MOSFET

- MOSFET or metal-oxide semiconductor field effect transistor is the most important device in microprocessors and solid state memories
- MOSFET is a general term, and most modern MOSFET devices do not have a metal gate but rather the gate made from poly-crystalline silicon
- Sometimes they are referred to as IGFET (i.e., insulated gate field effect transistor)

MOSFET

- MOSFET is a four terminal device
- It is fabricated on a $p$-type substrate (for $n$-channel device), sometimes referred to as bulk or body
- The device consists of two heavily doped $n^+$ regions forming the source and drain terminals
- The metal contact on top of the oxide is referred to as gate
- The gate electrode is now often made from heavily doped (conductive) poly-crystalline silicon
- The fourth terminal is an ohmic contact to the substrate
The characteristic parameters for a MOSFET:
- The channel length, $L$, the distance between the two $n^+ - p$-junctions.
- The width, $W$ (sometimes $Z$ is used).
- Oxide thickness, $d$ (sometimes $t_{ox}$).
- Junction depth, $r_j$.
- Substrate doping concentration, $N_A$.

When no voltage is applied to the gate source and drain are like two back-to-back $p - n$-junctions. Only a small saturation current flows.

When high enough gate voltage is applied the region below the gate is inverted and a “channel” of charge carriers is formed between the two highly doped $n^+$-regions.

Source and drain are then connected by a conducting path.

The conductivity of the channel is determined by the voltage applied to the gate.

The threshold voltage $V_{TH}$ is the minimum gate voltage for the channel to become conducting.

The most common is that the channel is not conducting when no voltage is applied and starts to conduct as high enough gate voltage is applied (enhancement-type, normally off).

In other cases the channel is conducting when no voltage is applied. For $n$-channel device a negative gate voltage will turn the channel off (depletion-type, normally on).

For the following discussion we assume the source is grounded.

Assume that the gate voltage is just above the threshold voltage.

When a small voltage is applied to drain the electrons flow from source to drain (and current flows from drain to source) through the conducting path.

The channel acts like a resistor and the drain current $I_D$ is proportional for the applied drain voltage — This is referred to as linear region.
Voltage controlled resistor

- In this region the MOSFET can be taken as a voltage controlled (gate voltage) resistor
- The gate and the substrate define a parallel plate capacitor where the oxide is the dielectric
- When the gate voltage is positive and higher than the threshold voltage the electrons (negative charge) are attracted to the Si/SiO$_2$ interface and create a conducting channel if the substrate is assumed to be p-type
- When the channel connects the n$^+$ regions of source and drain we have an n-type resistor
- When the gate voltage is increased further more electrons participate in the channel and the conductance increases

Pinch-off

- At first the drain current increases linearly with increased conductance
- However, increased current leads to increased voltage drop across the channel
- The potential is zero at the source and increases to become the drain voltage at the other end of the channel
- The potential difference between gate and channel decreases from $V_G$ at the source to ($V_G - V_D$) at the drain

Pinch-off

- Since the channel depth depends on this voltage, we note that the channel is not uniform in depth, rather the channel is tapered
- As the drain voltage is increased such that $(V_G - V_D) = V_{TH}$, the channel depth at the drain side decreases to almost zero, and the channel is said to be **pinched-off**
- If the drain voltage is increased further ($V_{Dsat}$) the pinch-off moves closer to the source e.g. location $L_1$
- When $V_D > V_G - V_{TH}$ no inverted channel exists between $L_1$ and $L$ close to the source

Pinch-off

- Electrons that reach the end of the channel will experience the high electric field in the depletion region surrounding the drain junction and are rapidly swept into the drain terminal
- The drain voltage $V_D$ no longer affects the current significantly and the MOSFET acts as a constant current source
- The drain current is said to be in saturation
**$I_D - V_D$ Characteristics**

- When the voltage applied to the gate is low, $V_G > V_{TH}$ ($V_D$ small), the substrate of the semiconductor is inverted

