Electronics Module 5D: Transistor Circuits

0. Review so far

In this Module we have considered switching and amplifying circuits, with an eye toward using our circuits to encode and transmit information. The essential problem is that we have information encoded in some time-dependent input signal $V_{in}(t)$, and want to use that signal to drive some load (a light bulb, a speaker, an alarm, a piece of machinery, or what have you). That load could be a significant distance away from the source, or may require far more power than is contained in the input signal. This is where switching and amplifying circuits come in.

In Module 5B we introduced the simplest such switching circuit, the electromechanical relay. It works well at what it does, but has a significant number of limitations. It can only be used for wholly switching a signal on and off, not amplification of an analog signal; relays are useful for telegraphy, but not for transmission of voice, music or images. They have moving parts, which means they have a finite lifetime and a limited switching rate. They also have a limited dynamic range, in that the input signal must be strong enough to power an electromagnet that latches the output switch closed.

What made the 20th century possible was the development of analog switching/amplifying devices. The first was the so-called triode vacuum tube, developed in 1915, in which a variable voltage applied to a terminal called the grid could be used to control a much larger current flowing between the plate (or anode) and cathode. But these too had some limiting factors, chiefly large power consumption and limited lifetime, although they persist to this day in certain niche applications.

The big breakthrough came in 1947 with the development of the transistor, whose properties we described in Module 5C. While there are several types of transistors, the most common is the bipolar junction transistor, or BJT. Others include field-effect transistors (FETs) and point-contact transistors. BJTs come in two varieties, NPN and PNP, and in construction are effectively two diodes (which themselves are PN junctions) connected back-to-back, as depicted at right, but their behavior is much richer than this description suggests. All BJTs have three terminals, the base (B), collector (C) and emitter (E). To tell the difference between the two, remember that the little emitter arrow on NPN transistors Never Points iN, and on PNP transistors it Points iN Proudly.

In all the following we shall assume NPN transistors; the behavior of PNP transistors is identical, but with the signs of voltages and directions of currents reversed.

The general picture is that currents $I_B$ and $I_C$ flow into the base and collector, respectively, and a current $I_E = I_C + I_B$ flows out of the emitter, as depicted at right. For a PNP transistor, simply reverse the directions of all three currents.
The operation of the BJT is governed by its biasing, the voltage levels of its three terminals. These are given by $V_B$, $V_C$, and $V_E$, and the voltage differences between the terminals are given by $V_{BE}$, $V_{CE}$, and $V_{CB}$. Also note that $V_{CB} = V_{CE} - V_{BE}$.

Therefore, there are exactly four independent parameters – two of the currents, and two of the voltage differences – which completely describe the behavior of the transistor! This is quite a bit more complicated than was the case for resistors, capacitors, inductors or diodes, which only had two terminals, and at most two independent parameters describing their behavior (i.e. $V$ and $I$). So analysis of transistors is going to be more complicated than it was for previous circuit elements.

Now let’s consider what happens when different voltages are applied to the three terminals.

When $V_B > V_C, V_E$, both “diodes” are in forward bias, and current flows freely, limited only by factors external to the transistor (the voltage sources, resistors, etc). This is known as saturation.

When $V_B < V_C, V_E$, both “diodes” are in reverse bias, and no current (except for very small leakage current) flows, and the transistor is said to be in cutoff.

When $V_E < V_B < V_C$, the B-E “diode” is in forward bias, and the C-B “diode” is in reverse bias. In this case, which is known as normal operating mode, the B-E circuit acts very much like a normal diode, with an in-flowing current $I_B$ flowing toward the emitter, and a base-emitter voltage drop $V_{BE}$ typical for a diode of the same material as the transistor (i.e. $V_{BE} \approx 0.7V$ for a Si transistor); a typical base-emitter characteristic is shown at right. However, a much larger current $I_C$ flows into the collector towards the emitter; in the region between cutoff and saturation, the current $I_C$ is related to $I_B$ via the collector-base characteristic depicted at below right. In the “linear regime” $I_C$ is closely given by:

$$I_C = \beta I_B$$

where $\beta$ is referred to as the gain. The emitter current is the sum of the base and collector currents:

$$I_E = I_C + I_B = (\beta+1)I_B = [(\beta+1)/\beta]I_C = (1/\alpha)I_C$$

where $\beta/(\beta+1) = \alpha$. For typical transistors, $\beta$ ranges from around 50 to several hundred, whereas $\alpha$ is always a number very slightly less than 1.

