NMOS Transistor

- NMOS Transistor
  - N-Channel MOSFET
  - Built on p-type substrate
  - MOS devices are smaller than BJTs
  - MOS devices consume less power than BJTs
NMOS Transistor - Layout

Top View

Cross Section
MOS Regions of Operation

Resistive

\[ V_{GS} > V_T \]
\[ V_{DS} \text{ small} \]

Triode

Nonlinear

\[ V_{GS} > V_T \]
\[ V_{DS} < (V_{GS} - V_T) \]

Active

Saturation

\[ V_{GS} > V_T \]
\[ V_{DS} \geq V_{GS} - V_T \]
MOS Transistor Operation

• As $V_G$ increases from zero
  – Holes in the p substrate are repelled from the gate area leaving negative ions behind
  – A depletion region is created
  – No current flows since no carriers are available

• As $V_G$ increases
  – The width of the depletion region and the potential at the oxide-silicon interface also increase
  – When the interface potential reaches a sufficiently positive value, electrons flow in the “channel”. The transistor is turned on

• As $V_G$ rises further
  – The charge in the depletion region remains relatively constant
  – The channel current continues to increase
MOS – Triode Region - 1

\[ I_D = \mu \frac{W}{L} C_{ox} \left[ (V_{GS} - V_T) V_{DS} \right] \]

\[ V_{DS} \ll \left( V_{GS} - V_T \right) \]

\[ C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3.9 \varepsilon_o}{t_{ox}} \]

\text{\(C_{ox}\): gate oxide capacitance}
\text{\(\mu\): electron mobility}
\text{\(L\): channel length}
\text{\(W\): channel width}
\text{\(V_T\): threshold voltage}
MOS – Triode Region

FET is like a linear resistor with

\[ r_{ds} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)} \]
\[ V_{GS} > V_T \]
\[ V_{DS} < (V_{GS} - V_T) \]

- Charge distribution is nonuniform across channel
- Less charge induced in proximity of drain
MOS – Active Region

Saturation occurs at pinch off when

\[ V_{DS} = (V_{GS} - V_T) = V_{DSP} \]

\[ V_{GS} > V_T \]

\[ V_{DS} > (V_{GS} - V_T) \] (saturation)

\[ I_D = \mu_n C_{ox} \frac{W}{2L} (V_{GS} - V_T)^2 \]
MOS Threshold Voltage

The value of $V_G$ for which the channel is “inverted” is called the threshold voltage $V_T$ (or $V_t$).

- Characteristics of the threshold voltage
  - Depends on equilibrium potential
  - Controlled by inversion in channel
  - Adjusted by implantation of dopants into the channel
  - Can be positive or negative
  - Influenced by the body effect
MOS – Active Region

- **Saturation**
  - Channel is pinched off
  - Increase in $V_{DS}$ has little effect on $i_D$
  - Square-law behavior wrt $(V_{GS}-V_T)$
  - Acts like a current source
Body Effect

- The body effect
  - $V_T$ varies with bias between source and body
  - Leads to modulation of $V_T$

Potential on substrate affects threshold voltage

\[
V_T (V_{SB}) = V_{To} + \gamma \left[ \left( 2|\phi_F| + V_{SB} \right)^{1/2} - \left( 2|\phi_F| \right)^{1/2} \right]
\]

\[
|\phi_F| = \left( \frac{kT}{q} \right) \ln \left( \frac{N_a}{n_i} \right)
\]

Fermi potential of material

\[
\gamma = \frac{(2qN_a \varepsilon_s)^{1/2}}{C_{ox}}
\]

Body bias coefficient
With depletion layer widening, the channel length is in effect reduced from $L$ to $L-\Delta L \Rightarrow$ Channel-length modulation

This leads to the following I-V relationship

$$i_D = \frac{1}{2} k_n \frac{W}{L} (V_{GS} - V_T)^2 \left(1 + \lambda V_{DS}\right)$$

Where $\lambda$ is a process technology parameter
Channel-Length Modulation

Channel-length modulation causes $i_D$ to increase with $v_{DS}$ in saturation region
Gate Capacitance

- Capacitance
  - Depends on bias
  - Fringing fields are present
  - Account for overlap C

\[ V_{GT} < 0 \]

\[ V_{GT} > 0, \ V_{DS} \text{ small} \]

\[ V_{GT} > 0, \ V_{DS} \text{ large} \]
Capacitance

• Gate Capacitance
  – $C_G$ determines the amount of charge to switch gate
  – Several distributed components
  – Large discontinuity as device turns on
  – At saturation capacitance is entirely between gate and source

$$C_{gs} = C_{gso} + \frac{2}{3}WLC_{ox} \left[ 1 - \left( \frac{1 - X}{2 - X} \right)^2 \right]$$