- A small voltage applied to drain leads to a current from source to drain and the channel acts as a resistor

\[ I_D \propto V_D \]

**Charge - Triode region**

- The total charge induced in the semiconductor side per unit area, $Q_s$, at a distance $y$ from the source

\[ Q_s(y) = -[V_G - \psi_s(y)]C_{ox} \]

where $\psi_s$ is the surface potential at $y$ and $C_{ox} = \epsilon_{ox}/d$ is the gate capacitance per unit area
Charge - Triode region

- The charge in the inversion layer per unit area is
  \[ Q_n(y) = Q_s(y) - Q_{sc}(y) \]
  or
  \[ Q_n(y) = -[V_G - \psi_s(y)]C_{ox} - Q_{sc}(y) \]
- The surface potential \( \psi_s \) at inversion can be approximated by
  \[ 2\psi_b + V(y) \]
  and \( V(y) \) is the reverse bias between the point \( y \) and the source electrode (which we assumed to be grounded)

Conductance

- The charge within the surface depletion region was given previously as
  \[ Q_{sc} = -qN_Ax_d \approx -[2\varepsilon_s qN_A(V(y) + \psi_b)]^{1/2} \]
- The charge within the surface depletion
  \[ Q_n(y) \approx -[V_G - V(y) - 2\psi_s(y)]C_{ox} + [2\varepsilon_s qN_A(V(y) + \psi_b)]^{1/2} \]
- The conductivity of the channel conductance at position \( y \) can be approximated by
  \[ \sigma(x) = qn(x)\mu_n(x) \]

Resistivity - Current

- For a constant mobility the channel conductance is then given by
  \[ g = \frac{W}{L} \int_0^{x_i} \sigma(x)dx = \frac{W\mu_n}{L} \int_0^{x_i} qn(x)dx \]
- The integral
  \[ \int_0^{x_i} qn(x)dx \]
  corresponds to the total charge per unit area in the inversion layer and is therefore equal to \( |Q_n| \) or
  \[ g = \frac{W\mu_n}{L} |Q_n| \]
- The channel resistance of an elemental section \( dy \) is
  \[ dR = \frac{dy}{gL} = \frac{dy}{W\mu_n |Q_n|} \]

Resistivity - Current

- The voltage drop across \( dy \) is then
  \[ dV = I_d dR = \frac{dy}{gL} = \frac{I_d dy}{W\mu_n |Q_n|} \]
  Here \( I_d \) is the drain current which is independent of \( y \)
- Substituting and integrating from the source \( (y = 0, V = 0) \) to drain \( (y = L, V = V_D) \)
  \[ I_d \approx \frac{W}{L\mu_n C_{ox}} \left\{ \left[ V_G - 2\psi_b - \frac{V_D}{2} \right] V_D - \frac{2}{3} \sqrt{2\varepsilon_s qN_A} \right\} \left[ (V_D + 2\psi_b)^{3/2} - (2\psi_b)^{3/2} \right] \]
**$I_D - V_D$ Characteristics**

- For a given $V_G$ the drain current first increases linearly with drain voltage (the **linear region**) and then gradually levels off (the **triode region**), approaching a saturated value (the **saturation region**).
- The dashed line indicates the locus of the drain voltage $V_{D_{\text{sat}}}$ at which the current reaches a maximum value.

**$I_D - V_D$ Characteristics**

- For a small $V_D$

\[ I_D = \frac{1}{2} \frac{W}{L} \mu_n C_{\text{ox}} \left( 2(V_G - V_{TH})V_D - V_D^2 \right) \]

which can be approximated

\[ I_D \approx \frac{W}{L} \mu_n C_{\text{ox}} (V_G - V_{TH}) V_D \]

for $V_D \ll (V_G - V_T)$ and $V_{TH}$ is the threshold voltage given as

\[ V_{TH} \approx \sqrt{\frac{2 \epsilon_s q N_A (2 \psi_b)}{C_o}} + 2 \psi_b \]