(It is possible to bias an NPN transistor such that $V_C < V_B < V_E$, known as inverse mode, which is very similar to normal operating mode but with the collector and emitter terminals reversed. However, the collector and emitter terminals have different levels of doping, so while this setup works, it generally does not work as well as the normal operating mode. So don’t do it!)

![BJT Base Characteristic Curve](image-url)
In the normal operating mode, the transistor maintains this collector current $I_C$ by enforcing a variable voltage difference $V_{CE}$ between the collector and emitter terminals; this voltage adjusts up and down to maintain $I_C$ at the appropriate value. In cutoff, $V_{CE}$ rises to a level large enough to prevent current flow, and in saturation $V_{CE}$ approaches zero. When $V_{CE}$ is somewhere between those extremes, $I_C = \beta I_B$ holds, and the transistor is said to be in the linear regime. This is depicted in the base-collector characteristic at right, showing the linear regime surrounded by the cutoff and saturation regimes to either side.

The symbiotic relationship between $I_C$ and $V_{CE}$ is shown more clearly in the collector-emitter characteristic, depicted again at right.

To understand this chart, we imagine setting $I_B$ to some particular value, which places us along one of the curves depicted. Now, the transistor would “like” to have $I_C = \beta I_B$, which corresponds to a horizontal line across the chart; indeed, the traces are relatively close to horizontal lines across most of the diagram. It does this by adjusting the level of $V_{CE}$ (the horizontal axis) such that $I_C$ is held to a relatively constant level irrespective of the external voltage applied to the collector-emitter circuit.

If the value of $I_B$ changes we move on to a different curve, which closely approximates a horizontal line at $I_C = \beta I_B$.

Saturation corresponds to the case where the collector-emitter circuit is unable to provide enough current to get $I_C = \beta I_B$. This corresponds to the grey region at left; here both $I_C$ and $V_{CE}$ are limited by the external elements (i.e. voltage sources, resistors) driving the collector-emitter current, with $I_C$ at some amount less than $\beta I_B$. The transistor “does its best” to maximize $I_C$ by reducing the collector-emitter voltage to a level approaching zero, but $I_C$ still falls short of $\beta I_B$, as it is limited by the external circuit elements. That was the earlier definition of saturation.

Cutoff corresponds to $I_B = 0$, and would correspond to a curve for that $I_B$ value; it would simply follow the x-axis, since $I_C = 0$. In this case, $V_{CE}$ adjusts itself up and down to a level which prevents any collector current $I_C$ from flowing, so that $I_C = 0$ irrespective of the applied collector voltage.

These three characteristic curves – the emitter-base characteristic, the collector-base characteristic, and the collector-emitter characteristic – together completely describe the behavior of a given transistor. These behaviors can be approximated by the Ebers-Moll mathematical model described in the previous Module, but in most cases the graphical representations are more useful.
I. Biasing: Switching vs. Amplifying

Now that we have characterized the behavior of transistors in terms of $I_B$, $I_C$, $I_E$, $V_{BE}$, $V_{CE}$ and $V_{CB}$ (only four of which are independent) using our three characteristic curves, and have defined the transistor’s operating regimes (linear, cutoff and saturation), we’re ready to think about placing them in circuits.

The first question that comes up is whether we want to use the transistor as a “switch” with on/off switching of some sort of “digital” signal, or as an “amplifier” which produces an amplified copy of the input signal.

