Metal-oxide-semiconductor field effect transistors (2 lectures)

MOS physics (brief in book)

Current-voltage characteristics
- pinch-off / channel length modulation
- weak inversion
- velocity saturation
- metrics extraction

Capacitances

Small-signal model

Reading: (Sedra, Smith, 7th edition)
5.1, 5.2, 5.3
5.4.3, 5.4.4
6.2.1 small signal model DC
9.2.1 small signal model AC
n-type MOSFET layout

Source

\( n^{++} \)

\( n^{++} \)

\( \text{p-type Si} \)

Body (\( V_{SB} \))

\( \text{SiO}_2 \)

Gate (\( V_{GS} \))

Drain (\( V_{DS} \))

n-MOSFET

p-MOSFET

n-MOSFET with body contact

depletion mode n-MOSFET

Modern Electronics: F4,5 MOSFET
Energy bands 1 – flat band

- Gate oxide has large band gap → no current. $E_F$ constant in semiconductor.

The diagram shows the energy bands within a semiconductor and a gate oxide. The energy levels are labeled as follows:

- $E_C$: Conduction band edge
- $E_i$: Valence band edge
- $E_F$: Fermi level
- $E_V$: Valence band edge

The oxide has a large band gap, resulting in no current flow.
Energy bands 2 - depletion

\[ X = \sqrt{\frac{2 \cdot \varepsilon_s \cdot \varepsilon_0 \cdot \phi}{q \cdot N_A}} \]

Charge per area:
\[ Q = X \cdot N_A \cdot q \]

\[ V_{ox} = \frac{Q}{C_{ox}} \]

\[ V_{GS} = V_{ox} + \phi \]
Energy bands 3 – intrinsic surface

\[ X = \sqrt{\frac{2 \cdot \varepsilon_s \cdot \varepsilon_0 \cdot \phi}{q \cdot N_A}} \]

\[ Q = X \cdot N_A \cdot q \]

\[ V_{ox} = \frac{Q}{C_{ox}} \]

\[ V_{GS} = V_{ox} + \phi \]
Energy bands 4 – weak inversion

\[
X = \sqrt{\frac{2 \cdot \varepsilon_s \cdot \varepsilon_0 \cdot \phi}{q \cdot N_A}}
\]

\[
Q = X \cdot N_A \cdot q
\]

\[
V_{ox} = \frac{Q}{C_{ox}}
\]

\[
V_{GS} = V_{ox} + \phi
\]
Energy bands 5 – onset of strong inversion ($V_{GS} = V_t$)

Many free electrons at surface

\[ n_p \gg p_p \]

\[ n_p = N_A \]

\[ X = \sqrt{\frac{2 \cdot \epsilon_s \cdot \epsilon_0 \cdot \phi}{q \cdot N_A}} \]

\[ Q = X \cdot N_A \cdot q \]

\[ V_{ox} = \frac{Q}{C_{ox}} \]

\[ V_{GS} = V_{ox} + \phi \]
Strong inversion $V_{GS} > V_t$

$V_{ox}$

$\phi$

$E_C$

$E_i$

$E_F$

$E_V$

$n_p = N_A$

$n_p >> N_A$

Conducting channel at surface!

Modern Electronics: F4,5 MOSFET
depletion / weak inversion
only acceptors contribute ->
slow increase in Q (increase $X$)

\[ Q = \text{const} \times \sqrt{\phi} \]

strong inversion
mainly electrons contribute ->
exponential increase in Q

\[ Q = \text{const} \times \exp(\phi) \]

Threshold for channel formation
Surface potential variation with gate voltage

\[ V_{GS} = V_{ox} + \phi = \frac{Q}{C_{ox}} + \phi \]

- \( V_{GS} \) falls over both oxide and semiconductor
- \( \Phi = 2\Phi_f \)
- \( V_{GS} \) falls mainly over oxide (parallel-plate capacitor)
MOS physics: capacitance

\[ C = \frac{dQ}{dV} \]

\[ \rho \]

\[ \rho \]

\[ \text{SiO}_2 \]

\[ \text{p-type semiconductor} \]

\[ C_{gs} \text{ (F m}^{-2}\text{)} \]

\[ V_{GS} \text{ (V)} \]

\[ \max(C_{gs}) = C_{ox} \]

\[ \min(C_{gs}) \text{ related to } X \]

\[ \text{can get } N_A \]

accumulation
depletion
inversion

Low frequency (10 Hz)
High frequency (>1 MHz)

