MOSFET Device Operation

Enhancement-mode $n$MOS transistor cross-section

$Holes$ are repelled from the gate by positive $V_{GS}$ ($n$MOSFET)

At the onset of INVERSION, $electrons$ attracted under the gate to form channel.

For a depletion-mode $n$MOS, area under gate is actually a lightly doped $n$-type material so that threshold voltage is $< 0V$. 
MOSFET Structure versus Bias

Cross-section (a): potential in channel same everywhere because $V_{GS} = V_{GD}$, channel "depth" same everywhere since $V_{GS} > V_T$ and $V_{GD} > V_T$

Cross-section (b): Depth of channel varies somewhat linearly with $V_{GS}$ and $V_{DS}$. As $V_{DS}$ is increased, the drain-side of channel (just beneath the gate) becomes "pinched" because $V_{GD}$ becomes less and less.

Cross-section (c): Here the current depends only on $V_{GS}$ and not $V_{DS}$ (if we neglect channel-length modulation) and the channel becomes completely pinched-off near the drain. With $V_{DS} > V_{GS} - V_T$ but $V_S = 0V$, then $V_D > V_G - V_T$ and hence, $V_T > V_{GD}$, i.e., $V_{gate-to-drain}$ is less than the threshold voltage.

How does conduction occur after "pinch-off"? Electrons enter channel from source, then are swept across depletion region near drain by the positive drain voltage with respect to source ($V_{DS}$).
**MOSFET Threshold Voltage**

\[ V_T = V_{T-MOS} + V_{fb} \]  
\((V_{T-MOS} \text{ is positive for } n\text{MOS, negative for } p\text{MOS})\)

\(V_{T-MOS}\) — ideal threshold voltage for a MOS capacitor (the capacitor formed between the gate and substrate)

\(V_{fb}\) — Flatband voltage

\[ V_{T-MOS} = 2\phi_b + \frac{Q_b}{C_{ox}} \]  
(Note: "\(Q_b\)" sometimes referred to as "\(Q_{bo}\)"

\[ \phi_b = \frac{kT}{q} \ln \left( \frac{N_A}{n_i} \right) \]  
\(\phi_b\) — bulk Fermi potential

\[ C_{ox} = \text{oxide capacitance, inversely proportional to oxide thickness} \left( \frac{C_{ox}}{t_{ox}} = \frac{\varepsilon_{ox}}{t_{ox}} \right) \]

\[ Q_b = \sqrt{2\varepsilon_{si} \cdot q \cdot N_A \cdot 2\phi_b} \]  
\(Q_b\) — bulk charge term (total charge stored in depletion layer), \(p\)-substrate in this case

Bulk potential — potential difference between Fermi level in intrinsic semiconductor and Fermi level in doped semiconductor

Fermi level is the average energy level in a material. For intrinsic materials, it is halfway between the valence band and conduction band.

\(p\)-type \(\Rightarrow\) Fermi level closer to valence band

\(n\)-type \(\Rightarrow\) Fermi level closer to conduction band

Other Constants (see text for values):
\(k\) = Boltzmann's constant (eV/K, J/K)
\(q\) = Electronic charge (coulombs)
\(T\) = temperature (˚K)
\(N_A\) = carrier density in doped semiconductor
\(n_i\) = intrinsic carrier concentration in Silicon
\(\varepsilon_{si}\) = permittivity of Silicon = \(\varepsilon_r \cdot \varepsilon_0\)
\(\varepsilon_r = 11.7\) (relative Silicon permittivity)
\(\varepsilon_0\) (permittivity of free space)
MOSFET Threshold Voltage (continued)

\[ V_{fb} = \phi_{ms} - \frac{Q_{fc}}{C_{ox}} \quad (\phi_{ms} = \text{gate work function}, \ Q_{fc} \ \text{sometimes referred to as} \ Q_{ss}) \]

\[ Q_{fc} \Rightarrow \text{fixed charge due to surface states which arise due to imperfections in silicon oxide interface and doping} \]

\[ \phi_{ms} \Rightarrow \text{gate work function which is the work function difference between the gate material and substrate} \]

\[ \phi_{ms} = -\left(\frac{E_g}{2q} + \phi_b\right) \]

\[ E_g \Rightarrow \text{Bandgap energy of Silicon (temperature dependent)} \]

\[ \phi_b \Rightarrow \text{bulk Fermi potential} \]

Note: \ E_g \ is actually in electron volts, 1eV = 1q \cdot 1V, so "q"'s in \ \phi_{ms} \ expression cancel out.
Two common techniques for increasing the native threshold voltage of a MOS device:

(1) Vary the doping concentration at the silicon-insulator interface through ion implantation (in process step called "threshold adjustment")

⇒ affects $Q_{fc}$ ($Q_{ss}$, surface state charge)

(2) Use different insulating material for gate

⇒ affects $C_{ox}$

Between transistors, use very thick oxide ($\gg t_{ox}$) to increase threshold voltage so that substrate surface does not become inverted through normal circuit voltage (obviously you do not want signal wire voltages and $V_{DD}$ lines inverting substrate). This keeps transistors electronically isolated from each other.

