Department of Engineering

Bipolar Junction Transistors in Detail

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

1

In this video series we're going to fill in a few more details about the bipolar junction transistor, which will allow us to make a small signal model of the BJT. Once we have a small signal model, we can use it to build an amplifier, our first analog circuit, which is our next big goal.

Department of Engineering

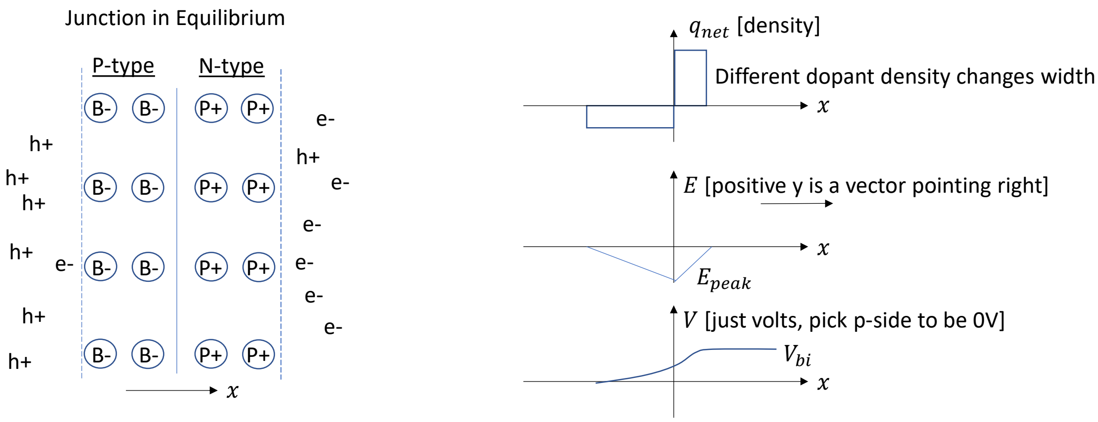
Base Width Modulation and the Early Voltage

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

2

In this video we're going to look at a detail of PN junction operation, which is that the width of the depletion region varies with the bias applied to a junction. This changing depletion width affects the current in bipolar junction transistors, so we'll modify our IV curves and our models to accommodate it.

Junction Depletion Width Changes with Bias



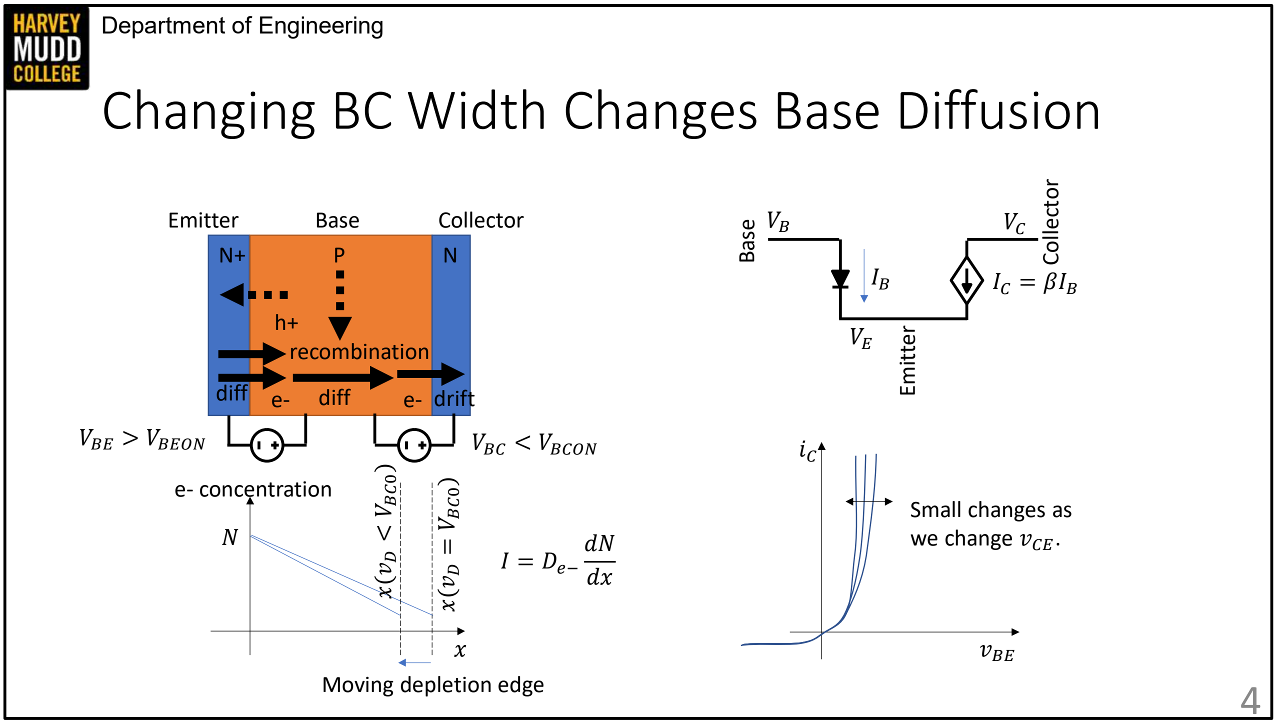
$$x_n = \sqrt{\frac{2\epsilon}{q}\frac{N_a}{N_d}\frac{1}{N_a + N_d}(V_{BI} - v_D)} \qquad x_p = \sqrt{\frac{2\epsilon}{q}\frac{N_d}{N_a}\frac{1}{N_a + N_d}(V_{BI} - v_D)}$$

