Dynamic Models of BJTs

Matthew Spencer

Harvey Mudd College
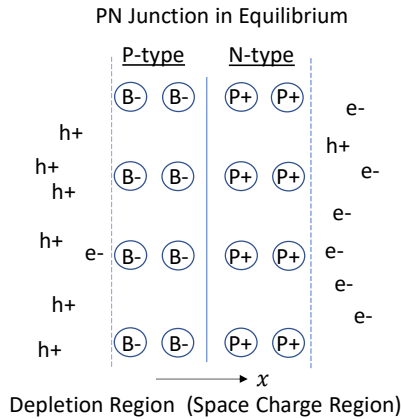
E151 – Analog Circuit Design

1

In this video series we're taking a big, life-changing step in the class: we're introducing capacitors! This is the start of our study of amplifier dynamics, which is a big and interesting problem that makes circuit design much richer and ties it closely to control theory in fun ways. We're going to start by uncovering the physical origins of capacitance in this video series, then we'll look at how that capacitance affects amplifiers in a few videos after that.
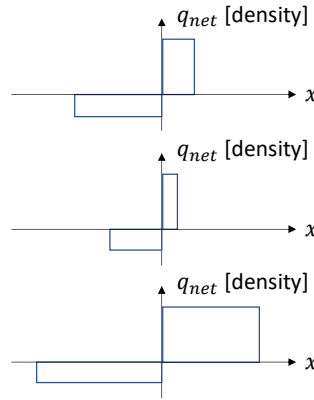
PN Junctions have Capacitance

Matthew Spencer

Harvey Mudd College

E151 – Analog Circuit Design

2

In this video we're going to observe that PN junctions always have some capacitance associated with them.

Space Charge Changes with Depletion Width

PN Junction in Equilibrium

P-type   N-type

(B-) (B-) (P+) (P+)   e-
h+   h+
h+ (B-) (B-) (P+) (P+)   e-
h+
          e-
h+ e- (B-) (B-) (P+) (P+) e-
          e-
h+   e-
h+ (B-) (B-) (P+) (P+)

→ x

Depletion Region (Space Charge Region)

q_{net} [density]
→ x
Equilibrium

q_{net} [density]
→ x
Forward Bias:
$W \downarrow \rightarrow Q \downarrow \rightarrow dQ/dV$
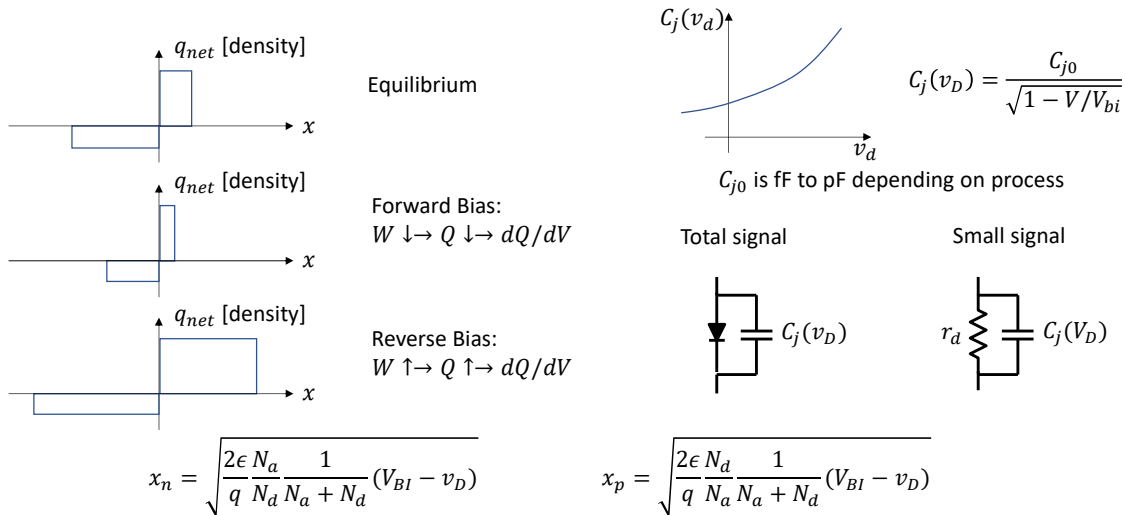
q_{net} [density]
→ x
Reverse Bias:
$W \uparrow \rightarrow Q \uparrow \rightarrow dQ/dV$

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

3

I've dragged out our old picture of a PN junction on the left side of this slide, and I've included the two equations we used to calculate the width of the PN junction on the bottom of the slide. These equations indicate that the width of a PN junction changes as vD, the bias voltage across it changes. That means forward biasing the junction by increasing vD results in a narrower depletion region with less total charge stored inside of it. Reverse biasing the junction by decreasing vD does the opposite, it results in a wider junction with more total charge stored in it.

