MOSFETs

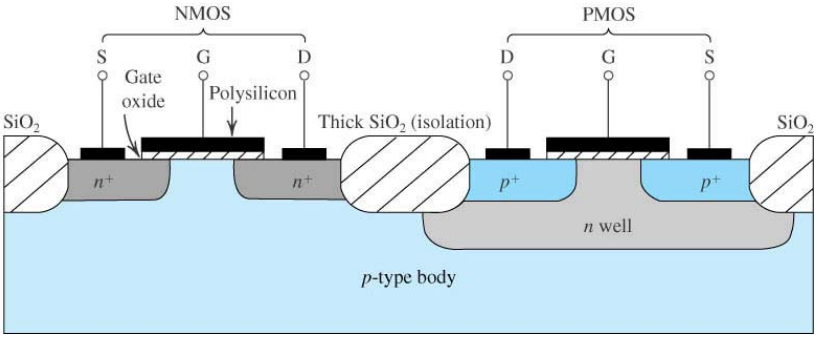
Reading

S&S (5ed): Sec. 4.1 - 4.3
S&S (6ed): Sec. 5.1 - 5.3

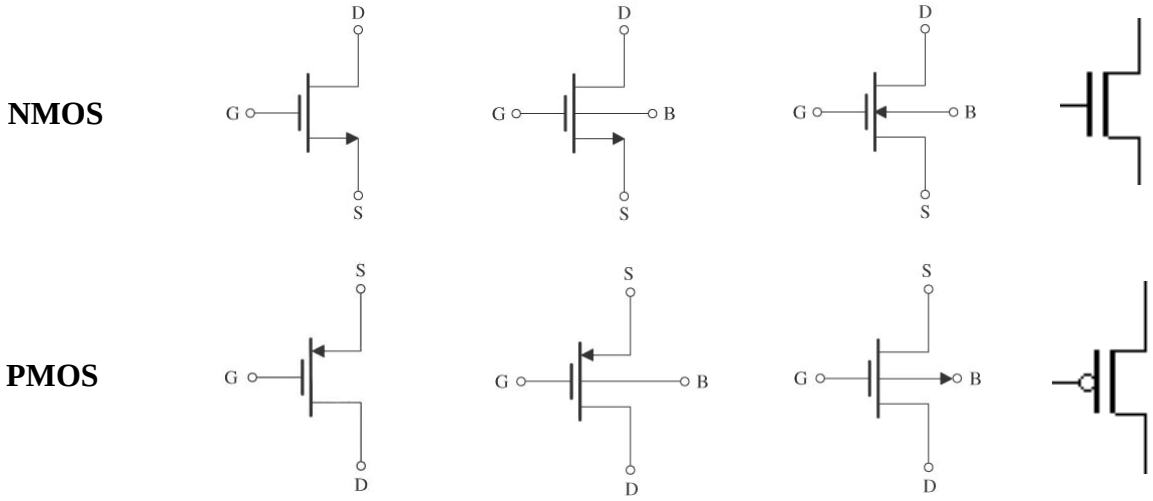
Basics

MOS field effect transistors (MOSFETs) are 4-terminal active elements.

G: gate
S: source
D: drain
B: body (well or substrate)



nMOS pMOS



Unlike BJTs, the MOSFET is symmetric (ignoring effects of advanced device fabrication). Which terminal is the S or D is determined entirely by the applied voltages.

For the NMOS: $v_S \leq v_D$
For the PMOS: $v_S \geq v_D$

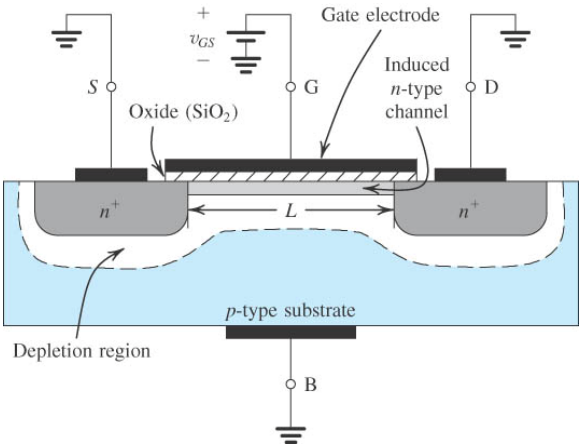
For correct operation, the voltage of the B of the NMOS cannot be higher than that of the S. For the PMOS, the B voltage cannot be lower than that of the S. (The reason is that otherwise, the S-B and D-B diodes would be forward biased and this conduct). This can be accomplished in two ways: (1) connect the B to the S, or (2) connect it to the lowest or highest voltage in the system (for NMOS and PMOS respectively).
For now, we will assume that the B is always connected to the S, in which case we can ignore the effect of the B (otherwise, the threshold voltage will depend on the B voltage). As a result, we essentially treat the MOSFET as a 3-terminal device (having S, D and G).

