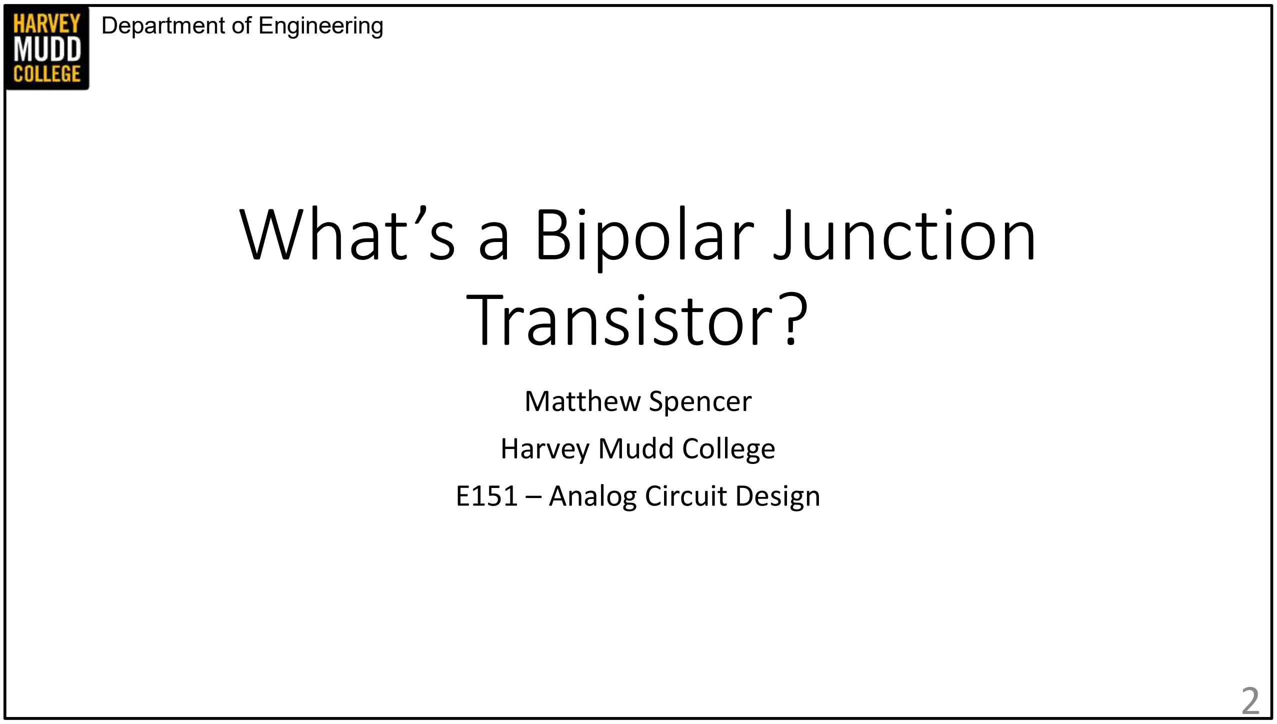
Bipolar Junction Transistors

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

In this video series we’re going to be learning about bipolar junction transistors, which are our first transistors of the course. All of modern electronics design is based on collections of transistors, so understanding the basics of transistors in these videos will be an important step towards building your own circuits. We’ll continue refining our BJT models in the next video set, and then we’ll be ready to dive right into amplifiers!

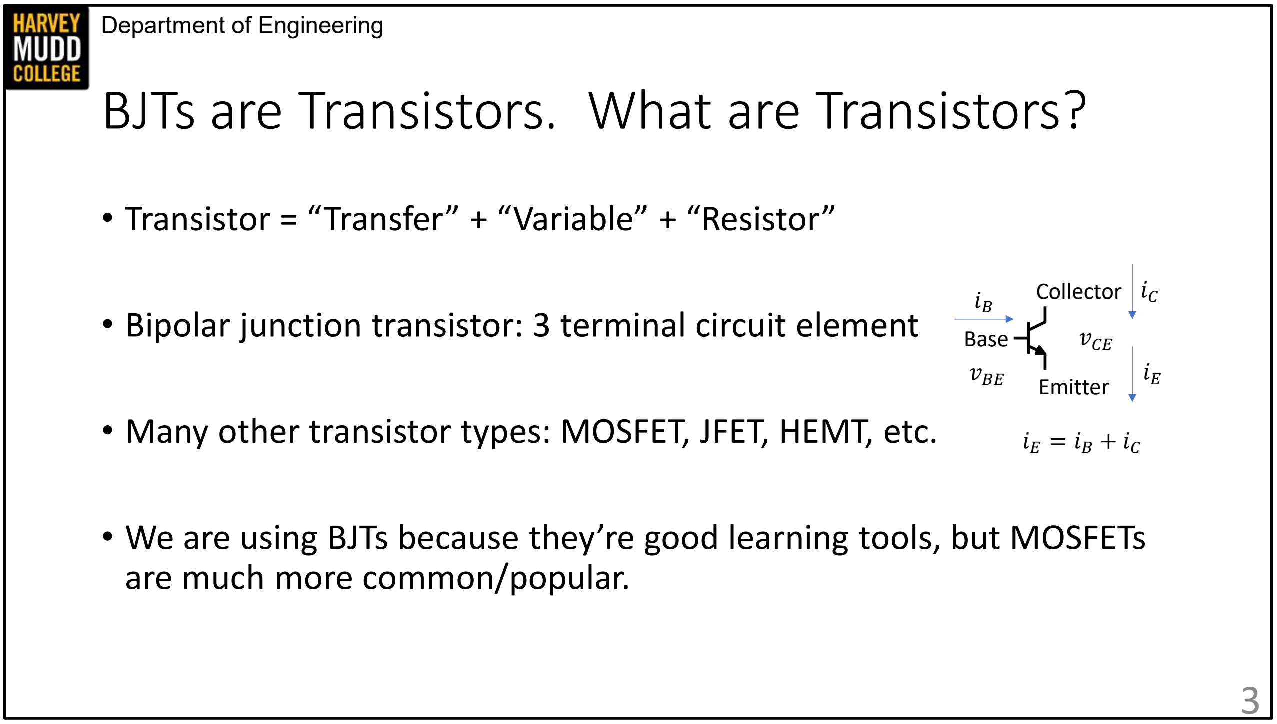


Department of Engineering

What’s a Bipolar Junction Transistor?

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

In this video we’re going to make a basic definition of bipolar junction transistors, often abbreviated BJT, and we’re going to talk about their construction.



Department of Engineering

BJTs are Transistors. What are Transistors?

- Transistor = "Transfer" + "Variable" + "Resistor"
- Bipolar junction transistor: 3 terminal circuit element
- Many other transistor types: MOSFET, JFET, HEMT, etc.
- We are using BJTs because they're good learning tools, but MOSFETs are much more common/popular.

$$i_E = i_B + i_C$$

Transistor is the most important word to latch onto in "bipolar junction transistor", because transistors of various types are ubiquitous in modern electronics. All transistors are electrically controlled switches, that let a voltage or a current affect the voltage or current somewhere else. That behavior is actually baked into the name transistor, which is a terrible portmanteau of transfer and varistor, where varistor is itself a terrible portmanteau of transfer and resistor. Taken together, this means a transistor causes a resistance somewhere to change in response to a signal elsewhere.