When a transistor is operating in switch mode, we would like “on” to correspond to saturation (so that the collector-emitter connection acts like a “short” with $V_{CE}$ nearly zero and a free-flowing $I_C$ limited only by external circuit parameters), and “off” to correspond to cutoff (so that $V_{CE}$ rises to a level high enough to staunch any current flow in the collector-emitter circuit, which acts like an “open” circuit). This is important because it’s best for small changes in the input to not have any effect on the output; when the transistor is in cutoff or saturation a small change in the input will have no effect on the output, since it’s either clamped to zero (cutoff), or limited only by circuit elements external to the transistor (saturation). This is exactly what gives digital modes the ability to propagate information without error! To ensure that this is the case, we must see to it that the circuit built around the transistor places the transistor firmly into saturation when $I_B$ is “on,” and firmly into cutoff when $I_B$ is in its “off” state.

When a transistor is being used as an amplifier, though, we need to ensure that we remain solidly within the linear regime at all times. This is done by providing appropriate DC bias voltages so that the signal “starts off” near the center of the linear regime. This “starting off” point is referred to as the quiescent point, or Q-point, and we must build a circuit around the transistor such that the Q-point is near the center of the linear regime.

Our task now becomes the construction of a biasing circuit to set the Q-point at an appropriate location for the transistor being used.

II. Gain considerations

While the collector-base characteristic curve seems to indicate that the transistor has a gain of $\beta$, we found with the operational amplifier that we could set the gain to whatever value we wanted, in lieu of the “standard” open-loop gain $A_{OL}$, by building a circuit around it. We could also adjust the input and output impedance, to more efficiently couple to source and load. The same is true here.

We also must make ourselves aware of the fact that $\beta$ isn’t a particularly “good” gain – it varies from one transistor to another, and even with a single transistor, the ratio of $I_C/I_B$ fluctuates with temperature, with $V_{CE}$, and several other factors. So a circuit whose properties depend on $\beta$ is not a particularly good circuit; it’s much better to have a gain which depends mostly on external circuit parameters, which can be controlled much more tightly; such a circuit is known as a $\beta$-independent circuit. $\beta$ independence is a good thing!
III. The Ridiculously Simple Transistor Switch

Depicted at right is a simple transistor switch. In this circuit, toggling the switch on and off will toggle the light bulb on and off. The advantage here is that the current going through the switch is a factor of $\beta$ (typically 100-300) times smaller than the current going through the light bulb.

When the switch is open, $I_B$ is manifestly zero. This means the transistor is in cutoff mode, and hence $I_C = I_E = 0$. How does the transistor maintain $I_C = 0$? It must do so by dropping the entirety of $V_0$ across itself. Therefore, $V_{CE} = V_0$. We can identify this point on our characteristic curves; it is the point at which $I_B = I_C = 0$, and hence $V_{BE} = 0$ and $V_{CE} = V_0$.

Now, let us imagine closing the switch. A current begins flowing into the base, using the now-forward-biased base-emitter junction to get to ground. $V_{BE}$ rises to its “usual” value of, say, 0.7V (for a silicon transistor). Hence,

$$I_B = \frac{V_0 - V_{BE}}{R}$$

To keep the numbers simple, let’s set $V_0 = 10.7V$, $V_{BE} = 0.7V$, and $R = 100 \, k\Omega$. From this,

$$I_B = \frac{(10 \, V)}{(100 \, k\Omega)} = 100 \, \mu A$$

Now that a current $I_B = 100 \, \mu A$ is flowing into the base, a current $I_C$ will begin flowing into the collector. And, of course, a current $I_E = I_C + I_B$ will flow from the emitter to ground. (We may be tempted to say $I_C = \beta I_B$, but can’t do so just yet, because that’s only true in the linear regime – we might not be there!)

When the current $I_C$ flows through the light bulb (which we’ll model as a resistor with resistance $R_{bulb} = 1k\Omega$), the bulb lights up, and a voltage $V_{b} = I_C R_{bulb}$ is dropped across it. This current then flows through the transistor, where an additional $V_{CE}$ is dropped, and from thence to ground. Kirchhoff’s Voltage Law applied on the right-hand path gives us

$$V_0 - I_C R_{bulb} - V_{CE} = 0.$$

Hence $V_{CE} = V_0 - I_C R_{bulb}$.