Physical Operation

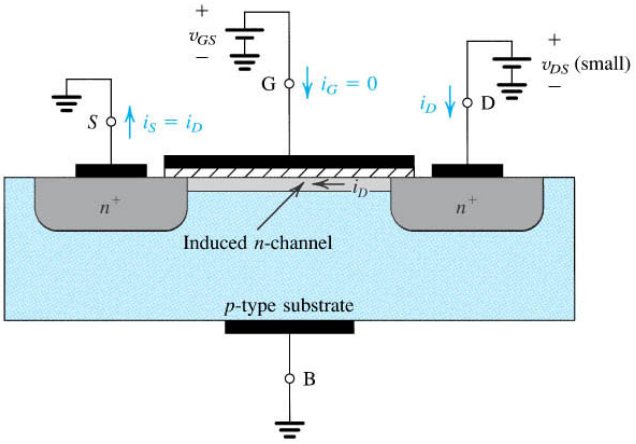
MOSFETs are non-linear devices. Although the physical operation of MOSFET devices is fairly complex when studied in detail, we can still gain valuable insights by looking at a very high-level view of the processes that govern their behavior. As an example, consider an nMOS.

Initially, we apply zero volts between S and D, i.e. $v_{DS} = 0$. The G and B together form a capacitor. When a small positive voltage is applied to the gate (as a result, positive charge accumulates on the gate electrode), the negative charge on the B-side of the capacitor comes from a depletion region.

[In an nMOS, the B is p-type, i.e. silicon doped with atoms that take bond with electrons, thereby leaving excess holes, while these doping atoms themselves become negative ions (note that the total charge is still neutral). In the depletion layer, the holes are repelled by the electric field, leaving only the negative ions behind to provide the negative charge on the capacitor.]

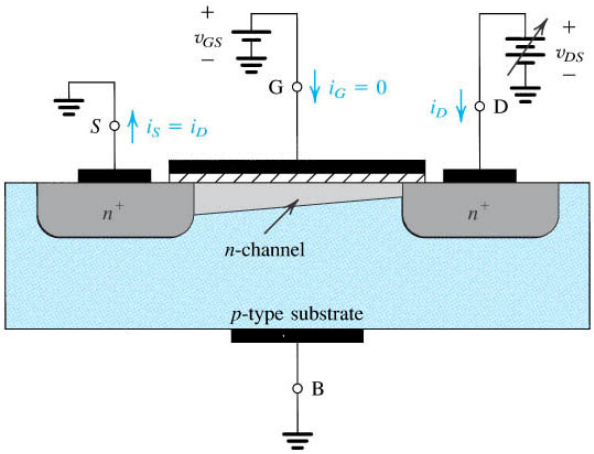


When the gate voltage is increased, the depletion layer grows. However, as the gate voltage keeps on increasing, at some point, the incremental negative charge mirroring the incremental positive charge on the gate electrode will be due to free electrons (remember a p-type material still has some free electrons, although they are minority carriers, as they are outnumbered by the holes). This happens at a gate voltage value called the threshold voltage V_{th}. At his point, a channel consisting of electrons (n-channel) is induced just below the gate dielectric, essentially forming a conductive layer between the S and D. The thickness of this channel is proportional to $(V_{GS} - V_{th})$.

