MOSFET: Introduction

Metal Oxide Semiconductor Field Effect Transistor (MOSFET) is widely used to implement digital designs

- High integration density
- Relatively simple manufacturing process

It is therefore possible to realize $10^6-7$ transistors on an Integrated Circuit (IC) economically
Heavily doped n-type source and drain regions are implanted (diffused) into a lightly doped p-type substrate (body)

A thin layer (approx. 50 Å) of silicon dioxide (SiO₂) is grown over the region between source and drain and is called thin or gate oxide
• Gate oxide is covered by a conductive material, often polycrystalline silicon (polysilicon) and forms the gate of the transistor

• MOS transistors are insulated from each other by thick oxide (SiO2) and reverse biased p-n+ diode

• Adding p+ field implant (channel stop implant) makes sure a parasitic MOS transistor is not formed

**MOS Transistor as a switch**

\[ V_{in} > V_T : \text{a conducting channel is formed between source and drain and current flows} \]

\[ V_{in} < V_T : \text{the channel does not form and switch is said to be Open} \]

\[ V_{in} > V_T : \text{current is a function of gate voltage} \]
In an NMOS transistor, current is carried by electrons (from source, through an n-type channel to the drain)
Different than diode where both holes and electrons contribute to the total current
Therefore, MOS transistor is also known as unipolar device

Another MOS device can be formed by having p+ source and drain and n-substrate (PMOS)
Current is carried by holes through a p-type channel

A technology that uses NMOS (PMOS) transistors only is called NMOS (PMOS) technology
In NMOS or PMOS technologies, substrate is common and is connected to +ve voltage, GND (NMOS) or VDD (PMOS)
In a complementary MOS (CMOS) technology, both PMOS and NMOS transistors are used. NMOS and PMOS devices are fabricated in isolated regions from each other (i.e., no common substrate for all devices).

MOS transistor is a 4 terminal device, if 4th terminal is not shown it is assumed to be connected to appropriate voltage.
Static Behavior

Only the NMOS transistor is discussed, however, arguments are valid for PMOS transistor as well.

The threshold voltage
Consider the case where $V_{gs} = 0$ and drain, source and bulk are connected to ground.
Static Behavior

Under these conditions (no channel), source and drain are connected by back to back diodes having 0 V bias (no conduction)

Hence, high resistance between source and drain ($10^7 \, \Omega$)

If now the gate voltage ($V_{GS}$) is increased, gate and substrate form plates of a capacitor with oxide as dielectric

+ve gate voltage causes +ve charge on gate and -ve charge on the substrate side

In substrate it occurs in two steps (i) depletion of mobile holes, (ii) accumulation of -ve charge (inversion)
At certain $V_{gs}$, potential at the interface reaches a critical value, where surface inverts to n-type (start of strong inversion).

Further VGS increase does not increase the depletion width but increases electrons in the inversion layer.

**Threshold Voltage**

where

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

where

$$\gamma = \frac{\sqrt{2q\varepsilon_{si}N_A}}{C_{OX}}$$

$V_T$ is +ve for NMOS and -ve for PMOS devices.
Current-Voltage Relationship

When $V_{GS} > V_T$

Let at any point along the channel, the voltage is $V(x)$ and gate to channel voltage at that point is $V_{GS} - V(x)$
If the $V_{gs} - V(x) > V_T$ for all $x$, the induced channel charge per unit area at $x$

$$Q_i(x) = -C_{OX}[V_{gs} - V(x) - V_T]$$

Current is given by

$$I_D = -\nu(x)Q_i(x)W$$

The electron velocity is given by

$$\nu_n = -\mu_n E(x) = \mu_n \frac{dV}{dx}$$

Therefore,

$$I_D dx = \mu_n C_{OX}W (V_{gs} - V - V_T) dV$$

Integrating the equation over the length $L$ yields

$$I_D = K'_n \frac{W}{L} [(V_{gs} - V_T)V_{ds} - \frac{V_{ds}^2}{2}]$$

or

$$I_D = K_n [(V_{gs} - V_T)V_{ds} - \frac{V_{ds}^2}{2}]$$
$K'_n$ is known as the process trans-conductance parameter and equals

$$K'_n = \mu_n C_{OX} = \mu_n \frac{\mathcal{E}_{ox}}{t_{ox}}$$

If the $V_{GS}$ is further increased, then at some $x$, $V_{gs} - V(x) < V_T$ and that point the channel disappears and transistor is said to be **pinched-off**