To find $I_C$ and $V_{CE}$, let’s use a graphical solution, noting that $I_C$ and $V_{CE}$ are constrained by $R_{bulb}$.

For example, if $I_C = 0$, then $V_{CE} = V_0 = 10.7 \, V$.

If $V_{CE} = 0$, then $I_C = V_0/R_{bulb} = (10.7 \, V) / (1k\Omega) = 10.7 \, mA$.

Let’s assume that this particular transistor obeys the collector-emitter characteristic curve posted at right,
and I draw those two points on the diagram. For any intermediate values of $I_C$, $V_{CE}$ will vary linearly with $I_C$:

$$V_{CE}(I_C) = V_0 - I_C R_{bulb}$$

so I can connect the two points with a line, which gives the value of $I_C$ that must exist for any value of $V_{CE} = V_0 - V_{bulb}$. This line is known as the “DC load line” since it is formed by the constraint between $I_C$ and $V_{CE}$ imposed by the load resistor $R_{bulb}$.

Since $I_b = 100 \mu A$, our solution is the intersection of the $I_b = 100 \mu A$ curve with the DC load line, in analogy with the way we graphically solved diode circuits. This is at (my eyeball estimate)

$$V_{CE} = 2.3 \text{ V}$$
$$I_C = 8.5 \text{ mA}$$

Therefore,

$$V_{bulb} = V_0 - V_{CE} = 8.4 \text{ V}$$

Let’s see if that makes sense. Does this solution comport with Ohm’s Law for the bulb? Since $R = 1k\Omega$ and $V_{bulb} = 8.4 \text{ V}$, then we expect $I_{bulb} = 8.4 \text{ mA}$. That’s pretty close to the 8.5 mA I read off the graph.

How about the “gain?” The ratio of collector current to base current is

$$I_C / I_b = (8.5 \text{ mA}) / (100 \mu A) = 85.$$

Hold on a second – I never told you what $\beta$ was! Looking at the characteristic above, it appears that $\beta$ is right around 100, since most of the curves have their “flatline” with $I_C \approx 100 I_b$. However, in this case, $I_C$ is only 85 times $I_b$. This is because the transistor is just getting into the saturation regime, where $I_C = \beta I_b$ no longer necessarily holds. (This is why I warned you against doing that a few paragraphs back!)

This is a pretty nifty transistor switch.

**IV. The Common-Emitter Transistor Amplifier: Base Bias**

Now that we’ve seen how to use a transistor for a switch, let’s make an amplifier. The first thing we need to do is to set up the biasing circuit so that our Q-point (which we’ll find using the DC load line) is near the midpoint of the linear regime. Let’s imagine setting up the circuit shown at right.

The emitter is connected to ground, so $V_E = 0$.

A current $I_b$ flows from $+V_{CC} = 20 \text{ V}$ to the base, and from thence to the emitter. $V_{BE}$ takes on the usual value of ~0.7V or so, meaning that
\[ I_B = (V_{CC} - V_{BE}) / R_b \]

Now, if the transistor is in its linear regime (and we will soon choose \( R_c \) and \( R_b \) to guarantee this), then

\[ I_C = \beta I_B \]

This means that the resistor \( R_c \) drops a voltage

\[ V_{Re} = I_C R_c = \beta I_B R_c = \beta (V_{CC} - V_{BE}) R_c / R_b \]

Since \( V_{CC} = 20V \) and \( V_{BE} \sim 0.7V \), \( V_{CC} \gg V_{BE} \), so we can simply say

\[ V_{Re} = \beta V_{CC} R_c / R_b \]

And the transistor has a collector-emitter voltage

\[ V_{CE} = V_{CC} - V_{Re} \]