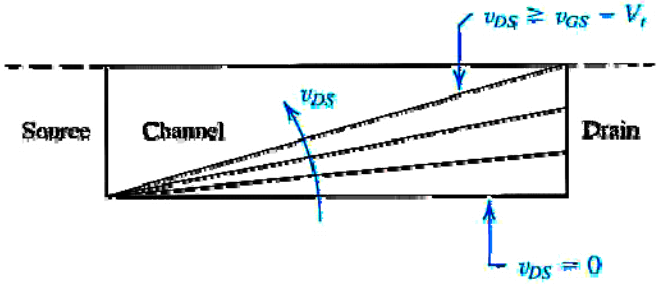


Now when a small voltage is applied at the D, a current will flow between S and D, proportional to V_{DS} and the thickness of the channel. Note that there is never a current at the gate (i.e., $i_G = 0$), and all the current flows between the S and D terminals (i.e., $i_S = i_D$).

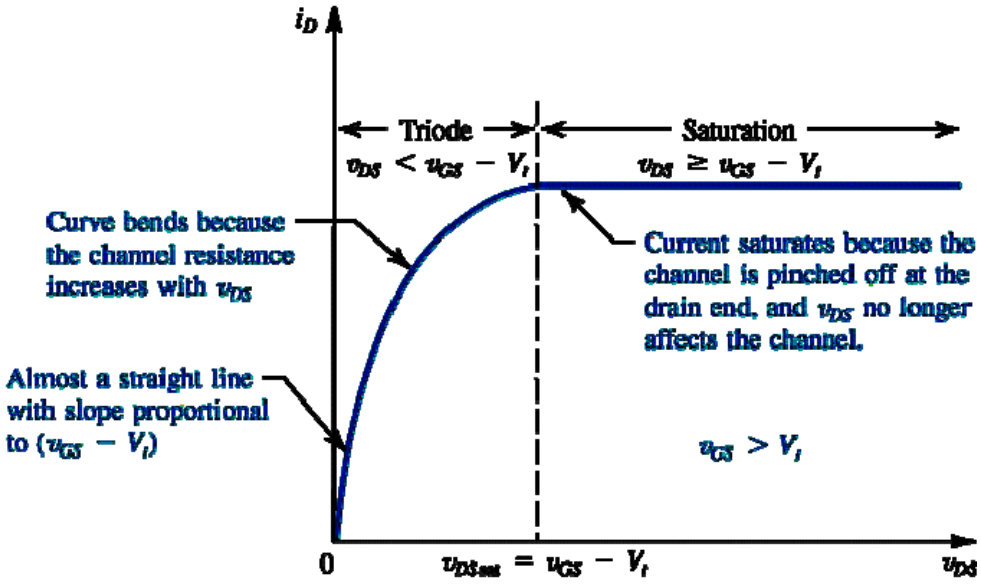
However, as the D voltage increases, the channel becomes narrower at the drain (resulting in increased resistance). The reason is that the voltage drop along the capacitor plate is lower there: it is $V_G - V_D$ instead of $V_G - V_S$ on the S side. This tapered nature of the channel causes the current between S and D to grow less than linear with V_{DS}.



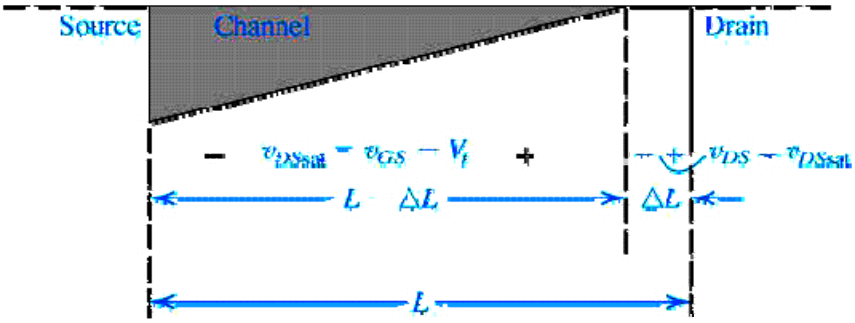
As soon as the voltage drop between the two capacitor plates at the drain side is such that no more free electrons are available there, the channel is pinched off. This happens when $V_G - V_D = V_{th}$. Or equivalently, $V_D = V_G - V_{th}$. Or equivalently, $V_D - V_S = V_G - V_S - V_{th}$. This value is called $V_{DSsat} = V_{GS} - V_{th}$, and thus depends on the gate voltage. When the drain voltage is increased further, the channel remains pinched off, and the current no longer increases (it does not go to zero: the channel is still there, except for the last miniscule part near the drain, and electrons 'jump' that little gap). When V_D increases further, the extra voltage drop will be over this little gap, and (as a first order approximation) the current between S and D remains the same.

