltage on the p-type slab) were applied to the n-type slab, few, if any electrons or holes would cross the junction. In that case, the diode is said to be *reverse biased*.

A forward-biased diode is *not* a linear device. That means that the current that flows through it is not proportional to the applied voltage. That is shown in Fig. 2. Note that when the applied voltage is small, the current rises very slowly. As the voltage increases, however, the current flow begins to increase rapidly—in fact, it increases so much faster than the voltage that a slight increase in applied voltage will cause a large increase in current.

But what happens when the voltage across the diode is very small? In that situation, so little current passes through the diode that for all intents and purposes, it is said to be cut-off. It is only after the voltage exceeds a critical point, called the *threshold voltage*, that the rapid increase in current we discussed occurs. The threshold voltage can be found easily by examining the diode's

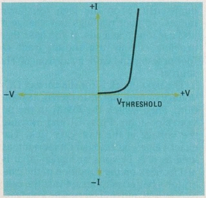

FIG. 2—A FORWARD-BIASED DIODE is considered "off" until the voltage across it exceeds the threshold voltage.

characteristic curve (a typical forward-bias characteristic curve is shown in Fig. 3). All you have to do is extend the straight portion of the curve, by drawing a dashed line, to the V (voltage) axis. The point at which the line crosses the V axis is the threshold voltage. Generally, that voltage falls between 0.6 and 0.8 volts for diodes made of silicon, and between 0.1 and 0.3 volts for germanium diodes. Knowing the threshold voltage of a diode is important, because the device will not conduct at voltages below that point.

As we all know, voltage, current, and resistance are related through Ohm's Law. Because of that, our discussion of the voltage and current characteristics of a diode leads us to another important diode characteristic—its resistance. At any point on the diode's characteristic curve, diode resistance is equal to the voltage applied to the diode divided by the resulting current flow. Finding the DC resistance of a diode at any point is fairly easy. To show how it is done, let's first pick a point on the curve and label it A. The voltage can be found by drawing a line from point A to the voltage (V) axis; that line should be perpendicular to the axis as shown in Fig. 3. The point at which the line crosses the axis is the voltage across the diode; let's call it V_1. As shown in Fig. 3, the current that flows through the diode can be found in the same manner; we'll call the point at which the line crosses the current (I) axis I_1. The diode's DC resistance at point A is, by Ohm's Law, the ratio of the voltage to the current, or $R = V_1/I_1$.

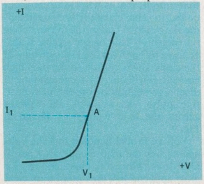

FIG. 3—THE RESISTANCE OF A DIODE at any point can be easily determined by finding the voltage across the device and the current through it, and applying Ohm's law.

The AC resistance of a diode will usually vary from its DC resistance. That is because the AC voltage is not a constant, fixed value, but varies with time. Consequently, the AC current must also vary. Since the diode is not a linear device, its resistance varies with applied voltage. AC resistance (sometimes called dynamic resistance, R_d) is therefore a change in voltage, V V, divided by the change in current, Δ I. The method for determining AC resistance is shown in Fig. 4.

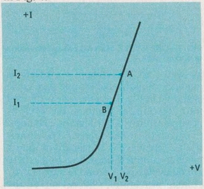

FIG. 4—FINDING THE AC RESISTANCE of a diode. Here, the voltages and currents at two points are used.

For example, let's find the AC resistance of the diode when the AC voltage across the device varies from V_1 to V_2. Here $\Delta V = V_2 - V_1$ and $\Delta I = I_2 - I_1$. Dividing, the AC resistance, R_{AC}, is $\Delta V / \Delta I = (V_2 - V/(I_2 - I_1)$. That is the resistance of the diode when the current varies between I_1 and I_2. If you were interested in the diode's resistance at lower current levels, you would find that the resistance of the diode rises as the current drops. Thus, the AC resistance will be the lowest when the current flow is a maximum; that is at the "top" of the curve.

The forward characteristics of a diode are quite important, but it is the combination of the diode's forward and reverse characteristics that makes it so useful. A diode is reverse-biased by applying a positive voltage to the n-type semiconductor material, so that it is made more positive than the p-type semiconductor material. If that were done, and the diode were an *ideal* device, it would not conduct any current. But, of course, no electronic device is ideal, and diodes *do* conduct a small amount of current. That current, called reverse or leakage current, flows because electrons and holes in the diode's depletion region are transferred back to their respective n-type and p-type slabs when reverse bias is present. That transfer generates heat in the diode, which reduces its resistance and allows an increasing amount of current to flow. Incidentally, that's the reason why manufacturer's data sheets give operating characteristics at a specific ambient temperature—those characteristics will vary somewhat with temperature.