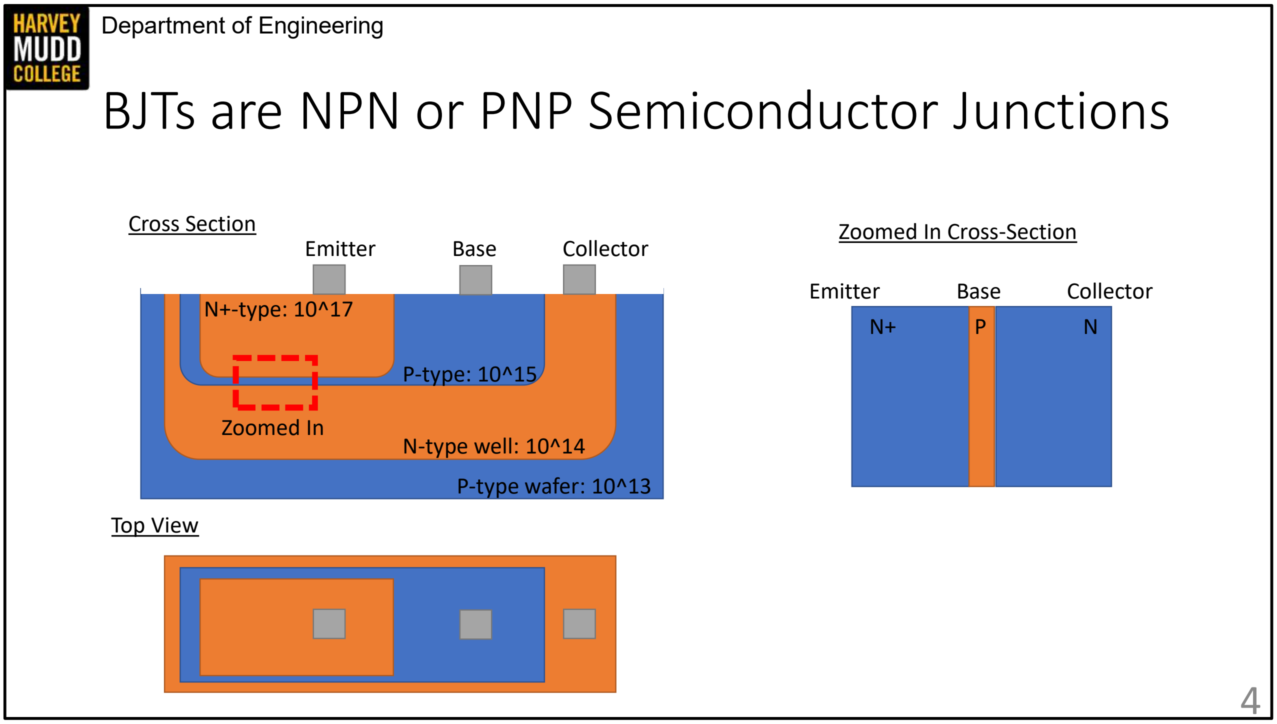
We can draw a circuit model to make this more concrete. Bipolar junction transistors are three terminal circuits, and a BJT circuit symbol is pictured on the right of this slide. The terminals are called the emitter, the collector and the base. There are several voltages and currents we use to monitor the behavior of BJTs, including the base-emitter voltage vBE, the collector-emitter voltage vCE, the base current iB, the collector current iC and the emitter current iE. Charge can't build up inside of the BJT, so iE has to be equal to iB+iC by Kirchoff's Current Law. It's worth noting that there are two types of transistors and this variety is called an NPN transistor. We'll talk about the other kind, called PNP transistors, in future videos.

As I mentioned before, there are lots of types of transistors, each of which has slightly different behaviors. These include metal-oxide-semiconductor field effect transistors, or

MOSFETs, junction field effect transistors, or JFETs, high electron mobility transistors or HEMTs and others. BJTs are somewhat odd ducks in this transistor menagerie because they aren’t used often in modern chips. However, BJTs are great learning tools and they let us do math in the lab easily, so we’re going to focus on them as we build up the basics of circuit analysis. Out in the real world you’re much more likely to use MOSFETs, which are the most common type of transistor by a wide margin.

Department of Engineering

BJTs are NPN or PNP Semiconductor Junctions



The other basic information that’s worth knowing about BJTs is how their built. BJTs are intricate combinations of PN junctions, specifically they are NPN junctions or PNP junctions. You make them by creating successively heavier doped regions in semiconductors as pictured on the left. The N-well is a construction that effectively converts the substrate to N-type in a limited region on the chip, which lets you make NPN transistors by implanting P-type and then N-type dopants. It’s worth noting that the emitter region is very heavily doped, so it is denoted as N+ doping instead of standard N-type. Note also that the base region is very thin, which is an important part of BJT design.

You can see a top-view on the bottom of the slide, which is the view a circuit designer would interact with while making a chip. In general, circuit designers are in charge of the X and Y design of chips while device designers have exclusive privilege over the Z dimension. You can see there’s a lot of wasted space in this particular BJT layout, so it might be possible to make a more condensed BJT by squeeze each of these areas left and right.

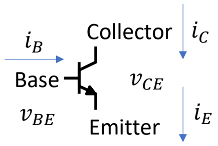
The region of the BJT surrounded by dashed lines is the most important region, so we have a zoomed in version of that region on the right side. Note that it has been rotated by 90 degrees. We’re going to zoom in further on a future slide, so I thought it was important to highlight how this picture related to the construction of the transistor.

Summary

- Bipolar junction transistors are a type of transistor, which means they change a current in response to a signal on another terminal.

