FEATURE

Basic Electronics #2B

by Ron C. Johnson

Finally! For those of you who have been following this series (and Basic Electricity before it) and wondering when you would find out something about transistors, we are finally there. (Unfortunately, now that we are there I have to tell you that transistors are on their way out . . . Not entirely, but it is true: there isn't much being designed and built now that uses discrete transistors. Everything has gone to integrated circuits, ASIC's, etc.) Regardless, lots of transistor stuff is still around and they are interesting devices. So let's get into it and see what can be done.

In our last segment we explored the area of semiconductor junctions and how they operate in diodes, rectifiers and such. Transistors involve taking another step forward with the same materials used in a different configuration.

Remember the two blocks of doped semiconductor material, (P and N types), which we brought together to form a PN junction which operated as a diode? With a transistor we take two pieces of N material and sandwich a narrow piece of P material in between them. Actually we can sandwich a piece of N material between two P materials as well. (See Figure 1). The first type is called (logically) an NPN transistor, and the second is a PNP transistor. Both are used and their primary difference involves the polarity of the power supply applied.

But I digress.

Let's talk about the NPN transistor because it is a bit more commonly used. Bipolar transistors, as we call them, are a three terminal device: we connect leads to both of the pieces of N type material and one to the P type. Just like the diodes we discussed last month, when P and N type materials are placed side by side, a buffer zone is created as electrons from the N type material migrate into the p type and recombine

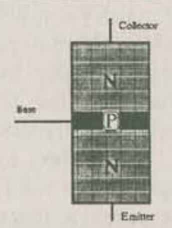

Figure 1. NPN Transistor Composition

to create a neutral area. This creates a PN junction, or diode, through which current will flow (normally) in only one direction, and then only after an initial knee voltage is overcome. In this case two back-to-back PN junctions are

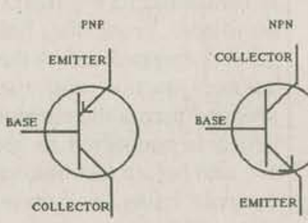

Figure 2. Transistor Symbols

formed with a connection, (which we call the base), between them. The connections to the N material are called the collector and emitter. See Figure 2.

What can be done with two back-to-back PN junction with a base connection between them? That depends on how they are connected in a DC powered circuit. Every useful transistor circuit requires that the diode created by the base-emitter PN junction be forward biased. Forward bias, as with diodes and rectifiers we have already talked about, means that a DC voltage is applied with the positive polarity connected to the P type material and the negative polarity connected to the N type. After the knee voltage (.7 volts) is overcome current will flow. The interesting thing about transistors is that, when the base-emitter junction is forward biased, and, if the collector-base junction is reverse biased (See Figure 3), current will flow from the collector, to the emitter. Moreover, a small amount of base to emitter current will control a much larger amount of collector to emitter current.

In order to understand this we must think in terms of electron flow rather than the conventional current flow we are used to. When the base emitter junction is forward biased the depletion, (or neutral), area is overcome and electrons can flow from emitter to base, but since the P type material in the base is very small, and we have biased the collector to be more positive than the base, electrons are drawn through the P type material and flow into the collector. In fact, large amounts of electrons can travel through this way, but the quantity is controlled by the much smaller amount that flow out the base lead. So if we vary the base current we can control a much larger level of collector current. This is called amplification or gain.

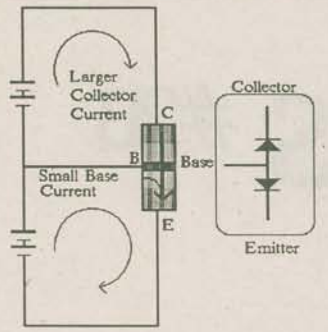

Figure 3. Transistor Bias

By varying the level of the base current we can control the collector current between zero and some maximum value determined by the power supply voltage and the resistance connected between the power supply and the collector. If we have zero current flowing through the transistor we say that it is "cutoff," whereas, if the maximum current is flowing we call it "saturation." These two extremes are used if the transistor is being used as a switch. By applying a small current to the base-emitter junction of the transistor we can switch much larger currents to turn on and off devices such as relays, lamps, LED's, etc. See Figure 4.

If we set the base-emitter current to some nominal value such that the collector current is in the middle of its range (between cutoff and saturation), we are in the "active region" of operation of the device. The circuit can be used as an amplifier because any small change in the base current will be reflected in the larger collector current.

