Lecture 2: MOS Transistor Modeling
Announcements

• If you haven’t already, turn in your 0.18um NDA form ASAP

• Lab 1 starts Jan 31

• Current Reading
  • Razavi Chapters 2 & 17
Agenda

• MOS Transistor Modeling
  • Physical Structure
  • Threshold Voltage, $V_T$
  • DC I-V Equations
  • Body Effect
  • Subthreshold Region
NMOS Physical Structure
CMOS Physical Structure

[NMOS]

S
G
D

SiO₂
n+

n+

SiO₂

PMOS

D
G
S

p+
p+

SiO₂

n-well

p−

[Karsilayan]
Threshold Voltage, $V_T$

- Applying a positive voltage to the gate repels holes in the p-substrate under the gate, leaving negative ions (depletion region) to mirror the gate charge.

- Before a “channel” forms, the device acts as 2 series caps from the oxide cap and the depletion cap.

- If $V_G$ is increased to a sufficient value the area below the gate is “inverted” and electrons flow from source to drain.
**$V_T$ Definition**

- The threshold voltage, $V_T$, is the voltage at which an “inversion layer” is formed.
- For an NMOS this is when the concentration of electrons equals the concentration of holes in the $p^-$ substrate.

$$V_T = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}} = \Phi_{MS} + 2\Phi_F + \gamma \sqrt{2\Phi_F}$$

$\Phi_{MS}$ is the difference between the work functions of the polysilicon gate and the silicon substrate.

$\Phi_F$ is the Fermi potential, $\Phi_F = \frac{kT}{q} \ln \left( \frac{N_{sub}}{n_i} \right)$

$N_{sub}$ is substrate doping density, $n_i$ is undoped silicon electron density.

$Q_{dep}$ is the depletion region charge, $Q_{dep} = \sqrt{4q\varepsilon_{si}|\Phi_F|N_{sub}}$

$C_{ox}$ is the gate cap/area, $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

Note, $\gamma$ will be defined later.
MOSFET in Accumulation Mode

- If a negative gate voltage is applied w.r.t. the source, then positive charge “accumulates” below the gate.
- In this Accumulation Mode, no current flows and the device is often used as a capacitor with the drain shorted to the source.
- This capacitor consists of parallel plate capacitance below the gate and overlap/fringing capacitance near the drain/source regions.

\[ C_{G,acc} = W L_{eff} C_{ox} + 2 W C_{ov} \]
MOSFET in Inversion Mode

**Subthreshold**

0 < \( V_G < V_T \)
\( V_{DS} > 0 \)

- **N-type transistor**
  - Subthreshold (extremely low-voltage low-power applications)

**Triode/ Linear**

\( V_G > V_T \)
Small \( V_{DS} \)

- Linear region
- Voltage controlled resistor, linear OTA's, multipliers, switches

**Saturation**

\( V_G > V_T \)
\( V_{DS} > V_{GS} - V_T \)

- Saturation region (Amplifiers)
MOS Equations in Triode Region (Small $V_{DS}$)

Current from Source to Drain: \[ I = \frac{dQ}{dt} = \frac{dQ}{dx} \frac{dx}{dt} = Q_d(x) \nu \]

Incremental Charge Density: \[ Q_d(x) = -C_{OX} W (V_{GC}(x) - V_T) \]

Gate-to-Channel Voltage: \[ V_{GC}(x) - V_T = V_G - V_{CS}(x) - V_T \]

Electron Velocity: \[ \nu = -\mu_n E(x) = \mu_n \frac{dv(x)}{dx} \]

\[ I_{DS} = -I = C_{OX} W (V_G - V_{CS}(x) - V_T) \mu_n \frac{dv(x)}{dx} \]

\[ \int_{0}^{L} I_{DS} dx = \int_{0}^{L} \mu_n C_{OX} W (V_G - V_T - V_{CS}(x)) dv(x) \]

\[ I_{DS} L = \mu_n C_{OX} W (V_G - V_T) V_{DS} - \mu_n C_{OX} W \left( \frac{1}{2} V_{DS}^2 \right) \]

\[ I_{DS} = \mu_n C_{OX} \frac{W}{L} \left( V_G - V_T - \frac{1}{2} V_{DS} \right) V_{DS} \]

Capacitance per unit gate area: \[ C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \]

