Introduction To Bipolar Transistors

Ken Kundert
Designer’s Guide Consulting, Inc.
1 Introduction

There are two types of transistors, bipolar junction transistors (BJT) and field-effect transistors (FET). This paper only discusses bipolar transistors.

There are two types of bipolar transistors, NPN and PNP. They both work the same but have opposite polarities. Only NPN transistors will be discussed.

The transistor is made on a silicon substrate. It has three regions, the emitter, the base, and the collector, as shown on the right. For an NPN transistor, the emitter and collector are made from N-type material and the base is made from P-type material.

2 Silicon

Intrinsic silicon forms a crystal where each atom bonds with 4 others (it has 4 electrons in its valance and it needs 8 to fill the band, so it shares its electrons with the 4 neighbors, and in return each neighbor shares one of its own, so that all atoms have full valence bands), as shown in Figure 1. In this situation, all of the electrons are firmly bound in the lattice, and so intrinsic silicon acts as an insulator (there are no free electrons to act as current carriers).

N-type material has a small number of its silicon atoms replaced by atoms such as phosphorus or arsenic that have 5 valence electrons. As it sits in the lattice it will have the opportunity to form stable bonds using four of its valence electrons with the neighboring 4 silicon atoms. At this point the valence band is full and so the fifth electron does not form a bond. It then is only loosely bound to the arsenic atom and will have a tendency to drift away. It will only be held to the vicinity of the arsenic atom by the extra proton in the nucleus of the arsenic atom. This is not a strong connection and so this extra electron will tend to move about from atom to atom. This is shown in Figure 2. In this case, the extra unbound electrons are available to carry current, so N-type material acts as a conductor. It is this ability for silicon to act like an insulator in its intrinsic (or un-doped) state and act like a conductor when small amount of impurities (dopants) are introduced that is the genesis of the term semiconductor to describe silicon.
The P-type material is similar, except that rather than using atoms with 5 valence electrons, atoms with only 3 valence electrons, such as boron or aluminum, are used. Now you get a situation like the one shown in Figure 3. An atom likes to form bonds until its valence shell is full (it takes 8 electrons to fill this shell). Thus, in this lattice, the boron atom and the silicon atom to its right are unhappy because they are unable to fill their valence shells. This is referred to as a ‘hole’. Any spare electrons that happen to be drifting through the lattice will tend to fill this hold. In fact, bound electrons from nearby atoms will leave their bond to fill the hole, and this in turn acts to form another hole. In this way, it appears as if the hole is moving, but in a direction opposite to that of the movement of the electron. This movement of holes allows P-type material to carry current and so P-type material also acts as a conductor.

It is useful to remind ourselves at this point that by convention current flows from positive potentials to more negative potentials. Holes also flow in this direction, but electrons flow in the opposite direction. It is important to keep these distinctions in mind in the following discussion.

3 PN Junctions

Now consider a PN junction. This is where N-type material is abutted to P-type material. This situation is illustrated at the atomic level in Figure 4. The N-type material will have unbound electrons that will move across the junction and fill holes near the junction. This causes a depletion region to form that is void of carriers (the free carriers in the N-type material (the electrons)
bind with the free carries in the P-type material (the holes), leaving no free carriers near the junction). This results in the PN junction acting like an insulator. However, this insulating nature is conditional, as we will see in a moment.

**FIGURE 4**  *The atomic level view of a PN junction. The free electrons from the arsenic atom has filled the hole of the boron atom.*

When the free electrons from the N-type material migrate into the P-type material to fill holes, they fill the valence shells of the atoms in the lattice, but in doing so they leave a charge imbalance near the junction. The arsenic atom ends up with a positive charge because it has one more proton that it has electrons. Similarly, the boron atom has a net negative charge because it has one fewer protons that it has electrons. As a result, an electric field builds up that limits the number of electrons that move from the N-type material to fill holes in the P-type material. This electric field continues to grow as more electrons cross the junction to fill holes until the desire on the part of the electrons to fill valence shells in the P-type material is balanced by their desire to move towards the excess of positive charge in the N-type material, at which point an equilibrium is reached. This causes a built-in potential to form at the junction, which for silicon is approximately 700mV. This is illustrated in Figure 5.