As the reverse bias is increased, the reverse current also increases, but very gradually. However, if the reverse bias is increased to a critical point V_B, the *breakdown* voltage, an interesting phenomenon takes place. What happens is that reverse-current flow suddenly increases to a high level. In fact, it increases so much that, no matter how much more reverse bias is applied, the voltage across the diode will not change. The area of that increased current is called the *breakdown* or *avalanche* region. Reverse current is usually specified at a reverse bias just before breakdown occurs. That reverse bias, V_R, is the highest reverse bias that can be applied without breakdown occurring. A typical reverse-characteristic curve, showing the points we've discussed, is shown in Fig. 5.

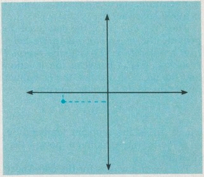

FIG. 5—REVERSE-BIASED CHARACTERISTIC of a diode. In manufacturer's specifications, reverse current (also called leakage current) is usually defined at a point just before breakdown.

Diode ratings

The first step in choosing a diode for a specific application is to obtain the manufacturer's data sheets. Those data sheets provide both the ratings and characteristics of the diode. The ratings are the maximum allowable values given to the diode before damage results. The characteristics are the normal operating parameters of the diode.

Unfortunately, there isn't an industry standard for diode specifications; you may find that the same specification has different names, depending on which manufacturer the data sheet comes from. Looking at a typical data sheet, you are likely to find a rating called *peak reverse voltage* (PRV) or *peak inverse voltage* (PIV). Both of those terms refer to the maximum allowable instantaneous reverse voltage that can be applied across the diode before breakdown occurs. Another rating is the *maximum reverse DC voltage* (V_{RDC}) or *working inverse voltage* (WIV). Those terms refer to the maximum allowable reverse-DC voltage. That rating is usually less than the peak rating. Those ratings are important, because a standard junction diode should never be operated in the breakdown region.

Other specifications found on the data sheets include the *maximum continuous forward current* (I_F). That refers to the maximum allowable forward current at a stated temperature. The *forward-voltage drop* (V_F) refers to the voltage drop across the diode at a specified forward current, at a given temperature. The *reverse-leakage current* (I_R) refers to the reverse current flow at a specified reverse voltage.

Another important specification refers to the maximum power dissipation, and it depends on the operating temperature. Manufacturer's data sheets will provide the information necessary, so that you can determine the maximum power dissipation at any given operating temperature. That information will be provided in either of two ways.

The first way is for the data sheet to provide the *maximum allowable junction temperature* (T_J) and the *thermal resistance* (θ_{JA}) between the junction and the air. To calculate the maximum power dissipation use the equation:

$$T_J = T_A + \theta_{JA} \times P_D$$

where T_A is the ambient temperature and P_D is the maximum power dissipation.

The second way is for the data sheet to provide the maximum power dissipation (P_D) at a specified temperature and also provide a derating factor expressed in milliwatts-per-°C. For example, suppose that a diode can dissipate 50 mW at 25°C, has a derating factor of 1 mW-per-°C, and the surrounding air temperature is 50°C. The maximum power dissipation will be 50 mW − 1 mW-per-°C × (50°C − 25°C), or 25 mW.

Junction diode applications

Junction diodes have a wide variety of applications. One of those is in power supplies. While we do not want to get into a detailed discussion of power-supply design at this point (that topic will be discussed later in this series), let's see how the diode can be used to convert AC into DC.

For discussion, we'll take a close look at a simple circuit, called a *half-wave rectifier*. This circuit is shown in Fig. 6.

Before we get too much farther, this is a good time to take a close look at an AC-voltage waveform; that information will be useful in other parts of this series as well. An AC-voltage waveform is made up of many cycles; one such cycle is shown in Fig. 7. Each cycle is divided into 360 angular degrees; the cycle is positive from 0 to 180 degrees, and negative from 180 to 360 degrees. The period, τ, of a sinewave is the length of time it takes to complete one cycle. Since there are 60 cycles-per-second in a 60-Hz AC voltage waveform, the period of the waveform is 1/60 of a second.

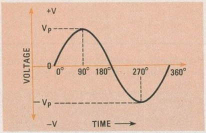

FIG. 7—ONE CYCLE of an AC voltage waveform is shown here.