Minority carriers don’t have time to be generated

\[ \text{Minority carriers don’t have time to be generated} \]

\[ \text{depletion in inversion} \]

\[ \text{max}(C_{gs}) = C_{ox} \]

\[ \text{min}(C_{gs}) \text{ related to } X \]

\[ \text{can get } N_A \]
Metal and semiconductor work functions usually not the same (as we assumed before).

1. Sketch band diagram (without bias) for metal with higher work function in MOS.
2. How does the higher work function impact $V_t$ (threshold voltage needed for strong inversion)?

\[ \Delta \Phi_{ms} = ??? \]
Inversion layers creates channel in MOSFET

Modern Electronics: F4,5 MOSFET
Geometry for drain current calculation (5.1.4 - 5.1.6)

\[ Q(y) = \text{electrons per unit area at position } y \]
\[ V(y) = \text{voltage at } y \text{ w.r.t. source} \]
\[ V_{GS} - V(y) = \text{gate-to-channel voltage at } y \]
\[ \varepsilon(y) = \text{electric field at } y \]

\[ Q(y) = C_{ox} (V_{GS} - V(y) - V_t) \]
Current and channel shape – linear (triode) region

\[ I_D = 0 \quad \text{for} \quad V_{GS} < V_t \]

\[ I_D = \frac{\mu_n C_{ox} W}{2 L} \left[ 2(V_{GS} - V_t)V_{DS} - V_{DS}^2 \right] \quad \text{for} \quad V_{GS} > V_t \quad \text{and} \quad V_{DS} < V_{GS} - V_t \]

\[ Q(y) = C_{ox} \left( V_{GS} - V_t - V(y) \right) \]
Current and channel shape – saturation region (pinch-off)

\[ V_{DS} = V_{GS} - V_t \]

\[ I_{D,sat} = \frac{\mu_n C_{ox} W}{L} (V_{GS} - V_t)^2 \]

Pinch-off: \( I_D \) independent of \( V_{DS} \)

\[ V_{DS} < V_{GS} - V_t \]
\[ Q = C_{ox} (V_{GS} - V_t - V(y)) \]
\[ Q(0) = C_{ox} (V_{GS} - V_t) \]
\[ Q(L) \approx 0 \]

\[ V_{DS} = V_{GS} - V_t \]
\[ Q(L) \approx 0 \]

\[ V_{DS} > V_{GS} - V_t \]

Modern Electronics: F4,5 MOSFET
Pinch-off – band structure and electron velocity

Pinch-off point:
Electrons swept into drain by E-field
Channel length modulation (5.2.4)

\[ Q = C_{ox} (V_{GS} - V_t - V(y)) \]

\[ I_{D,\text{sat}} = \frac{\mu_n C_{ox} W}{2L} (V_{GS} - V_t)^2 (1 + \lambda V_{DS}) \]

- A non-inverted region is formed beyond pinch-off i.e. the channel length is reduced with \( V_{DS} \)
- Gives output resistance \( r_0 = \left( \frac{dI_D}{dV_{DS}} \right)^{-1} = \frac{1}{\lambda I_D} \)
- Output conductance \( g_d = 1/r_0 \)
- \( \lambda \approx 0.005-0.05 \text{ V}^{-1} \)
Which of these is correct for channel length modulation?

a) Stronger impact for longer channels
b) Lower impact for longer channels
c) No dependence on channel length
d) Stronger impact with higher channel doping
e) Lower impact with higher channel doping
f) No dependence on channel doping
Example: MOSFET with channel length modulation

\[ V_t = 1.0 \text{ V} \]
\[ C_{ox} = 3 \text{ mF/m}^2 \]
\[ \mu_n = 0.135 \text{ m}^2/\text{Vs} \]
\[ L = 5 \mu\text{m} \]
\[ W = 50 \mu\text{m} \]
\[ \lambda = 0.02 \text{ V}^{-1} \]