**FIGURE 5**  *The atomic level view of a PN junction.*
3.1 Junction in Forward Bias

If we apply a voltage to the PN junction such that the positive voltage is applied to the P-type material, electrons will be injected into the N-type material and electrons will be removed from the P-type material. Removing an electron from the P-type material is the same as injecting a hole. Increasing the applied voltage causes carriers to be injected into the depletion region making it smaller. Once the built-in potential is reached, the depletion region essentially disappears and current flows freely through the junction. Current flows because the potential causes electrons to drift through the N-type material towards the junction and causes holes to drift through the P-type material in the opposite direction but also towards the junction. These carriers meet at the junction and combine. In this way, current flows freely across the junction.

Applying a voltage in this way, so that current flows through the junction, is referred to as forward-biasing the junction.

3.2 Junction in Reverse Bias

Now consider the opposite case, where the voltage is applied so that N-type material is at a more positive voltage than the P-type material. For current to flow in this situation, electrons must be pulled from the N-type material and injected into the P-type material. The electrons injected into the P-type material would fill the holes. In this case, electrons would need to flow left from the junction and holes would have to flow right from the junction. This would act to further remove carriers from the region of the junction and cause the depletion region to grow. The end result is that there would be no carriers left near the junction to carry the current. With the carriers (the electrons and holes) depleted from the junction, no current flows.

Applying a voltage in this way is known as ‘reverse biasing’ the junction.

Thus, a PN junction acts as a diode. The current can flow in one way, but not in the other. This is a bit of a simplification as shown in Figure 6. The current that flows as a function of applied voltage is given by

\[ I_D = I_s \left( \frac{V_D}{V_T} - 1 \right) \]

where \( I_s \) is referred to as the saturation current, and it is very small, typically around 1fA. Thus, when the junction is reverse biased very little current flows, and when forward biased beyond the built-in potential, a great deal of current flows. The ratio between these currents is often 12-15 orders of magnitude, so a PN junction makes a really good diode.

---

1. \( I_s \) is referred to the saturation current and \( V_T = kT/q \) is the thermal voltage (about 26mV). Here \( k \) is Boltzmann’s constant, \( T \) is absolute temperature, and \( q \) is the charge of an electron.
4 The Bipolar Junction Transistor

A transistor combines two junctions and has three terminals as shown in Figure 7.

In normal operation the base-emitter junction is forward biased and the base-collector junction is reverse biased. For an NPN transistor, this means that the collector has the most positive voltage, followed by the base, and then the emitter. Thus, the circuit shown in Figure 8 would put a transistor in its normal region of operation if $V_{CC} > V_{BB}$. This region is referred to as the forward-active region.

2. The conceptual drawings of a BJT shown here are symmetric, so you would think that you could reverse the emitter and collector and have the transistor operate properly. Such a region of operation has a name as well, which is ‘reverse active’. However, transistors are not fully symmetric when they are made and so while both forward and reverse active regions can be used, generally the forward active regions gives substantially better performance.
A transistor looks like two back-to-back diodes, but there is a critical difference between back-to-back diodes and a transistor. The base region in a transistor is very thin. To see why this is important, consider first the base-emitter junction. When forward biased the electrons in the emitter will be drawn to the base hoping to combine with a hole. However, since the base is so thin, many will simply pass right through the base into the N-type region of the collector. At this point the will be pulled to the positive voltage at the collector terminal. Thus when you apply a forward bias to the base-emitter junction it will not only cause current to flow in the base, but it also causes currents to flow in the collector. If the base is very thin, then most of the electrons moving out of the emitter will actually end up in the collector. The ratio of electrons that end up in the collector to number that end up flowing in the base is named $\beta$ and is referred to as the current gain of the transistor. It often has a value of around 100, meaning that roughly 99% of the electrons that leave the emitter end up in the collector.

