

In this video series we're taking another aside to introduce MOSFETs. We aren't using MOSFETs extensively in this class even though they're very important out in the world: most circuits are made with MOSFETS. However, MOSFETS are a bit tougher to bias and do math with than BJTs, so we're using BJTs to learn about circuit concepts. So consider this video series as a bridge to future designs! This video series is also a really good review of all the steps we took to analyze BJTs, so we'll get to see another take at device physics, large signal analysis, making small signal models and single stage amplifiers.



In this video we're going to do a gross overview of a very rich field. We're going to try to introduce the basics of MOSFET device construction and operation in about ten minutes. People get whole PhDs in just this video, so be aware that we're skimming over quite a bit.



MOSFET stands for metal oxide semiconductor field effect transistor, and I've included schematic symbols for MOSFETs on the left of this slide. Like BJTs, there are two flavors of MOSFETs, which are called NMOS and PMOS. MOSFETs have three terminals, the gate, the drain and the source. The gate-to-source voltage and the drain-to-source voltage control the current in the MOSFET, which is labeled as the drain current. There are two ways that MOSFETs are commonly drawn, so I've included a set of alternate symbols on the bottom of the left diagram. Nominally the alternate symbols are more associated with analog design, because they mimic BJTs, and the top ones are more associated with digital design, because the PMOS circle represents and inversion, but I haven't seen any consistency in their application.

The right side of the slide shows how an NMOS and a PMOS would be built on a chip. There's a cross section on the top of the slide and a top view of MOSFET layout on the bottom. NMOS devices are made of two N+ regions source/drain regions in a P-type substrate with this mysterious gate-oxide and gate metal structure sitting on top of them. PMOS devices are similar, but they live in N-type wells and have P+ dopants regions. We'll talk about what the gate structure does in a second, but just looking at the MOSFET cross section can reveal a few interesting facts. The first of these is that the MOS in MOSFETS comes from the cross-section, because you can see a metal-oxide-semiconductor, stackup of materials in the gate structure. Pause the video and try to write down another few interesting things you see in a MOSFET.

OK, the first thing that I see is that the gate is separated from any charge carriers that flow between the source and the drain by an insulator. That means that basically no current can flow into the gate! Another thing I noticed is that MOSFETs are symmetric, so the drain and the source seem to be labeled arbitrarily, or by convention. That convention carries weight in many discrete MOSFETs, so don't plug things upside down at random, but it's interesting to note that these are quite different than BJTs. Finally, the voltage in the P-type wafer probably affects how the MOSFET operates, so there's a hidden fourth terminal in this device called the bulk or body. Finally, I noticed that there are a lot of PN junctions in MOSFETs, which probably makes current conduction pretty tricky. It seems important to keep all of them at the right voltage too, if the body were at a higher voltage than the N-well we could get a lot of current flowing in ways we didn't expect. That type of behavior is called latch-up, and it's usually destructive.

One last note before we leave the slide, MOSFETs have two really important dimensions in their layout that affect how they work. One is the channel length, which I've labeled L in the layout view. Another is the channel width, which I've labeled W in the layout view. When you hear that a semiconductor company builds 90nm transistors, that dimension sort of refers to the channel length, though marketing in recent years has obfuscated that connection.



Describing MOSFET operation is easiest to do if we apply a few voltages to terminals and see what happens. We're going to start by applying a voltage between the gate and the body. We're also going to short the body and the source because by convention we measure all voltages relative to the source in a MOSFET.

When the gate-body voltage, which is the same as the gate-source voltage, is zero, then everything in the device behaves like a normal semiconductor: the gate isn't having any special effect on the behavior of carriers. There is a depletion region around the PN junctions at the edges of the drain and the source regions. I've outlined that in the cross section, and I've drawn the drain and source depletion regions merging because they're close together. This region of operation is called cutoff, and we stay in it as long as the gate-source voltage is low enough.

CLICK As we start increasing Vgs, we remain cutoff, but start seeing some carriers moving around. Electric field lines that pass through the gate oxide terminate on fixed negative phosphorous charges and and on minority electron carriers in the P-type wafer. Holes move along the field lines, so they flee from the area under the gate. Electrons do the opposite, and they start building up under the gate. This process is called depletion because we are depleting the P-type carriers underneath the gate. As we do this, the depletion region around the source and drain grows because we're removing positive

charge carriers, so the region needs to expand so that we have more P- atoms for the E field through the gate to terminate on.