Now, the idea behind selecting \( R_b \) and \( R_c \) is to find values that will put us almost halfway between cutoff and saturation; this guarantees that we have the maximum amount of linear range on either side! At cutoff, \( V_{CE} = V_{CC} - 20V \). At saturation, \( V_{CE} = 0 \). Therefore, let’s try to set \( V_{CE} = (1/2) V_{CC} = 10V \):

\[ (1/2) V_{CC} = V_{CE} = \beta V_{CC} R_c / R_b \]

Solving this, and assuming \( \beta = 100 \), I find:

\[ R_c / R_b = 1/(2\beta) = 1/200 = 0.005 \]

Finally, let’s decide how much current I need in the collector circuit. Let’s say I need 10 mA. Since half of \( V_{CC} \) is dropped across \( V_{CE} \), the other half is dropped across \( R_c \): \( V_{Re} = (1/2) V_{CC} = 10V \). Then

\[ I_c / R_c = (1/2) V_{CC}/R_c = 10 \text{ mA} \]

\[ R_c = V_{Re} / I_c = (10 \text{ V})/(10 \text{ mA}) = 1 \text{ k}\Omega \]

Finally,

\[ R_b = 2*\beta*R_c = 200 R_c = 200 \text{ k}\Omega \]

I can now solve for the base current (remembering that \( V_{BE} \) is negligibly small so we can ignore it):

\[ I_B = V_{CC} / R_b = (20 \text{ V}) / (200 \text{ k}\Omega) = 100 \mu\text{A} \]

Since \( \beta = 100 \), if we’re in the linear regime, \( I_c \) should be 100 times larger than \( I_B \). And, indeed, it is!

On the collector-emitter I-V characteristic shown again at right (I added in the relevant DC load line for \( V_{CC} = 20V \) and \( R_c = 1 \text{ k}\Omega \)),

![Collector Current vs. Collector-Emitter Voltage](image-url)
we’re sitting at the point marked in yellow. It’s not exactly where I calculated it to be, because $I_C = \beta I_B$ isn’t exactly right (another reason to be very careful about using that relation!).

That point marked in yellow is the $Q$-point, (where the circuit sits in its steady state equilibrium) and it’s almost exactly midway between saturation (the grey region at left) and cutoff ($I_C = 0$). That’s exactly where we want to be!

V. The Common-Emitter Transistor Amplifier: bringing in the signal!

At this point, you’d be justified in scratching your head and wondering what the point of this whole exercise is. All I’ve done is set up a circuit that sits there and does nothing. A current $I_B$ comes into the base, and this causes a current $I_C$ to be drawn into the collector; $I_C$ is about $\beta = 100$ times larger than $I_B$, as expected. I’ve sized the resistors so that $V_{CE}$ is about half of the supply voltage $V_{CC}$, which puts us almost exactly halfway between saturation and cutoff, right smack in the middle of the linear regime.

But as we agreed way back in Module 5A, if there’s a signal that contains information, it has to modulate in time! Now we’ll think about adding that in.

Consider the circuit at right, which is essentially the same thing as before, except that I’ve coupled an input voltage $V_{in}(t)$ to the base (through a capacitor $C_1$) and an output voltage $V_{out}(t)$ from the collector (through a capacitor $C_2$). $V_{in}(t)$ is some small voltage (relative to ground) produced by a source (such as a motor controller, or the electromagnetic waves pulled from the air by an antenna), and $V_{out}(t)$ is a voltage (relative to the same ground) I wish to couple to a load (e.g. a speaker, a light, an alarm, a motor).