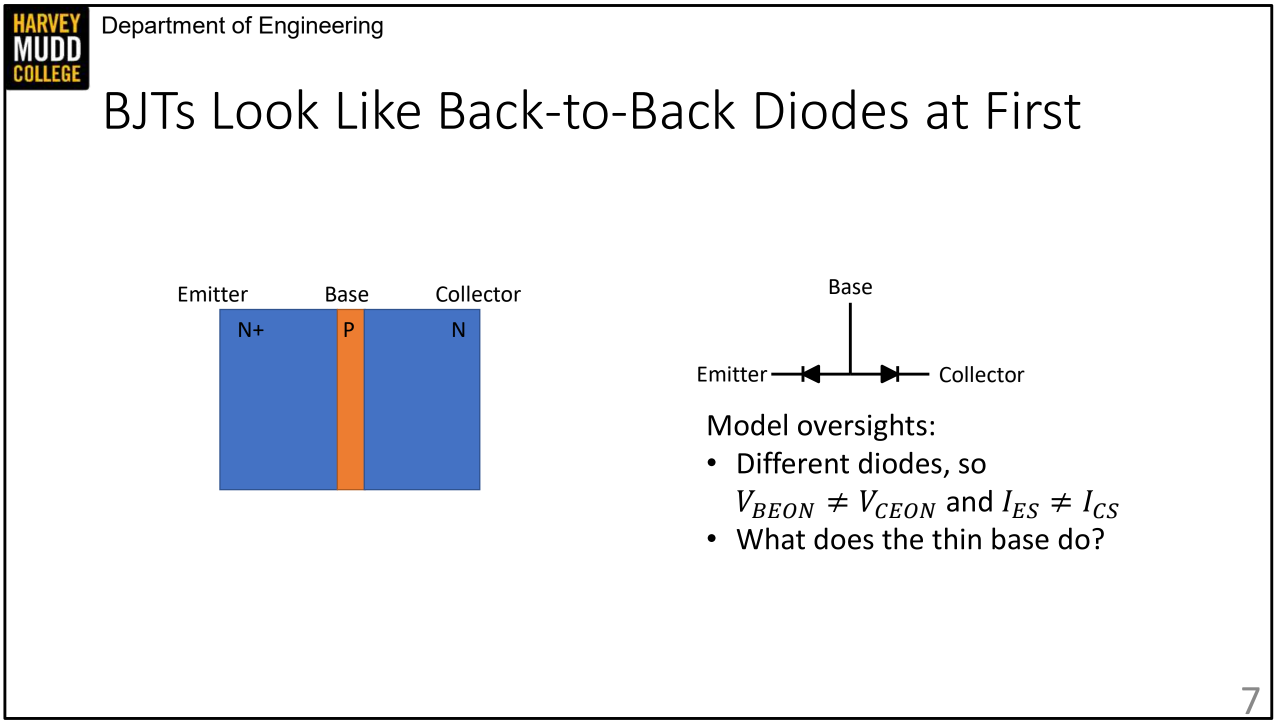
- Transistors have three terminals: emitter, collector and base

- Transistors are NPN/PNP junctions with specific doping levels and widths



$$i_E = i_B + i_C$$

In this video we’re going to illustrate how carriers move around in BJTs, which will give us a qualitative sense of how transistors achieve the magic of transferring electrical signals from one port to another.



Department of Engineering

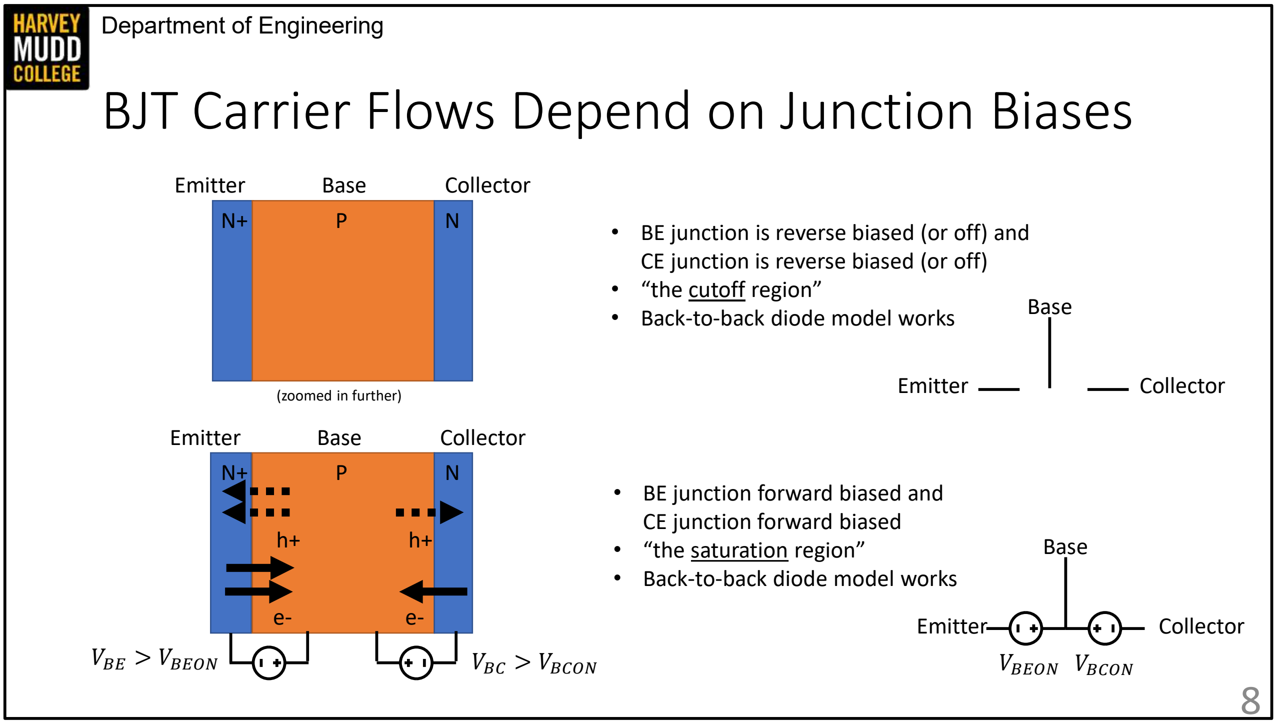
BJTs Look Like Back-to-Back Diodes at First

Model oversights:

- Different diodes, so $V_{BEON} \neq V_{CEON}$ and $I_{ES} \neq I_{CS}$
- What does the thin base do?

If you look at a BJT, your first instinct might be to model it as two back-to-back diodes. After all, there are two PN junctions that make up a BJT. I'd like you to take a minute to write down any details you notice about BJTs that make it different than two back-to-back diodes.

OK, first, it's worth noting that the doping is different on the left PN+ junction and the right PN junction. That means the two diodes we include in this model would need to have slightly different IS and VON values, which we call VBEON and IES for the base-emitter junction, and VCEON and ICS for the base-collector junction. But that seems like a reasonable modification that makes the back-to-back diode model better and not worse. The more interesting detail that this model overlooks is the very thin base region. That suggests the two diodes might be able to interact. We'll see that is the case under specific bias conditions.



