Chapter 6  Physics of MOS Transistors

- 6.1 Structure of MOSFET
- 6.2 Operation of MOSFET
- 6.3 MOS Device Models
- 6.4 PMOS Transistor
- 6.5 CMOS Technology
- 6.6 Comparison of Bipolar and CMOS Devices
<table>
<thead>
<tr>
<th>Operation of MOSFETs</th>
<th>MOS Device Models</th>
<th>PMOS Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>• MOS Structure</td>
<td>• Large-Signal Model</td>
<td>• Structure</td>
</tr>
<tr>
<td>• Operation in Triode Region</td>
<td>• Small-Signal Model</td>
<td>• Models</td>
</tr>
<tr>
<td>• Operation in Saturation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• I/V Characteristics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The MOS structure can be thought of as a parallel-plate capacitor, with the top plate being the positive plate, oxide being the dielectric, and Si substrate being the negative plate. (We are assuming P-substrate.)
This device is symmetric, so either of the n+ regions can be source or drain.
The gate is formed by polysilicon, and the insulator by Silicon dioxide.
Formation of Channel

First, the holes are repelled by the positive gate voltage, leaving behind negative ions and forming a depletion region. Next, electrons are attracted to the interface, creating a channel ("inversion layer").
The inversion channel of a MOSFET can be seen as a resistor. Since the charge density inside the channel depends on the gate voltage, this resistance is also voltage-dependent.
As the gate voltage decreases, the output drops because the channel resistance increases.

This type of gain control finds application in cell phones to avoid saturation near base stations.
MOSFET Characteristics

- The MOS characteristics are measured by varying $V_G$ while keeping $V_D$ constant, and varying $V_D$ while keeping $V_G$ constant.
- (d) shows the voltage dependence of channel resistance.

The MOS characteristics are measured by varying $V_G$ while keeping $V_D$ constant, and varying $V_D$ while keeping $V_G$ constant.

(d) shows the voltage dependence of channel resistance.
Small gate length and oxide thickness yield low channel resistance, which will increase the drain current.
Effect of W

- As the gate width increases, the current increases due to a decrease in resistance. However, gate capacitance also increases thus, limiting the speed of the circuit.

- An increase in W can be seen as two devices in parallel.
Since there’s a channel resistance between drain and source, and if drain is biased higher than the source, channel potential increases from source to drain, and the potential between gate and channel will decrease from source to drain.
Channel Pinch-Off

As the potential difference between drain and gate becomes more positive, the inversion layer beneath the interface starts to pinch off around drain.

When $V_D - V_G = V_{th}$, the channel at drain totally pinches off, and when $V_D - V_G > V_{th}$, the channel length starts to decrease.
The channel charge density is equal to the gate capacitance times the gate voltage in excess of the threshold voltage.

\[ Q = W C_{ox} (V_{GS} - V_{TH}) \]
Charge Density at a Point

Let $x$ be a point along the channel from source to drain, and $V(x)$ its potential; the expression above gives the charge density (per unit length).

$$Q(x) = WC_{ox}[V_{GS} - V(x) - V_{TH}]$$

- Let $x$ be a point along the channel from source to drain, and $V(x)$ its potential; the expression above gives the charge density (per unit length).
The current that flows from source to drain (electrons) is related to the charge density in the channel by the charge velocity.

\[ I = Q \cdot v \]
Drain Current

\[ v = +\mu_n \frac{dV}{dx} \]

\[ I_D = WC_{ox}\left[V_{GS} - V(x) - V_{TH}\right]\mu_n \frac{dV(x)}{dx} \]

\[ I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L}\left[2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2 \right] \]
By keeping $V_G$ constant and varying $V_{DS}$, we obtain a parabolic relationship.

The maximum current occurs when $V_{DS}$ equals to $V_{GS} - V_{TH}$. 

$$
\frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2
$$
$I_D - V_{DS}$ for Different Values of $V_{GS}$

$I_{D,max} \propto (V_{GS} - V_{TH})^2$
Linear Resistance

At small $V_{DS}$, the transistor can be viewed as a resistor, with the resistance depending on the gate voltage.

It finds application as an electronic switch.