In this way a small current on the base controls a large current in the collector, which is the basis of an amplifier. For example, consider the circuit of Figure 9. Assume that $I_{BB}$ is a fixed current of 1mA, $V_{DD}$ is a fixed voltage of 5V, and $I_{in} = 100\mu$A $\cos(\omega t)$. Then if $\beta = 100$ the output current is

$$I_C = \beta I_B = \beta (1mA + 100\mu A \cos(\omega t)) = 1mA\beta + 100\mu A\beta \cos(\omega t)$$ (2)

$I_{BB}$ is provided as a bias to the input to make sure that it never goes negative. This is important because the base-emitter junction of the transistor is a diode and it will not allow current to flow in the negative direction (out of the base). This constant current is referred to as the ‘bias’ current and it is distinguished from $I_{in}$, which we assume to be out signal current. You can see that the signal current is amplified at the output by a factor of $\beta$.

This example demonstrates that a transistor can amplify current. This ability to amplify current can converted into an ability to amplify voltage. Consider the circuit shown in Figure 10.

For this circuit we will make the simplifying assumption that the forward biased base-emitter junction can be modeled with a fixed 700 mV voltage source (the 700 mV represents the built-in potential of the junction). We will refine that approximation later, but for now it is a adequate model as long as the transistor remains in the forward-active region. With this simplification, we can now build a simple model for the transistor in forward-active region, shown in Figure 11.
Replace the transistor with our model, as shown in Figure 12. Analyzing our voltage amplifier circuit with our transistor model is made easier if we notice that circuit can be partitioned into two pieces, the base circuit and the collector circuit, and that nothing in the base circuit is affected by the collector circuit, so we can analyze it first. Do so by writing an equation for the base current,

$$I_B = \frac{V_{in} - 700mV}{R_B}. \quad (3)$$

Now we can analyze the collector circuit,
To assure proper operation we must assume that $V_{in}$ consists of a bias component as well as a signal component, and that the bias component is larger than the signal component. However, if we ignore the constant components and just look as the signal components, (4) can be simplified to

$$v_{out} = -\frac{R_c}{R_B} \beta v_{in}.$$ (5)

Here we use lower case $v$ to represent incremental quantities (the signal components). This is rewritten as a voltage gain as

$$a_v = \frac{R_C}{R_B} \beta.$$ (6)

The direct dependence of gain to $\beta$ is worrisome in practical amplifiers because $\beta$ is often only known approximately. Thus, it is desirable to change the circuit to make the gain largely independent of $\beta$. To do so, move the base resistor to the emitter, as shown in Figure 13.

**FIGURE 13** *Another transistor voltage amplifier.*

The simplified model of this circuit is shown in Figure 13. Start by writing an equation that represents Kirchhoff’s Voltage Law in the base circuit.

$$V_{in} - 700\text{mV} - V_{RE} = 0$$ (7)

$$V_{in} = 700\text{mV} + R_E (I_B + I_C).$$ (8)

From our transistor model, we have

$$I_C = \beta I_B.$$
This can be solved for $I_B$, 

$$I_B = \frac{V_{in} - 700 \text{mV}}{(\beta + 1)R_E}. \quad (10)$$

Now you can compute the output voltage as before.

$$V_{out} = V_{CC} - R_C I_C$$

$$= V_{CC} - R_C \beta I_B$$

$$= V_{CC} - R_C \beta \left(\frac{V_{in} - 700 \text{mV}}{(\beta + 1)R_E}\right)$$

$$= V_{CC} - \frac{R_C}{R_E \beta + 1} (V_{in} - 700 \text{mV}) \quad (11)$$

Neglecting constant values gives

$$v_{out} = \frac{R_C}{R_E \beta + 1} v_{in} \quad (12)$$

Generally $\beta >> 1$, and so we can further approximate the incremental output voltage with

$$v_{out} = \frac{R_C}{R_E} v_{in}, \quad (13)$$

which leads to an incremental gain of

$$a_{out} = \frac{R_C}{R_E}. \quad (14)$$
5 Common Amplifier Configurations

There are three single-transistor amplifier configurations that are commonly used. These configurations will be described and analyzed, but before doing so we will make two changes to our transistor model. First, we will be focusing our attention on incremental quantities. Meaning that we will assume that the bias voltages will be chosen appropriately to make sure the transistor remains well within the forward active region and we will instead be focusing our attention on the small-signal characteristics of our circuit. Second, we will supplement the model of Figure 11 with an input resistance. If you recall, we assumed that the input voltage was a constant 700 mV. We can improve that model using the actual diode equation, (1), here with the terms re-labeled for the transistor

\[
I_B = I_s \left( \frac{V_{BE}}{V_T} - 1 \right). \tag{15}
\]