HARVEY MUDD COLLEGE Department of Engineering

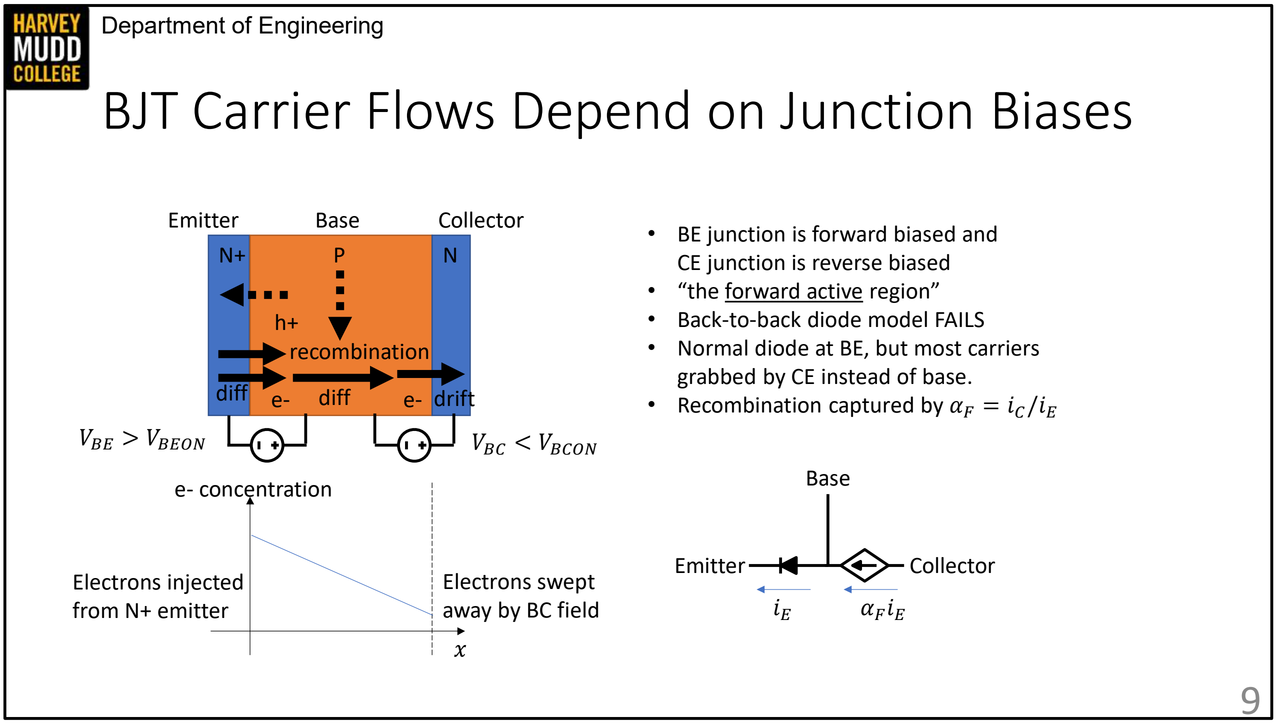
BJT Carrier Flows Depend on Junction Biases

- BE junction is reverse biased (or off) and CE junction is reverse biased (or off)
- “the cutoff region”
- Back-to-back diode model works

- BE junction forward biased and CE junction forward biased
- “the saturation region”
- Back-to-back diode model works

The first bias condition we’re going to examine is one in which both PN junctions are either unbiased or reverse biased. As you might expect, that means little current flows in either PN junction. We call this bias condition “cutoff”, and we model it with two open circuits in both the large and the small signal regime. Notice, by the way, that we’ve super-zoomed in on the base in this picture because most of the exciting behavior in BJTs happens in the base.

CLICK The second bias condition we’re going to examine is one in which both the base-emitter and the base-collector juntions are forward biased. I’ve indicated that in the diagram with external voltage sources attached between the base and the other terminals. This results in current flowing out of the base into both the emitter and the collector just like two forward biased diodes. This bias condition is called saturation, and we model it with two voltage sources in the large signal region, which is just like the back-to-back diode model. Notice that we’re using different notation for the electron and hole flows. This notation is a bit overloaded, but the main thing to note is the double arrow used for the injection of electrons from the N+ region. The size of a carrier flow is proportional to the doping of the region injecting the carriers, so the N+ region injects the lion’s share of the carriers in the BE junction, while the holes are more prevalent in the BC junction. I used the double arrow to show where most of the carriers are coming from.



Department of Engineering

BJT Carrier Flows Depend on Junction Biases

- BE junction is forward biased and CE junction is reverse biased
- “the forward active region”
- Back-to-back diode model FAILS
- Normal diode at BE, but most carriers grabbed by CE instead of base.
- Recombination captured by $\alpha_F = i_C/i_E$

9

The next bias condition we’re going to look at is one where the base-emitter junction is forward biased while the base-collector junction is reverse biased. You can see that the base-collector bias has reversed direction which will mean that the collector is at a higher voltage than the base, and the base is at a higher voltage than the emitter.

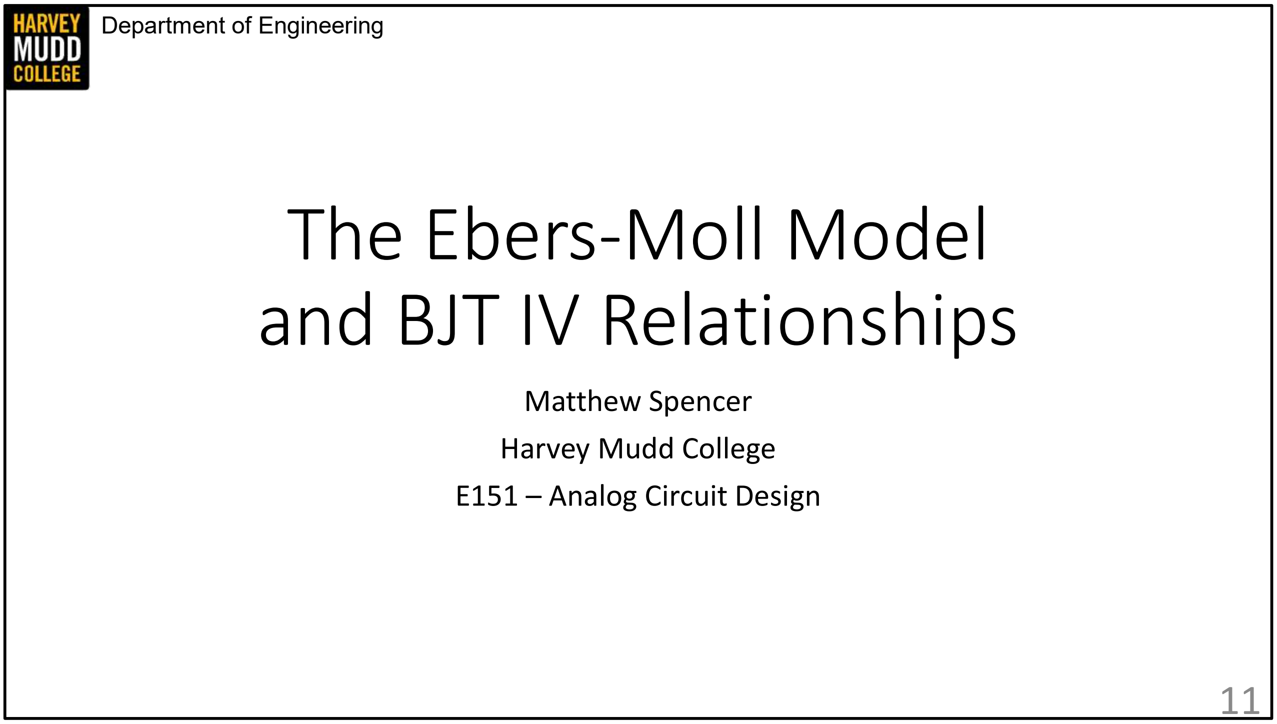
CLICK the emitter side of this BJT works like a normal diode. Holes are injected from the base to the emitter and many more electrons are injected from the heavily doped emitter into the base. Because the junction is forward biased, these carriers are primarily driven by diffusion.