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})}$$

CH 6  Physics of MOS Transistors  20
In a cordless telephone system in which a single antenna is used for both transmission and reception, a switch is used to connect either the receiver or transmitter to the antenna.
To minimize signal attenuation, $R_{on}$ of the switch has to be as small as possible. This means larger W/L aspect ratio and greater $V_{GS}$. 
Different Regions of Operation

\[ \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 \]

- Triode Region
- Saturation Region
How to Determine ‘Region of Operation’

- When the potential difference between gate and drain is greater than $V_{TH}$, the MOSFET is in triode region.
- When the potential difference between gate and drain becomes equal to or less than $V_{TH}$, the MOSFET enters saturation region.

\[
\frac{1}{2} \mu_n C_ox \frac{W}{L} (V_{GS} - V_{TH})^2
\]
When the region of operation is not known, a region is assumed (with an intelligent guess). Then, the final answer is checked against the assumption.
The original observation that the current is constant in the saturation region is not quite correct. The end point of the channel actually moves toward the source as $V_D$ increases, increasing $I_D$. Therefore, the current in the saturation region is a weak function of the drain voltage.

\[
I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS})
\]
Unlike the Early voltage in BJT, the channel-length modulation factor can be controlled by the circuit designer.

For long L, the channel-length modulation effect is less than that of short L.
Transconductance

Transconductance is a measure of how strong the drain current changes when the gate voltage changes.

It has three different expressions.

\[ g_m \propto \sqrt{I_D} \]

\[ g_m \propto V_{GS} - V_{TH} \]

\[ g_m \propto \frac{W}{L} \]

\[ g_m \propto \frac{1}{V_{GS} - V_{TH}} \]
Doubling of $g_m$ Due to Doubling W/L

- If $W/L$ is doubled, effectively two equivalent transistors are added in parallel, thus doubling the current (if $V_{GS} - V_{TH}$ is constant) and hence $g_m$. 
Velocity Saturation

Since the channel is very short, it does not take a very large drain voltage to velocity saturate the charge particles.

In velocity saturation, the drain current becomes a linear function of gate voltage, and $g_m$ becomes a function of $W$.

$$I_D = v_{sat} \cdot Q = v_{sat} \cdot WC_{ox} \left( V_{GS} - V_{TH} \right)$$

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = v_{sat} WC_{ox}$$
Body Effect

\[ V_{TH} = V_{TH0} + \rho \left( \sqrt{2 \phi_F} + V_{SB} - \sqrt{2 \phi_F} \right) \]

- As the source potential departs from the bulk potential, the threshold voltage changes.
Based on the value of $V_{DS}$, MOSFET can be represented with different large-signal models.
Example: Behavior of $I_D$ with $V_1$ as a Function

Since $V_1$ is connected at the source, as it increases, the current drops.

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left(V_{DD} - V_1 - V_{TH}\right)^2$$
When the bias point is not perturbed significantly, small-signal model can be used to facilitate calculations.

To represent channel-length modulation, an output resistance is inserted into the model.

\[ r_o \approx \frac{1}{\lambda I_D} \]
PMOS Transistor

Just like the PNP transistor in bipolar technology, it is possible to create a MOS device where holes are the dominant carriers. It is called the PMOS transistor.

It behaves like an NMOS device with all the polarities reversed.
PMOS Equations

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

\[
I_{D,tri} = \frac{1}{2} \mu_p C_{ox} \frac{W}{L} \left[ 2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2 \right]
\]

\[
I_{D,sat} = \frac{1}{2} \mu_p C_{ox} \frac{W}{L} \left( |V_{GS}| - |V_{TH}| \right)^2 \left( 1 + \lambda |V_{DS}| \right)
\]

\[
I_{D,tri} = \frac{1}{2} \mu_p C_{ox} \frac{W}{L} \left[ 2(|V_{GS}| - |V_{TH}|)|V_{DS}| - V_{DS}^2 \right]
\]
The small-signal model of PMOS device is identical to that of NMOS transistor; therefore, $R_X$ equals $R_Y$ and hence $(1/gm)||r_o$. 
It is possible to grow an n-well inside a p-substrate to create a technology where both NMOS and PMOS can coexist. It is known as CMOS, or “Complementary MOS”.
### Comparison of Bipolar and MOS Transistors

<table>
<thead>
<tr>
<th>Bipolar Transistor</th>
<th>MOSFET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential Characteristic</td>
<td>Quadratic Characteristic</td>
</tr>
<tr>
<td>Active: ( V_{CB} &gt; 0 )</td>
<td>Saturation: ( V_{DS} &gt; V_{GS} - V_{TH} )</td>
</tr>
<tr>
<td>Saturation: ( V_{CB} &lt; 0 )</td>
<td>Triode: ( V_{DS} &lt; V_{GS} - V_{TH} )</td>
</tr>
<tr>
<td>Finite Base Current</td>
<td>Zero Gate Current</td>
</tr>
<tr>
<td>Early Effect</td>
<td>Channel-Length Modulation</td>
</tr>
<tr>
<td>Diffusion Current</td>
<td>Drift Current</td>
</tr>
<tr>
<td>-</td>
<td>Voltage-Dependent Resistor</td>
</tr>
</tbody>
</table>

- Bipolar devices have a higher \( g_m \) than MOSFETs for a given bias current due to its exponential IV characteristics.