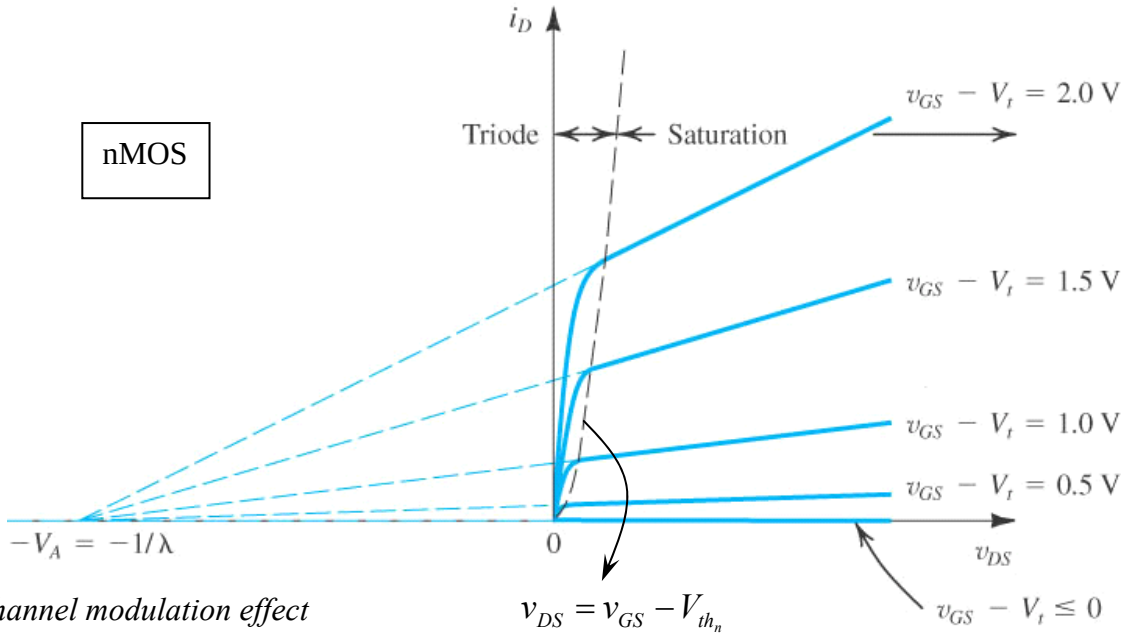


The resulting curve of current versus V_{DS} is shown below, for one value of V_{GS}. Before the channel gets pinched off, the MOSFET is said to be in triode. On the other hand, once $V_{DS} \geq V_{DSsat}$, the MOSFET is in saturation. When there is no channel induced, i.e. $V_{GS} < V_{th}$, the MOSFET does not conduct current and is said to be in cutoff.



In saturation, we said that the current between source and drain does not vary anymore with V_{DS}, since the channel is pinched off. However, in practice, the extra V_{DS} beyond V_{DSsat} manifests itself as a small decrease in the effective length of the channel. This effect is called "channel length modulation". It results in a slight increase in $i_S = i_D$ with V_{DS} when the MOSFET is in saturation.





Channel modulation effect

As explained above, MOSFETs can be considered as operating in a set of different *modes*. To model their overall behavior, we do this for each of these modes separately.