CLICK The collector side is reverse biased, so any electrons that stumble into this depletion region will drift to the collector. Note that this creates a current travelling from left-to-right, out of the collector because electrons are negative charge carriers. Note also that this drift behavior is relatively insensitive to vBC. As long as the junction stays reverse biased, then this carrier flow will be dominated by drift.

CLICK These two carrier flows mean that the concentration of electrons is going to be very different at the emitter junction, where many electrons are being injected, and the collector, where they are being swept away by drift. That means there will be a gradient of minority charge carriers across the base, which results in diffusion from the emitter to the collector. Put another way, the Base-emitter diode injects a lot of electrons into the base, but many of those electrons get stolen by the collector instead of passing out through the base.

CLICK Not all of the minority carriers injected into the base make it to the collector, some balance the small hole current that is flowing into the emitter, and some recombine with majority carriers. Those two effects explain why BJTs are designed with very asymmetric doping profiles and very thin bases, the thin base means carriers have less chance to recombine and the asymmetric doping ensures there are more than enough carriers to cancel out the flow of holes from the base. We describe how effective a BJT is at conveying electrons from emitter to the collector using the ratio of iC to iE, which is called alpha and describes how many electrons are lost in the base. In well designed BJTs, alpha is often close to 99.7%.

CLICK These complicated effects can be summarized with this large signal model, where the emitter current is determined by the base-emitter voltage, but most of that current is provided by a current source between the collector and a base. The portion of current that is drawn from the base is (1-alpha)*iE. Notice that current runs from the collector the emitter in this model because current flows the opposite direction from electrons. Notice also that the emitter emits electrons in an NPN device, so it sinks current. The name can be confusing, it doesn't emit current in an NPN device.

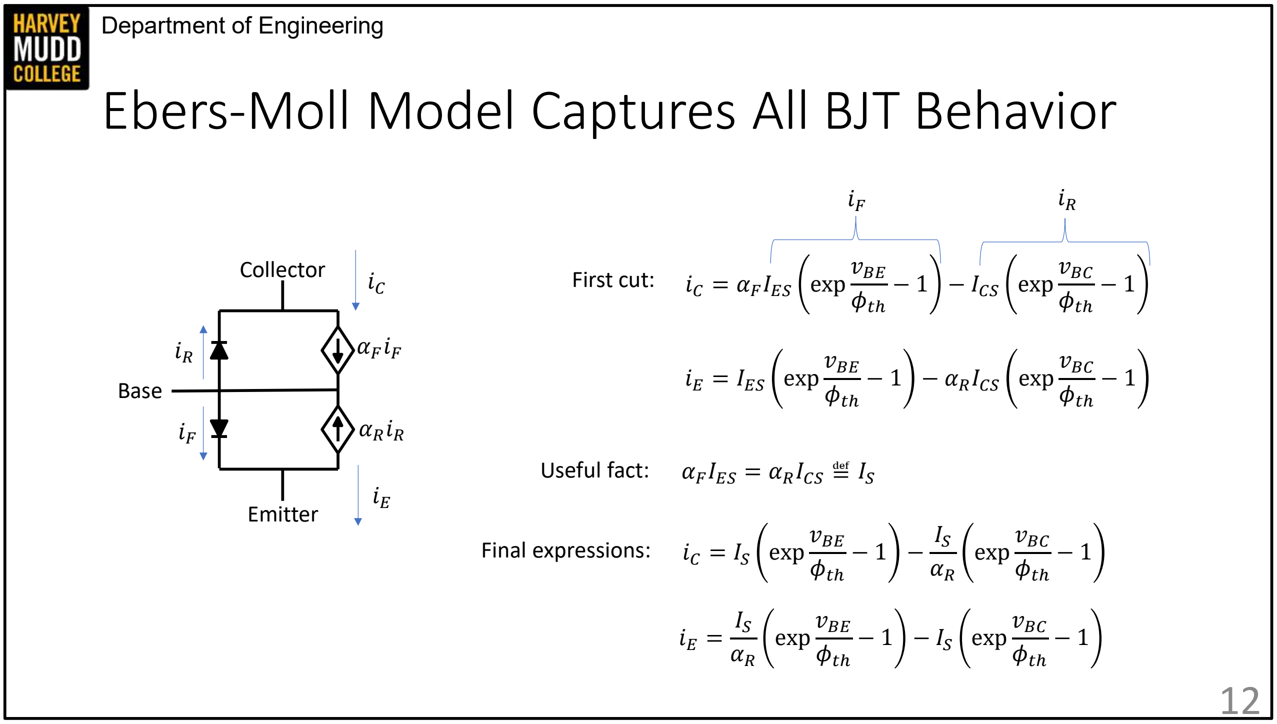


Department of Engineering

The Ebers-Moll Model and BJT IV Relationships

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

In this video we are going to stitch together our models into one continuous model that describes the operation of a BJT in all bias regions. The model is called the Ebers-Moll model. This is a good tool to have because we can take derivatives of it and graph it, but it’s too complicated for use in circuits. We’ll derive circuit-usable models in a later video.



Department of Engineering

Ebers-Moll Model Captures All BJT Behavior

First cut:

$$i_C = \overbrace{\alpha_F I_{ES}\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right)}^{i_F} - \overbrace{I_{CS}\left(\exp\frac{v_{BC}}{\phi_{th}} - 1\right)}^{i_R}$$

$$i_E = I_{ES}\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right) - \alpha_R I_{CS}\left(\exp\frac{v_{BC}}{\phi_{th}} - 1\right)$$

Useful fact: $\alpha_F I_{ES} = \alpha_R I_{CS} \overset{\text{def}}{=} I_S$

Final expressions:

$$i_C = I_S\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right) - \frac{I_S}{\alpha_R}\left(\exp\frac{v_{BC}}{\phi_{th}} - 1\right)$$

$$i_E = \frac{I_S}{\alpha_R}\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right) - I_S\left(\exp\frac{v_{BC}}{\phi_{th}} - 1\right)$$

12

We can start to make this model by combining the models from our cutoff, saturation, and forward active regions. The back-to-back diodes work for cutoff and saturation, and we add in this current source to represent forward active. However, this looks a little lopsided; the model contains an asymmetry where the forward current can trigger a current source, but the reverse current can't.