Define

$$X = \frac{V_{DS}}{V_{GS} - V_T}$$

$$C_{gd} = C_{gdo} + \frac{2}{3}WLC_{ox} \left[ 1 - \left( \frac{1}{2 - X} \right)^2 \right]$$
MOS Capacitances

- Expect capacitance between every two of the four terminals.
PMOS Transistor

- All polarities are reversed from nMOS
- \( v_{GS}, v_{DS} \) and \( V_t \) are negative
- Current \( i_D \) enters source and leaves through drain
- Hole mobility is lower \( \Rightarrow \) low transconductance
- nMOS favored over pMOS
Complementary MOS

- CMOS Characteristics
  - Combine nMOS and pMOS transistors
  - pMOS size is larger for electrical symmetry
CMOS

• **Advantages**
  – Virtually, no DC power consumed
  – No DC path between power and ground
  – Excellent noise margins ($V_{OL}=0$, $V_{OH}=V_{DD}$)
  – Inverter has sharp transfer curve

• **Drawbacks**
  – Requires more transistors
  – Process is more complicated
  – pMOS size larger to achieve electrical symmetry
  – Latch up
Voltage Transfer Characteristics (VTC)

The static operation of a logic circuit is determined by its VTC

• In low state: noise margin is $NM_L$

$$NM_L = V_{IL} - V_{OL}$$

• In high state: noise margin is $NM_H$

$$NM_H = V_{OH} - V_{IH}$$

• An ideal VTC will maximize noise margins

$V_{IL}$ and $V_{IH}$ are the points where the slope of the VTC=-1

Optimum: $$NM_L = NM_H = V_{DD} / 2$$
Switching Time & Propagation Delay

Input

Output

\[ v_I \]
\[ v_{IL} \]
\[ (v_{IL} + v_{IH})/2 \]
\[ v_{IH} \]

\[ v_o \]
\[ v_{oL} \]
\[ (v_{oL} + v_{oH})/2 \]
\[ v_{oH} \]

\[ t_r \]
\[ t_{f} \]
\[ t_{pHL} \]
\[ t_{pLH} \]
\[ t_{THL} \]
\[ t_{TLH} \]
Switching Time & Propagation Delay

\[ t_r = \text{rise time (from 10\% to 90\%)} \]
\[ t_f = \text{fall time (from 90\% to 10\%)} \]
\[ t_{pLH} = \text{low-to-high propagation delay} \]
\[ t_{pHL} = \text{high-to-low propagation delay} \]

Inverter propagation delay:

\[ t_p = \frac{1}{2} \left( t_{pLH} + t_{pHL} \right) \]
NMOS Switch

![NMOS Switch Diagram]
CMOS Switch

CMOS switch is called an inverter

The body of each device is connected to its source \( \Rightarrow \text{NO BODY EFFECT} \)
CMOS Switch – Input Low

**NMOS**

\[ V_{GSN} < V_{TN} \implies OFF \]

\[ r_{dsn} \text{ high} \]

**PMOS**

\[ r_{dsp} = \frac{1}{k_p \left( \frac{W}{L} \right) \left( V_{DD} - |V_{TP}| \right)} \]

\[ r_{dsp} \text{ is low} \]
CMOS Switch – Input High

\[ r_{dsn} = \frac{1}{k'_n \left( \frac{W}{L} \right)_n (V_{DD} - V_{TN})} \]

\( r_{dsn} \) is low

\( V_{GSP} > V_{TP} \implies OFF \)

\( r_{dsp} \) high
CMOS Inverter

\[ r_{dsn} = \frac{1}{k_N \left( \frac{W}{L} \right)_n (V_{DD} - V_T)} \]

\[ r_{dsp} = \frac{1}{k_P \left( \frac{W}{L} \right)_p (V_{DD} - V_T)} \]

Short switching transient current ➔ low power
CMOS Inverter

Advantages of CMOS inverter

- Output voltage levels are 0 and $V_{DD}$ → signal swing is maximum possible
- Static power dissipation is zero
- Low resistance paths to $V_{DD}$ and ground when needed
- High output driving capability → increased speed
- Input resistance is infinite → high fan-out