Calculate \( I_D \) for

1) \( V_{DS} = 0.5, 1, 2 & 3 \text{V} \) if \( V_{GS} = 0 \text{ V} \)

2) \( V_{DS} = 0.5, 1, 2 & 3 \text{V} \) if \( V_{GS} = 2.5 \text{ V} \)

**Linear region**

\[ V_{DS} < V_{GS} - V_t \]

\[ I_D = \frac{\mu_n C_{ox} W}{2 L} \left[ 2(V_{GS} - V_t)V_{DS} - V_{DS}^2 \right] \]

**saturation**

\[ V_{DS} > V_{GS} - V_t \]

\[ I_{D,sat} = \frac{\mu_n C_{ox} W}{2 L} \left( V_{GS} - V_t \right)^2 \left( 1 + \lambda V_{DS} \right) \]
In modern MOSFETs, resistance in the source/drain cannot always be neglected. How does the output characteristics change if a significant source/drain resistance is added?
Subthreshold current (weak inversion) (not in book)

- For $V_{GS} < V_t$ (but high enough for weak inversion): channel charge $<<$ inversion charge
- Drift current is small, mainly diffusion current (as NPN BJT)
- Exponentially increasing injection over barrier with $V_{GS}$.

\[ I_D \propto \exp\left(\frac{V_{GS} - V_t}{nV_T}\right) \cdot \left(1 - \exp\left(-\frac{V_{DS}}{V_T}\right)\right) \]

Forward biased np-junction

Modern Electronics: F4,5 MOSFET
Velocity saturation in short-channel MOSFETs (5.4.4)

- Short channel length, high $V_{DS}$ -> high electric field.
- At critical field $\varepsilon_c$ (1.5*10^6 V/m) electron velocity saturates due to optical phonon emission.

\[
\nu_d = \frac{\mu_n \varepsilon}{1 + \varepsilon / \varepsilon_c} \begin{cases} 
\varepsilon \ll \varepsilon_c \rightarrow \nu_d \approx \mu_n \varepsilon \\
\varepsilon \gg \varepsilon_c \rightarrow \nu_d \approx \mu_n \varepsilon_c = \nu_{scl}
\end{cases}
\]

\[
\lim_{E_C \to 0} I_D = WC_{ox}(V_{GS} - V_t)\nu_{scl}
\]

Without velocity sat. ($V_{GS}^2$)

With velocity sat. ($V_{GS}$)
Detrimental effects

- **Junction breakdown** – For large $V_{DS}$ of 20-150V (long channel) the drain-substrate pn-junction is reversed biased enough to cause avalanche breakdown.

- **Punchthrough** – For large $V_{DS}$ (short channel) the drain depletion region can touch the source depletion region. Slower current increase than junction breakdown. Increasing channel doping makes depletion regions thinner.

- **Hot carriers** – high electric fields give electrons enough energy to be injected into gate oxide. Can give higher gate current and/or shift the threshold voltage.

- **Oxide breakdown** – For high ($>6\times10^6$ V/cm) vertical electric fields (from gate) the gate oxide may break.
Summary – MOSFET DC characteristics

- **Linear (triode)**
  \[ V_{DS} < V_{GS} - V_t \]

- **Saturation (active, pinch-off)**
  \[ V_{DS} > V_{GS} - V_t \]
  included channel length modulation

- **Velocity saturation**

\[
I_D = \frac{\mu_n C_{ox} W}{2} \left[ 2(V_{GS} - V_t)V_{DS} - V_{DS}^2 \right]
\]

\[
I_{D,\text{sat}} = \frac{\mu_n C_{ox} W}{2} \left( V_{GS} - V_t \right)^2 \left( 1 + \lambda V_{DS} \right)
\]

\[
\lim_{E_C \to 0} I_D = WC_{ox}(V_{GS} - V_t)v_{scl}
\]
How can a MOSFET operate at GHz when MOSCAP only responds at frequencies < 1 MHz in inversion?
Small-signal model (6.2.1)

\[ i_D(t) = I_D + i_d(t) \]

\[ v_{gs}(t) = V_{GS} + v_{gs}(t) \]

\[ V_{GS} + v_{gs}(t) = v_{GS}(t) \]

\[ I_D + i_d(t) = i_D(t) \]