This is interesting because we know capacitance is a relationship between voltage and stored charge. So we're seeing some dQ/dV both as we increase and as we decrease the bias on the junction.

dQ/dV is Capacitance, so Junctions have Cap

q_{net} [density]

Equilibrium

$C_j(v_d)$

$$C_j(v_D) = \frac{C_{j0}}{\sqrt{1 - V/V_{bi}}}$$

v_d

C_{j0} is fF to pF depending on process

q_{net} [density]

Forward Bias:
$W \downarrow \rightarrow Q \downarrow \rightarrow dQ/dV$

Total signal

Small signal

q_{net} [density]

Reverse Bias:
$W \uparrow \rightarrow Q \uparrow \rightarrow dQ/dV$

$C_j(v_D)$

r_d $C_j(V_D)$

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

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

4

You're probably used to the linear capacitance formula Q=CV, which says there's a linear relationship between voltage and stored charge. Charge stored in a depletion region isn't linear, the width varies as the square root of voltage, so we need to use a differential definition of capacitance, C=dQ/dV. If we do that, we can see that capacitance is large when the depletion region is forward biased, because the slope of the width function is steeper when (VBI-vd) is close to zero. Similarly, we can see that capacitance will fall off as the junction is reverse biased.

CLICK If you draw a picture of that behavior, you find the capacitance vs. diode voltage curve that I've included on the right. The capacitance of the junction increases as the total signal diode voltage is increased. The junction cap is rather hard to model, but one common equation for it is given on the far right of the slide, where Cj0 is some reference capacitance that is in the fF to pF range depending on the manufacturing process, and Vbi is the built in voltage that we got from integrating junction charge way back in the day. Other versions of this model will use a cube root on the bottom instead of a square root, and modern transistors are modeled by measuring them carefully and looking up the junction cap in a table.

One way students like to remember this capacitance curve is that capacitance is high when the depletion width is narrow, like a linear capacitor, which gets bigger when the separation

between plates is small. That's fine as a mnemonic, but remember that the physics of this junction are very different than a linear cap. The charge isn't stored on either side of the depletion region, it's stored _in_ the depletion region, and capacitance is related to the depletion region's changing width.

CLICK We can convert this capacitance curve into a large and a small signal model for a diode. Large signal, a diode looks like the diode we know and love in parallel with the capacitance curve I've drawn at the top of the slide. In a small signal sense, we pretend the capacitor isn't changing much, and just say that the capacitance is set by the large signal bias across the junction.
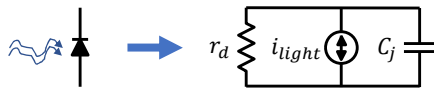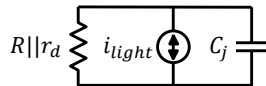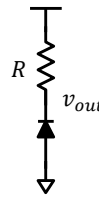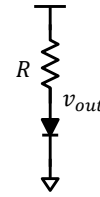
Alright, let's put this model to the test. We're going to use it to figure out if we should forward or reverse bias photodiodes. Photodiodes are light sensitive diodes that are often used as receivers in telecommunications systems. As a result, you really care how fast you can change the light signal at the input of the diode and still measure a signal, because that sets the data rate of your communication system. We can make a small signal model for photodiodes as a current source in parallel with our normal diode model, which is a differential resistance in parallel with our junction cap.

Two quick ways you could imagine making a receive circuit are pictured on the right of the slide: one has the diode reverse biased, and other has the diode forward biased. In both cases, we're biasing the diode with a resistor, and hoping the light current source drives into this resistance, R, to make a voltage signal. Pause the video and figure out the -3dB bandwidth of the response to ilight and write down whether you think we should use the forward or reverse biased setup.