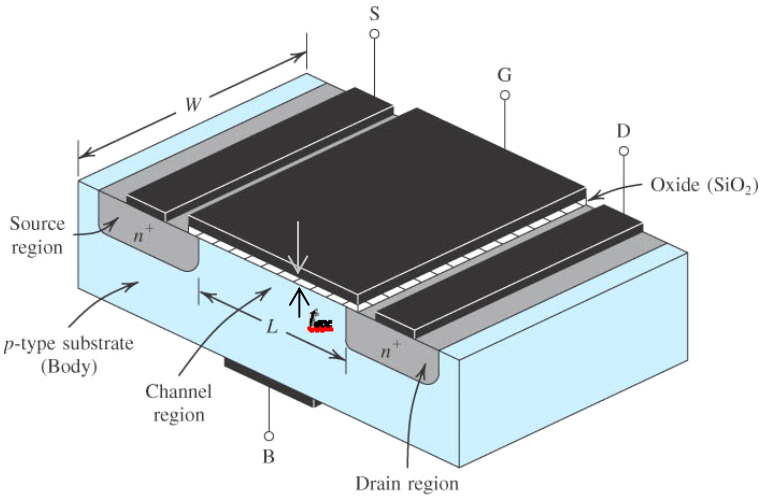
Different models have been proposed for each of these modes (triode, cutoff, saturation), with varying levels of complexity. For hand-calculations, we will mainly use a set of models, which are introduced in the next few pages.

To solve circuits with MOSFET devices, these models lend themselves to the following solution strategy

1. Hypothesis: assume one of the modes of operation for the MOSFET
2. Solve: Use the equations for the selected mode to solve the circuit
3. Check: at the end perform the check for the selected mode to verify the hypothesis
4. Redo: if the hypothesis check fails, try another hypothesis and start over

We will now introduce the different modes of operation of the MOSFETS and the models. For each mode, we will list the 'equations' (used in step 2 above) and the 'check' (used in step 3).

There actually exist a number of slightly different versions of these models. It is possible that you may have learned one of these other versions in a prior course. However, we will list the models as you will need them in this course (which you need to use on all quizzes and exams). Also, use the terms **saturation, triode and cut off** for the different modes!



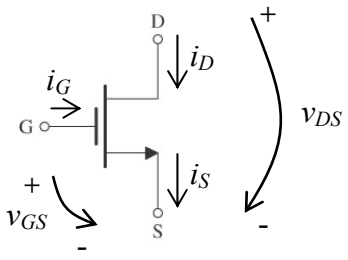
The models use the following parameters, which relate to the physical dimensions of the device or material properties.

W	Width of the device
L	Length of the channel
$\frac{W}{L}$	Aspect ratio
$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$	Oxide capacitance (capacitance per unit area)
t_{ox}	Oxide thickness
ε_{ox}	Oxide permittivity (physical material constant)
μ_n	Electron mobility (physical material constant)
μ_p	Hole mobility (physical material constant)
V_{th_n}	Threshold voltage of nMOS
V_{th_p}	Threshold voltage of pMOS
$k'_n = \mu_n \cdot C_{ox}$	
$k'_p = \mu_p \cdot C_{ox}$	

Sometimes, you find equations with parameter $K = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} = \frac{k'_n}{2} \cdot \frac{W}{L}$ or $K = \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} = \frac{k'_p}{2} \cdot \frac{W}{L}$.

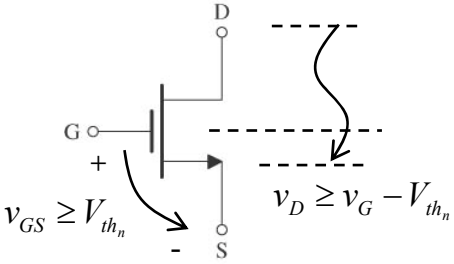
Operational Modes of nMOS

Voltage and current conventions:



Saturation (also called **active**)

equations:
$$\begin{cases} i_G = 0 \\ i_S = i_D = \dfrac{k_n'}{2} \cdot \dfrac{W}{L} \cdot (v_{GS} - V_{th_n})^2 \cdot (1 + \lambda \cdot v_{DS}) \end{cases}$$



check:
$$\begin{cases} v_{GS} \geq V_{th_n} \\ v_{DS} \geq v_{GS} - V_{th_n} \quad \text{or equivalently } v_D \geq v_G - V_{th_n} \end{cases}$$

Triode (also called **ohmic**)

equations:
$$\begin{cases} i_G = 0 \\ i_S = i_D = k_n' \cdot \dfrac{W}{L} \cdot \left[(v_{GS} - V_{th_n}) \cdot v_{DS} - \dfrac{v_{DS}^2}{2} \right] \end{cases}$$

check:
$$\begin{cases} v_{GS} \geq V_{th_n} \\ v_{DS} < v_{GS} - V_{th_n} \quad \text{or equivalently } v_D < v_G - V_{th_n} \end{cases}$$

