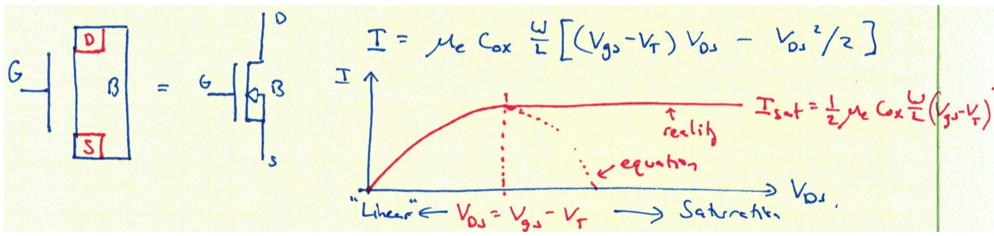
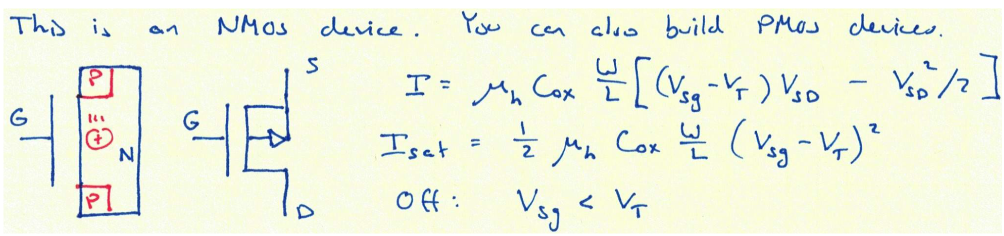
Lecture Notes 14: PMOS, Large and Small Signal Analysis

Recall that we derived the MOSFET $i - v$ equation:



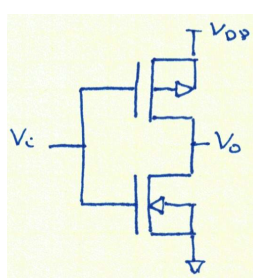
This is called an **NMOS** device because, when it conducts, current flow from an N region (the drain) to an N region (the inverted channel) to an N region (the source), and the charge carriers are electrons.

You can also build a **PMOS** device by swapping the P and N regions and applying positive v_{sg} (instead of v_{gs} and positive v_{sd} (instead of v_{ds}). You get exactly the same equations:



Note that some people use different notation for PMOS devices. They continue to express things in terms of v_{gs} and v_{ds}. You would apply a negative v_{gs} and the threshold voltage itself would be negative in this formulation. The drain-source voltage v_{ds} would then be negative (flowing from S to D). I find this way of expressing things confusing.

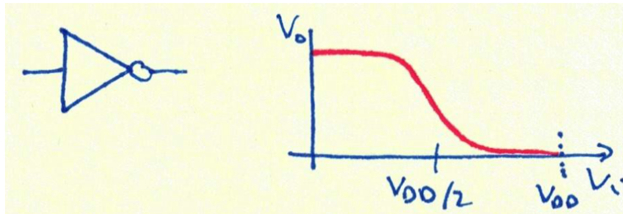
The mixed use of PMOS and NMOS devices is called **CMOS**, where the C stands for Complimentary. Consider, for example, a PMOS device stacked on top of an NMOS device, with a voltage source V_{DD} applied from the top of the stack to the bottom of the stack (ground). (The voltage source is usually not shown; the top of the stack is simply labeled V_{DD}). The two gates are tied together and called the "input;" the two drains are tied together and called the "output."



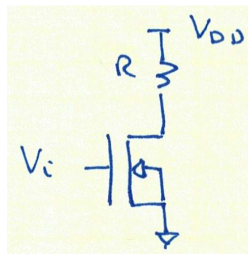
When $v_i \approx V_{DD}$, the PMOS turns off since its $v_{sg} < V_T$ while the NMOS is on since its $v_{gs} > V_{th}$. The output voltage $v_o \to 0$, which means the NMOS device is deeply in the linear region.