CLICK I've drawn the small signal model that represents both circuits here and calucalted vout as a function of ilight. I found this relationship by noting that ilight was going to drive into the parallel combination of rd, R and Cj, which form the input impedance of the photodiode. That means the -3dB bandwidth is given by one over the time constant, or 1/(Cj*(R||rd)) in rad/sec. I contend that reverse biasing the junction makes more sense

here. It will make Cj smaller, which increases our bandwidth, it will make rd bigger, which increases our gain, and there's no DC current. However, if you have a slowly varying signal, sometimes forward biasing the diode makes sense so that your output voltage is referenced against 0V instead of the supply. That was a pretty common configuration in E80, which had very slowly varying light signals.

Summary

- Depletion width changes with applied voltage, so does stored charge.

- $C = dQ/dV$, so there is a capacitance C_j at junctions.

- Junction cap at the junction increases with forward bias
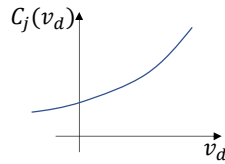
$$C_j(v_D) = \frac{C_{j0}}{\sqrt{1 - V/V_{bi}}}$$

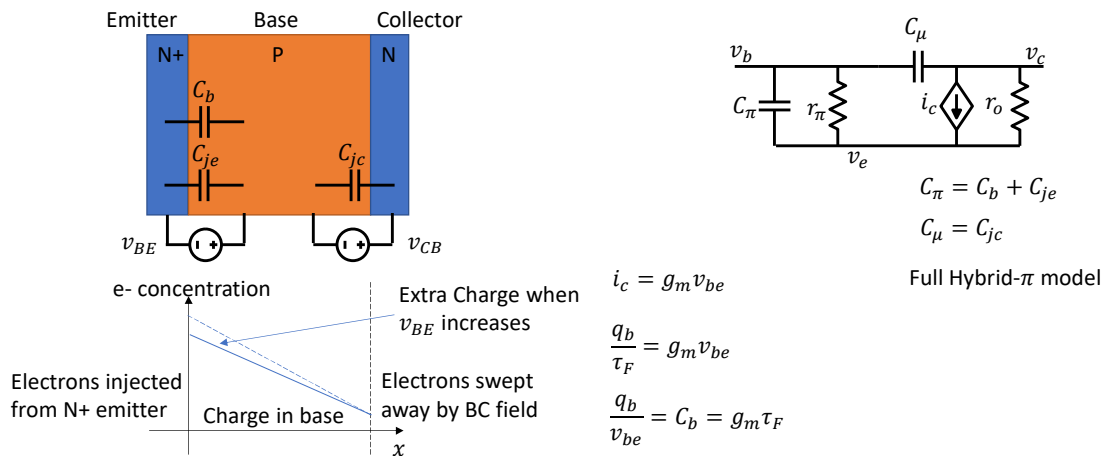# Dynamic Model of the BJT

Matthew Spencer

Harvey Mudd College

E151 – Analog Circuit Design

7

In this video we're going to make a model of our BJT with capacitors in it by identifying the parasitic capacitances in a BJT.

BJTs have Two Junction Caps + Base Charge

Emitter   Base   Collector

N+   P   N

C_b

C_{je} C_{jc}

v_{BE} v_{CB}

e- concentration

Extra Charge when v_{BE} increases

Electrons injected from N+ emitter

Charge in base

Electrons swept away by BC field

x

v_b C_μ v_c

C_π r_π i_c r_o

v_e

$C_\pi = C_b + C_{je}$

$C_\mu = C_{jc}$

Full Hybrid-π model

$i_c = g_m v_{be}$

$\dfrac{q_b}{\tau_F} = g_m v_{be}$

$\dfrac{q_b}{v_{be}} = C_b = g_m \tau_F$

8

I've pulled out our cross section of the base of a BJT on the left of this slide, and we can immediately identify two junction capacitors at the base-emitter junction and the base collector junction. We'll call these Cje and Cjc. Cje is usually forward biased when we're in forward active, so it's pretty big, while Cjc is usually reverse biased in forward active, so it's usually pretty small. However, Cjc is often in a feedback position in amplifiers because it connects the collector and the base, we'll see that it's quite influential in future videos.

CLICK Charge hides in one other place in a BJT, which results in an additional base-to-emitter capacitor we call Cb, for base transit capacitance.