We are only interested in the incremental behavior, which means that we can assume that our signals represent small variations about the bias point. This allows us to simplify the equation by replacing it with its Taylor series expansion and discard all but the first order terms (in our first model we discarded all but the zeroth order terms). To do so, we replace (15) with its derivative. In effect we are linearizing our model about the operating point (replacing a curved model with a linear model that represents the tangent to the curve at the operating point). As before, we will use lower case variable names to represent small-signal (incremental) quantities.

\[
\frac{dI_B}{dV_{BE}} = \frac{I_s}{V_T} \frac{V_{BE}}{V_T} = \frac{I_B}{V_T}. \tag{16}
\]

The derivative \(dI_B/dV_{BE}\) represents the small-signal input conductance for the transistor.

\[
g_{BE} = \frac{dI_B}{dV_{BE}} = \frac{I_B}{V_T}. \tag{17}
\]

Conductance is the reciprocal of resistance, and this resistance is traditionally known as \(r_\pi\).

\[
r_\pi = \frac{V_T}{I_B} = \frac{\beta V_T}{I_C}. \tag{18}
\]

Our new small-signal model of the transistor is shown in Figure 15.

**FIGURE 15** A small-signal model of a bipolar junction transistor in the forward-active region.
5.1 Common Emitter Amplifier

The common emitter amplifier is shown in Figure 16. It gets its name because its emitter is connected to ground (and so is common to both the input circuit and the output circuit). Also shown in this figure is the small signal equivalent model for the common emitter amplifier. Notice that since this is an incremental model, all of the DC or constant valued voltages are gone. This includes $V_{CC}$, which is replaced by ground. It may seem odd to replace $V_{CC}$ with ground, but this is perfectly consistent with the idea that we are building an incremental model. In an incremental model the large constant valued bias signals are ignored. in which case $V_{CC}$ is nothing more than a constant valued voltage source connected to ground. Once we ignore its constant voltage, it is no different from ground itself.

The output voltage of the common emitter amplifier as a function of the input voltage is

$$v_{out} = R_C \beta i_B = R_C \beta v_{in} \frac{I_C}{r_p} = R_C \beta v_{in} \frac{I_C}{V_T} = R_C v_{in} \frac{I_C}{V_T}$$

(19)

Here $I_C$ is the large signal bias current flowing through the collector and $V_T$ is the thermal voltage (a constant that depends only on temperature). The ratio $I_C/V_T$ will come up a lot in these models and has special significance that can be seen if you look at the large signal model for the transistor collector current. We can derive it from (15) by recalling that in forward-active region that $I_C = \beta I_B$. Then

$$I_C = \beta I_B = \beta I_s \left( \frac{V_{BE}}{V_T} - 1 \right)$$

(20)

From this equation we can compute the incremental transconductance of the transistor.
Thus, the voltage gain of the common emitter amplifier is simply

$$a_v = -g_m R_C. \quad (22)$$

This is a very simple result. $g_m$ is the transconductance of the transistor, which its voltage to current gain. To get the gain of the circuit we just multiply this by the output resistance, $R_C$. This is as you should expect. The applied input voltage is converted to a current with a conversion factor of $g_m$, and the output resistor converts it back to a voltage, with a conversion factor of $R_C$. The composite conversion factor would be the product of these two.

The current gain of the common emitter amplifier is usually calculated by ignoring $R_C$,

$$a_i = \frac{i_C}{i_B} = \beta. \quad (23)$$

The input resistance is

$$r_{in} = \frac{v_{in}}{i_B} = r_\pi. \quad (24)$$

You can calculate the output resistance by applying a voltage source to the output and computing the resulting current that flows. The output resistance will then be $\frac{v_{test}}{i_{test}}$, where $v_{test}$ is the applied voltage and $i_{test}$ is the resulting current in the newly applied voltage source that is driving the output. In this case the current through the collector is completely unaffected by the output voltage, so you can compute the output resistance by inspection to be

$$r_{out} = R_C. \quad (25)$$

### 5.2 Degenerated Common Emitter Amplifier

It is very common to modify the common emitter amplifier by adding an emitter degeneration resistor. This will have the effect of reducing the gain, but it increases some other performance metrics such as bandwidth and linearity by the same factor, allowing you to trade off gain for these other factors. The degenerated common emitter amplifier is shown in Figure 17.