Close to drain no channel exists, the pinched-off condition in the vicinity of drain is $V_{GS} - V_{DS} \leq V_T$

Under these conditions, transistor is in the **saturation region**

If a complete channel exists between source and drain, then transistors is said to be in **triode** or **linear region**

Replacing $V_{ds}$ by $V_{gs} - V_T$ in the current equation we get,

MOS current-voltage relationship in saturation region

$$I_D = \frac{K'_n \ W}{2L} (V_{gs} - V_T)^2$$
This equation is not entirely correct, the position of pinch-off point and hence the effective channel length is a function of $V_{ds}$, a more accurate equation is given as

$$I_D = \frac{K' n W}{2} \left( V_{gs} - V_T \right)^2 \left[ 1 + \lambda V_{ds} \right]$$

where is an empirical constant parameter called channel length modulation factor

(a) $I_D$ as a function of $V_{DS}$

(b) $\sqrt{I_D}$ as a function of $V_{GS}$ (for $V_{DS} = 5V$).
Dynamic Behavior
MOS transistor is a unipolar (majority carrier) device, therefore, its dynamic response is determined by time to (dis)charge various capacitances.

**MOS capacitances**

**Gate oxide capacitance:** $C_{ox} = \text{per unit area}$, for a transistor of width, $W$ and length, $L$, the $C_{gate} = WL \frac{\varepsilon_{ox}}{t_{ox}}$

From current equation it is apparent that Cox should be high or gate oxide thickness should be small.

Gate capacitance consists of several components.

Source and drain diffusions extend below the thin oxide (lateral diffusion) giving rise to overlap capacitance.
MOS Overlap Capacitance

Source and drain diffusions extend below the thin oxide (lateral diffusion) giving rise to overlap capacitance

\[ C_{gsO} = C_{gdO} = C_{ox} X_d W = C_o W \]
Average Gate Capacitance

Different distributions of gate capacitance for varying operating conditions

<table>
<thead>
<tr>
<th>Operation Region</th>
<th>$C_{gb}$</th>
<th>$C_{gs}$</th>
<th>$C_{gd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutoff</td>
<td>$C_{ox}W_{L_{eff}}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Triode</td>
<td>0</td>
<td>$C_{ox}W_{L_{eff}}/2$</td>
<td>$C_{ox}W_{L_{eff}}/2$</td>
</tr>
<tr>
<td>Saturation</td>
<td>0</td>
<td>$(2/3)C_{ox}W_{L_{eff}}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Most important regions in digital design: saturation and cut-off
Diffusion Capacitance

\[ C_{\text{diff}} = C_{\text{bottom}} + C_{\text{sw}} = C_j \times \text{AREA} + C_{jsw} \times \text{PERIMETER} \]

\[ = C_j L_s W + C_{jsw} (2L_s + W) \]
The Sub-Micron MOS Transistor
Short Channel Effects

• Threshold Variations
• Parasitic Resistances
• Velocity Sauturation and Mobility Degradation
• Subthreshold Conduction
• Latchup
Threshold Variations

Long-channel threshold

Threshold as a function of the length (for low $V_{DS}$)

Drain-induced barrier lowering (for low $L$)
Parasitic Resistances

With transistor scaling, junctions are made shallower & contacts windows are made smaller while their depth is increased. Technology and design objective is to reduce source-drain resistance. Often source drain regions are covered by titanium or tungsten (silicidation) to reduce the resistance.
Variation in I-V Characteristics

\[ E_t (V/\mu m) \]

\[ \mu_n (cm^2/Vs) \]

\[ \nu_{sat} = 10^7 \]

Constant mobility (slope = \( \mu \))

Constant velocity

\[ E_{sat} = 1.5 \]

(a) Velocity saturation

(b) Mobility degradation
Variation in I-V Characteristics

- While developing the I-V equation we assumed that carrier velocity is proportional to $E$
  However, as $E = E_{\text{sat}}$ (approx. 104/micron), the carrier velocity saturates, as a consequence

  $$I_{DSAT} = \nu_{sat} C_{ox} W (V_{gs} - V_{Dsat} - V_T)$$

- In long channel MOSFET we also assumed that there is no vertical electric field
  However, as transistor scales, $E_{\text{vertical}}$ cannot be ignored
  Carrier mobility is decreased as vertical electric field is increased
Variation in I-V Characteristics

(a) $I_D$ as a function of $V_{DS}$

(b) $I_D$ as a function of $V_{GS}$ (for $V_{DS} = 5$ V).

