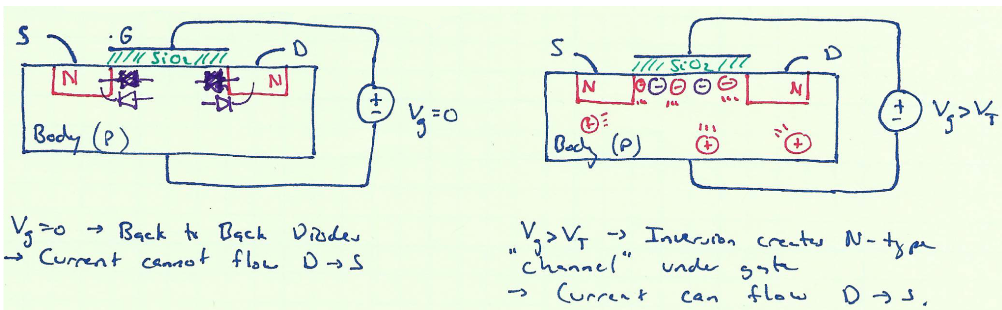
Lecture Notes 13: MOSFETs

We now have all of the building blocks to make a controllable transistor.

1. We know that conductance is governed by several factors, especially the concentration of mobile charges.
2. We know that current will not flow from N-type to P-type materials.
3. We know that we can change P-type material to N-type material using an electric field ("inversion")

Consider the following strategy:



An MIS capcitor is formed between a metal plate (the **gate**) separated from a P-type semiconductor (the **body**) using an insulator (usually SiO_2). Two N-type regions (the **drain** and **source**) are formed on the surface of the semiconductor adjacent to (but not directly under) the gate. The area directly under the gate, which is still P-type, is called the **channel**.

We want to examine whether it's possible for current to flow from D to S. Clearly, if no voltage is applied to the gate, no charge can flow from $D \to S$ since the drain-channel junction is $N \to P$. Charge couldn't flow from $S \to D$ either since the source-channel junction is also $N \to P$. These back-to-back diodes prevent current from flowing when the gate-to-body voltage $V_{gb} = 0$. (We will often take the body to be grounded, and simply refer to "the" gate voltage V_g).

However, if we apply voltage to the gate, the electric field will push all the holes out of the channel. Once the gate-body voltage exceeds the threshold voltage V_{th}, mobile electrons will be attracted into the channel. At this point, the channel is N-type, and there is no PN junction between drain-channel nor between source-channel. Current can now flow freely from the drain, through the channel, all the way to the source (or vice versa).

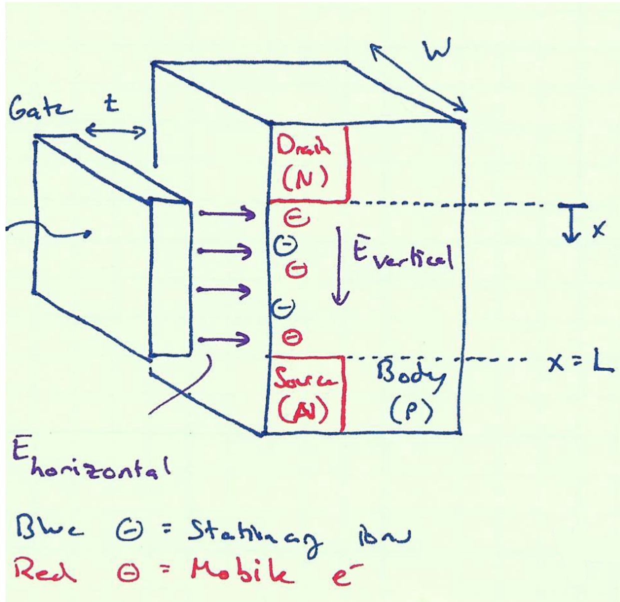
This is a **MOSFET**: a (M)etal (O)xide (S)emiconductor (F)ield (E)ffect (T)ransistor. Let's turn the MOSFET on its side so it looks like its symbol:

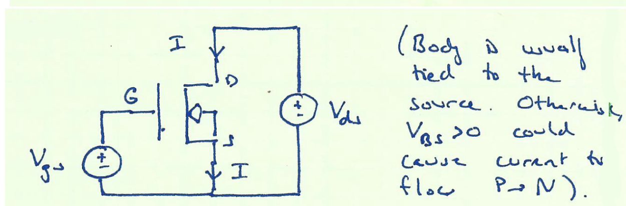
Let's also assign a coordinate system such that x points vertically down from the edge of the drain. The length of the channel from the edge of the drain to the edge of the source is L, the thickness of the insulator or oxide is t, and the MOSFET's width (into the page) is W.