Electron mobility: \( \mu_n \)
Triode or Linear Region

- Channel depth and transistor current is a function of the overdrive voltage, $V_{GS} - V_T$, and $V_{DS}$
- Because $V_{DS}$ is small, $V_{GC}$ is roughly constant across channel length and channel depth is roughly uniform

$$V(x) = V_{GS} - V(x) = V_{GS} - V_{DS} \frac{x}{L}$$

$$V_{GC}(x) = V_{GS} - V(x) = V_{GS} - V_{DS} \frac{x}{L}$$

$$I_{DS} = \frac{W}{L} \mu_n C_{ox} (V_{GS} - V_{Tn} - 0.5V_{DS}) V_{DS}$$

For small $V_{DS}$, $R_{DS} \approx \frac{1}{W} \frac{1}{L} \mu C_{ox} (V_{GS} - V_{Tn})$
MOS Equations in Triode Region (Large $V_{DS}$)

Drain current: Expression used in SPICE level 1

$$I_{DS} = \frac{W}{L} \mu_n C_{OX} (V_{GS} - V_T - 0.5V_{DS})V_{DS}$$

This doesn't really happen

$$V_{DSsat} = V_{GS} - V_{Tn}$$
Triode Region Channel Profile

- Recall that the channel charge density is \( \propto (V_{GC}(x) - V_T) \)

\[
V_{GC}(x) - V_T = V_{GS} - V(x) - V_T = V_{OV} - V_{DS} \frac{x}{L}
\]

- If \( V_{GC} \) is always above \( V_T \) throughout the channel length, the transistor current obeys the triode region current equation

[Sedra/Smith]
When $V_{DS} \geq V_{GS} - V_T = V_{OV}$, $V_{GC}$ no longer exceeds $V_T$, resulting in the channel “pinching off” and the current saturating to a value that is no longer a function of $V_{DS}$ (ideally).

$$V_{GC}(x) - V_T = V_{GS} - V(x) - V_T = V_{OV} - V_{DS} \frac{x}{L}$$
Saturation Region

- Channel “pinches-off” when $V_{DS} = V_{GS} - V_T$ and the current saturates
- After channel charge goes to 0, the high lateral field “sweeps” the carriers to the drain and drops the extra $V_{DS}$ voltage

$$I_{DS} = \mu_n C_{OX} \frac{W}{L} \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS} \bigg|_{V_{DS} = V_{GS} - V_T}$$

$$I_{DS} = \frac{\mu_n C_{OX} W}{2} \left( V_{GS} - V_T \right)^2$$
NMOS $I_D - V_{DS}$ Characteristics

$$I_{DS} = \frac{W}{L} \mu_n C_{OX} (V_{GS} - V_{Tn} - 0.5V_{DS}) V_{DS}$$

$$V_{OV} = V_{GS} - V_{TN}$$

$$I_{DS} = \frac{W}{2L} \mu_n C_{OX} (V_{GS} - V_{Tn})^2$$

- **Triode** ($v_{DS} \leq V_{OV}$)
- **Saturation** ($v_{DS} \geq V_{OV}$)

Current saturates because the channel is pinched off at the drain end, and $v_{DS}$ no longer affects the channel.

$$v_{GS} = V_I + V_{OV}$$

Circuit bends because the channel resistance increases with $v_{DS}$

Almost a straight line with slope proportional to $V_{OV}$
MOS “Large-Signal” Output Characteristic

\[ i_D = \frac{1}{2} k_n \left( \frac{W}{L} \right) V_{OV4}^2 \]

\[ \frac{1}{2} k_n \left( \frac{W}{L} \right) V_{OV3}^2 \]

\[ \frac{1}{2} k_n \left( \frac{W}{L} \right) V_{OV2}^2 \]

\[ \frac{1}{2} k_n \left( \frac{W}{L} \right) V_{OV1}^2 \]

\[ v_{DS} \leq v_{OV} \]

Triode region

Saturation region

\[ v_{DS} = v_{OV} \]

\[ i_D = \frac{1}{2} k_n \left( \frac{W}{L} \right) v_{DS}^2 \]

\[ v_{GS} = V_t + V_{OV4} \]

\[ v_{GS} = V_t + V_{OV3} \]

\[ v_{GS} = V_t + V_{OV2} \]

\[ v_{GS} = V_t + V_{OV1} \]

\[ v_{GS} \leq V_t \text{ (Cutoff)} \]

Note: \( V_{OV} = V_{GS} - V_T \) and \( k_n' = \mu_n C_{ox} \)
What about the PMOS device?

- The current equations for the PMOS device are the same as the NMOS EXCEPT you swap the current direction and all the voltage polarities.