We talked about the width of depletion regions previously and mentioned that the charge in the P region had to balance the charge in the N region. That means doping densities will change the relative width of the P and the N region. A heavier doped region will have more charge per unit of width, so it doesn’t need to be as big to balance the charge on the lighter doped side.

However, we haven’t yet spoken about the total width of the depletion region. We can use constraints that are similar to charge balance that we derive from the E field and voltage plots to calculate the width of the total depletion region. I’ve skipped straight past the derivation to the results on the bottom of the slide. The size of the depletion region in the P and N parts of the diode are determined by Na, the number of acceptors (dopants in the P region), Nd, the number of donors (dopants in the N region), the permittivity of the material, the elemental charge, the built-in voltage and the applied voltage.

Looking at these formulas, we can see that forward biases cause the size of the depletion region to shrink while reverse biases cause them to grow. We can also see that higher doping will shrink the depletion region. These are great facts to remember, and I often do so by thinking of a forward bias as pushing on the squishy walls of the depletion region. That image is physically inaccurate, but it helps me remember.

Though my pneumonic is fun, having a physically accurate model to rederive these qualitative effects – forward bias and heavier doping shrinks the region, reverse bias grows the region – is worthwhile. I think the electric field is the best way to make a united theory of these changes. The width of the depletion region determines the height of the peak E field because you integrate over the depletion region to make the E field curve. This relation cuts the other way too, if you know the peak E field then you have to know the depletion width because the E field is a one-to-one function with the charge you integrate over. Forward bias reduces your peak E field, which implies that you have to have a narrower depletion region to produce the E peak. Reverse bias increases your peak E field, which implies that you have to have a wider depletion region to produce the higher E peak. Heavier doping on either side of the junction is trickier to model because it both increases the peak E field and the charge density on either side of the junction. However, it turns out that the peak E field grows as the square root of the doping density while the charge grows linearly with doping density, so the depletion width needs to shrink proportionally to the square root of the doping density.



The CE voltage in a BJT in the forward active region can move around significantly without affecting the operation of the transistor. That's captured by our large signal model, which has a current source between the collector and emitter to indicate the current doesn't care about vCE. It's also captured by our semiconductor picture of BJT operation, because the BC junction's function just depends on it capturing electrons with drift, which is true for a wide range of reverse biases and even for slightly forward biases as long as the BC diode doesn't "turn on".

We see the purpose of this reverse biased BC junction in our picture of electron concentration. The reverse biased junction sweeps electrons away, which keeps the concentration of electrons low at the BC junction. That low concentration is important to drive current through the BJT because the rate of diffusion in the steady state is given by the concentration gradient as shown by the equation on the right, which is Fick's first law applied to electrons.

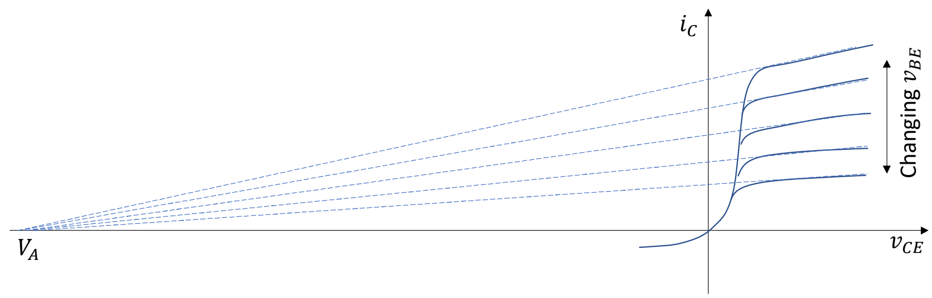
CLICK However, changing depletion widths on the BC junction can affect the total width of the depletion region. This makes dx smaller under higher reverse bias, resulting in more current. It also makes dx bigger under a lesser revers bias, resulting in less current.