Linear Dependence on $V_{GS}$
Sub-Threshold Conduction

\[ \ln(I_D) (A) \]

\[ V_{GS} (V) \]

- Linear region
- Subthreshold exponential region

\[ V_T \]
Latchup

(a) Origin of latchup

(b) Equivalent circuit
SPICE MODELS

Level 1: Long Channel Equations - Very Simple

Level 2: Physical Model - Includes Velocity Saturation and Threshold Variations

Level 3: Semi-Emperical - Based on curve fitting to measured devices

Level 4 (BSIM): Emperical - Simple and Popular
## MAIN MOS SPICE PARAMETERS

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Symbol</th>
<th>SPICE Name</th>
<th>Units</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPICE Model Index</td>
<td></td>
<td>LEVEL</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Zero-Bias Threshold Voltage</td>
<td>VT0</td>
<td>VT0</td>
<td>V</td>
<td>0</td>
</tr>
<tr>
<td>Process Transconductance</td>
<td>k'</td>
<td>KP</td>
<td>A/V2</td>
<td>2E-5</td>
</tr>
<tr>
<td>Body-Bias Parameter</td>
<td>g</td>
<td>GAMMA</td>
<td>V0.5</td>
<td>0</td>
</tr>
<tr>
<td>Channel Modulation</td>
<td>1</td>
<td>LAMBDA</td>
<td>I/V</td>
<td>0</td>
</tr>
<tr>
<td>Oxide Thickness</td>
<td>tox</td>
<td>TOX</td>
<td>m</td>
<td>1E-7</td>
</tr>
<tr>
<td>Lateral Diffusion</td>
<td>xd</td>
<td>LD</td>
<td>m</td>
<td>0</td>
</tr>
<tr>
<td>Metallurgical Junction Depth</td>
<td>xj</td>
<td>XJ</td>
<td>m</td>
<td>0</td>
</tr>
<tr>
<td>Surface Inversion Potential</td>
<td>2</td>
<td>F</td>
<td></td>
<td>PHI</td>
</tr>
<tr>
<td>Substrate Doping</td>
<td>NA,ND</td>
<td>NSUB</td>
<td>cm-3</td>
<td>0</td>
</tr>
<tr>
<td>Surface State Density</td>
<td>Qss/q</td>
<td>NSS</td>
<td>cm-3</td>
<td>0</td>
</tr>
<tr>
<td>Fast Surface State Density</td>
<td>NFS</td>
<td></td>
<td>cm-3</td>
<td>0</td>
</tr>
<tr>
<td>Total Channel Charge Coefficient</td>
<td>NEFF</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Type of Gate Material</td>
<td>TPG</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Surface Mobility</td>
<td>m0</td>
<td>U0</td>
<td>cm2/V-sec</td>
<td>600</td>
</tr>
<tr>
<td>Maximum Drift Velocity</td>
<td>umax</td>
<td>VMAX</td>
<td>m/s</td>
<td>0</td>
</tr>
<tr>
<td>Mobility Critical Field</td>
<td>xcrit</td>
<td>UCRIT</td>
<td>V/cm</td>
<td>1E4</td>
</tr>
<tr>
<td>Critical Field Exponent in Mobility Degradation</td>
<td>UEXP</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Transverse Field Exponent (mobility)</td>
<td>UTRA</td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
# SPICE Parameters for Parasitics