Load driving capability of CMOS is high. Transistors can sink or source large load currents that can be used to charge and discharge load capacitances.
CMOS inverter can be made to switch at specific threshold voltage by appropriately sizing the transistors.

\[
\left( \frac{W}{L} \right)_p = \frac{\mu_n}{\mu_p} \left( \frac{W}{L} \right)_n
\]

Symmetrical transfer characteristics is obtained via matching equal current driving capabilities in both directions (pull-up and pull-down).
CMOS Dynamic Operation

- Exact analysis is too tedious
- Replace all the capacitances in the circuit by a single equivalent capacitance $C$ connected between the output node of the inverter and ground
- Analyze capacitively loaded inverter to determine propagation delay
CMOS Dynamic Operation

\[ t_p = \frac{1}{2} (t_{PHL} + t_{PLH}) \]

- Components can be equalized by matching transistors
- \( t_p \) is proportional to \( C \) \( \Rightarrow \) reduce capacitance
- Larger \( V_{DD} \) means lower \( t_p \)
- Conflicting requirements exist
CMOS – Dynamic Power Dissipation

In every cycle
- $Q_N$ dissipate $\frac{1}{2} CV_{DD}^2$ of energy
- $Q_P$ dissipate $\frac{1}{2} CV_{DD}^2$ of energy
- Total energy dissipation is $CV_{DD}^2$

If inverter is switched at $f$ cycles per second, dynamic power dissipation is: $P_D = fCV_{DD}^2$
Digital Logic - Generalization

De Morgan’s Law

\[
A + B + C + \ldots = \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \ldots
\]

\[
A \cdot B \cdot C \cdot \ldots = \overline{A + B + C} + \ldots
\]

Distributive Law

\[
AB + AC + BC + BD = A(B + C) + B(C + D)
\]

• **General Procedure**
  1. Design PDN to satisfy logic function
  2. Construct PUN to be complementary of PDN in every way
  3. Optimize using distributive rule
Pull-Down and Pull-Up

**Truth Tables**

\[ Y_{DP} = \overline{A + B} \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>( Y_{DP} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ Y_{US} = \overline{AB} \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>( Y_{US} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Two-Input NOR Gate

\[ Y = \overline{A + B} = \overline{AB} \]
Pull-Down and Pull-Up

PDN-Series
NMOS

PUN-Parallel
PMOS

Truth Tables

\[ Y_{DS} = \overline{A}B \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>( Y_{DS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ Y_{UP} = \overline{A} + \overline{B} \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>( Y_{UP} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Two-Input NAND Gate

\[ Y = \overline{A} \overline{B} = \bar{A} + ar{B} \]
CMOS Logic Gate Circuits

- **Two Networks**
  - Pull-down network (PDN) with NMOS
  - Pull-up network (PUN) with PMOS

- **PDN and PUN utilize devices**
  - In parallel to form OR functions
  - In series to form AND functions

PUN conducts when inputs are low and consists of PMOS transistors

PDN consists of NMOS transistors and is active when inputs are high
Basic Logic Function

<table>
<thead>
<tr>
<th>Basic Function</th>
<th>INVERTER</th>
<th>NOR</th>
<th>NAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td><img src="image1" alt="INVERTER Symbol" /></td>
<td><img src="image2" alt="NOR Symbol" /></td>
<td><img src="image3" alt="NAND Symbol" /></td>
</tr>
<tr>
<td># Devices PUN</td>
<td>1 PMOS</td>
<td>2 PMOS-Series</td>
<td>2 PMOS-Parallel</td>
</tr>
<tr>
<td># Devices PDN</td>
<td>1 NMOS</td>
<td>2 NMOS-Parallel</td>
<td>2 NMOS-Series</td>
</tr>
<tr>
<td>Truth Table</td>
<td><img src="image4" alt="INVERTER Truth Table" /></td>
<td><img src="image5" alt="NOR Truth Table" /></td>
<td><img src="image6" alt="NAND Truth Table" /></td>
</tr>
</tbody>
</table>
Example

Implement the function

$$\bar{Y} = A\bar{B} + C$$

Pull down

$$Y = \overline{A\overline{B} + C} = \overline{A\overline{B}} \cdot \overline{C} = (\overline{A} + B) \cdot \overline{C}$$
Exclusive-OR (XOR) Function

\[ Y = A\bar{B} + \bar{A}B \quad \bar{Y} = (\bar{A} + B)(A + \bar{B}) \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>