What’s the purpose of the capacitors? Well, in the steady-state situation, capacitors look like “opens” to DC. Any non-varying DC voltage coming from $V_{CC}$ cannot get through those capacitors, meaning that this DC voltage does not spill over into the input, or to the output. On the other hand, $V_{in}(t)$ is some rapidly-varying voltage; at its frequency, $C_1$ is sized such that $X_{C1} = 1/\omega C_1$ is small enough to permit current to pass through to the base. Therefore, $V_{in}(t)$ very easily goes through the capacitor, and adds to or detracts from $I_B$:

$$I_B(t) = I_{B0} + I_{in}(t)$$

Where $I_{B0} = 100 \, \mu A$ is the current calculated in the previous section, the current supplied by $V_{CC}$ through $R_b$. $I_{in}(t)$ is the additional current that gets through $C_1$ and heads into the base, coming from the input voltage $V_{in}(t)$. Let’s say we size $C_1$ such that $X_{C1}$ at the frequency of interest is $10 \, k\Omega$, and that the source $V_{in}(t)$ has a very low output impedance. Assuming $V_{BE} = 0$, the capacitor $C_1$ and the base-emitter junction provide a path from $V_{in}(t)$ to ground, with
\[ I_{in}(t) = V_{in}(t) / X_{C1} \]

If the source \( V_{in}(t) \) has a finite output impedance, then the current will generally be something less than that, and we should carefully select \( C_2 \) in order to efficiently couple \( V_{in}(t) \) into the base (remembering all that stuff about the maximum power transfer theorem, etc).

Let's say that \( V_{in}(t) \) is a sinusoidal voltage, varying from +100 mV to -100 mV (it’s a small signal!). Then \( I_{in}(t) \) varies from +10 \( \mu \)A to -10 \( \mu \)A, since \( X_{C1} = 10 \text{k}\Omega \).

This \( I_{in}(t) \) adds to \( I_{B0} \) to give us a time-varying base current \( I_{B}(t) \). In the limit that \( I_{in}(t) \) is small relative to \( I_{B0} \) (the “small-signal approximation,” which is valid here, as \( I_{B0} \) is ~10x larger than the signal), this can be modeled as a small perturbation: \( I_{B}(t) \) varies from 90 to 110 \( \mu \)A as a function of time.

This means that the collector current \( I_{C} \) is also a function of time:

\[ I_{C}(t) = \beta I_{B}(t) \]

If \( I_{B} \) varies from 90 to 110 \( \mu \)A, then \( I_{C} \) varies from 9 to 11 mA, since \( \beta = 100 \).

When \( I_{C} = 9 \text{mA} \), the voltage dropped across \( R_{bulb} \) is

\[ V_{bulb} = (9 \text{mA}) \times (1 \text{k}\Omega) = 9 \text{V} \]

When \( I_{C} = 11 \text{mA} \), the voltage dropped across \( R_{bulb} \) is

\[ V_{bulb} = (11 \text{mA}) \times (1 \text{k}\Omega) = 11 \text{V} \]

\( V_{CE} \), in turn, must vary up and down in order to satisfy Kirchhoff’s Voltage Law:

\[ V_{CE} = V_{CC} - V_{bulb} \]

So \( V_{CE} \) varies up and down from 9 to 11 V. Since \( V_E = 0 \) (emitter is connected to ground), \( V_C = V_{CE} \), and hence the point at which \( V_{out} \) couples varies up and down from 9 V to 11 V as a function of time.

\( C_2 \) serves a similar purpose to \( C_1 \) – it acts as an infinite capacitive reactance to DC frequencies, preventing any DC component of the voltage from escaping. At the frequency of the input signal, though, it has a finite reactance, so this signal couples through.

At the terminal \( V_{out} \), I measure a sinusoidal voltage which oscillates between +1 V and -1V (the amount left over after the \( V_{DC} = 10 \text{V} \), the Q-point value of \( V_{CE} \), is “eaten” by \( C_2 \)).

Finally, we may ask what is the gain of this circuit?

In terms of voltage, the amplitude of the input signal was 100 mV. The amplitude of the output signal is 1 V. So the voltage gain, \( A_V \), is

\[ A_V = V_{out} / V_{in} = 10. \]
That’s not huge, but by choosing different values for the R’s and C’s, I could in principle make it bigger.