When $v_i \approx 0$, the PMOS is on since its $v_{sg} > V_{th}$ and the NMOS is off since its $v_{gs} < V_{th}$. The output voltage $v_o \to V_{DD}$.

This is a device that takes in a logical `1` and outputs a logical `0` and vice versa $\Rightarrow$ an **inverter**!



Here we operate the MOSFETs as switches. Other circuits take advantage of the saturation region, like this **common-source amplifier** below.



If the MOSFET is in saturation (which we have to check!), then $I = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_i)^2$. The output voltage right above the MOSFET is then

$$v_o = V_{DD} - R \times \frac{1}{2}\mu_n C_{ox}\frac{W}{L}v_i^2$$

As v_i goes up or down a little, v_o goes down/up a lot. This is an amplifier! (and an inverting one, since the output goes the opposite direction from the input).

How do we know which region the MOSFET will be in? We don't, and there's often no *a priori* way to say. Instead, we use the **method of assumed states**, which is dumber than it sounds. We first *assume* that the MOSFET is in a certain state (or region), solve the problem with that assumption, and then check to see if all of the results are consistent with the assumption. If the results are consistent with the assumption, then we were right. If the results are not consistent, then we were wrong and we need to assume a different state.

For example, consider a common-source amplifier with

$$\begin{array}{cccc} \mu_n = 1000\text{cm}^2/V\cdot s & C_{ox} = 5\mu\text{F}/\text{cm}^2 & W/L = 10 & V_{th} = 0.4 \\ v_i = 1.0\text{V} & V_{DD} = 5\text{V} & R = 550\Omega & \end{array}$$

Step 1: Assume a state and solve. Let's assume that the NMOS will be in the saturation region. If this is true, then $I = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_i)^2 = 9\text{mA}$ and $v_o = V_{DD} - RI = 0.05\text{V}$.

Step 2: Examine whether $v_{ds} > v_{gs} - V_{th}$ (consistent with our assumption of saturation region) or whethe $v_{ds} < v_{gs} - V_{th}$ (inconsistent with our assumption). In this case, $0.05\text{V} < 1\text{V} - 0.4\text{V} = 0.6\text{V}$. Therefore, the solution is not consistent with our assumption. We guessed wrong.

Step 3: If the assumed state were consistent with the answer, we would be done. If the assumed state isn't consistent with the answer, then we need to start over with a new assumption. This time, let's try the NMOS being in the linear region. Then $I = \mu_n C_{ox}\frac{W}{L}[(v_i - V_{th})v_o - v_o^2/2]$ and $v_o = V_{DD} - R\mu_n C_{ox}\frac{W}{L}[(v_i - V_{Th})v_o - v_o^2/2]$. Taking everything to one side gives us a quadratic equation,

$$\underbrace{\left(\frac{1}{2}R\mu_n C_{ox}\frac{W}{L}\right)}_{13.75/V} v_o^2 - \underbrace{\left(R\mu_n C_{ox}\frac{W}{L}(v_i - Vth) + 1\right)}_{17.5} v_o + \underbrace{V_{DD}}_{5V} = 0$$

Therefore,

$$v_o = \frac{17.5 \pm \sqrt{17.5^2 - 4 \times 13.75 \times 5}}{2 \times 13.75} = \begin{cases} 0.84\,\text{V} & \text{(inconsistent with assumed state)} \\ \underline{\underline{0.43\,\text{V}}} & \text{(consistent with assumed state)} \end{cases}$$

(By contrast, if R had been 200Ω, then the saturation guess would have yielded $v_o = 5 - 9mA \times 200\Omega = 3.2V$ which would have been $> 0.6V$ and therefore consistent.)

The equations we've derived are valid for the full range of MOSFET operation (**large signal** equations). But they're nonlinear, so our intuition isn't as good and we lose powerful analysis tools like superposition that only apply to linear systems.

One of the most common things we do to understand nonlinear systems is to develop a linear system that is approximately equal to the nonlinear system. The linear approximation will only be exactly true at one point and only approximately true for small deviations away from that point (**small signal**). Nevertheless, the advantages of working with a linear system make it worth it.