The incremental output voltage is

$$v_{out} = R_C i_C = R_C \beta i_B. \quad (26)$$

Now the input current $i_B$ is computed by writing Kirchhoff’s Voltage Law around the base loop.

$$v_{in} - r_\pi i_B - R_E (i_B + i_C) = 0 \quad (27)$$

$$v_{in} = r_\pi i_B + R_E (1 + \beta) i_B \quad (28)$$
The voltage gain is
\[
a_v = \frac{v_{out}}{v_{in}} = \frac{R_C\beta}{r_\pi + R_E(1 + \beta)} \approx \frac{R_C}{R_E}.
\]  
(30)

This last approximation assumes of course that $\beta >> 1$ and that $R_E >> r_\pi/\beta = 1/g_{m}$, both of which are most likely true.

The current gain will again be
\[
a_i = \frac{i_C}{i_B} = \beta.
\]  
(31)

Again, the current gain neglects $R_C$.

From (29) the input resistance is
\[
r_{in} = \frac{v_{in}}{i_B} = r_\pi + R_E(1 + \beta).
\]  
(32)

Notice that the input resistance is the input resistance of the un-degenerated common emitter stage plus roughly $\beta$ times the emitter resistance. Thus, the input resistance of this stage is relatively high.
The output resistance is again found by inspection to be

\[ r_{out} = R_C. \]  (33)

### 5.3 Common Base Amplifier

The common base amplifier is sometimes referred to as a current follower because the current gain is very close to one, meaning that the output current will follow the input current. The common base amplifier is shown in Figure 18.

The output voltage is

\[ v_{out} = -R_C \beta i_B, \text{ where} \]  (34)

\[ i_B = \frac{v_{in}}{r_\pi}. \]  (35)

\[ v_{out} = R_C \beta \frac{v_{in}}{r_\pi}. \]  (36)

As such, the voltage gain is

\[ a_v = \frac{R_C \beta}{r_\pi} = g_m R_C. \]  (37)

The current gain is (again, neglecting \( R_C \))

\[ a_i = \frac{i_C}{i_E} = -\frac{i_C}{-i_B - i_C} = \frac{\beta i_B}{i_B + \beta i_B} = \frac{\beta}{1 + \beta} \approx 1. \]  (38)
The input resistance is
\[ r_{\text{in}} = \frac{v_{\text{in}}}{i_E} = \frac{v_{\text{in}}}{-i_B - i_C} = \frac{-v_{\text{in}}}{\frac{i_B}{1 + \beta}} = \frac{-v_{\text{in}}}{i_B(1 + \beta)} = \frac{r_\pi}{(1 + \beta)} \]  
(39)

That last transformation follows from (35).

The output resistance is again found by inspection to be
\[ r_{\text{out}} = R_C. \]  
(40)

### 5.4 Common Collector Amplifier

The common collector amplifier is sometimes referred to as a voltage follower because the voltage gain is very close to one, meaning that the output voltage will follow the input voltage. The common collector amplifier is shown in Figure 19.

**FIGURE 19** A common-collector amplifier and its small-signal equivalent model.

The output voltage is
\[ v_{\text{out}} = R_E \beta i_B, \]  
(41)

Now the input current \( i_B \) is computed by writing Kirchhoff’s Voltage Law around the base loop.
\[ v_{\text{in}} - r_\pi i_B - R_E (i_B + i_C) = 0 \]  
(42)
\[ v_{\text{in}} = r_\pi i_B + R_E (1 + \beta) i_B \]  
(43)
\[ i_B = \frac{v_{\text{in}}}{r_\pi + R_E (1 + \beta)}. \]  
(44)
Now the voltage gain is
\[ a_v = \frac{v_{\text{out}}}{v_{\text{in}}} = \frac{R_E \beta i_B}{(r_\pi + R_E(1 + \beta))i_B} = \frac{R_E \beta}{r_\pi + R_E(1 + \beta)} \approx 1. \quad (45) \]

This last approximation assumes of course that \( \beta >> 1 \) and that \( R_E >> \frac{r_\pi}{\beta} = \frac{1}{g_m} \), both of which are most likely true.