CLICK The trick to identifying the capacitance is to recall that BJTs work by injecting a lot of extra minority charge carriers into the base. The concentration gradient of those carriers drives current conduction from the emitter to the collector, but while those carriers are in transit they're all stored in the base. Here the area under the concentration curve represents the total charge currently stored in the base.

CLICK So when we raise vBE to inject more electrons, this results in more total charge stored in the base. Here, the dotted line is the concentration gradient with a higher vbe, and the area between the dotted line and the original line is additional charge that we had to shoot into the base to achieve that gradient. This is a change in charge with voltage, so

it's a capacitance. And because all the extra charge gets injected from the emitter, the capacitance has to be connected between the base and the emitter.

This capacitance can seem a bit cryptic, but here's one analogy students like: imagine the collector to emitter current like a river of charge, and the base is some little chuck of the river we're scrutinizing. If we raise the water level of the river to increase the flow rate, then we're storing more water in the little section of the base we're looking at.

CLICK One natural question is "how big is the base cap anyway"? We can calculate that using our differential small signal relations. First, we note that we've decided that ic is equal to gm*vbe. Second, we note that if the base transit time is linear (which it more or less is in a small signal model), then the current in the base can be calculated as the total charge in the base qb, divided by the base transit time tau_F. You can get that equation by imagining all the charge in the base is at once side and asking how long it takes to get to the other side, which is a super coarse approximation of what's going on, but which works OK. Rearranging slightly, small signal charge divided by small signal voltage is like dQ/dV, so we call it Cb and note that it's equal to transit time times gm.

CLICK We can put this together into a dynamic model for the BJT. We add two caps to the model we had before, Cpi from the base to the emitter, and Cmu from the base to the collector. Cpi is equal to Cb + Cje, and Cmu is just equal to Cjc. This model is called the full hybrid pi model of the BJT, because the capacitor across the top makes it look like the letter pi, I guess.

Summary

- BJTs have two junction capacitances and charge stored in the base

- C_{je} is forward biased (and large), C_{jc} is reverse biased (and small)

- Base charge is "charge in transit" and increasing current adds more base charge by injecting more e- at source.

$$C_\pi = C_b + C_{je}$$
$$C_\mu = C_{jc}$$
$$C_b = g_m \tau_F$$

9
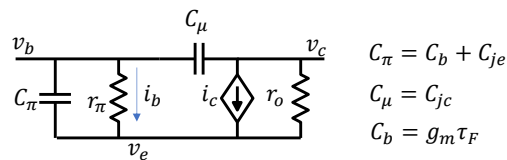
Transition Frequency (f_T), a BJT Speed Limit

Matthew Spencer

Harvey Mudd College

E151 – Analog Circuit Design

10

In this video we're going to learn about a transistor property that is used to describe the effects of transistor capacitances. The property is called the transition frequency, or fT, and it is a commonly reported measure of how fast a transistor can go.

Transistor Current Gain is 1 at f_T
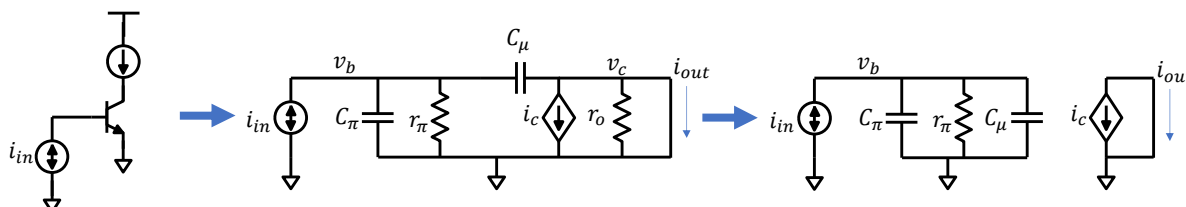
$$i_{out} = g_m v_b = g_m \frac{r_\pi}{r_\pi (C_\pi + C_\mu) s + 1} i_{in}$$

$$a_i = \frac{\beta}{\beta \left(\frac{C_\pi + C_\mu}{g_m} \right) s + 1} i_{in}$$

$$f_T \approx \frac{1}{2\pi} \frac{g_m}{C_\pi + C_\mu}$$

11

So, fT is a type of speed limit that is often reported for transistors, though it's a bit of a weird one. It's the frequency where the transistor's current gain drops to 1. So I've drawn a schematic on the left that we can use to measure the current gain of a transistor without any external elements messing with our measurement.