CLICK We can import this effect into our iC-vBE curve as slight fluctuations in the exponential behavior as vCE is varied. As usual, though, we'll be able to get a better look at this behavior when we plot iC vs. vCE and don't have our changes swamped by a big exponential.

However,

Department of Engineering

Early Voltage Captures this Diffusion Change



$$i_C \approx I_S \left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right)\left(1 + \frac{v_{CE}}{V_A}\right)$$ in forward active

5

Here I've redrawn our iC-vCE curves with base width modulation included. This results in a linear increase in current as vCE is increased which is pretty exaggerated on this plot. Often this increase is barely perceptible. The iC-vCE dependence is linear because the depletion width only changes a little bit, so we can safely Taylor expand our equations for depletion width to make a good linear approximation to the change in current.

CLICK One detail that might not be immediately obvious is that each of these slopes in the forward active regions points to the same intersection point at (0,-VA). That intersection point is called the Early Voltage, and it's usually in the range of 50-200 volts depending on the BJT.

CLICK We can use the Early Voltage to modify our equation for collector current in the forward active region. This is pretty easy, we just multiply our existing collector current expression by 1+vCE/VA, which creates a small linear dependence of current on vCE. This term will be important when we make a small signal model of the transistor.

Department of Engineering

Summary

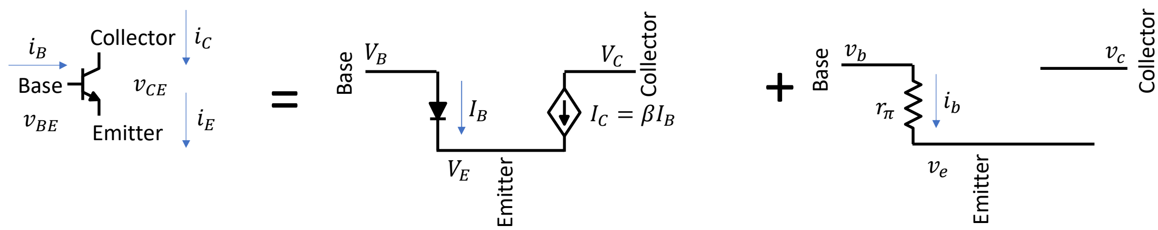
- Changing bias on the BC junction will change the depletion width and in turn the base width, which causes small changes in i_C.
- This changes the shape of the IV curve, most notably by adding a linear dependence on v_{CE}
- All $i_C - v_{CE}$ curves converge at a point $(0, -V_A)$ called the Early Voltage. V_A describes the severity of base-width modulation.

6

In this video we're finally going to start making a small signal model of the NPN BJT in the forward active region. We've been building towards this for a while, but we're going to make the model the exact same way we did for a diode: we're going to calculate local derivatives to find the slope of the dependency between variables we care about.

HARVEY MUDD COLLEGE Department of Engineering

Base-to-Emitter Diode is a Small Signal R



Quick justification: BE junction is a diode, and diodes are small signal resistors.

$$i_B = \frac{i_C}{\beta} \approx \frac{I_S}{\beta}\left(\exp\frac{v_{BE}}{\phi_{th}} - 1\right)\left(1 + \frac{v_{CE}}{V_A}\right)$$

$$\left.\frac{\partial i_B}{\partial v_{be}}\right|_{V_{BE}} = \frac{1}{\beta}\frac{1}{\phi_{th}} I_S \exp\frac{V_{BE}}{\phi_{th}} \approx \frac{I_C}{\beta\phi_{th}}$$

$$r_\pi = \frac{\beta\phi_{th}}{I_C} = \frac{\phi_{th}}{I_B}$$

8

Here I'm showing that our diode symbol can be replaced with a large signal circuit, which we've derived before, plus a small signal circuit that is currently blank. Note that I'm using large signal notation on the large signal circuit to show that it's our bias point, and I'm using small signal notation for the small signal circuit.