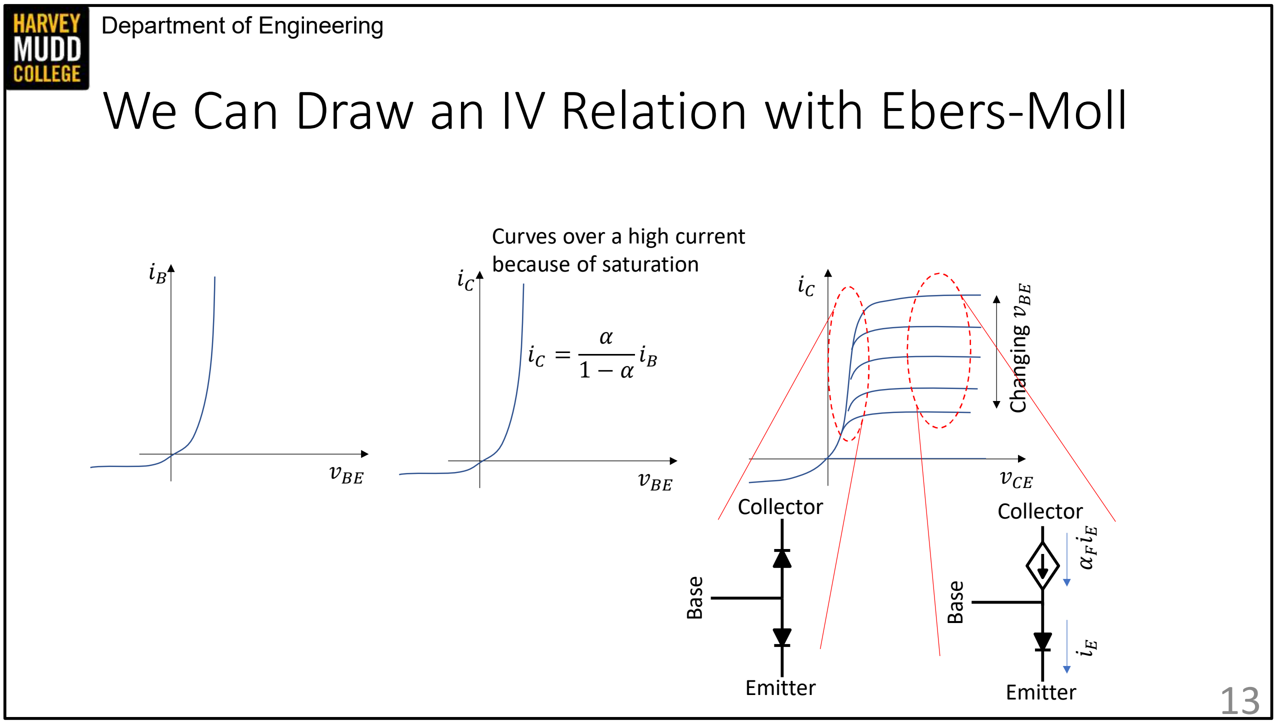
CLICK That lopsidedness represents a hole in our model. The BJT is a somewhat symmetric device, so the forward active behavior we described has a reverse active analog, where the collector injects carriers into the base and the emitter absorbs them. Reverse active conduction works very poorly because of the reversed doping profile, and it is difficult to induce in the lab, even by accident, so we rarely consider it in circuit models. However, it helps us to paint a complete picture of carrier flows in the BJT. Now the forward diode and current source describe forward active behavior, the reverse diode and current source describe reverse active behavior, and the two diodes can describe saturation and cutoff because the current sources are linearly dependent on the diode behavior. So if there's no diode current then there's no current source current, and if current flows in both diodes then the current sources both conduct with the effect of changing the IS coefficients in the diode relationships.

CLICK We can use this circuit diagram to make equations for the collector and emitter current. I’ve substituted expressions in for the forward and reverse current in these equations as indicated at the top of the slide. So we can see that the collector current is alpha times iF minus iR, and the emitter current is iF minus alphaR, a new parameter that describes carrier recombination in reverse active, times iR.

CLICK Because the recombination rate in the base is constant, we know that alphaF times IES has to be the same as alphaR times ICS. We define that quantity as IS, which let sus simplify our notation for the BJT.

CLICK Using that useful fact, we can rewrite our expressions in terms of IS. These final equations are called the Ebers-Moll model. There are lots of things to notice about this, but one is that both the collector and the emitter voltage depend on both vCE and vBE. That means the IV curves are likely to be somewhat complicated.

… note that current source is TRULY LINEAR in diode current, so you get an actual linear gain in small signal … next video



Department of Engineering

We Can Draw an IV Relation with Ebers-Moll

Curves over a high current because of saturation

$i_C = \frac{\alpha}{1-\alpha} i_B$

We'll tackle that complexity by rendering the IV curves in a few different ways. On the left we see a picture of base current vs. base-emitter voltage, and this just behaves like a diode. That matches our Ebers-Moll model, which has a diode connected between the base and the emitter. Changing the collector-emitter voltage won't affect this much.

CLICK The collector current is linear in the base current, so it is also an exponential when plotted against vBE. This curve is the same shape as the base current, but it has been multiplied by a large coefficient because alpha of the emitter current comes from the collector while (1-alpha) of it comes from the base. At very high current levels the iC curve will bend to horizontal lines as the BJT saturates. However, it's difficult to represent that behavior on a plot with vBE on the x-axis because they change in iC is so rapid.

CLICK As a result, it's common to plot iC vs. vCE parametrized in vBE.
CLICK The left side of this plot shows an exponential increase because it is the saturation region of the BJT bias plane. In this region vBE is large and vCE is small, as indicated by the low x values.
CLICK The right side of this plot shows the current remaining stable as vCE changes, which is consistent with our forward active model. The current source keeps the same current through it even as vCE is changed. The level at which the current flattens out is determined by vBE because vBE sets the base-emitter current. This is actually a beautiful part of the

graph because it’s the key to transistor operation: changing the base will change the collector current, but changing the collector won’t, so we’ve transferred the signal on the base into current at the collector. We will almost always seek to operate our transistors in the forward active region when doing analog design.

… base width modulation and early voltage next time

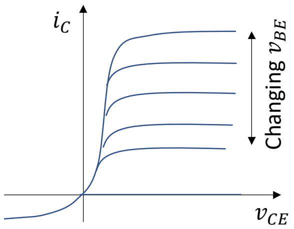
Summary

- The Ebers-Moll model is a continuous model that represents all regions of BJT operation.