Our goal is to find current, I, as a function of v_{gb} and v_{ds}. The most basic and most common configuration is to have the body electrically connected to the source with a wire, so we usually talk about the gate-to-source voltage v_{gs} instead of the gate-to-body voltage v_{gb}. Thus, we are looking for $I(v_{gs}, v_{ds})$.

Current is the derivative of charge, which we can split using the chain rule into the charge per unit length multiplied by velocity:

$$I = \frac{dq}{dt} = \frac{dq}{dx}\frac{dx}{dt} = \lambda(\text{linear charge density}) \times \text{velocity}$$



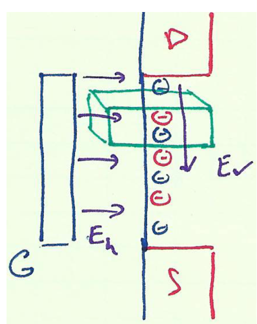


Velocity is easy;

$$\frac{dx}{dt} = \underbrace{-\mu_e E_v(x)}_{\text{neg b/c electrons}} = \underbrace{+\mu_e \frac{dv(x)}{dx}}_{\text{pos b/c } E = -dv/dx}$$

where E_v is the vertical electric field (the electric field pointing from drain to source caused by v_{ds}.

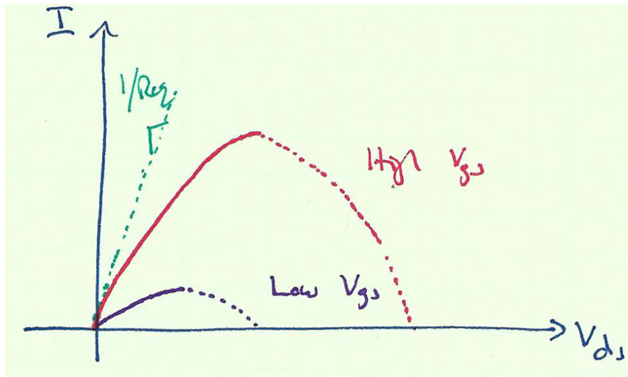
The charge distribution $\lambda(x)$ is a little trickier. To find it, think of "dq" as a tiny bit of charge contained in a tiny slice of the channel "dx." Do Gauss's law on that tiny dx alone:



$$\oint \epsilon E \cdot dA = \underbrace{-\epsilon_{ox} E_h \times dx \times W}_{\text{left side, neg b/c E } entering} + \underbrace{0}_{\text{right, b/c E}_\text{h}=0 \text{ past surface}} + \underbrace{0}_{\text{front/back}} + \underbrace{0}_{\text{top/bottom cancel out}}$$

where ϵ_{ox} is the permittivity of the oxide (or other insulator). Therefore,

$$dq(x) = -\epsilon_{ox} E_h(x) W dx$$



where we have rigorously done Gauss' law to obtain the negative sign, but we also knew the charges would be negative from the beginning.

If the potential at any given location of the channel is $v(x)$, then $E_h(x) = v_{\text{gate-to-channel}}/t = (v_g - v(x))/t$. Therefore, $dq(x) = -\epsilon_{ox} W dx \times (v_g - v(x))/t$.

> Now is the time to be very careful. The dq we just calculated came from Gauss's law and therefore includes *all* of the charge at the semiconductor surface. But we don't want to know the amount of stationary charge (which won't contribute to conduction) – we only want the mobile charges.

We know how much stationary charge there is because *voltages up to V_{th} result in stationary ions*, the amount of stationary charge is $-\epsilon_{ox} W dx \times V_{th}/t$. Whatever's left must be mobile charge, so

$$\boxed{\underbrace{dq(x)}_{\text{mobile e}^- \text{ only}} = \underbrace{-\epsilon_{ox}\frac{v_g - v(x)}{t} W dx}_{\text{total charge}} - \underbrace{\left[-\epsilon_{ox}\frac{V_{th}}{t} W dx\right]}_{\text{stationary charge}} = -\epsilon_{ox}\frac{v_g - v(x) - V_{th}}{t} W dx} \tag{6}$$

Put it all together:

$$I = \frac{dq}{dx}\frac{dx}{dt} = \left[-\epsilon_{ox}\frac{v_g - v(x) - V_{th}}{t} W\right] \times \left[+\mu_e \frac{dv(x)}{dx}\right] = -\mu_e \frac{\epsilon_{ox}}{t}\frac{W}{L}(v_g - v(x) - V_{th})\frac{dv}{dx}$$