1\textsuperscript{st} order Taylor expansion - linearization

\[ f(x_0 + \delta x) \approx f(x_0) + \frac{df(x)}{dx}\bigg|_{x=x_0} \delta x + \ldots \]

\[ i_D \left( V_{GS} + v_{gs} \right) \approx I_D \left( V_{GS} \right) + \frac{dI_D}{dV_{GS}} \cdot v_{gs} = I_D \left( V_{GS} \right) + g_m \cdot v_{gs} \]

\[ i_D \left( V_{DS} + v_{ds} \right) \approx I_D \left( V_{DS} \right) + \frac{dI_D}{dV_{DS}} \cdot v_{ds} = I_D \left( V_{DS} \right) + \frac{1}{r_0} \cdot v_{ds} \]
Capacitances (9.2.1)

- Linear region:
  \[ C_{gs} = C_{gd} = \frac{1}{2} WLC_{ox} \]
  \[ C_{ox} [\text{F/\mu m}^2] \]

- Saturation region:
  \[ C_{gs} = \frac{2}{3} WLC_{ox} \]
  \[ C_{gd} = 0 \]

- Overlap capacitance:
  \[ C_{ov} = W L_{ov} C_{ox} \]

source/drain-to-body capacitance ….
(simple) Small-signal model - saturation

- Transconductance – controls the current source
- Output resistance – channel length modulation with increasing $V_{DS}$
- gate-source capacitance
- Input resistance – infinite due to gate oxide

**DC**

\[
g_m v_{gs} \quad r_0
\]

**AC**

\[
g_m v_{gs} \quad r_0
\]

**Equations**

\[
g_m = \sqrt{2 \cdot \mu_n \cdot C_{ox} \frac{W}{L} I_D}
\]

\[
r_o = \frac{1}{\lambda I_{DS}}
\]

\[
C_{GS} = \frac{2}{3} W L C_{ox}
\]
Small-signal model – more advanced

- drain-body / source-body capacitances ($C_{db}$/$C_{sb}$) – reversed biased pn-junctions
- gate overlap capacitances ($C_{ov}$)
- gate-body capacitance - outside active device area
- body transconductance – acting as 2nd gate
- source/drain resistances – (not included here)
Parameter extraction from output characteristics

- **on-resistance**: $R_{ON} \ (\Omega - \mu m)$
- **output resistance**: $r_d \ (\Omega - \mu m)$
- **on-current**: $(mA/\mu m)$
  $$I_D (V_{GS} = V_{DS} = V_{DD})$$
- **transconductance**: $g_m \ (\mu S/\mu m)$
  $$g_m = \left. \frac{\Delta I_D}{\Delta V_{GS}} \right|_{V_{DS}}$$
Parameter extraction from transfer characteristics

Long channel

Short channel (velocity saturation)

\[ \sqrt{I_D} \quad I_D \quad (\text{mA/\mu m}) \]

off-current

\[ V_{TSAT} \quad V_{TLIN} \quad V_{DD} \]

\[ V_{DS} = V_{DD} \]

\[ V_{DS} = 0.05 \text{ V} \]

\[ I_{ON} \]
Parameter extraction from transfer characteristics (log)

- $\log_{10} I_D$ (mA/μm)
- $V_{DS} = V_{DD}$
- $I_{ON}$
- $V_{DS} = 0.05$ V
- DIBL (drain-induced barrier lowering) (mV/V)
- Subthreshold swing: (mV/decade)

Modern Electronics: F4,5 MOSFET
Summary - MOSFETs

• n-type (p-doped substrate) or p-type (n-doped substrate).

• Increasing $V_{GS}$ voltage beyond $V_t$ gives inversion layer (channel) under gate
  Linear (triode): $I_D$ depends on both $V_{GS}$ and $V_{DS}$.
  Saturation (active): $I_D$ depends only on $V_{GS}$ due to pinch-off of channel at drain.