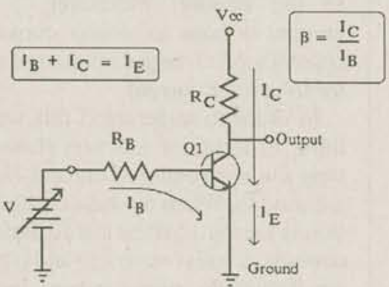

Figure 4. Transistor Current Relationships

The amount of amplification is given in the transistor specifications as ß, beta or h_{FE}.

As you can see, the bipolar junction transistor is versatile. Once you know how the base current controls the collector current and that it can be used for switching or amplification the rest is just a matter of learning how to use variations of the circuit to accomplish your particular application. Because transistors are somewhat temperature sensitive, have differing specs such as beta, maximum current and voltage, etc., the quality of the design is important in obtaining reliable and predictable operation. Transistors circuits are used for many applications from oscillators to amplifiers to digital circuits. Rather than plodding through a each configuration of transistor circuits from a design perspective, let's look at some practical circuits and see how they work.

Figure 5 shows you a typical small signal transistor amplifier. This one is called common emitter with voltage divider bias. That just relates to the configuration in which the transistor is connected. Remember, in a transistor amplifier we want to set up the base current so that the collector current is flowing somewhere near the centre of the range between cutoff and saturation. In this circuit we do that by using a voltage divider. Resistors R_1 and R_2, are connected in series between V_{cc}, (the power supply) and ground. The voltage set at the connection between them, (we'll call it V_B), is connected to the base of the transistor. From the base, .7 volts is dropped across the base emitter junction and the rest dropped across the resistor, R_E, which is connected to ground. We said before that base current controls collector current and that collector current flows down through the transistor from collector to emitter. If so, then both the base current and the collector current combine to flow through R_E. Since we know that the collector current is much greater than the base current the voltage drop across R_E will indicate, for the most part, the collector current. If we have set the emitter to some voltage this will also set the emitter current and thus, the collector current. This is also called the quiescent collector current because that is the DC level it is at without any AC signal on the circuit.

Having set the collector current, let's look at the voltage on the collector. R_c, the resistor between the power supply and the collector drops a voltage across it because of the collector current, I_c but it has no control over what the collector current is. That is controlled by the base current (which we have set in this case using the voltage divider). The voltage (with respect to ground) on the collector will be the power supply voltage, V_{cc}, less the voltage that has been dropped across the collector resistor, R_c. For this circuit the collector is our output point and the voltage there will be proportional to the collector current (but inverted as we will see shortly).

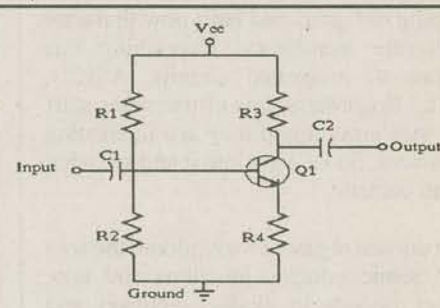

Figure 5. A Small Signal Amplifier

Now we can apply a small AC signal to the base through the capacitor shown, C_1, and this will vary the voltage out at the collector. Note, however, that if the signal applied goes positive, this increases V_b, (which controls the base current), which increases R_E, which increases I_c, which increases *the voltage drop across R_c*, which *decreases* the voltage output on the collector, V_c. All of this means that the output signal will be inverted with respect to the input signal.

This circuit is called a Class A small signal amplifier because it will take a small AC voltage signal (for example, a line out level from a cassette deck) and increase the amplitude of the voltage.

Class A means that the transistor is conducting all the time that the signal is applied, (360° of input cycle). This one would not be used to amplify the signal to drive a speaker for example because it is primarily designed to provide voltage gain, not current gain as would be required for a power amplifier. The actual gain of the amplifier can be calculated but we won't get into the theory necessary for that here.

Figure 6 shows a simple power amplifier design. This one uses a positive and negative dual power supply so that capacitive coupling is not necessary between the output and the speaker. This improves the low frequency response of the amplifier because the lower the frequency, the higher the impedance of a capacitor. This design also uses two transistors to drive the speaker: one is NPN and the other is PNP. The idea here is to use one transistor to provide current to the load (speaker, or whatever) during the positive excursion of the signal, and the other transistor to do the same on the negative excursion. More efficient than Class A, Class B would have an advantage except that it causes distortion because every time you turn on either transistor you have to overcome the .7 volt knee voltage before you start to amplify. This causes crossover distortion where the signal crosses over the zero point on the AC waveform. To alleviate this problem we operate the amp in Class AB mode, a hybrid of the