CLICK Finally, when the gate-source voltage reaches a critical level called the threshold voltage, which I've labeled VT, we accumulate enough electrons underneath the gate that they outnumber the holes in that region. That means this tiny strip of semiconductor has become N-type since free electrons outnumber free holes. When we create this N-type region is called inversion, and the region itself is called the channel. Connecting the drain and source with a N-type semiconductor means there are no PN junctions between them, which allows a lot more current to flow than when we don't have inversion. Finally, note that the depletion region has grown even more as the gate-source voltage has increased and produced more field lines.

By the way, the threshold voltage, VT, is why we went with the symbol phi\_t for the thermal voltage. This way we don't have two VT symbols running around the class.



Great, so we have this channel, we should try to run some current through it. We do so by introducing a drain-source voltage, which will cause electrons to get sucked from the source to the drain (because electrons move towards high voltages). Note that means a MOSFET is a majority carrier, drift device because electrons are moving in N-type semiconductor. This is the opposite of the base of a BJT, where minority carriers diffuse from the emitter to the collector. MOSFETS operating like this, with a channel connecting the source and the drain, are said to be operating in the linear region. This region is sometimes also called triode.

However, introducing the drain-source voltage causes some changes in the channel. I've used a sort-of overloaded notation where the thickness of the channel indicates the density of charge at different parts of the channel. On the left, by the source, the voltage in the channel is 0V, so charges at the left side feel a full VGS of voltage pulling on them. However, the right side of the channel is connected to VDS, so charges there only feel VGS-VDS, which we sometimes call VGD, of voltage pulling on them. That means less total charge will accumulate in the drain side of the channel because of the reduced voltage applied across the channel.

CLICK If you continue to increase VDS so that it rises above VGS-VT then VGD will fall below VT at the drain side of the channel, so that part of the channel is no longer inverted. This

phenomenon is called pinch-off, and it results in a tiny region at the end of the channel that has a very high field and very few carriers in it. This result seems bad at first glance, because we're getting less current through the channel than we would otherwise, but the pinched off channel also means that drain current doesn't really depend on VDS anymore because VDS just enhances the already high field in the pinched off region. That means a pinched off MOSFET starts to look like a current source, which is a behavior that we've also seen in BJTs. Confusingly, this desirable behavior is called the saturation region in MOSFETs, which shouldn't be confused with the undesirable saturation region in BJT operation.

As a final note, increasing VDS has made the depletion region asymmetric. We have some additional reverse bias on the drain PN junction that makes the depletion region expand.





In this video we're going to calculate the current associated with different operating regions of the MOSFET. This derivation is a bit gratuitous because we're not going to be doing a lot of device physics work in this class, but I think it's a really cool derivation. It goes from simple principles to complex expressions, and it helps illuminate some of the basics of how MOSFETs work.



The big trick to analyzing MOSFET current is to break current down into a charge density multiplied by a velocity. Doing that requires some sense of distance to measure velocity, so I've introduced a distance dimension, y, on the drawing on the left. The y dimension refers to distance along the channel. The source is at y=0 and the drain is at y=L. A tiny little sliver of the channel has width dy, and the voltage in that sliver is labeled V(y).

We start using our charge times velocity formulation in this first line on the right of this slide. Drain current is defined as the change in charge with respect to time, which seems fine, and then we break that into the change in charge with y multiplied by the speed of the charge: how fast the y dimension is changing with time. The Q(y) in our dQ/dt expression can refer to charge anywhere in the channel because we have to have current continuity at every point of the channel.

CLICK OK, finding the differential charge in the channel isn't so bad because the channel is just a funny looking capacitor plate. So we can find the charge on the plate using the capacitance density, Cox, multiplied by the x and y dimensions of our differential, which are W and dy. Recall that W refers to the width of the transistor, and it runs off into the page, hidden behind the cross section we're looking at. The dy in this dQ expression is kind of convenient, it will cancel with the dy that's already in our iD expression.

CLICK Next we need to find the speed at which charge is moving, dy/dt. Fortunately, we have a model for velocity of ballistic conduction, where carriers run down a material scattering off of obstacle, which is that carrier speed is given by their mobility times the electric field. Great, but we'd rather have this in terms of voltages than fields, so we replace the electric field with the derivative of channel voltage with respect to y.

CLICK We can substitute all of this together into a decent looking expression. However, this depends on channel voltages and a differential in y. We'd like to get rid of those so we have an expression in terms of terminal variables. We'll take some integrals to get rid of these field differentials on the next few slides.