CLICK We can draw the small signal circuit to set up our current gain calculation. Note that we have shorted the output, because current gain is measuring current through an output short from an input current source. That's the same as saying we test current gain with Rs=infinity and Rl=0.

CLICK That short at the output lets us simplify the small signal model quite a bit. ro is shorted on both sides, so it's removed from this simplified schematic, and Cmu is grounded on the right, so we just put it in parallel with Cpi and rpi.

CLICK the current through the output short is equal to the current in the gm generator, which is given by gm*vb. I substitute in for vb by multiplying the input current by the input impedance, rpi in parallel with the combined Cmu and Cpi capacitors.

CLICK Rearranging a bit, we get the current gain is equal to beta at DC, which makes sense, and that it rolls off as a first order system at a frequency given by rpi and the combined

capacitance. I've rewritten rpi in the first term of the denominator as beta over gm because it suggests an approximation.

CLICK If we pretend the first term of the denominator is bigger than the second, then we know ai will equal one when the s factor cancels out (Cpi+Cmu)/gm. That means our transition frequency is at gm/(Cpi+Cmu) in radians/second, or a factor of 2pi smaller in Hz, which is what I've written here.

I've mentioned a few times that ft is weird, and one way it's weird is that it weighs Cpi and Cmu equally. They're just added together to make the denominator. In practice, you often need to design much more carefully around Cmu than Cpi. fT also isn't the ultimate speed limit of a a transistor because you can still amplify signals even if you can't amplify current. A separate measure called fmax measures the frequency at which power gain drops to one, and fmax is a much more serious speed limit than fT. Even so, fT is reported quite often, and it's useful to be able to read it off of datasheets.

Summary

- Transistor Current Gain is 1 at f_T.

$$f_T \approx \frac{1}{2\pi} \frac{g_m}{C_\pi + C_\mu}$$

- This current gain test weights C_π and C_μ equally, for better for worse.

- f_{max} is the frequency where power gain becomes 1.

12

Dynamic Model of the MOSFET

Matthew Spencer

Harvey Mudd College

E151 – Analog Circuit Design

13

In this video we're going to create a dynamic model of a MOSFET by finding the parasitic capacitances in a MOS device.

Structure of MOSFET Suggests Capacitances
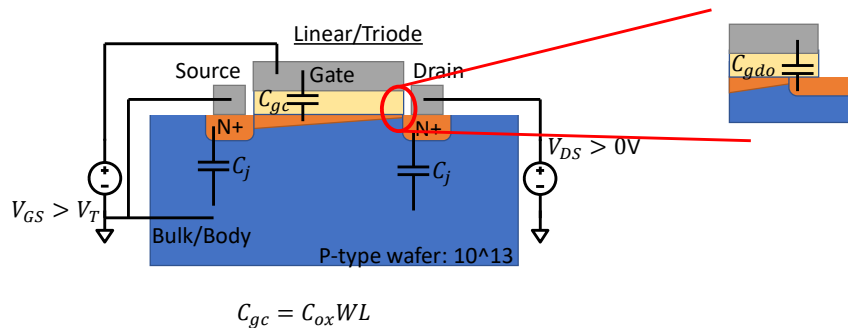
Looking at a MOSFET cross section should make a few capacitances jump out at us. Pause this video and write down capacitances that you see in this device cross section.

CLICK The first quick ones are junction capacitances between the drain/source and body. These are sometimes called diffusion capacitors in MOSFETs because the drain and source layers were created by diffusing material into the substrate in older MOS processes.

CLICK The second really important capacitor is one that we put in place on purpose: the gate-to-channel capacitance. This capacitor is how we created the electric field that makes the FET part of a transistor. We have calculated the gate capacitance previously when we were trying to figure out how much charge was in the channel, and its value was Cox, the gate capacitance density, times the MOSFET width times the MOSFET length. This capacitor is a bit weird for modeling because it doesn't connect to any of our device terminals like the drain or the source, but we'll tackle that on the next slide.

CLICK Finally, there's a sneaky capacitance that you may not have been able to see in the big drawing, but the gate overlaps the drain and source by a tiny amount. This happens because of a fabrication technique called self-aligned source/drain implantation, which uses the gate itself as a screen to prevent dopants from getting into the channel. These small capacitances are called extrinsic drain and source capacitances, and they are named

Cgdo for gate-drain overlap capacitance or Cgso for gate-source overlap capacitance.