The current gain will again be
\[ a_i = \frac{i_E}{i_B} = \beta + 1. \quad (46) \]

The current gain neglects \( R_E \).

From (44) the input resistance is
\[ r_{\text{in}} = \frac{v_{\text{in}}}{i_B} = r_\pi + R_E(1 + \beta) \approx \beta R_E. \quad (47) \]

Notice that the input resistance is roughly \( \beta \) times the emitter resistance. Thus, the input resistance of this stage is relatively high.

The output resistance is computed by connecting a voltage source \( v_{\text{test}} \) directly to the output and measuring the current \( i_{\text{test}} \) that flows in that source. In this case we assume that there is no input voltage, \( v_{\text{in}} = 0 \).

\[ r_{\text{out}} = \frac{v_{\text{test}}}{i_{\text{test}}} \], where
\[ i_{\text{test}} = \frac{v_{\text{test}}}{R_E} + \frac{v_{\text{test}}}{r_\pi} + \frac{\beta v_{\text{test}}}{r_\pi}. \quad (49) \]

Thus, the output resistance is
\[ r_{\text{out}} = \left( \frac{1}{R_E} + \frac{1}{r_\pi} + \frac{\beta}{r_\pi} \right)^{-1} = R_E \parallel r_\pi \parallel \frac{\beta}{r_\pi} \approx \frac{1}{g_m}. \quad (50) \]

This last approximation assumes that \( b >> 1 \), and that \( R_E >> r_\pi/\beta \).

### 5.5 Comparing the Configurations

The various configurations are compared in Table 1.

By inspecting this table you will see that ...

**Voltage Gain.** The common emitter and common base amplifiers both have high voltage gains, the common collector or voltage follower has a voltage gain of 1, and the degenerated common emitter is somewhere in between. You can also see that the common emitter inverts the signal whereas the others do not.

**Current Gain.** All amplifiers have current gain near that of the transistor except the common base or current follower amplifier, which has a gain of 1.
Input Resistance. The normal common emitter amplifier has an intermediate input resistance, where both the degenerated common emitter and the common collector amplifiers both have very high input resistances. The common base stage has a very low input impedance.

Output Resistance. All amplifiers have an output resistance set by the output resistor except for the common collector or voltage follower, which has a very low output impedance.

6 Conclusion

The basic operating mechanism of bipolar junction transistors is described and used to derive the small signal model of the transistor. It is then used to determine the characteristics of the most common amplifier configurations. In the process of doing this, we found several important quantities that characterize the small signal behavior of the transistor about its operating point. Those quantities were:

Current Gain. The current gain is the ratio of the collector current to the base current,

$$\beta = \frac{I_C}{I_B}.$$  (51)

It is a parameter of the transistor itself and typically has a value around 100.

Transconductance. The transconductance is the incremental gain that relates output current to input voltage,

$$g_m = \frac{dI_C}{V_{BE}} = \frac{i_C}{v_{BE}}.$$  (52)

It is derived in (21) to be

### Table 1

<table>
<thead>
<tr>
<th>Characteristics of the common amplifier configurations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common Emitter</strong></td>
</tr>
<tr>
<td>voltage gain</td>
</tr>
<tr>
<td>current gain</td>
</tr>
<tr>
<td>input resistance</td>
</tr>
<tr>
<td>output resistance</td>
</tr>
</tbody>
</table>
\[ g_m = \frac{I_C}{V_T} . \]  

(53)

**Base Resistance.** The base resistance is the incremental resistance across the base-emitter junction,

\[ r_\pi = \frac{dV_{BE}}{dI_B} \]  

(54)

It is derived in (18) to be

\[ r_\pi = \frac{V_T}{I_B} = \frac{\beta V_T}{I_C} \]  

(55)

**Relationships Between These Quantities.** These are related by

\[ r_\pi = \frac{\beta}{g_m} . \]  

(56)

Thus, if you know two, you can compute the third.

### 6.1 If You Have Questions

If you have questions about what you have just read, feel free to post them on the *Forum* section of *The Designer’s Guide Community* website. Do so by going to www.designers-guide.org/Forum.