I've copied the equation from the last slide here, and as I promised

CLICK we're going to take an integral to clean it up. I've moved the dy differential to the left side of this equation, and we'll get rid of it by integrating across the channel from y=0 to y=L.. The right side of the equation has a dV, so we need to integrate from the voltage at the left of the channel, which is 0 volts, to the voltage at the right side of the channel, which is volta.

CLICK Evaluating this, we get our first expression for drain current, and we find that it's parabolic.

CLICK Great, we have parabolic current voltage curves. That seems sort of fine, but it's a bit weird that our current decreases at high vDS values. However, note that we did this integral in the linear/triode region. Perhaps an integral in the saturation region will clarify what's going on.



So now let's see how the equation changes when we're in saturation. I copied our prototype equation from two slides ago, and we're going to do a different integral that captures saturation behavior. That integral is going to have different bounds because the voltage at the end of the pinch-off region has a fixed value. Technically it will also have a different value because the channel doesn't stretch all the way to y=L too, but we'll talk about that more later.

CLICK We rearrange the equation in the same way, but the bounds of our integral are different. We know that the highest voltage that a channel can achieve is VGS-VT because that's the voltage at which it falls out of inversion. So we integrate along L on the left side of this equation and from 0 to VGS-VT on the right side.

CLICK Evaluating that integral we get this expression for iD, which is now totally independent of vDS. Technically vDS can slip back into this expression through the length of the pinch-off region, but to first order we just control drain current with gate voltage in saturation. It's also interesting that this expression depends quadratically on vGS-VT, which is called the overdrive voltage. One factor of overdrive comes from charge accumulation on the capacitor, and the other comes from the velocity calculation because the channel-end voltage is pinned at vGS-VT.

CLICK That means our overall expectation for MOSFET devices is that the iD-vDS curves will follow a parabolic behavior at low vDS until they saturate and become independent of vDS. Changing vGS will move us to different curves in this family.



I've put everything together on this slide. Equations for iDS in all three regions of operation appear at the top of the slide, and those regions of operation are illustrated on the iD-vDS curves on the bottom of the slide. The equivalent circuit model is pretty simple: it's just a dependent current source between the drain and source nodes. It's controlled by the voltage between the gate and source nodes.

So this is our large signal model. I do want to add one quick reminder that we did this analysis assuming the source and the body were shorted. If the body voltage isn't the same as the source voltage, then the threshold voltage changes slightly because charge in the channel sees a different total potential. That change is called the body effect.





Department of Engineering

## MOSFET Small Signal Model

Matthew Spencer Harvey Mudd College E151 – Analog Circuit Design

In this video we're going to convert our MOSFET large signal model into small signal models.

13



We need to add one detail to our large signal model in order to make that conversion, and that detail is called channel length modulation.

When we did our analysis of MOSFET current in saturation, we assumed that the channel length wasn't changed much by pinch-off. That meant we integrated iD from 0 to L. In reality, pinch-off makes the channel shorted, so it would have been more appropriate to integrate from 0 to L-DeltaL, where Delta L is the size of the pinched off region, and it's a function of vDS. I've included that correction in the first line on the right.

CLICK This Delta L factor is annoying and non-linear, so I've factored it into a correction term over on the right of this expression. Our normal current without channel length modulation appears on the left, and it's multiplied by a correction factor to account for Delta L on the right. The Delta L / L expression in this term is guaranteed to be much less than 1.

CLICK That means we can linearize this term using a binomial expansion. Doing so gives us this expression, where vDS causes a small linear change in iD. This looks a lot like our output resistance expression in BJTs, where the current depends on vCE divided by the Early Voltage.

CLICK In fact, the lambda parameter is one over the Early Voltage for a MOSFET. You can also calculate it directly if you can find how Delta L changes with vDS, but that's quite tough to model.



As usual, we're going to take a bunch of derivatives to find our small signal dependence on different voltages. I'll be filling those relationships into our small signal in the lower right as we find them. We can get a quick start on that modeling process by recalling that there's no gate current in our model, so we won't see any elements connected from the gate node to either the drain or the source.

CLICK Noting that, we're going to continue finding elements by ignoring channel length modulation to find the drain current dependence on the gate-source voltage. I've included our expression for drain current in saturation here as a reminder. We take a derivative of it with respect to vgs and evaluate it at our bias point, and that gives us a kind of ugly expression that's linear with overdrive. Cool, but hard to use unless we have the mu\_n\*Cox\*W/L coefficient memorized. I've also included another expression for gm that involves rearranging the expression for iD to substitute for the overdrive. We used that rearrangement trick a few times with BJTs to make easier expressions for gm, but no luck here. (Though it's fun to note that gm is proportional to the square root of the bias current.)

