Lecture 5: CMOS Transistor Theory

Deming Chen

Slides based on the initial set from David Harris
Outline

- Introduction
- MOS Capacitor
- nMOS I-V Characteristics
- pMOS I-V Characteristics
- Gate and Diffusion Capacitance
- Nonideal Transistor Behavior
  - High Field Effects
  - Channel Length Modulation
  - Threshold Voltage Effects
  - Leakage
- Process and Environmental Variations
- Reading 2.1-2.4
Introduction

- So far, we have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
  - Depends on terminal voltages
  - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
  - \( I = C \left( \frac{\Delta V}{\Delta t} \right) \rightarrow \Delta t = \frac{C}{I} \Delta V \)
  - Capacitance and current determine speed
MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes
  - Accumulation
  - Depletion
  - Inversion
Terminal Voltages

- Mode of operation depends on $V_g$, $V_d$, $V_s$
  - $V_{gs} = V_g - V_s$
  - $V_{gd} = V_g - V_d$
  - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$

- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence $V_{ds} \geq 0$

- nMOS body is grounded. First assume source is 0 too.

- Three regions of operation
  - Cutoff
  - Linear
  - Saturation
nMOS Cutoff

- No channel
- \( I_{ds} \approx 0 \)
nMOS Linear

- Channel forms
- Current flows from d to s
  - e\(^{-}\) from s to d
- \(I_{ds}\) increases with \(V_{ds}\)
- Similar to linear resistor
nMOS Saturation

- Channel pinches off
- $I_{ds}$ independent of $V_{ds}$
- We say current "saturates"
- Similar to current source
I-V Characteristics

- In Linear region, $I_{ds}$ depends on
  - How much charge is in the channel?
  - How fast is the charge moving?
Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversions
  - Gate – oxide – channel

- $Q_{\text{channel}} = CV$

- $C = C_g = \varepsilon_{\text{ox}} \frac{W}{t_{\text{ox}}} = C_{\text{ox}} \frac{W}{L}$

- $V = V_{gc} - V_t = \left( V_{gs} - \frac{V_{ds}}{2} \right) - V_t$

- SiO$_2$ gate oxide (good insulator, $\varepsilon_{\text{ox}} = 3.9$)

- $C_{\text{ox}} = \frac{\varepsilon_{\text{ox}}}{t_{\text{ox}}}$
Carrier velocity

- Charge is carried by e-
- Electrons are propelled by the lateral electric field between source and drain
  - \( E = \frac{V_{ds}}{L} \)
- Carrier velocity \( v \) proportional to lateral E-field
  - \( v = \mu E \quad \mu \) called mobility
- Time for carrier to cross channel:
  - \( t = \frac{L}{v} \)
Now we know
- How much charge $Q_{\text{channel}}$ is in the channel
- How much time $t$ each carrier takes to cross

\[
I_{ds} = \frac{Q_{\text{channel}}}{t} = \mu C_{ox} \frac{W}{L} \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} = \beta \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds}
\]

$\beta = \mu C_{ox} \frac{W}{L}$
nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta \left( V_{gs} - V_t - \frac{V_{dsat}}{2} \right) V_{dsat}$$

$$= \frac{\beta}{2} \left( V_{gs} - V_t \right)^2$$
nMOS I-V Summary

- Shockley 1\textsuperscript{st} order transistor models

\[
I_{ds} = \begin{cases} 
0 & V_{gs} < V_{t} \\
\beta \left( V_{gs} - V_{t} - \frac{V_{ds}}{2} \right) V_{ds} & V_{ds} < V_{dsat} \\
\frac{\beta}{2} \left( V_{gs} - V_{t} \right)^2 & V_{ds} > V_{dsat}
\end{cases}
\]
cutoff  
linear  
saturation
Example

- Use 0.6 μm process
  - From AMI Semiconductor
  - $t_{ox} = 100 \text{ Å}$
  - $\mu = 350 \text{ cm}^2/\text{V} \cdot \text{s}$
  - $V_t = 0.7 \text{ V}$

- Plot $I_{ds}$ vs. $V_{ds}$
  - $V_{gs} = 0, 1, 2, 3, 4, 5$
  - Use $W/L = 4/2$ λ

\[
\beta = \mu C_{ox} \frac{W}{L} = (350) \left( \frac{3.9 \times 8.85 \cdot 10^{-14}}{100 \cdot 10^{-8}} \right) \left( \frac{W}{L} \right) = 120 \frac{W}{L} \mu \text{A/V}^2
\]
pMOS I-V