Let us do this for the MOSFET in saturation. Consider each variable to be an **operating point** (the point at which the linear approximation will be exactly true) plus small deviations away from that operating point.

$$i \to I + \Delta i \tag{7}$$
$$v_{gs} \to V_{gs} + \Delta v_{gs} \tag{8}$$
$$v_{ds} \to V_{ds} + \Delta v_{ds} \tag{9}$$

In the saturation region, the large signal (always true) equation is

$$I + \Delta i = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}[V_{gs} + \Delta v_{gs} - V_{th}]^2$$

Now expand the right hand side and collect together (1) terms with only operating-point values, (2) terms with only *Delta*-values, and (3) other terms.

$$I + \Delta i = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})^2 + \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(\Delta v_{gs})^2 + \mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})\Delta v_{gs}$$

Remember that the original large-signal equation is obeyed exactly at the operating point. This means that $I = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})^2$ and these terms cancel from the equation:

$$\cancel{I} + \Delta i = \cancel{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})^2} + \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(\Delta v_{gs})^2 + \mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})\Delta v_{gs}$$
$$\Delta i = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(\Delta v_{gs})^2 + \mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})\Delta v_{gs}$$

So far, we actually haven't made any approximations. This equation is an *exactly true* equation for Δi (the small deviation in current away from its operating point). Our goal is to get a linear system and $(\Delta v_{gs})^2$ is the offending nonlinear term. Fortunately, as long as the deviations away from the operating point (Δ anything) are small, then terms like $(\Delta v_{gs})^2$ are *very* small and can be approximated as zero. Doing this yields

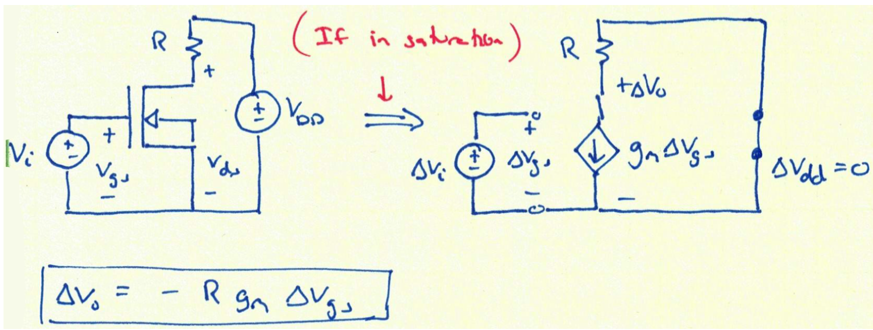
$$\Delta i = \underbrace{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(\Delta v_{gs})^2}_{\text{approximated as very small}} + \mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})\Delta v_{gs}$$

$$\Rightarrow \boxed{\Delta i \approx \mu_n C_{ox}\frac{W}{L}(V_{gs} - V_{th})\Delta v_{gs} = g_m \Delta v_{gs}}$$

where the constant terms mushed together convert a voltage in one place (gate-to-source) to a current in another place (drain-to-source) and are therefore called the **transconductance g_m**.

This is the small-signal approximation for the MOSFET. It does not tell us about i and v. It only tells us about their deviations from the operating point, Δi and Δv. But it's linear!

The small signal model for the MOSFET is a dependent current source with value $g_m \Delta v_{gs}$. If we apply small signal modeling to an entire circuit, we obtain:



The constant voltage source V_{DD} never deviates (even if v_{gs} is squiggled), so ΔV_{DD} is always zero, which means that the small signal model for a constant voltage source is a short-circuit. The small signal model for a resistor is a resistor ($\Delta v_R = R\Delta i_R$). The small signal model for the (possibly varying) gate-source voltage source is just a voltage source of value Δv_{gs}. Including the MOSFET model, we get a very simle circuit, and its solution is found extremely rapidly:

$$\Delta v_o = -Rg_m \Delta v_{gs}$$