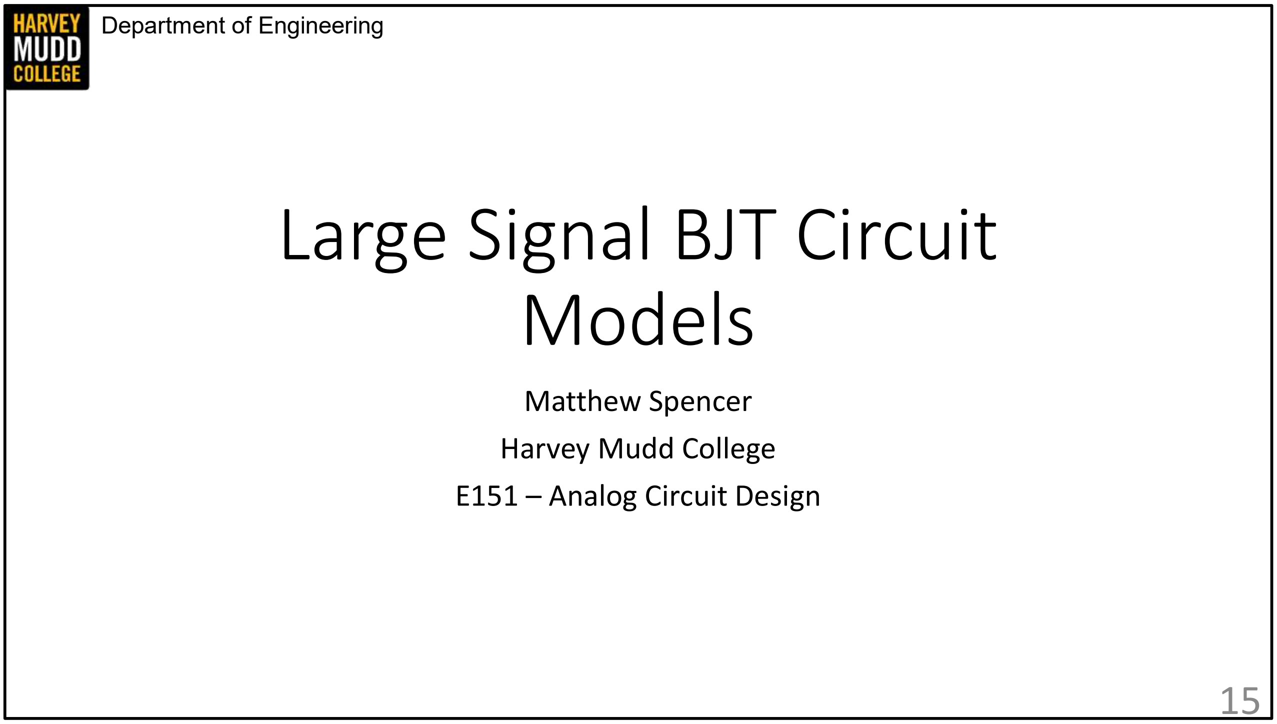
$$i_C = I_S\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right) - \frac{I_S}{\alpha_R}\left(\exp\frac{v_{BC}}{\phi_{th}} - 1\right)$$

$$i_E = \frac{I_S}{\alpha_R}\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right) - I_S\left(\exp\frac{v_{BC}}{\phi_{th}} - 1\right)$$

- You can use the Ebers-Moll model to draw BJT IV curves, which are usually $v_{CE} - i_C$ parametrized in v_{BE}



- We'll add one detail to our model then use it to calculate small signal models.



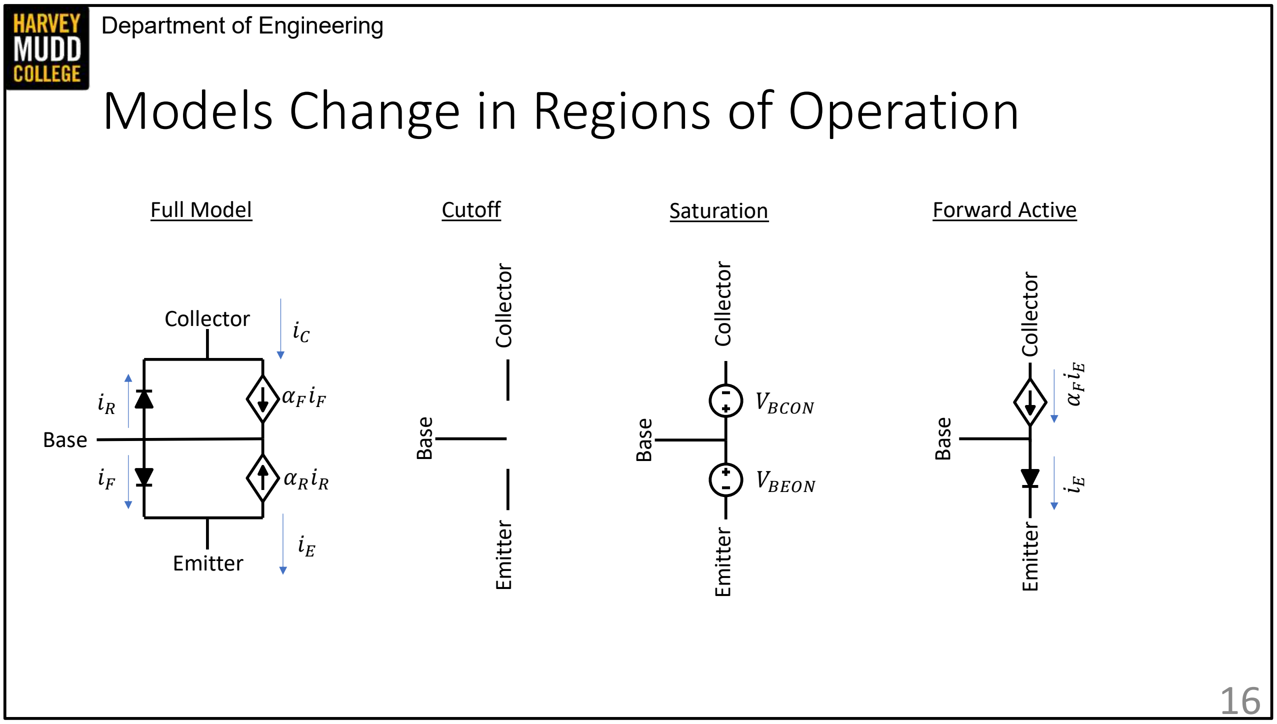
Large Signal BJT Circuit Models

Matthew Spencer

Harvey Mudd College

E151 – Analog Circuit Design

In this video we’re going to simplify the Ebers-Moll model so that we can use it to do large signal analysis.



Department of Engineering

Models Change in Regions of Operation

Fortunately, you've seen these simplifications before. I'll rederive each of them quickly though:

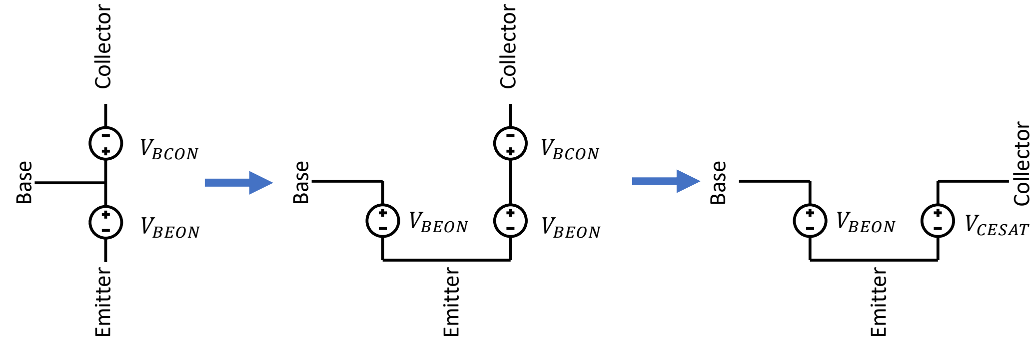
(1) If we use a switch-voltage-source model for the diodes in the Ebers-Moll model, then the iR and iF currents are zero in cutoff, which also zeros the current sources resultin in no current flowing. That means the transistor can be well represented by open circuits.

(2) If we keep using the switch-voltage-source model, then currents in iR and iF are both big in saturation. That will change the shape of our diode curves, but who cares if the diode curves are almost vertical anyway. We replace the diodes and current sources with voltage sources that have values of VBEON and VBCON.