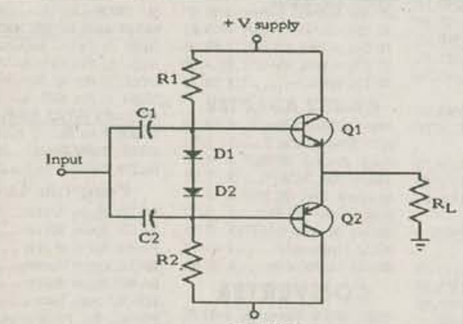

Figure 6. Class AB Power Amplifier

two where we keep both transistors slightly turned on all the time (.7 volts on the base-emitter junctions) and then apply the signal in such a way that they alternately provide current drive as with Class B. The two diodes shown between the base leads of the transistors, being forward biased by the power supply through a resistor, provide the voltage that keeps the transistors on.

There are several categories of transistor amplifiers: common emitter, common base, common collector, Class A, B, AB, C, and a few others we won't mention. Beyond this there are literally hundreds of variations on these

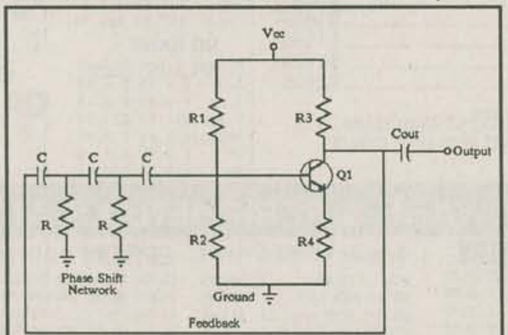

Figure 7. A Phase Shift Oscillator

categories. Rather than bore you to tears with a lot of them we'll go on to other transistor applications. In a future segment we'll build a simple amplifier to see how it works.

Meanwhile, let's look at some other circuits with transistors. Oscillators might be a good one. You may already know that an oscillator is a circuit which has an output but no input. Actually, within the circuit itself, the output of an amplifier is fed back to its own input. The circuit generates its own signal as a result of circuit action.

How does it do that? Well, there are actually two criteria which must be met in order for any system (whether it be electronic, mechanical or otherwise) to oscillate. (If you're into it these are called the Barkhausen Criterion.) First, the system must have an overall gain of greater than one. Second, the feedback signal must be returned to the input *in-phase* with the output signal. One way to do that is shown in Figure 7. The circuit is called a phase shift oscillator, and is quite simple to set up or build (although it has some drawbacks as well.) Using the same small signal amplifier we talked about earlier we route the output back to the input through a capacitor/resistor network designed to provide exactly 180° of phase shift.

Remember that a common emitter amplifier inverts the signal when amplifying it. (Inversion is the same as 180° phase shift.) Adding another 180° brings the signal back to 360° or 0°, which is what we need. The approximate frequency of the oscillation can be predicted using the formula $f_r = 1/(2\pi 6RC)$, where both resistors and all three capacitors are the same value. There is also a predictable amount of attenuation in this circuit: $B = 1/29$. In order to meet the criterion of a gain of one, then, the amplifier must provide a gain of 29 (multiply the amplitude of its input by 29 times).

If the appropriate values for components are chosen this circuit will produce a sine wave at the frequency designed for. One of the drawbacks of the circuit is that it is temperature sensitive and tends to drift, but if you need a "cheap and dirty" oscillator for testing it is quite simple to build. In fact, we might just do that in a future segment...

We have looked at a couple of linear applications using transistors. What about an application using it as a switch? The circuit shown in Figure 8 is just such an application. In digital circuits, where the only two possible outputs of a gate are zero volts or 5 volts (TTL logic), a useful piece of test equipment is a logic probe. Logic probes usually have two leads (to connect to the 5 volt power supply and ground) and a probe tip used to contact various points in the circuit you want to check out. On the body of the probe itself there are two LED's. One indicates a zero voltage (logic level 0) at the probe tip, the other a 5 volt level (logic level 1) at the probe tip. This can be used to track down problems if you know what logic levels should be where.

The circuit shown is a simple one and doesn't have all the features that a professional logic probe would provide but it will do the job for most applications. It also serves to illustrate a couple of transistor concepts which I want to talk about.