CLICK The first behavior we're going to capture in our small signal circuit is small deviations in the base current as vBE is changed. We represent these changes using the symbol rpi, which I've written all in lower case to show that it's a small signal resistor that doesn't exist in the large signal circuit.

CLICK The quickest way to understand why we choose a resistor to show the relationship between small signal ib and small signal vbe is that our large signal model has a diode connected between the base and the resistor, and we've already gone through the exercise of finding the small signal model of a diode. We found it was a resistor with value n*phi_th/ID, so we'd expect something similar for this particular junction.

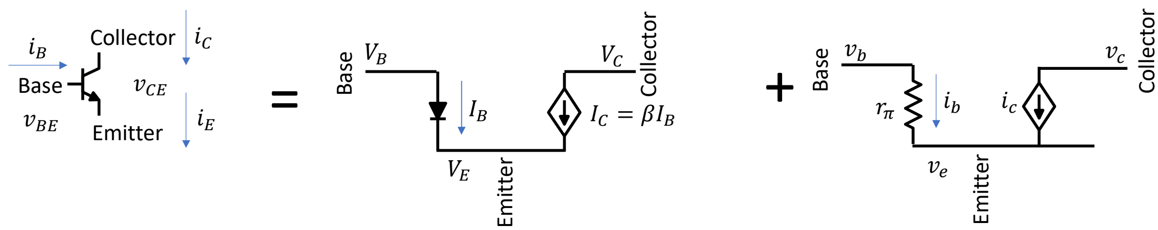
CLICK We can show that formally by taking the derivative of total signal iB with respect to small signal vBE. We need to find iB first, and here I'm just representing it as the collector current divided by beta, which works well in forward active.
CLICK So we take our derivative at the large signal VBE bias point, which gives us an

expression that contains IS*exp(VBE/phi_th). Just like our other model, we'll say that quantity is about the same as large signal IC, so our local derivative is given by IC over beta time phi_th.
CLICK However, that derivative is a conductance because it's a current over a voltage. We let that conductance be defined as one over rpi, then flip it around to find rpi is equal to beta phi_th over IC. That's the expression I use most often to calculate rpi, but if we sub in IC=beta*IB, then we can also see that rpi is equal to phi_th/IB. That last expression is exactly the differential resistance of the base-emitter diode presuming the non-ideality factor is one.

If we didn't know anything about the inner workings of a BJT, we would also have to take a derivative of iB with respect to vce to see if it was appropriate to include an element that represented iB changes with iC. Taking that derivative would reveal that base-width modulation has a small effect on base current, but that the small effect is divided by beta. Because the effect of base width modulation is already small before dividing by beta, we ignore it's effect on the base current. However, it's worth noting that when you're making general small signal models, you need to consider the effect of each variable on each other because they define a N-dimensional surface that determines the device's behavior.

Department of Engineering

Collector-to-Emitter Current Source is Linear



Possible i_c control laws

$i_c = \beta i_b$ — Because large signal $i_C = \beta i_B$

$i_c = g_m v_{be}$ where $g_m = \beta / r_\pi$ — Because we can sub v_{be}/r_π into βi_b

OR where $g_m = I_C/\phi_{th}$ — Because $\frac{\partial i_C}{\partial v_{be}}|_{V_{BE}} = \frac{1}{\phi_{th}} I_S \exp \frac{V_{BE}}{\phi_{th}} \approx \frac{I_C}{\phi_{th}}$