**NMOS**

Linear:  \[ I_{DS} = \frac{W}{L} \mu_n C_{OX} (V_{GS} - V_{Tn} - 0.5V_{DS})V_{DS} \]

Saturation:  \[ I_{DS} = \frac{W}{2L} \mu_n C_{OX} (V_{GS} - V_{Tn})^2 \]

**PMOS**

Linear:  \[ I_{SD} = \frac{W}{L} \mu_p C_{OX} \left( V_{SG} - |V_{Tp}| - 0.5V_{SD} \right)V_{SD} \]

Saturation:  \[ I_{SD} = \frac{W}{2L} \mu_p C_{OX} \left( V_{SG} - |V_{Tp}| \right)^2 \]
PMOS $I_D - V_{SD}$ Characteristics

$V_{OV} = V_{SG} - |V_{TP}|$

[Karsilayan]
Body Effect

- If the body and source potential are equal, a certain $V_G = V_{T0}$ is required to form an inversion layer
  
  $$V_{T0} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep0}}{C_{ox}} = \Phi_{MS} + 2\Phi_F + \gamma \sqrt{2\Phi_F}$$

- As $V_S$ becomes positive w.r.t. $V_B$, a larger depletion region forms, which requires a higher $V_G$ to form a channel

- The net result is that $V_T$ increases due to this “body effect”

- Note, it also works in reverse, as if you increase $V_B$ w.r.t. $V_S$, then $V_T$ lowers

$$V_T = V_{T0} + \gamma \left( \sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F} \right)$$

Body effect coefficient, $$\gamma = \frac{\sqrt{2qe_{si}N_{sub}}}{C_{ox}}$$

$\gamma$ typically ranges from 0.3 to 0.4$V^{1/2}$
MOS MODEL: SPICE LEVEL-II

NMOS: \( I_{DS} = \frac{W}{L} \mu_n C_{OX} (V_{GS} - V_{Th} - 0.5 V_{DS}) V_{DS} \)

• Drain current, Triode region

PMOS: \( I_{SD} = \frac{W}{L} \mu_p C_{OX} \left( V_{SG} - |V_{Tp}| - 0.5 V_{SD} \right) V_{SD} \)

• Drain Current, Saturation region

NMOS: \( I_{DS} = \frac{W}{2L} \mu_n C_{OX} (V_{GS} - V_{Th})^2 \)

PMOS: \( I_{SD} = \frac{W}{2L} \mu_p C_{OX} (V_{SG} - |V_{Tp}|)^2 \)

• Threshold voltage (zero bias)

\( V_{T0} = \Phi_{MS} + 2\Phi_F + \gamma \sqrt{2\Phi_F} \)

• Threshold voltage

\( V_T = V_{T0} + \gamma \left[ \sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F} \right] \Rightarrow V_{T0} \mid_{V_{SB}=0} \)

• KP and \( \gamma \) (Spice Model)

\( KP = \mu C_{OX}; \quad \gamma = \frac{\sqrt{2q\varepsilon_{si}N_{sub}}}{C_{OX}} \)
Subthreshold Region

- So far we have assumed that $I_D=0$ when $V_{GS}<V_T$
- However, in reality an exponentially decreasing current exists for $V_{GS}<V_T$

In subthreshold region: $I_D = I_0 \exp\left(\frac{V_{GS}q}{\zeta kT}\right)$

where $I_0$ is a scale current

$\zeta > 1$ is a nonideality factor

The steepest subthreshold slope is 1 dec./60mV with $\zeta = 1$

- $V_T$ values are often set by extrapolating above threshold data to current values of zero or infinite $R_{on}$
- A rough value often used is the $V_{GS}$ which yields $I_D/W = 1\mu A/\mu m$

[Razavi]
Subthreshold Current & $V_T$ Scaling

• This subthreshold current prevents lowering $V_T$ excessively

• Assuming $V_T=300$ mV and has an 80 mV subthreshold slope, then the $I_{on}/I_{off}$ ratio is only on the order of $10^{(300/80)}=5.6e3$

• Reducing $V_T$ to 200 mV drops the $I_{on}/I_{off}$ ratio to near 316

• If we have a large number of “off” transistors on our chip, these subthreshold currents add up quickly, resulting in significant power dissipation

• This is a huge barrier in CMOS technology scaling and one of the main reasons Vdd scaling has slowed
Next Time

• MOS Transistor Modeling
  • Small-Signal Model
  • Spice Models