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Symbol</th>
<th>SPICE Name</th>
<th>Units</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source resistance</td>
<td>$R_S$</td>
<td>$RS$</td>
<td>$\Omega$</td>
<td>0</td>
</tr>
<tr>
<td>Drain resistance</td>
<td>$R_D$</td>
<td>$RD$</td>
<td>$\Omega$</td>
<td>0</td>
</tr>
<tr>
<td>Sheet resistance (Source/Drain)</td>
<td>$R_o$</td>
<td>$RSH$</td>
<td>$\Omega/\epsilon$</td>
<td>0</td>
</tr>
<tr>
<td>Zero Bias Bulk Junction Cap</td>
<td>$C_{j0}$</td>
<td>$CJ$</td>
<td>F/m$^2$</td>
<td>0</td>
</tr>
<tr>
<td>Bulk Junction Grading Coeff.</td>
<td>$m$</td>
<td>$MJ$</td>
<td>-</td>
<td>0.5</td>
</tr>
<tr>
<td>Zero Bias Side Wall Junction Cap</td>
<td>$C_{jsw0}$</td>
<td>$CJSW$</td>
<td>F/m</td>
<td>0</td>
</tr>
<tr>
<td>Side Wall Grading Coeff.</td>
<td>$m_{sw}$</td>
<td>$MJSW$</td>
<td>-</td>
<td>0.3</td>
</tr>
<tr>
<td>Gate-Bulk Overlap Capacitance</td>
<td>$C_{gbo}$</td>
<td>$CGBO$</td>
<td>F/m</td>
<td>0</td>
</tr>
<tr>
<td>Gate-Source Overlap Capacitance</td>
<td>$C_{gs0}$</td>
<td>$CGSO$</td>
<td>F/m</td>
<td>0</td>
</tr>
<tr>
<td>Gate-Drain Overlap Capacitance</td>
<td>$C_{gd0}$</td>
<td>$CGDO$</td>
<td>F/m</td>
<td>0</td>
</tr>
<tr>
<td>Bulk Junction Leakage Current</td>
<td>$I_S$</td>
<td>$IS$</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>Bulk Junction Leakage Current Density</td>
<td>$J_S$</td>
<td>$JS$</td>
<td>A/m$^2$</td>
<td>1E-8</td>
</tr>
<tr>
<td>Bulk Junction Potential</td>
<td>$\phi_0$</td>
<td>$PB$</td>
<td>V</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Ever since ICs were invented, dimensions are scaled to
❖ Integrated more transistors in the same area
❖ Allow higher operational speed
■ Scaling has profound impact on many aspects of ICs
■ Constant Voltage Scaling
❖ All device dimensions are scaled by a factor $S$
❖ Voltage (i.e., VDD) after the scaling is same as before
❖ This method of scaling is followed till 0.8 micron
❖ However for lower geometries, higher electric field resulted in poor device reliability
# Technology Evolution

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel length (μm)</td>
<td>0.4</td>
<td>0.3</td>
<td>0.25</td>
<td>0.18</td>
<td>0.13</td>
<td>0.1</td>
</tr>
<tr>
<td>Gate oxide (nm)</td>
<td>12</td>
<td>7</td>
<td>6</td>
<td>4.5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$V_{DD}$ (V)</td>
<td>3.3</td>
<td>2.2</td>
<td>2.2</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>$V_T$ (V)</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>NMOS $I_{D_{Sat}}$ (mA/μm) (@ $V_{GS} = V_{DD}$)</td>
<td>0.35</td>
<td>0.27</td>
<td>0.31</td>
<td>0.21</td>
<td>0.29</td>
<td>0.33</td>
</tr>
<tr>
<td>PMOS $I_{D_{Sat}}$ (mA/μm) (@ $V_{GS} = V_{DD}$)</td>
<td>0.16</td>
<td>0.11</td>
<td>0.14</td>
<td>0.09</td>
<td>0.13</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Devices parameters vary between runs and even on the same die!

Variations in the process parameters, such as impurity concentration densities, oxide thicknesses, and diffusion depths. These are caused by non-uniform conditions during the deposition and/or the diffusion of the impurities. This introduces variations in the sheet resistances and transistor parameters such as the threshold voltage.

Variations in the dimensions of the devices, mainly resulting from the limited resolution of the photolithographic process. This causes \((W/L)\) variations in MOS transistors and mismatches in the emitter areas of bipolar devices.
Impact of Device Variations

Delay of Adder circuit as a function of variations in $L$ and $V_T$