9

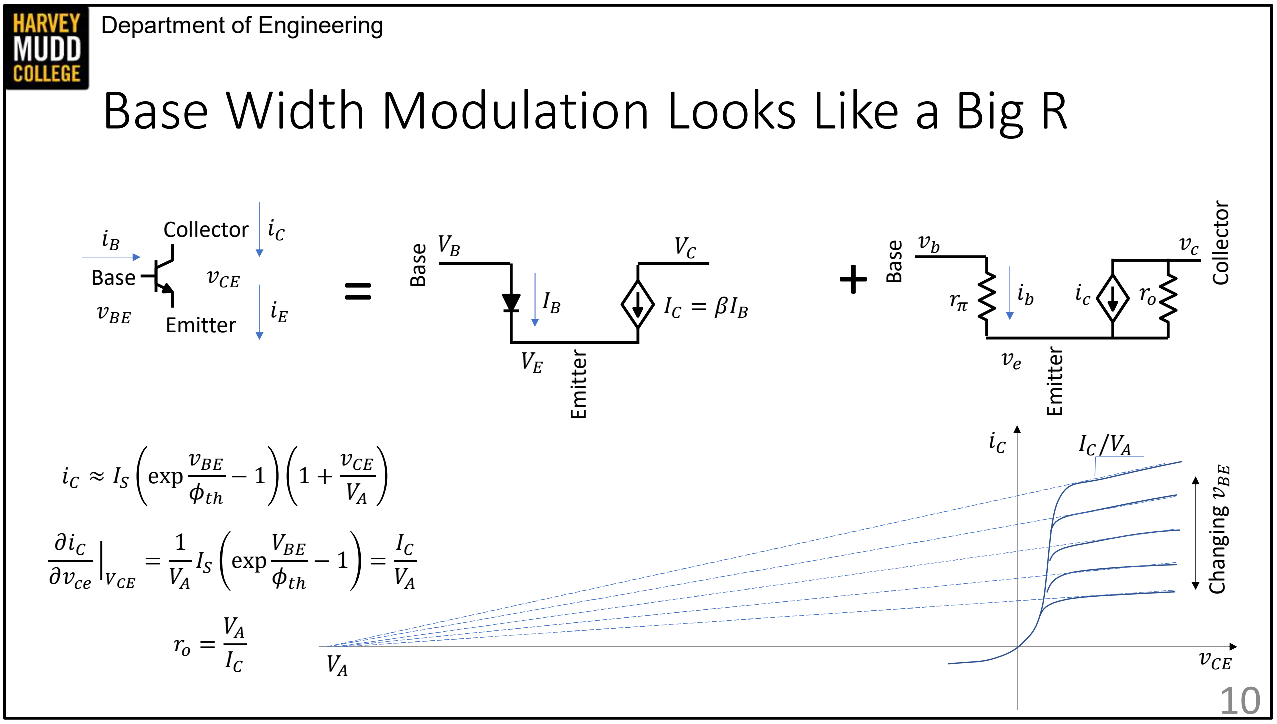
Next we're going to consider the effect of the base-emitter voltage on the collector current CLICK and we know that the base-emitter junction creates a current between the collector and the emitter in our modified U model. So I've included a dependent current source in our model to get started.

CLICK We know that dependent current source is there because large signal IC is linear in large signal IB, so the same has be to true for small signals too. Just like we keep resistors in small signal circuits, any linear element will respond the same to both large and small signals with no approximations required.

CLICK However, it's common to think of amplifiers as having voltage inputs, so it's similarly common to express iC as the product of a transconductance, gm, and vbe. We can find the transconductance by noting that gm*vbe must be equal to beta ib, and that vb is equal to ib times rpi. We can substitute those quick equations together to note that gm is beta/rpi.

CLICK That's a nice way to calculate gm using small signal parameters. We can also find gm using large signal parameters by taking a derivative of iC with respect to vbe like we have with all of our other small signal elements. Chasing that through, including substituting in IC in place of IS*exp(VBE/phi_th), reveals that gm is IC/phi_th. I use this formula to find gm slightly more often than the other formula, but both are worth remembering. Also, for the formula that represents gm in terms of small signal quantities, I think it's easier to remember that gm*rpi=beta.

… note that we took partial diC/dvBE to find this current contribution. In general N-dimensional small signal models require N^2 derivatives to represent the slope in every direction in the iC-iB-vBE-vCE space.



Finally, we need to find the effect of vce on ice.
CLICK This current is going to be represented by a resistor between the collector and emitter named ro, which suggests that current linearly increases with vce. It's worth noting that ro is usually quite large, around 100k or higher, so the slope of the ic-vce curve will be shallow.
CLICK We are going to take a derivative of the iC expression as usual to find the value of ro, so I've copied it in here.
CLICK Taking the derivative, we bring down a factor of the early voltage, and then don't need to approximate when we substitute IC back in place of IS*exp(VBE/phi_th-1).
CLICK This derivative is a conductance, so we define 1/ro as being equal to that slope giving us a resistance of VA/IC, which is a large number divided by a small one, about 100V over one milliamp, giving us hundreds of kilo-ohms of resistance.
CLICK This equation captures a feature that we can see on our IC/VCE graph. If we assume that VA is much bigger than the large signal VCE, then the slope of lines in our forward active region is given by IC, the rise, divided by VA, the run. This slope is 1/ro. In general, each of our small signal models correspondes to some slope on our IV curves, and it's worth remembering that IV curves are an exact representation of the underlying Ebers-Moll model. We're going to play a trick with IV curves in the next video, so I wanted to emphasize that IV curves are the same thing as the underlying model, and the derivatives we take to make a small signal model show up in IV curves.