Now let's integrate. If we take dx to the left hand side, we can integrate both sides. The left hand side integral is over x, and let's integrate it from the drain to the source, i.e. from $x = 0$ to $x = L$. We must also integrate the right hand side from the drain to the source; this means integrating from $v(x) = v_{ds}$ to $v(x) = 0$. Therefore:

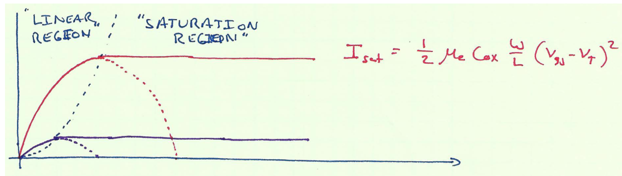
$$\int_0^L I dx = -\mu_e \frac{\epsilon_{ox}}{t} W \int_{v_{ds}}^0 (v_g - v(x) - V_{th}) dv$$

$$\boxed{I = +\mu_e \frac{\epsilon_{ox}}{t}\frac{W}{L}\left[(v_{gs} - V_{th})v_{ds} - \frac{v_{ds}^2}{2}\right]}$$

This final equation is sometimes written with $\epsilon_{ox}/t \rightarrow$"$C_{ox}$".

For a given v_{gs}, this equation is quadratic in v_{ds}. The current starts at 0 and quadratically rises to a peak. In principle, the current then falls again, but we'll learn in just a moment that doesn't really happen. The equation is good before the peak, though.

Before we look at what happens past the peak, let's note two important facts.



1. For small v_{ds}, I is approximately linear, with

$$\boxed{I \approx \mu_e C_{ox} \frac{W}{L} (v_{gs} - V_{th}) v_{ds}}$$

   This is called the **linear region** or **triode region**. The MOSFET behaves like a gate-controlled resistance with

$$R_{eq} = \frac{1}{\mu_e C_{ox} \frac{W}{L} (v_{gs} - V_{th})}$$

2. The current maxes out at $v_{ds} = v_{gs} - V_{th}$, and it maxes out at a value of $I_{max} = \frac{1}{2} \mu_e C_{ox} \frac{W}{L} (v_{gs} - V_{th})^2$.

Now, the model predicts that current should go down for $v_{ds} > v_{gs} - V_{th}$, but thid doesn't make intuitive sense – why should we get less current with more voltage and more electric field? $\Rightarrow$ **Either the model breaks down or our intuition is bad**.

In this case, the model breaks down. Recall that the linear charge density is

$$\frac{dq}{dx} = \lambda = -\epsilon_{ox} \frac{v_{gs} - v(x) - V_{th}}{t} W$$

Right at the problematic voltage, $v_{ds} = v_{gs} - V_{th}$, the charge density at the drain (where $v(x) = v_{ds}$ hits zero.

Aha! The model from before predicts that current should go down for $v_{ds} > v_{gs} - V_{th}$ because it thinks that the charge density near the drain will start to become positive. But this clearly won't happen! Instead, the model breaks down at $v_{ds} = v_{gs} - V_{th}$ (aka, **pinchoff**), and we need a new model that is valid beyond that point.

After pinchoff, the current can't increase but it doesn't decrease either. The correct model is that the current **saturates** at its maximum value and stays there for all $v_{ds} > v_{gs} - V_{th}$.

We therefore have the linear region (actually quadratic) for $v_{ds} < v_{gs} - V_{th}$ and $I < \frac{1}{2} \mu_e C_{ox} \frac{W}{L} (v_{gs} - V_{th})^2$. Then, we have a **saturation region** for $v_{ds} > v_{gs} - V_{th}$ with $I = \frac{1}{2} \mu_e C_{ox} \frac{W}{L} (v_{gs} - V_{th})^2$. Because the peak current is also the saturation current, is known as $\boldsymbol{I_{sat}}$.

The MOSFET can thus be operated in two ways:

1. As a switch. Apply high v_{gs} to turn the MOSFET into a small resistance, which will make its v_{ds} small. Apply $v_{gs} = 0$ to turn the MOSFET "off" (the device people call this **cutoff**). You now have a controllable switch! MOSFETs are most often used this way in digital and power circuits.

2. As a voltage-controlled current source. Apply a modest v_{gs} and high-ish v_{ds} such that the MOSFET is in the saturation region. Now the MOSFET drain-source current does not depend on v_{ds} (unless you exit the saturation region), i.e. it acts as a current source. The amount of current is controllable by adjusting v_{gs}. MOSFETS are most often used this way in analog and rf circuits.