- We can obtain the value of $V_{D_{\text{sat}}}$ using the condition $Q(L) = 0$

\[ V_{D_{\text{sat}}} \approx V_G - 2 \psi_b + K^2 \left[ 1 - \left( 1 + \frac{2V_G}{K^2} \right)^{1/2} \right] \]

where

\[ K \equiv \frac{\sqrt{\epsilon_s q N_A}}{C_{\text{ox}}} \]

- The saturation current is

\[ I_{D_{\text{sat}}} \approx \frac{W \mu_n C_{\text{ox}}}{2dL} (V_G - V_{TH})^2 \]

**$I_D - V_D$ Characteristics**

- In the linear region the channel conductance is

\[ g_D \equiv \frac{\partial I_D}{\partial V_D} \bigg|_{V_G=\text{constant}} \approx \frac{W}{L} \mu_n C_{\text{ox}} (V_G - V_{TH}) \]

and the transconductance $g_m$ is given as

\[ g_m \equiv \frac{\partial I_D}{\partial V_G} \bigg|_{V_D=\text{constant}} \approx \frac{W}{L} \mu_n C_{\text{ox}} V_D \]

- When the drain voltage is increased to the point that the charge in the inversion layer at $y = L$ becomes zero, $Q(L) = 0$ the number of mobile electrons in the drain are reduced drastically.

- This point is called pinch-off and the drain voltage and drain current are designated as $V_{D_{\text{sat}}}$ and $I_{D_{\text{sat}}}$, respectively.

- For drain voltage larger than $V_{D_{\text{sat}}}$ we have the saturation region.
For an idealized MOSFET in the saturation region, the channel conductance is zero and the transconductance is given by

\[ g_m = \frac{\partial I_D}{\partial V_G} \bigg|_{V_D=\text{constant}} \approx \frac{W \mu_n \epsilon_{\text{ox}}}{L} (V_G - V_{\text{TH}}) \]

and \( g_m \) is linearly proportional to \( W/L \) for a given \( V_G - V_{\text{TH}} \) and \( g_m \) is linearly proportional to \( V_G - V_{\text{TH}} \) for a given \( W/L \).

The drain current versus gate voltage is given as

\[ I_{D\text{sat}} \approx \frac{W \mu_n \epsilon_{\text{ox}}}{2dL} (V_G - V_{\text{TH}})^2 = \frac{W \mu_n C_{\text{ox}}}{2L} (V_G - V_{\text{TH}})^2 \]

where \( V_{\text{TH}} \) is the threshold voltage.

We note that the point at which the channel vanishes \( L_1 \) in fact moves toward the source as the drain voltage increases.

The value of \( L_1 \) varies with \( V_D \) to some extent.

This phenomena is referred to as channel length modulation.
This phenomenon yields a larger drain current as $I_{\text{Dsat}} \propto 1/L_1$.

Similar to the Early effect in bipolar devices, channel-length modulation results in a finite output impedance.

To account for channel-length modulation, we introduce a parameter $\lambda$ so that $L$ is a constant, but multiply the right-hand side by a corrective term

$$I_{\text{Dsat}} \approx \frac{W \mu_n C_o}{2L} (V_G - V_{\text{TH}})^2 (1 + \lambda V_D)$$

For a long channel the relative change in $L$ and hence $I_{\text{Dsat}}$ is smaller than for a short channel.

Example 5.3.

At high electric fields, carrier mobility degrades, eventually leading to a constant velocity.

Owing to the very short channels (< 100 nm) modern MOSFET devices experience velocity saturation even with drain-source voltage of 1 V.

The drain current is then

$$I_D = v_{\text{sat}} WC_o (V_G - V_{\text{TH}})$$

has a linear dependence on $V_G - V_{\text{TH}}$ and independent of $L$.

We also see that

$$g_m = \frac{\partial I_D}{\partial V_G} = v_{\text{sat}} WC_o$$

Example 5.4.