• Channel length modulation: Increasing $V_{DS}$ beyond pinch-off shortens channel.
  Increasing $I_D$ with $V_{DS}$.

• Small signal model:
  – Infinite input resistance due to gate oxide.
  – Output resistance ($r_0$) due to channel length modulation.
  – Gate-source capacitance ($C_{GS}$) dominates
  – Can add more capacitances and resistances to get more accurate (and complicated) model.
Electrostatics

- \( \lambda = \) screening length, reduced by higher gate dielectric constant or thinner channel

- \( L_g > 5 \lambda \) to avoid short channel effects

- More "wrapping" of the channel reduces \( \lambda \) -> enables \( L_g \) scaling

- Progression to FinFETs and Nanowire FETs

\[
\lambda_1 \approx \sqrt{\frac{\varepsilon_{ch}}{\varepsilon_{ox}}t_{ox}t_{ch}} > \lambda \approx \sqrt{\frac{\varepsilon_{ch}}{2\varepsilon_{ox}}t_{ch}t_{ox} \left(1 + \frac{\varepsilon_{ox}}{4\varepsilon_{ch}}t_{ch}t_{ox}\right)} > \lambda \approx \sqrt{\frac{\varepsilon_{ch}}{4\varepsilon_{ox}}t_{ch}t_{ox} \left(1 + \frac{\varepsilon_{ox}}{4\varepsilon_{ch}}t_{ch}t_{ox}\right)}
\]
## BJT vs MOSFET

<table>
<thead>
<tr>
<th></th>
<th>BJT</th>
<th>MOSFET</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Terminals</strong></td>
<td>Emitter, base, collector</td>
<td>Source, gate, drain, (body)</td>
</tr>
<tr>
<td><strong>Symmetric</strong></td>
<td>no (higher doping in emitter)</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Transport mechanism</strong></td>
<td>Diffusion</td>
<td>Drift</td>
</tr>
<tr>
<td><strong>Current formula (active region)</strong></td>
<td>$I_C = I_S \left(1 + \frac{V_{CE}}{V_A}\right) e^{V_{BE}/V_T}$</td>
<td>$I_{D,\text{sat}} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{GS} - V_t)^2 (1 + \lambda V_{DS})$</td>
</tr>
<tr>
<td><strong>Transconductance</strong></td>
<td>$g_m = \frac{I_C}{V_T}$</td>
<td>$g_m = \sqrt{2 \cdot \mu_n \cdot C_{ox} \frac{W}{L} I_D}$</td>
</tr>
<tr>
<td><strong>Input resistance ($r_0$)</strong></td>
<td>Finite due to base current</td>
<td>Infinite due to insulating gate oxide</td>
</tr>
</tbody>
</table>
Modern transistors

MOSFET
- Strain enhance mobility
- High-k gate dielectric gives higher $C_{ox}$ without leakage current
- Tri-gate enables channel length reduction due to better electrostatic control (less short-channel effects)
- (III-V semiconductors would give higher mobility)

High electron mobility transistor (HEMT)
- III-V semiconductors to get high mobility
- Move doping away from channel to avoid scattering

Heterojunction bipolar transistor (HBT)
- Combine (III-V) semiconductors with different $E_g$. Valence band barrier gives less backinjection from base to emitter. Can increase base doping to lower base resistance but keep same gain.
Analog electronics

- HEMTs, Si MOSFETs, HBTs
- Improvement due to gate length scaling
Modern MOSFET – (semi)ballistic III-V nanowires

- $g_m = 3.3 \text{ mS/µm}$
- $f_T = 280 \text{ GHz}$
- $f_{\text{max}} = 312 \text{ GHz}$

$V_{DS} = -0.4, ..., 1 \text{ V}$
$\Delta V_{DS} = 0.2 \text{ V}$

$L_g = 32 \text{ nm}$
$L_g = 200 \text{ nm}$

$g_m = 0.2 \text{ fS/s/µrad}$
$g_s = 0.53 \text{ mS}$
$\tau_1 = 52 \text{ ps}$
$\tau_s = 63 \text{ ps}$