However, there is also a current gain $A_i = \frac{I_{out}}{I_{in}}$. I cannot yet calculate this, since I need to know the impedances of the source and the load. However, through judicious choices of the R’s and C’s, I could optimally couple $V_{in}(t)$ to the base, and independently optimally couple $V_{out}(t)$ to whatever load I’m using. That means that, at a minimum, I’m going to couple $V_{in}(t)$ to $V_{out}(t)$ more efficiently than if I did not have a transistor amplifier. So my current gain is going to be reasonably high as well.

Finally, there is a power gain $A_p$, which (neglecting for phase differences between voltage and current) is the product of $A_V$ and $A_i$:

$$A_p = A_V \cdot A_i$$

Since both $A_V$ and $A_i$ are large, $A_p$ is going to be even higher.

The amplifier depicted above is known as a common-emitter amplifier, since the emitter terminal is shared between the input and output. Common-emitter amplifiers generally have reasonably high voltage gain and current gain, and thus very high power gain.

It is also possible to design amplifiers where the collector terminal is shared (known as a common-collector amplifier, or an emitter-follower); in this case the voltage gain is 1, but the current gain can be very high. One can also design an amplifier where the base is shared between the input and output (known as a common-base amplifier); this amplifier has a relatively high voltage gain, but the current gain is only 1.

The common-emitter amplifier, then, is the “All-American” well-rounded star. But emitter followers have some uses, particularly since they can have very high input impedances, so they can be efficiently coupled to “squishy” sources without fear of affecting them greatly.

VI. $\beta$-dependence, the bane of our existence

Our results from the previous two sections had one very significant (and undesirable) feature: the output depended critically on the value of $\beta$. Different transistors have different $\beta$ values, and even the same transistor can have a $\beta$ that changes with temperature, or $V_{CE}$.

It is imperative, therefore, to come up with a way to bias a transistor amplifier such that the gain is independent of $\beta$.

There are a number of ways to do this, including emitter feedback bias, collector feedback bias, and voltage divider bias.

Depicted at right is a voltage divider bias amplifier. In this circuit, $R_1$ and $R_2$ act as a voltage divider, holding the base at a constant voltage $V_B$. If $I_1$
and $I_2$ are much greater than $I_b$ (and the sizes of $R_1$ and $R_2$ can be chosen to guarantee this), then $I_1 \approx I_2$, and

$$I_1 = I_2 = \frac{V_{CC}}{R_1 + R_2}$$

And

$$V_B = I_2 R_2 = V_{CC} \cdot \frac{R_2}{R_1 + R_2}$$

Therefore, I choose $R_1$ and $R_2$ to hold $V_B$ at whatever voltage I'd like.

The base-emitter junction acts like a “normal diode” with $V_{BE} \approx 0.7V$, and, so $V_E$ is fixed at $V_B - V_{BE}$.

Since the emitter voltage $V_E$ is held fixed, the emitter current $I_E$ is also fixed by the value of the emitter resistor $R_E$:

$$I_E = \frac{V_E}{R_E} = \frac{V_B - V_{BE}}{R_E}$$

At this point, the collector current is also fixed since $I_C = I_E - I_B$. $I_B$ is so much smaller than $I_E$ that it cannot cause $I_C$ to vary very much!

The collector voltage $V_C$ is fixed as well, this time by the collector resistor $R_C$:

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

This also fixes $V_{CE}$, since

$$V_{CE} = V_{CC} - I_C R_C - I_E R_E$$

Now, here’s the cool part. Suppose that this transistor dies, and you have to replace it with another transistor that has a different value of $\beta$.

Hold on – I never mentioned anything about the value of $\beta$, and in fact the results don’t depend on it!

I put in a transistor with a different value of $\beta$. $I_C$ and $I_E$ cannot change; $I_E$ is fixed by the fact that $V_B$ (and thus $V_I$) are fixed by the $R_1/R_2$ voltage divider, and if $I_E$ cannot change, $I_C$ can only change a very tiny bit (since $I_B/I_C \sim 1/\beta$ is very small whether $\beta = 100$ or 500).