- All dopings and voltages are inverted for pMOS
  - Source is the more positive terminal
- Mobility $\mu_p$ is determined by holes
  - Typically 2-3x lower than that of electrons $\mu_n$
  - 120 cm$^2$/V•s in AMI 0.6 $\mu$m process
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$
Capacitance

- Any two conductors separated by an insulator have capacitance
- Gate to channel capacitor is very important
  - Creates channel charge necessary for operation
- Source and drain have capacitance to body
  - Across reverse-biased diodes
  - Called diffusion capacitance because it is associated with source/drain diffusion
Gate Capacitance

- Approximate channel as connected to source
- \( C_{gs} = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL = C_{\text{permicron}}W \)
- \( C_{\text{permicron}} \) is typically about 2 fF/\( \mu \text{m} \)

![Diagram of MOSFET with gate oxide and body](image)
Diffusion Capacitance

- $C_{sb}$, $C_{db}$
- Undesirable, called \textit{parasitic} capacitance
- Capacitance depends on area and perimeter
  - Use small diffusion nodes
  - Comparable to $C_g$ for contacted diff
  - $\frac{1}{2} C_g$ for uncontacted
  - Varies with process
Ideal Transistor I-V

- **Shockley long-channel transistor models**

\[
I_{ds} = \begin{cases} 
0 & V_{gs} < V_t \\
\beta \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right)V_{ds} & V_{ds} < V_{dsat} \\
\frac{\beta}{2} \left( V_{gs} - V_t \right)^2 & V_{ds} > V_{dsat}
\end{cases}
\]

- **cutoff**
- **linear**
- **saturation**
65 nm IBM process, $V_{DD} = 1.0$ V

- Velocity saturation & Mobility degradation: $I_{on}$ lower than ideal model predicts
- Channel length modulation: $I_{on} = 747$ mA @ $V_{gs} = V_{ds} = V_{DD}$
- Saturation current increases with $V_{gs}$
- Velocity saturation & Mobility degradation: Saturation current increases less than quadratically with $V_{gs}$
ON and OFF Current

- $I_{on} = I_{ds} @ V_{gs} = V_{ds} = V_{DD}$
  - Saturation

- $I_{off} = I_{ds} @ V_{gs} = 0, V_{ds} = V_{DD}$
  - Cutoff
Electric Fields Effects

- Vertical electric field: \( E_{\text{vert}} = \frac{V_{gs}}{t_{ox}} \)
  - Attracts carriers into channel
  - Long channel: \( Q_{\text{channel}} \propto E_{\text{vert}} \)
- Lateral electric field: \( E_{\text{lat}} = \frac{V_{ds}}{L} \)
  - Accelerates carriers from drain to source
  - Long channel: \( v = \mu E_{\text{lat}} \)
Coffee Analogy

- Tired student runs from ECE425 lab to ECE cafe
- Freshmen are pouring out of the ECE 120 lecture hall
- $V_{ds}$ is how long you have been up
  - Your velocity = fatigue $\times$ mobility
- $V_{gs}$ is a wind blowing you against the glass ($\text{SiO}_2$) wall
- At high $V_{gs}$, you are buffeted against the wall
  - Mobility degradation
- At high $V_{ds}$, you scatter off freshmen, fall down, get up
  - Velocity saturation
  - Don’t confuse this with the saturation region
Threshold Voltage Effects

- $V_t$ is $V_{gs}$ for which the channel starts to invert
- Ideal models assumed $V_t$ is constant
- Really depends (weakly) on almost everything else:
  - Body voltage: *Body Effect*
  - Drain voltage: *Drain-Induced Barrier Lowering*
  - Channel length: *Short Channel Effect*
Body Effect

- Body is a fourth transistor terminal
- $V_{sb}$ affects the charge required to invert the channel
  - Increasing $V_s$ or decreasing $V_b$ increases $V_t$
  
  $$V_t = V_{i0} + \gamma \left( \sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right)$$

- $\phi_s = \text{surface potential}$ at threshold
  
  $$\phi_s = 2\nu_T \ln \frac{N_A}{n_i}$$
  - Depends on doping level $N_A$
  - And intrinsic carrier concentration $n_i$

- $\gamma = \text{body effect coefficient}$
  
  $$\gamma = \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{2q\varepsilon_{si}N_A} = \frac{\sqrt{2q\varepsilon_{si}N_A}}{C_{ox}}$$
Body Effect Cont.