(3) Finally, in forward active the collector voltage is higher than the base voltage, so iR is small, while the base is above the emitter, so iF is large. That means iR is near zero, removing the reverse diode and current source. We're left with only the forward components of the Ebers-Moll model.

I need to add one pedantic aside about the bias conditions that lead to the forward active condition: the collector doesn't need be above the emitter in forward active. The collector can be reverse biased, such that VBC is less than VBCON, even when the collector dips below the base. That relation is a little bit awkward to picture in the circuit though because it requires us to reference the collector voltage against the base voltage, while the base voltage is usually referenced against the emitter voltage. It might be nice to simplify our model for saturation so that we can reference both the collector and the base against the emitter.

Saturation is Easy to Identify Using V_{CESAT}



Where: $V_{CESAT} = V_{BEON} - V_{BCON}$

Saturated if: $v_{BE} > V_{BEON}$ and $v_{BC} > V_{BCON}$

$v_{BE} - v_{CE} > V_{BEON} - V_{CESAT}$

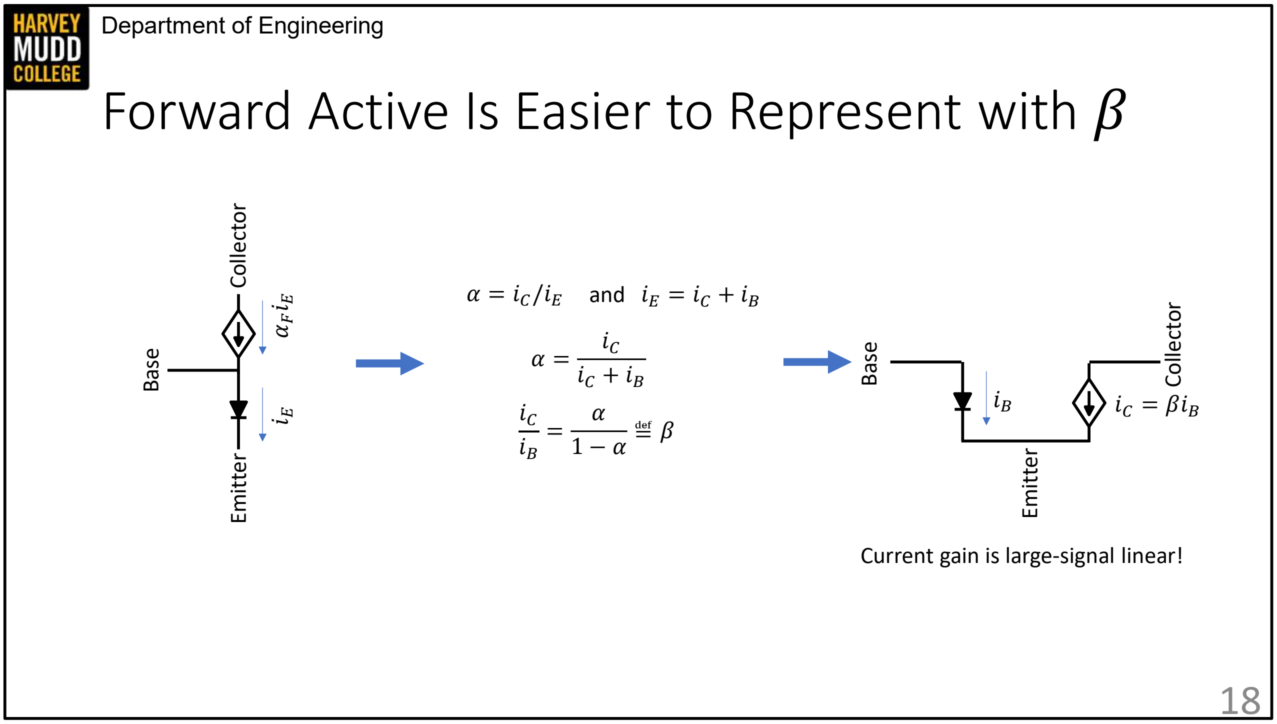
$v_{CE} < V_{CESAT}$

We can make that simplification by performing a series of circuit transformations. First, we create an imaginary node that has the same potential as the base by making a copy of VBEON that is also attached to the emitter. If voltage sources can pass infinite current, then it doesn't matter whether we attach the collector the base copy of VBEON, or this new copy of VBEON. We choose to hook it to the new version.

Then we can take a final step of merging VBEON and VBCON in series to create a new, smaller voltage source called VCESAT.

CLICK VCESAT can be useful to us by helping us redefine the conditions that we use to identify saturation. The first condition that we use to identify saturation is that the base-emitter junction is forward biased, and that one is easy enough to reason about. The second is that the base-collector junction is forward biased, which is trickier. We can represent vBC as a combination of vBE and vCE, and we can similarly replace VBCON using our definition of VCESAT. Finally, we notice that vBE is equal to VBEON in satuation, which let us eliminate vBE from both sides. Dividing by negative one gets us to our final expression, which is that we're in saturation if vCE ever falls below vCESAT. This is handy, we can think of both the base and emitter as being higher than the emitter, and we know how high each voltage needs to be: more than VBEON for the base and less than VCESAT for the collector.

I refer to this model as a U-model, which isn't its official name. It's technically the beginning of something called a hybrid-pi model, and we'll see more about that later.



Department of Engineering

Forward Active Is Easier to Represent with β

$\alpha = i_C / i_E$ and $i_E = i_C + i_B$

$$\alpha = \frac{i_C}{i_C + i_B}$$

$$\frac{i_C}{i_B} = \frac{\alpha}{1-\alpha} \overset{\text{def}}{=} \beta$$

$i_C = \beta i_B$

Current gain is large-signal linear!

We can make another useful U-model by imagining that we have current flowing directly from the collector to the emitter. We describe this transformation using a little bit of algebra in the middle of the slide, though it's worth noting that you can also derive this using a series of circuit transformations like the previous slide.

That said, we start by noting that alpha is defined as the ratio of iC and iE, and that we can express iE in terms of iC and iB. Substituting the second relation into the first shows us that alpha is iC over iC+iB. You can rearrange that expression through some modest algebra to find iC over iB, which is equal to alpha over 1-alpha. This number is important enough that we define it as beta. Technically this is the forward beta, but we use revrese beta so rarely that I've decided to drop the subscript on this parameter. Beta tends to range from 100-300, so it's quite large in general.

We can use this relation to draw a U-model, where both base current and collector current flow into the emitter. We achieve this by calculating the collector current as beta times the base current. This model will give us exactly the same terminal currents as the model on the left, but it reveals something really important: the BJT has a current gain that is linear, even in a large signal model. This is remarkable in a transistor, which is a wildly non-linear (often exponential) device. That relationship will help us in small signal modeling.

Summary

- Large signal cutoff, saturation and forward active models are special cases of Ebers-Moll with some diodes on or off.

- The U-model of saturation shows $V_{CE} = V_{CESAT}$ and you can identify saturation with $v_{BE} > V_{BEON}$ and $v_{CE} < V_{CESAT}$.

- The U-model of forward active shows $i_C = \beta i_B$, which is a truly linear current gain relationship.