Department of Engineering

Example of Analyzing a BJT in a Circuit

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

12

In this video we're going to test out our model in a circuit. This is going to be an example problem and it's very instructional to wrestle with it for a bit. So please pause the video when asked and give the problem a try.

Department of Engineering

Here's an Example BJT Circuit

- Let V_{BEON}=0.7V, β=100 and V_{CESAT}=0.2V
- Assuming this circuit is in Forward Active, What is I_C?
- How big can R_C be before the device enters the saturation region?
- Calculate g_m and r_π

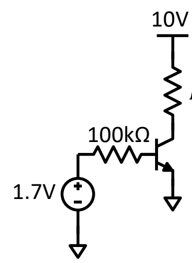
$I_B = \frac{V_{IN}-V_B}{R_B} = \frac{V_{IN}-V_{BEON}}{R_B}$ =1V/100kΩ=10μA

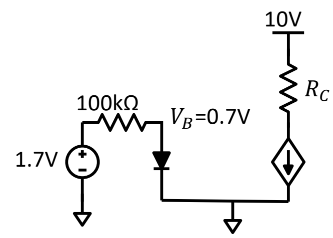
$I_C = \beta I_B$ =100*10μA=1mA

$V_{CC} - I_C R_{C,max} = V_{CESAT}$ (b/c then $V_C < V_{CESAT}$)

$\frac{V_{CC}-V_{CESAT}}{I_C} = R_{C,max} \rightarrow R_{C,max}$ =9.8kΩ

$g_m = \frac{I_C}{\phi_{th}}$ = 40mS, $r_\pi = \frac{\beta}{g_m}$ =2.5kΩ





13

A circuit is pictured below. I'd like you to find the large signal IC, figure out the largest RC that you can use in this circuit and still have it be in forward active, and calculate a few small signal parameters. Paus the video and give it a try.

CLICK To start solving this, I've drawn the large signal equivalent circuit in place of the BJT in this circuit on the side. I don't usually draw large signal circuits out like this when I'm solving problems, but it's good to draw them out when you're just getting started with BJT analysis. You can see from this circuit that we're assuming the base voltage VB is 0.7V because the VBE junction diode is on in the forward active region. One quick note about the circuit symbol for BJTs, the little arrow in the BJT gets replaced by a diode in the NPN and PNP BJT symbols, which is a handy way to remember where the diode is and which way it's pointing.

CLICK We calculate the base current by finding the current in the 100kohm base resistor. We find the voltage across it by subtracting the base voltage from the 1.7V input voltage. The base voltage is 10 microamps.

CLICK Because we're in forward active, IC is given by beta times IB, so IC is 1mA. Note that we can't find IC until we find IB because IC is completely determined by IB in forward active. The base sets the behavior of the collector. Also, though it doesn't matter in this problem, it's worth noting that the base current is much smaller than the collector current, which means the emitter current, the sum of the base and collector currents, is going to

be about the same as the collector current. It's common to approximate iE as equal to iC for hand analysis, though iC is really beta/(beta+1) times iE. For a beta around 100, that's less than 1% error in your calculations.
CLICK This circuit will enter saturation when the base-to-emitter voltage is greater than 0.7V, which is already true, and when the collector-to-emitter voltage is less than VCESAT, or 0.2V. So we need to find what value of RC will create 0.2V at the collector. Because we know IC is 1mA and the supply is 9.8V above 0.2V, we know 9.8kohm of resistance will drive VC low enough to cause saturation. As an aside, the supply is called VCC by tradition since it's often separated from NPN collectors by just one element.
CLICK Since we know IC, it's easy to use our large signal relation to calculate gm. 1mA divided by 25mV, which is a convenient number to round the thermal voltage to, is 40 mS. We can use that and beta to find rpi by remembering that gm*rpi=beta. rpi turns out to be 2.5kOhm. We could also find rpi as beta*phi_th/IC, which would be 100*25mV/1mA or 2500 ohms.

That's the basics of large signal analysis for BJTs, and doing the large signal analysis lets us calculate small signal quantities.

Summary

- Calculate the bias point of BJT circuits by using the large-signal model and a switch-voltage source model of the BE diode.

- Double check that your assumption of your region of operation is correct.

- You enter saturation when V_{CE} drops below V_{CESAT}.