- For small source-to-body voltage, treat as linear

\[ V_t = V_{t0} + k_\gamma V_{sb} \]

\[ k_\gamma = \frac{\gamma}{2\sqrt{\phi_s}} = \frac{\sqrt{\frac{q\varepsilon_{si}N_A}{v_T \ln \frac{N_A}{n_i}}}}{2C_{ox}} \]
DIBL

- Electric field from drain affects channel
- More pronounced in small transistors where the drain is closer to the channel
- Drain-Induced Barrier Lowering
  - Drain voltage also affect $V_t$
  
  $$V'_t = V_t - \eta V_{ds}$$

- High drain voltage causes current to increase.
In small transistors, source/drain depletion regions extend into the channel
  - Impacts the amount of charge required to invert the channel
  - And thus makes $V_t$ a function of channel length

Short channel effect: $V_t$ increases with L
  - Some processes exhibit a reverse short channel effect in which $V_t$ decreases with L
Leakage

- What about current in cutoff?
- Simulated results
- What differs?
  - Current doesn’t go to 0 in cutoff

![Graph showing current vs. voltage for CMOS transistors with subthreshold and saturation regions, indicating leakage current and other phenomena.]
Leakage Sources

- Subthreshold conduction
  - Transistors can’t abruptly turn ON or OFF
  - Dominant source in contemporary transistors
- Gate leakage
  - Tunneling through ultrathin gate dielectric
- Junction leakage
  - Reverse-biased PN junction diode current
Subthreshold Leakage

- Subthreshold leakage exponential with $V_{gs}$

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_{10} + \eta V_{ds} - k_y V_{sb}}{n^v_T}} \left(1 - e^{-\frac{-V_{ds}}{v_T}}\right)$$

- $n$ is process dependent
  - typically 1.3-1.7

- Rewrite relative to $I_{off}$ on log scale

$$S \approx 100 \text{ mV/decade} @ \text{room temperature}$$
Gate Leakage

- Carriers tunnel through very thin gate oxides
- Exponentially sensitive to $t_{ox}$ and $V_{DD}$

$$I_{gate} = WA \left( \frac{V_{DD}}{t_{ox}} \right)^2 \ exp \left( -B \frac{t_{ox}}{V_{DD}} \right)$$

- $A$ and $B$ are tech constants
- Greater for electrons
  - So nMOS gates leak more
- Negligible for older processes ($t_{ox} > 20 \text{ Å}$)
- Critically important at 65 nm and below ($t_{ox} \approx 10.5 \text{ Å}$)

From [Song01]
Temperature Sensitivity

- Increasing temperature
  - Reduces mobility
  - Reduces $V_t$
- $I_{ON}$ decreases with temperature
- $I_{OFF}$ increases with temperature
So What?

- So what if transistors are not ideal?
  - They still behave like switches.
- But these effects matter for…
  - Supply voltage choice
  - Logical effort
  - Quiescent power consumption
  - Pass transistors
  - Temperature of operation
Parameter Variation

- Transistors have uncertainty in parameters
  - Process: $L_{\text{eff}}$, $V_t$, $t_{\text{ox}}$ of nMOS and pMOS
  - Vary around typical (T) values
- Fast (F)
  - $L_{\text{eff}}$: short
  - $V_t$: low
  - $t_{\text{ox}}$: thin
- Slow (S): opposite
- Not all parameters are independent for nMOS and pMOS
Environmental Variation

- $V_{DD}$ and $T$ also vary in time and space
- **Fast:**
  - $V_{DD}$: high
  - $T$: low

<table>
<thead>
<tr>
<th>Corner</th>
<th>Voltage</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1.98</td>
<td>0°C</td>
</tr>
<tr>
<td>T</td>
<td>1.8</td>
<td>70°C</td>
</tr>
<tr>
<td>S</td>
<td>1.62</td>
<td>125°C</td>
</tr>
</tbody>
</table>
Process Corners

- Process corners describe worst case variations
  - If a design works in all corners, it will probably work for any variation.
- Describe corner with four letters (T, F, S)
  - nMOS speed
  - pMOS speed
  - Voltage
  - Temperature
## Important Corners

- Some critical simulation corners include

<table>
<thead>
<tr>
<th>Purpose</th>
<th>nMOS</th>
<th>pMOS</th>
<th>$V_{DD}$</th>
<th>Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle time</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Power</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Subthreshold leakage</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>S</td>
</tr>
</tbody>
</table>
Summary

- Fundamental transistor theory
- Transistors are non-ideal
- Leakage can be a big concern while scaling down
- Also parameter variation is an issue

- Next lecture
  - DC & Transient Response
  - Readings 2.5, 4.1-4.3
Backup Slides
Mobility Degradation