When the gate voltage is below the threshold voltage and the semiconductor surface is only weakly inverted, the corresponding current is referred to as subthreshold current.

The subthreshold region is particularly important when the MOSFET is used as a low-voltage low-power device, such as switch in digital logic and memory applications.

Then the subthreshold current determines how the switch turns on and off.

Here the drain current is dominated by diffusion instead of drift.

Example 5.4.
Subthreshold current

- If we consider the MOSFET to be as an n-p-n (source-substrate-drain) bipolar transistor then we have

\[ I_D = -qAD_n \frac{dn}{dy} = qAD_n \frac{n(0) - n(L)}{L} \]

where \( A \) is the channel cross section of the current flow and \( n(0) \) and \( n(L) \) are the electron densities in the channel at source and drain, respectively

- The electron densities are given as

\[ n(0) = n_i \exp \left( \frac{q(\psi_s - \psi_b)}{kT} \right) \quad \text{and} \quad n(L) = n_i \exp \left( \frac{q(\psi_s - \psi_b - V_D)}{kT} \right) \]

where \( \psi_s \) is the surface potential at the source

Subthreshold current

- A typical measured curve for the subthreshold region
- An important parameter in this region is the subthreshold swing

\[ S = \left[ \frac{\partial \log(I_D)}{\partial V_G} \right]^{-1} \]

which is typically 70 – 100 mV/decade at room temperature

- To reduce the subthreshold current to a negligible value, we must bias the MOSFET a half-volt or more below \( V_{TH} \)

Subthreshold current

- Thus we have

\[ I_D = \frac{qAD_n n_i}{L} \exp \left( -\frac{q\psi_b}{kT} \right) \times \left( 1 - \exp \left( -\frac{qV_D}{kT} \right) \right) \exp \left( \frac{q\psi_s}{kT} \right) \]

- The surface potential is

\[ \psi_s \approx V_G - V_{TH} \]

and the drain current decreases exponentially when \( V_G \) falls below \( V_{TH} \)

\[ I_D \sim \exp \left( -\frac{q(V_G - V_T)}{kT} \right) \]

MOSFET

- There are basically four types of MOSFET
  - n-channel enhancement (normally off)
  - n-channel depletion (normally on)
  - p-channel enhancement (normally off)
  - p-channel depletion (normally on)
MOSFET

- There are basically four types of MOSFET
  - The channel conductance is low for \( V_G = 0 \) and positive voltage is applied to the gate to form the n-channel
  - The \( n- \) channel exists for \( V_G = 0 \) and negative voltage decreases the conductance of the channel
  - The channel conductance is low for \( V_G = 0 \) and negative voltage is applied to the gate to form the p-channel
  - The \( p- \) channel exists for \( V_G = 0 \) and positive voltage decreases the conductance of the channel

Large-signal model

- In the triode region
  \[
  I_D = \frac{1}{2} \frac{W}{L} \mu_n C_o \left( 2(V_G - V_T) V_D - V_D^2 \right)
  \]
  and in the saturation region
  \[
  I_{Dsat} \approx \frac{W \mu_n C_o}{2L} (V_G - V_T)^2 (1 + \lambda V_D)
  \]

Small-signal model

- When the transistor acts as voltage-controlled current source
  \[
  i_D = g_m v_G
  \]
  and the gate is like an open circuit
- Channel-length modulation is described by a resistor
  \[
  r_O = \left( \frac{\partial I_D}{\partial V_D} \right)^{-1} = \frac{1}{\frac{W}{L} \mu_n C_o} \approx \frac{1}{\lambda D}
  \]
Further reading

This discussion is based on Chapter 6 in Razavi (2008) and Chapter 2 of Razavi (2001). Some of the discussion is taken from Sze (2002) section 6.2. Similar discussion is also found in sections 6.4.1. and 6.5.1 - 6.5.4. in Streetman and Banerjee (2000).

References