CLICK Fortunately, we can find an expression for gm that makes our life easier. I've started with our gm expression from the left column, and in the second step I've multipled by overdrive over overdrive, which gives us a promising looking expression because the left

half of it looks almost like our expression for iD. Substituting that in, we find a relationship between ID, overdrive and gm that is easy to remember and usable. It's also a bit redundant because iD and overdrive are one-to-one functions of one another, but this is nice because we don't need to remember a weird constant to use it.

CLICK Regardless, this derivative indicates that vgs changes ids, and because a voltage at one place affects a current somewhere else, so we represent this behavior as a dependent source between the drain and source nodes that is controlled by vgs.

CLICK We add channel length modulation back in to find how our drain current varies with vds. That's because there would be no vDS dependence if we ignored channel length modulation. Taking a derivative of the modified iD expression, we're left with lambda times ID.

CLICK And this looks like an output resistance, so we add ro between the drain and the source. And that component covers all the derivatives in the equation we started with. So this is the small signal representation you'll use most of the time.

CLICK However, it's worth noting that there's one additional component we can add to capture the effect of the body being at a different voltage than the source. The body effect creates another gm generator. This gm generator is sometimes called the backgate effect.





In this video we're going to look at using MOSFETs in some amplifier circuits.



We need to take a careful look at MOSFET amplifiers because large signal analysis for them is trickier than BJTs. For instance, this common source amplifier on the left looks a lot like our common emitter amplifier, but in the common emitter, we could calculate a base current using VB, RS and VBEON. However, VGS doesn't have a fixed value like VBEON. That means the only way to find our drain current is this annoying, quadratic device equation. As a result, it's really common to use a simulator liberally while designing MOSFET circuits.

CLICK However, we can also go back to our usual tricks to analzye gain and swing and get some insight. A load line analysis will show us a qualitative bias point, reveal that we've got some gain that depends on the separation between these curves, show us our swing limits: we'll clip if we enter triode at low vDS or cutoff at high vDS.

CLICK There's another graphical tool that like a lot for MOSFET amplifiers called a transfer characteristic. This is a plot of large signal VOUT vs. large signal VIN. You can clearly see the swing on this plot, it's the range of voltage between the linear region at high VIN and the cutoff region at low VIN. You can also pick out the gain as the slope of the curve in the region between those swing boundaries. One interesting feature of this transfer characteristic is that we don't get pinned at a voltage in the linear region. The collector emitter voltage gets pinned at VCESAT when a BJT saturates, but MOSFET linear regions are

much slower changing than BJT saturation regions.

CLICK Finally, we can do some codesign with our small signal model to take advantage of our fancy new gm expression. Let's imagine that we want a gain of 20, a VOUT of 5V and that we're constrained to have a VDD of 10V and a threshold voltage of 1V. One quick aside, we use VDD with MOSFETS because they have drains, while we use VCC with BJTs because they have collectors.

CLICK OK, we know that the dropout across the resistor has to be 5V, and I know that the gain is gm\*RD. (Feel free to pause the video and prove that's the correct small signal gain for this amplifier, but I'm not going to go over the solution.) We substitute our fancy expression for gm into this expression, and all the sudden we have two pretty simple equations in three unknowns: RD, overdrive and ID.

CLICK I picked ID=1mA because it's a convenient number. That means we need RD to be 5k to meet our dropout constraint. Finally, we can pick a VB value that gets our gain correct: ID\*RD is 5V, so we need VB-VT to be 0.5 to double our gain.



Because the large signal relationship in MOSFETs is weaker than BJTs, it's harder to bias them. For instance, I've included our standard resistor divider bias on the common source with degeneration pictured on the left of this slide. On the upside, our divider equation works really well because there's no gate current. On the down side, when we write a KVL equation to find VGS, we're left with a quadratic equation. We could solve this, but I just go to a simulator because it's a mess.

On the upside, current mirrors are really nice for biasing this circuit. Because there's no gate current, RBIG isn't going to have any current running in it. That means that we can set the large signal VGS of our common source amplifier perfectly. We can do something even cooler by changing the width of our common source resistor relative to the diode connected device. Because its width is twice that of the diode connected device, we know that it will have double the current for the same VGS because of the W/L factor in the iD expression. That lets us bias multiple amplifier stages at different levels with one current source.