Assign Gate-Channel Cap to Drain and Source

Linear/Triode
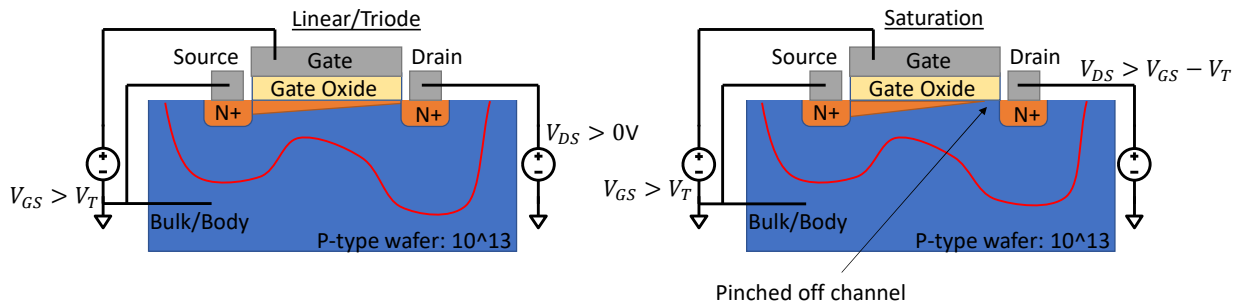
Split gate-channel cap in half
$$C_{gs} = C_{gso} + \frac{1}{2}C_{ox}WL$$
$$C_{gd} = C_{gdo} + \frac{1}{2}C_{ox}WL$$

Saturation

Reduced gate-channel cap, all to source
$$C_{gs} = C_{gso} + \frac{2}{3}C_{ox}WL$$
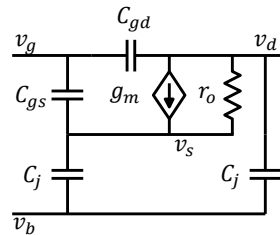$$C_{gd} = C_{gdo}$$

OK, we need to figure out how to model this gate-to-channel capacitance. We don't want to add a channel node to our small signal model, so we need to assign this capacitance to our device terminals. \

We do that in different ways in the linear region and the saturation region. In linear, the channel is low impedance and it connects to both the drain and the source, so we just split the gate capacitance in half between the drain the source as shown in the lower left. We say the total gate-source cap is the extrinsic Cgso plus the intrinsic capacitance, one half Cox*W*L. Cgd is similar.

In saturation, the channel doesn't connect to the source, so we assign all of the channel capacitance to the drain. Also, cutoff means the total size of the channel is a bit smaller, so we're going to assign some reduced fraction of the channel charge to the source instead of the full Cox*W*L. That fraction turns out to be 2/3 for old school transistors, though that fraction varies in more modern devices. The only capacitance assigned to the drain in saturation is the extrinsic overlap capacitance.

… capacitance matrix

Dynamic Small Signal Model of MOSFET

$$C_{gs} = C_{gso} + \frac{2}{3}C_{ox}WL$$

$$C_{gd} = C_{gdo}$$

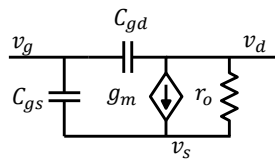$$f_T = \frac{g_m}{C_{gs} + C_{gd}}$$

16

We can combine all of these capactiors to make a small signal model for the saturation region. The gate-drain capacitor has been connected between the gate and drain, and the gate-source capacitor has been attached between the gate and the source.

CLICK One feature of note is that fT for MOSFET is exactly gm/(Cgs+Cgd) because our input is purely capacitive, there's no rpi in the way to make a tricky first order system at the input. Note that the approximation we used in the BJT video has been replaced with an equality here.

CLICK However, we've forgotten a terminal. Don't forget that we still have junction capacitances to our body terminal down here.

Summary

- Most MOSFET cap deliberately comes from gate-channel overlap

- Gate-channel cap gets split between drain and source
  (1/2 each in triode, 2/3 all to source in cutoff)

- Additional parasitic overlap caps from gate to drain and source

$$f_T = \frac{g_m}{C_{gs} + C_{gd}}$$

$$C_{gs} = C_{gso} + \frac{2}{3}C_{ox}WL$$

$$C_{gd} = C_{gdo}$$

17