- High $E_{vert}$ effectively reduces mobility
  - Collisions with oxide interface

$$
\mu_{\text{eff}}^{-n} = \frac{540 \ \text{cm}^2 \text{V} \cdot \text{s}}{1 + \left( \frac{V_{gs} + V_{t}}{0.54 \ \text{V} \text{nm} \ t_{ox}} \right)^{1.85}}
$$

$$
\mu_{\text{eff}}^{-p} = \frac{185 \ \text{cm}^2 \text{V} \cdot \text{s}}{1 + \left( \frac{V_{gs} + 1.5V_{t}}{0.338 \ \text{V} \text{nm} \ t_{ox}} \right)}
$$
Velocity Saturation

- At high $E_{\text{lat}}$, carrier velocity rolls off
  - Carriers scatter off atoms in silicon lattice
  - Velocity reaches $v_{\text{sat}}$
    - Electrons: $10^7$ cm/s
    - Holes: $8 \times 10^6$ cm/s
  - Better model

$$v = \begin{cases} \frac{\mu_{\text{eff}} E}{1 + \frac{E}{E_c}} & E < E_c \\ v_{\text{sat}} & E \geq E_c \end{cases}$$

$$E_c = \frac{2v_{\text{sat}}}{\mu_{\text{eff}}}$$

![Graph showing velocity saturation for electrons and holes with velocity as a function of electric field.](image)
Vel Sat I-V Effects

- Ideal transistor ON current increases with $V_{DD}^2$
  \[ I_{ds} = \mu C_{ox} \frac{W}{L} \left( \frac{V_{gs} - V_t}{2} \right)^2 = \frac{\beta}{2} \left( V_{gs} - V_t \right)^2 \]

- Velocity-saturated ON current increases with $V_{DD}$
  \[ I_{ds} = C_{ox} W \left( V_{gs} - V_t \right) v_{max} \]

- Real transistors are partially velocity saturated
  - Approximate with $\alpha$-power law model
  - $I_{ds} \propto V_{DD}^\alpha$
  - $1 < \alpha < 2$ determined empirically ($\approx 1.3$ for 65 nm)
$\alpha$-Power Model

\[ I_{ds} = \begin{cases} 
0 & V_{gs} < V_t \quad \text{cutoff} \\
I_{dsat} \frac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} \quad \text{linear} \\
I_{dsat} & V_{ds} > V_{dsat} \quad \text{saturation}
\end{cases} \]

\[ I_{dsat} = P_c \frac{\beta}{2} \left(V_{gs} - V_t\right)^{\alpha} \]

\[ V_{dsat} = P_v \left(V_{gs} - V_t\right)^{\alpha/2} \]
Channel Length Modulation

- Reverse-biased p-n junctions form a *depletion region*
  - Region between n and p with no carriers
  - Width of depletion $L_d$ region grows with reverse bias
  - $L_{\text{eff}} = L - L_d$
- Shorter $L_{\text{eff}}$ gives more current
  - $I_{ds}$ increases with $V_{ds}$
  - Even in saturation
Chan Length Mod I-V

\[ I_{ds} = \frac{\beta}{2} (V_{gs} - V_t)^2 (1 + \lambda V_{ds}) \]

- \( \lambda \) = *channel length modulation coefficient*
  - not feature size
  - Empirically fit to I-V characteristics
Reverse-biased p-n junctions have some leakage
- Ordinary diode leakage
- Band-to-band tunneling (BTBT)
- Gate-induced drain leakage (GIDL)
Diode Leakage

- Reverse-biased p-n junctions have some leakage
  \[ I_D = I_S \left( e^{\frac{V_D}{V_T}} - 1 \right) \]

- At any significant negative diode voltage, \( I_D = -I_S \)
- \( I_S \) depends on doping levels
  - And area and perimeter of diffusion regions
  - Typically < 1 fA/\( \mu \text{m}^2 \) (negligible)
Band-to-Band Tunneling

- Tunneling across heavily doped p-n junctions
  - Especially sidewall between drain & channel
    when *halo doping* is used to increase $V_t$
- Increases junction leakage to significant levels

\[
I_{\text{BTBT}} = WX_j \frac{E_j}{E_g^{0.5}} V_{dd} e^{- \frac{BE_g}{E_j}}
\]

\[
E_j = \sqrt{\frac{2qN_{\text{balo}}N_{\text{sd}}}{\varepsilon(N_{\text{balo}} + N_{\text{sd}})}} \left( V_{DD} + v_T \ln \frac{N_{\text{balo}}N_{\text{sd}}}{n_i^2} \right)
\]

- $X_j$: sidewall junction depth
- $E_g$: bandgap voltage
- $A, B$: tech constants
Gate-Induced Drain Leakage

- Occurs at overlap between gate and drain
  - Most pronounced when drain is at $V_{DD}$, gate is at a negative voltage
  - Thwarts efforts to reduce subthreshold leakage using a negative gate voltage