Example $V_T$ calculation: Calculate the native threshold voltage for an $n$-transistor at 300°K for a process with a Si substrate with $N_A = 1.80 \times 10^{-16}$ cm$^{-3}$, a SiO$_2$ gate oxide with thickness 200Å. (Assume $\Phi_{ms} = -0.9$V, $Q_{fc} = 0$C.)

$$\Phi_b = 0.02586 \ln \frac{1.80 \times 10^{-16}}{1.45 \times 10^{10}} = 0.36$V;$$

note $\frac{kT}{q} = 0.02586$V @ $T = 300^\circ$K

with

$$C_{ox} = \frac{3.9 \times 8.85 \times 10^{-14}}{0.2 \times 10^{-5}} = 1.726 \times 10^{-7}$$ Farads $\frac{\text{cm}^2}{\text{cm}^2}$

resulting in

$$V_T = \Phi_{ms} + \frac{\sqrt{2\varepsilon_{siq}N_A2\Phi_b}}{C_{ox}} + 2\Phi_b = (-0.9 + 0.384 + 0.72)V = 0.16$V

This device has a very low threshold voltage.
Substrate (bulk) bias effect on Threshold Voltage

For nMOS, substrate usually tied to ground. However, if $V_{SB}$ (source-to-bulk) $\neq$ 0V, the threshold equations become:

$$V_T = V_{fb} + 2\Phi_b + \frac{\sqrt{2}e_{si}qN_A(2\Phi_b + |V_{SB}|)}{C_{ox}}$$

$$V_T = V_{TO} + \gamma \left( \sqrt{2\Phi_b + |V_{SB}|} - \sqrt{2\Phi_b} \right)$$

where $V_{TO}$ is threshold voltage when $V_{SB} = 0V$ and $\gamma$ is a constant which describes substrate bias effect.

$$\gamma = \frac{t_{ox}}{e_{ox}} \sqrt{2e_{si}qN_A} = \frac{1}{C_{ox}} \sqrt{2e_{si}qN_A}$$

Values of $\gamma$ usually range from (0.4 to 1.2)V$^{1/2}$.

In SPICE, $\gamma = $ GAMMA, $V_{TO} = $ VTO, $N_A = $ NSUB, $\Phi_s = 2\Phi_b$ is PHI.

Example of substrate bias effect on threshold voltage: With $N_A = 3\times10^{16}$cm$^{-3}$, $t_{ox} = 200\AA$, $e_{ox} = 3.9\times8.85\times10^{-14}$F/cm, $e_{si} = 11.7\times8.85\times10^{-14}$F/cm, and $q = 1.6\times10^{-19}$Coulomb

$$\gamma = \frac{0.2\times10^{-5}}{3.9\times8.85\times10^{-14}} \sqrt{2\times1.6\times10^{-19}\times11.7\times8.85\times10^{-14}\times3\times10^{16}} = 0.57V^{1/2}$$

$$\Phi_b = 0.02586 \ln \frac{3\times10^{16}}{1.5\times10^{10}} = 0.375V$$

At a $V_{SB} = 2.5V$,

$$V_T = V_{TO} + 0.57\left(\sqrt{0.75 + 2.5} + \sqrt{0.75}\right)$$

$$V_T = V_{TO} + 0.53V$$

In analog designs it is quite common to use substrate bias to shift threshold voltage.
**Note:** When connecting devices in series, $V_T$ of top device will increase if $V_B$ tied to appropriate rail because $V_{SB}$ is not zero.

Actual shift in threshold voltage due to the above arrangement is very small.

---

**Revisit operation under Saturation**

Ideal equation $I_D = \frac{\beta_n}{2} (V_{GS} - V_{Tn})^2$ is not entirely accurate because pinch-off point under gate is influenced by $V_{DS}$. This influence of $V_{DS}$ on pinch-off *essentially* modifies the length of the channel (channel length modulation effect).
New equation for Saturation

\[ I_D = \frac{\beta}{2} (V_{GS} - V_{Tn})^2 (1 + \lambda V_{DS}) \]

\( \lambda \) in SPICE is called LAMBDA, and is the channel length modulation factor. Empirical values range from (0.02 to 0.005) V\(^{-1}\).

If we rewrite our current equation as

\[ I_D = \frac{K' W}{2 L} (V_{GS} - V_{Tn})^2 (1 + \lambda V_{DS}) \]

then when \( \lambda > 0 V^{-1} \), the effective channel length is reduced. Be careful not to confuse channel length with gate length. In saturated pinch-off, they are not equal!