While the sinewave is, as we said, divided up into 360 angular degrees, it is often more convenient to measure angles in other units, called *radians*. To convert from degrees to radians, or back again, the following formula is used:

$$\frac{\text{Angle in Radians}}{2\pi} = \frac{\text{Angle in Degrees}}{360°}$$

As you can see in Fig. 7, the level of an AC voltage varies with time. If you wanted to find the voltage at any particular instant of time (called the *instantaneous voltage*), the following formula should be used:

$$V = V_P \sin(2\pi f t)$$

or, since 2π is approximately equal to 6.28:

$$V = V_P \sin(6.28 f t)$$

Where V_P is the peak voltage, f is frequency, and t is time. As you probably noticed, we used radians in that equation.

Now let's refer back to the half-wave rectifier shown in Fig. 6. As you can see, that circuit is made up of just a transformer, diode, and capacitor. The transformer is needed because the voltage available from a wall outlet is usually about 117-volts AC; but, aside from requiring DC instead of AC, few devices will be able to use exactly that voltage for operation—some will require more, some will require less.

For the time being, let's ignore the presence of the capacitor and see how the half-wave rectifier works. When an AC voltage is input to the transformer, the output of the transformer is also an AC voltage, but at a different level. During the positive half of the cycle, the top of the transformer, as drawn, will be positive with respect to ground. Because of that, the diode in the circuit will be forward-biased, and thus conduct. Since there is very little voltage drop across the diode (no more than about 0.7 volts), the diode has little effect on circuit operation during the positive half of the cycle, and the output of the transformer appears across the load.

During the negative half of the cycle, however, the top of the transformer is negative with respect to ground. That, of course, means that the diode is reverse biased and will not conduct. Since no current will flow (actually, a *small* amount of current will flow but not enough to have any effect), the circuit is essentialy an open circuit and no voltage is placed across the load. The output voltage of the half-wave rectifier (again *ignoring* the presence of the capacitor) is shown in Fig. 8.

But we are not done yet. The rectifier's output as shown is called pulsed DC, and is not really that useful, especially in such things as radio circuits. Something must be done to smooth out that output; that is where the capacitor comes in. That capactior is charged during the positive half of the cycle; during the negative half of the cycle, it discharges through the load. If the value of the capacitor is chosen so that it is high enough, it will only have enough time to discharge slightly before the next positive pulse. As a result, the voltage across the load will be as shown in Fig. 9.

You've probably noted that the voltage across the load is still not "pure" DC. It is, however, close enough for most applications; the output from most wall-plug power-supplies is generally no smoother than what is shown here. The amount of the remaining *ripple* will depend upon the size of the capacitor used. How to find out how large a capacitor you need, as well as other types of rectifier circuits, will be covered when we take a closer look at power supplies later in this series.

Zener diodes

Let's take another look at the diode's reverse characteristic shown in Fig. 5. That figure shows that, in the breakdown region, the voltage across the diode remains constant, no matter how much reverse bias is applied. While operating a general-purpose diode in the breakdown region is not advisable (the result is likely to be at least a ruined diode), *Zener* diodes are designed to be operated in the breakdown region; they can do so without adverse results, up to

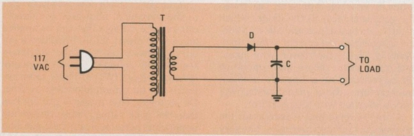

FIG. 6—A JUNCTION DIODE is used in this simple half-wave rectifier circuit. Other types of rectifiers will be discussed later in this series.

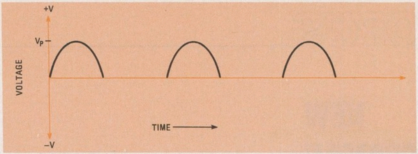

FIG. 8—THE OUTPUT of the rectifier shown in Fig. 6, ignoring the presence of the capacitor.

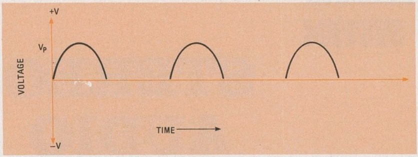

FIG. 9—WHEN THE CAPACITOR is added, the pulsing DC is "smoothed" out, making it more useful. The output from a wall-plug power supply is rarely smoother than this.

the device's rated limits. That property makes the Zener diode useful in a wide variety of voltage-regulator and voltage-reference applications.

A circuit using a Zener diode is shown in Fig. 10. Since it is the reverse characteristic that's important, the diode is connected so that it is reverse biased. The resistor is added so that the current through the Zener is limited to safe (i.e., within the device's rating) levels. Here, the Zener is used as a voltage "regulator." What happens is that, regardless of the level of the supply voltage, the voltage across the Zener will remain at V_B. Since the diode is in parallel with the load, the voltage across the load is equal to the voltage across the diode.