First of all, why are there two transistors connected together near the input to the circuit? Remember we said that the base current of a transistor is much lower than the collector current and that the ratio of those two current was expressed as beta, or h_{FE}? If we put two transistors together, one after another, those beta values are multiplied

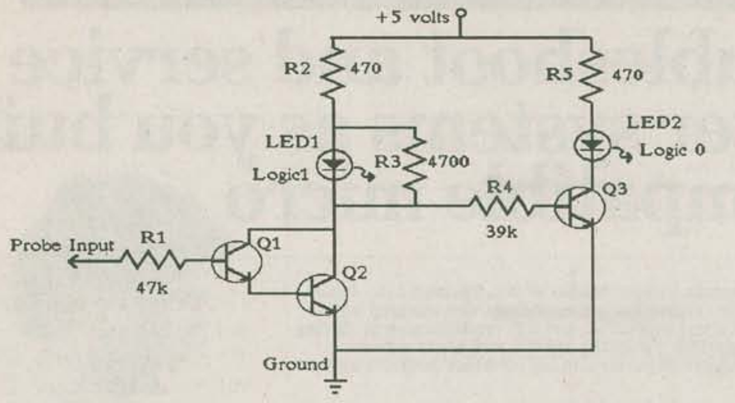

Figure 8. Schematic of a Logic Probe

together. If the beta of the first transistor is about 100 and the beta of the second is the same, the total beta, or current gain is 10,000. That means that very little base current is required at the first transistor for the second transistor to turn on and draw the necessary current through LED1. (That would be a maximum of 20 mA, so 20 mA/10,000 = 2 uA. Not very much.) That is a desirable situation because we do not want the logic probe to "load down" the circuit by drawing a lot of current from it. Another way of describing the benefit of this configuration, which is called a Darlington pair, is that it increases the input impedance of the circuit. (There is a way of calculating the actual input impedance but I'd rather stay away from the math for now.)

Now that we know why there is a Darlington pair in the circuit, how does it work overall? If the input probe is connected to the output of a digital gate, it should be either 0 volts or 5 volts. If it is 0 volts the Darlington pair will not turn on because the basic requirement of a transistor's operation is that the base-emitter junction must be forward biased with requires at least .7 volts across it. (An important fact to remember if you are troubleshooting transistor circuits.) In this case, between the input and circuit ground we have a 47k resistor and the base-emitter junctions of both transistors in series. That will require at least 1.4 volts to turn on the transistors. That means that anytime 1.4 volts or greater is applied to the circuit the Darlington pair will be turned on.

This circuit has been designed so that if the transistor is turned on it will saturate, or the maximum current possible, given the resistance in the collector circuit, will flow. The collector circuit has a 470 ohm resistor in series with an LED in it and is powered by the 5 volt supply of the circuit you are working on. Because the LED probably drops about 1.5 volts across it the total voltage possible across the 470 ohm resistor is about 3.5 volts. This gives about 7.5 mA through the LED. (A little low but there should be enough brightness to see it.) Assuming our estimation of the gain of the Darlington pair was correct we should have a gain of 10,000. Now let's see what happens if the input is connected to 5 volts. That voltage is dropped across the 47k resistor and two base-emitter junctions (5 volts - 1.4 volts = 3.6 volts) if we have 3.6 volts across the 47k resistor there must be about 76 Ua flowing through it. If the current gain is 10,000 then the collector current (current through the LED) should be: 76uA x 10,000 = 760mA. Of course this is ridiculous because we know the maximum that could ever flow through the LED is 7.5 Ma. This indicates to us the transistor is completely turned on, or saturated and the maximum, 7.5 mA is flowing. (By the way, when a transistor is saturated there should be .3 volts or less from collector to emitter— another troubleshooting hint.)

So if the input is 5 volts, LED1 turns on. We said that the collect to emitter voltage of the second transistor in the Darlington would be .3 volts at saturation. Since the emitter is connected to ground, the voltage at the collect is about .3 volts. That voltage is applied to the base of transistor, Q3, through R_4. Because .3 volts is not enough to turn on the base emitter junction of Q3, there is no current flowing through the transistor and so LED2 is off. If the input to the circuit is now changed from 5 volts to 0 volts the Darlington pair will turn off. With the second transistor turned off no current is flowing through LED1 so the voltage at the collector rises to 5 volts. (No voltage drops between the supply and collector means there must be 5 volts there. With 5 volts at the collector there is enough voltage to turn on Q3 which switches current through LED2.

Five volts in: LED1 is on. Zero volts in: LED2 is on. There is one other possibility: that there might be a series of pulses applied to the input of the circuit. In this case the LED's would turn on and off alternatively but, depending on the frequency of the pulse waveform, you would probably just see both of the LED's on.

Well, that's about all I can cram into this month's segment. Next month we'll set up some more circuits and take a walk through SCR's, triacs and some other good stuff. ❐