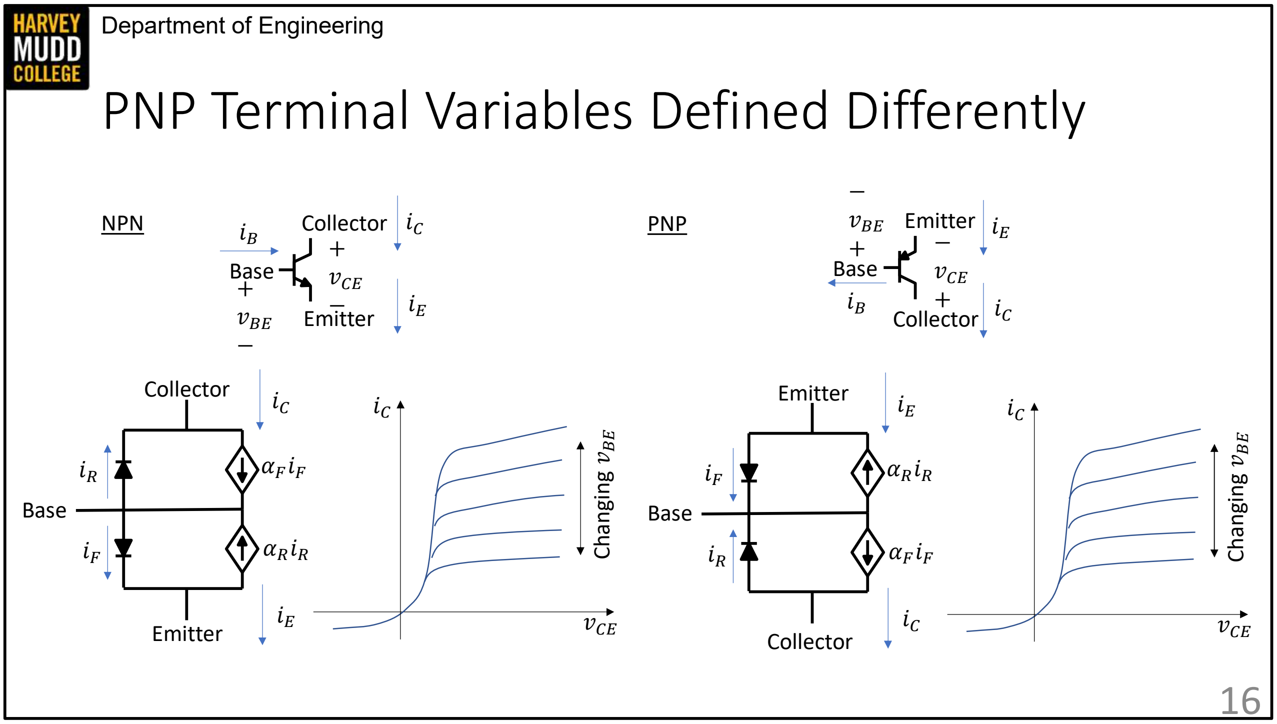
Department of Engineering

PNP BJTs

Matthew Spencer
Harvey Mudd College
E151 – Analog Circuit Design

15

In this video we're going to introduce PNP BJTs, which will leverage much of what we learned from NPN BJTs.



I've included three of our common representations of NPN BJTs on the left of this slide. I've left out the full Ebers-Moll equations, but both the Ebers-Moll diagram and the IV curve are standing in for them. We're going to slowly introduce PNP representations on this page and compare them against NPN representations.

CLICK The first thing to look at is the terminal assignments of the PNP BJT. You can see that they are upside down compared to the NPN BJT. I've drawn the emitter at the top, with an arrow running from the emitter to the base (indicating a diode), and the collector at the bottom. The currents have switched directions, so current flows into the emitter and out of the collector and base. And the base-emitter and collector-emitter voltages are now upside down as well because the terminals switched places. We pick this arrangement because the PNP transistor is turned on by lowering the base voltage below the emitter voltage rather than raising it above the emitter voltage. That's consistent with what we know of the construction of PNPs, the PN junctions inside PNP BJTs should be reversed, so you'd expect diodes to point the opposite way.

CLICK That is clearly true in our PNP version of the Ebers-Moll model. The diodes have switched direction because the base has to be below the collector or emitter to forward bias the PN diodes in a PNP junction. This means almost all the logic we applied making the NPN Ebers-Moll model has flipped around too, so the dependent current sources are now pointing out instead of in. In forward active, that means the collector will sink most of the current passing through the forward diode, while in the NPN the collector sources the current that would pass through the forward diode. It's also worth noting that the emitter in a PNP emits holes, so current flows the same way as carriers in a PNP device.