Zener diode applications

The circuit in Fig. 10 can be used to remove the ripple from DC output of the rectifier in Fig. 6. To do that, simply replace the battery with the output from the rectifier. The voltage across the Zener will remain at V_B regardless of how much ripple is in the rectifier's output. The output, which is taken across the Zener, is pure DC and will also be fixed at V_B.

Using what we now know about Zeners, we can now attempt some simple design work. Rather than design an entirely new circuit, however, let's see if we can calculate what values are required for the components used in the circuit shown in Fig. 10. We need to start off by defining some of the terms that we'll be using.

First of all, if the current that flows through the Zener is I_Z and the current that flows through the load is I_L, then the current that flows through the resistor is $I_Z + I_L$. That is the current that must be supplied by the battery. Additionally, the voltage supplied by the battery is V, and the circuit's output voltage is equal to the Zener's breakdown voltage, or V_B. Finally, the resistance of the load is R_L. Using those terms, we can now write a basic equation that describes the circuit:

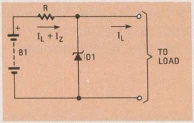

FIG. 10—A SIMPLE VOLTAGE REGULATOR using a Zener diode. This circuit will keep the voltage across the load at a constant level equal to the breakdown voltage of the Zener.

$$V_B = V - R(I_Z + I_L) \qquad (1)$$

That equation can be somewhat simplified, however. In designing Zener regulator circuits of this type, I_Z is generally assumed to be equal to 10 percent of I_L, or $I_Z = 0.1I_L$. Let's also make that assumption. Plugging that back into our circuit equation, we get:

$$V_B = V_{MIN} - 1.1 I_L R \qquad (2)$$

In the circuit we've been discussing, either the supply voltage or the load could vary; or both could vary. Let's look at what happens when only the supply voltage varies.

The first thing we need to calculate is a value for R. To do that, we need to rewrite equation 2 as:

$$R = \frac{V - V_B}{1.1 I_L} \qquad (3)$$

To solve for the maximum value that we can use for R, simply substitute V_{MIN} for V, or:

$$R = \frac{V_{MIN} - V_B}{1.1 I_L} \qquad (4)$$

One more thing we need to know about the resistor is how much power it will need to dissipate. The maximum current flows through the resistor when the supply voltage is at V_{MAX}, and is equal to $(V_{MAX} - V_B)/R$. Using the relationship $P = V^2/R$, the power rating of the resistor, in watts, must be greater than $(V_{MAX} - V_B)^2/R$.

Let's now turn our attention to the Zener. Because the maximum current flowing through the Zener ($I_{Z(MAX)}$) is equal to the maximum current flowing through the resistor, minus the current that flows through the load, then:

$$I_{Z(MAX)} = \frac{V_{MAX} - V_B}{R} - I_L \qquad (5)$$

and the power rating of the Zener must be greater than $V_B I_{Z(MAX)}$.

Now, let's consider what happens when the supply voltage remains constant, but the load varies. We can find the largest and smallest acceptable values for R in the same way we did before. However, the equation we used must be changed slightly, because we are now concerned with a variable load current and a fixed supply voltage (instead of a fixed load current and a variable supply voltage).

The maximum acceptable value of R is found by looking at the circuit when the current through the load is at a maximum. That can be found from the equation: $I_{L(MAX)} = V_B/R_{L(MIN)}$. Substituting into equation 4, we can find the maximum value of R from:

$$R = \frac{V - V_B}{1.1 I_{L(MAX)}} \qquad (6)$$

The power rating of the resistor is equal to $(V - V_B)/R$. The power rating of the diode is equal to $V_B I_{Z(MAX)}$, where:

$$I_{Z(MAX)} = \frac{V - V_B}{R} - I_{L(MIN)} \qquad (7)$$

Things get slightly more complicated when both the supply voltage and the resistance of the load vary. In that case, the maximum value of R is limited to

$$R = \frac{V_{MIN} - V_B}{1.1 I_{L(MAX)}} \qquad (8)$$

The power rating of the resistor is the same as it would be if just the supply voltage varied. To find the power dissipated by the Zener, substitute $I_{L(MIN)}$ for I_L in equation 5.

Even though we've learned a lot about diodes this month, there is still quite a bit that needs to be covered. But let's leave that for the next part of this series. Among the things we'll discuss then will be special-purpose diodes and how to use them. **R-E**