Cutoff

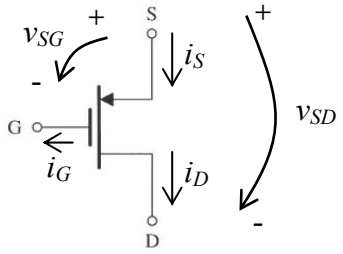
equations: $i_G = i_S = i_D = 0$

check: $v_{GS} < V_{th_n}$

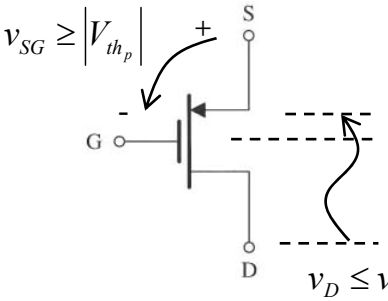
Note that in this course we only consider so-called ‘long channel’ devices. In practice, most integrated circuits contain ‘short channel’ devices (i.e. the channel length is small). In that case, there is an effect called velocity saturation, which results in different equations in saturation mode.

Operational Modes of pMOS

Voltage and current conventions:
$V_{thp} < 0$ (Enhancement pMOS)



Saturation (also called **active**)



equations:
$$\begin{cases} i_G = 0 \\ i_S = i_D = \dfrac{k'_p}{2} \cdot \dfrac{W}{L} \cdot (v_{SG} - |V_{th_p}|)^2 \cdot (1 + \lambda \cdot v_{SD}) \end{cases}$$

check:
$$\begin{cases} v_{SG} \geq -V_{th_p} = |V_{th_p}| \\ v_{SD} \geq v_{SG} - |V_{th_p}| \text{ or equivalently } v_D \leq v_G + |V_{th_p}| \end{cases}$$

$v_D \leq v_G + |V_{th_p}|$

Triode (also called **ohmic**)

equations:
$$\begin{cases} i_G = 0 \\ i_S = i_D = k'_p \cdot \dfrac{W}{L} \cdot \left[(v_{SG} - |V_{th_p}|) \cdot v_{SD} - \dfrac{v_{SD}^2}{2} \right] \end{cases}$$

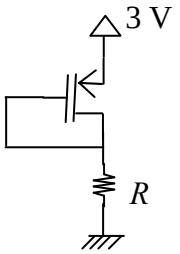
check:
$$\begin{cases} v_{SG} \geq -V_{th_p} = |V_{th_p}| \\ v_{SD} < v_{SG} - |V_{th_p}| \text{ or equivalently } v_D > v_G + |V_{th_p}| \end{cases}$$

Cutoff

equations: $i_G = i_S = i_D = 0$

check: $v_{SG} < -V_{th_p} = |V_{th_p}|$

Example



$$\begin{cases} V_{th_p} = -0.6 \text{ V} \\ k'_p = \mu_p \cdot C_{ox} = 200 \ \dfrac{\mu A}{\text{V}^2} \\ W = 4 \ \mu m \\ L = 0.8 \ \mu m \\ \lambda = 0 \end{cases}$$

Find R such that i_D = 80 μA.

GS-KVL: $v_{SG} = v_{SD}$
DS-KVL: $3 = + v_{SD} + Ri_D$

Since $i_D > 0$, pMOS is NOT in cut-off. Furthermore, since $v_{GS} = v_{SD}$, pMOS should be in saturation (prove it!). Then (since λ = 0)

$$i_S = i_D = \frac{k'_p}{2} \cdot \frac{W}{L} \cdot (v_{SG} - \left|V_{th_p}\right|)^2$$

$$80X10^{-6} = \frac{200X10^{-6}}{2} \cdot \frac{4}{0.8} \cdot (v_{SG} - 0.6)^2 = 500X10^{-6}(v_{SG} - 0.6)^2$$

$$v_{SG} = 0.6 \pm 0.4$$

Which lead to two roots of 1.0 and –0.2 V. Since v_{SG} has to be positive and > $|V_{thp}|$, v_{SG}=*1.0* V. Substituting in DS-KVL gives:

$3 = 1. + R \, X \, 80 \, X \, 10^{-6}$ or R = *2*5 kΩ