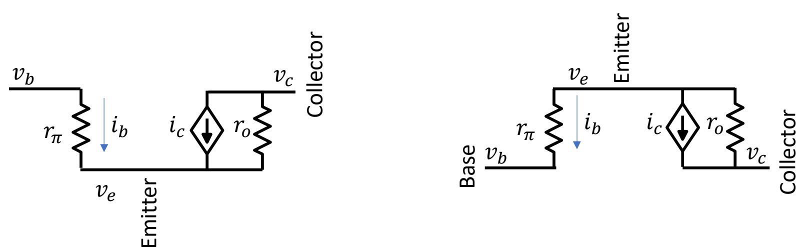
CLICK However, even though our Ebers-Moll model is backwards, the wacky redefinition of terminals that we did has a magical effect on our IV curve. The relation between iC and vCE is the same in a PNP as in an NPN, but we just need to remember that the emitter is at a high voltage and the collector is at a lower voltage when vCE is high. Similarly, high vBE corresponds to a high emitter and a low base. As an quick justification for how this IV curve can be the same as the NPN, you can see that reducing the base in the Ebers-Moll model will increase current just like increasing the base voltage in the NPN Ebers-Moll. Our funny terminal definitions for the PNP let us use the same IV curve to describe it (though often the Early Voltage and beta will be quite different than an NPN, usually lower).

… flipped upside down,
… note arrow points from E to B, tells you where the controlling diode is
… note that minority carriers in the base are holes, so current makes more sense

Department of Engineering

PNP Small Signal Model is the Same as NPN

NPN PNP



17

Having the same IV curve for the NPN and the PNP means that the underlying model is the same, so taking derivatives of it would yield the same behavior. That in turn means that the PNP small signal model is identical to the NPN small signal model. There is an rpi=beta*phi_th/IC between the base and the emitter, a dependent current source equalt to gm*vbe between the emitter and the collector where gm=IC/phi_th, and there is an ro=VA/IC between the emitter and the collector. The only difference is that we have to remember that the emitter of a PNP is often attached to a high voltage, while the emitter of an NPN is attached to a low voltage. However, because voltage sources in a small signal model are shorted, you'll often see the emitters of NPNs and PNPs shorted together.

To summarize, because we did footwork with our terminals, the PNP has the same IV curve as the NPN, which means the small signal model is the same as the NPN.

HARVEY MUDD COLLEGE Department of Engineering

Quick Example in a Circuit

- Let V_{BEON}=0.7V, β=50 and V_{CESAT}=0.2V
- Assuming this circuit is in Forward Active, What is I_C?
- How big can R_C be before the device enters the saturation region?

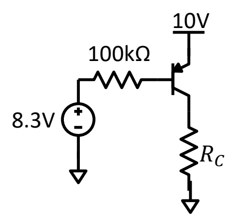
$V_B = V_{CC} - V_{BEON}$ =9.3V

$I_B = \frac{V_B - V_{IN}}{R_B}$ =(9.3V-8.3V)/100kΩ=10µA

$I_C = \beta I_B$ =50*10µA=500µA

$V_{C,max} = V_{CC} - V_{CESAT}$ =9.8V

$V_{C,max} = I_C R_{C,max} \rightarrow R_{C,max}$ =19.6kΩ



18

I’m going to run through a quick example of a PNP in a circuit. This is great practice, so feel free to pause the video and work through this problem by yourself.

CLICK We start by finding the base voltage, which is a diode drop below the supply voltage. The supply still called VCC because of a tradition associated with NPN transistors.
CLICK Once we know the base voltage, we can find the voltage difference across the base resistance and calculate the base current, which flows out of the base.
CLICK Since we’re in forward active, the collector current is beta times the base current, so we have 1mA flowing down this branch of the circuit.
CLICK The maximum collector voltage we can have before the transistor becomes saturated is 9.8V, which is one VCESAT below the supply.
CLICK We know the collector voltage is just IC times RC, so we can find the maximum RC as 19.6kohm.

… why’s it called VCC?

Summary

- PNPs have the emitter at high voltage and collector at low voltage.
- The large signal model of the PNP reverses the diodes and current sources compared to the large signal model of the NPN.
- However, reversing both the model and terminals means the IV curve and small signal models for the PNP device are the same as the NPN.