Therefore, any change in $\beta$ is compensated by a change in the base current $I_b$. $I_B$ goes up or down in order to maintain $I_C$, and $V_{CE}$, at the same values they had with the old transistor.

But since $I_b$ is very small (specifically, smaller than $I_1$ and $I_2$ so it has a negligible loading effect on the voltage divider), a change in $I_b$ has no significant effect on $V_B$. And thus it has no effect on $V_E$, and hence on $I_E$, and hence on $I_C$, and hence on $V_C$ or $V_{CE}$.

This means that the gain of the circuit is controlled wholly by the external resistors, and is nearly independent of the $\beta$ of the transistor!
Finally, I note that this circuit, once again, fails to couple any input or output. Well, that can be done just as we did last time, and is depicted at right: the input $V_{in}(t)$ and the output $V_{out}(t)$ are coupled to the base and collector, respectively, via capacitors which confine the DC voltages to the bias circuitry, but freely allow AC voltages in and out.

VII. More complicated amplifiers: from complexity comes simplicity

At this point, we’ve merely scratched the surface of what can be done with transistor amplifiers. Most real amplifiers use multiple transistors. They do this for many reasons, including

- higher gain
- better control over gain, via adjustable parameters such as potentiometers
- better input and output impedance characteristics
- gain which remains constant over a wide range of input amplitudes (this is known as linearity, and prevents distortion of a signal)
- gain which remains constant over a wide range of input frequencies: low-frequency signals tend to couple inefficiently to the amplifier, and high-frequency signals suffer from a whole host of problems (some of which were discussed in Module 3D). Frequency-dependent gain also leads to distortion of the input signal.
- higher power-dissipating capability by using multiple stages, or dividing the load among multiple transistors

Dealing with transistor amplifiers is, in general, messy. For that reason, it is usually wise to deal with pre-assembled amplifier circuits which may have complicated internal structures, but have properties that are amenable to much simpler forms of analysis.

The simplest form of amplifier is something that is described by a simple equation such as

$$V_{out} = AV_{in}$$

where $A$ is the voltage gain of the amplifier. (Similar statements could be made for current gain and power gain.) The reason for complex structure inside the amplifier is to ensure that for a wide range of inputs and outputs, the above relation is very nearly satisfied. That is, there is a very close linearity between $V_{out}$ and $V_{in}$ over a wide range of values of $V_{out}$ and $V_{in}$ over a wide (the wider the better) range of signal frequency.

This leads us to a rather astonishing conclusion: as our amplifier circuit becomes more and more complex on the inside, it becomes possible to describe its operation from the outside in a simpler and simpler way, as long as the complexity is added in a way that tends to cause the nonlinearities and other deviations to “cancel out.” Amplifiers are thus very amenable to “black box”-type reductionist analysis.
Indeed, most circuits in electronics are designed to work in relatively simple ways that belie their interior complexity. This is indeed the spiritual foundation of Thevenin’s and Norton’s theorems: complex circuits can be described by simple models. Many of the items we have worked with in the lab – the DC power supply, the oscilloscope and the function generator, for example – are excellent examples of this: their exterior operation and characteristics are remarkably simple, considering the horrendous amount of “stuff” in their interior.

The operational amplifier was a particularly simple amplifier to deal with, since the analysis only needed to know that the input impedances were huge, that $A_{OL}$ was huge, and that $V_{out} = A_{OL} (V_{+} - V_{-})$. Everything else flowed directly from those three statements. This makes it quite close to the “ideal amplifier model” we just discussed.

Well, here is what the IC 741 op amp looks like on the inside:

![IC 741 op amp diagram]

This device has 22 transistors and at least as many resistors. Other models of op amp have somewhat fewer, or somewhat more, transistors. While I pity the person who has to analyze the properties of such a circuit (although it’s mostly done with computer simulations nowadays), it’s amazing to realize that such a complex device can be understood with the very simple op amp model we used in Module 5B, and that such a wide variety of devices can be made from it.