E NEE 359a
Digital VLSI Circuits

P/N Junction, MOS Transistors, CMOS Inverter

Prof. Bruce Jacob
blj@eng.umd.edu

Credit where credit is due:
Slides contain original artwork (© Jacob 2004) as well as material taken liberally from Irwin & Vijay’s CSE477 slides (PSU), Schmit & Strojwas’s 18-322 slides (CMU), Wolf’s slides for Modern VLSI Design, and/or Rabaey’s slides (UCB). Device physics: http://hyperphysics.phy-astr.gsu.edu/hbase/solids/sselcn.html
Overview

- Electrons & holes, bands & band gaps, insulators, conductors, semiconductors
- Silicon crystal lattice & doping
- P/N junction & parasitic capacitance
- n-type/n-channel MOSFET
- Timing analysis of MOSFET, capacitance
- Body effect, series-connected FETs
- CMOS inverter: timing, switching threshold, transistor sizing
- Dynamic behavior (preview)
What Is Conductivity?

Perspective from Band Theory of Solids:

- Large band gap (not “Gap Band”) between valence and conduction bands in insulator material suggests that, at ordinary temperatures, no electrons can reach conduction band (i.e. material won’t conduct).

- In semiconductors, the band gap is small enough that thermal energy can bridge gap for small fraction of electrons.

- In conductors, there is no band gap (conduction and valence bands overlap).

Energy of electrons

INSULATOR

Valence Band

Fermi level

Conduction Band

SEMICONDUCTOR

Conduction Band

Valence Band

Fermi level

CONDUCTOR

Valence Band

Conduction Band
Silicon, Specifically

14 protons in nucleus
4 valence electrons
Si

Shared electrons of covalent bonds

Si

Si

Si
Silicon, Specifically

Silicon Lattice (artistic license exploited)
Silicon, Specifically

Silicon Lattice — It is a semiconductor
Silicon, Specifically

Semiconductor current: electron/hole flow
Silicon, Specifically

Perspective from Band Theory of Solids:

- Conductivity is non-zero; mobile electrons/holes in conduction/valence band; can be increased with doping
Silicon, Specifically

Doping: small % of foreign atoms in lattice

Breaks up regular lattice, produces dramatic changes in electrical properties

- **Donors**: pentavalent impurities (5 valence electrons) produce n-type semiconductors by adding electrons. E.g. antimony, arsenic, phosphorus
- **Acceptors**: trivalent impurities (3 valence electrons) produce p-type semiconductors by adding electron deficiencies (“holes”). E.g. boron, aluminum, gallium
Addition of acceptor impurities contributes hole energy levels low in the semiconductor band gap so that electrons can be easily excited from the valence band into these levels, leaving mobile holes in the valence band. This shifts the effective Fermi level to a point about halfway between the acceptor levels and the valence band. Electrons can be elevated from the valence band to the holes in the band gap with the energy provided by an applied voltage. Since electrons can be exchanged between the holes, the holes are said to be mobile. Holes are said to be the “majority carriers” for current flow in a p-type semiconductor.
Addition of donor impurities contributes electron energy levels high in the semiconductor band gap so that electrons can be easily excited into the conduction band. This shifts the Fermi level to a point about halfway between the donor levels and the conduction band. Electrons can be elevated to the conduction band with the energy provided by an applied voltage and move through the material. Electrons are said to be the “majority carriers” for current flow in an n-type semiconductor.
One Way to Think About It

• **P-type**: Conduction band is pulled down close to the valence band by the creation of available holes (willing acceptors of free electrons)

• **N-type**: Valence band is pushed up close to the conduction band by the addition of mobile electrons
The P/N Junction

<table>
<thead>
<tr>
<th>Acceptor side</th>
<th>Donor side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduction Band</td>
<td>Conduction Band</td>
</tr>
<tr>
<td>Valence Band</td>
<td>Valence Band</td>
</tr>
<tr>
<td>Extra hole energy levels</td>
<td>Extra electron energy levels</td>
</tr>
</tbody>
</table>

- **P-type**: extra holes in band gap allow excitation of valence-band electrons, leaving mobile holes in valence band
- **N-type**: electron energy levels near the top of the band gap allow easy excitation of electrons into conduction band
The P/N Junction

Accepter side

- Conduction Band
- Valence Band

Extra hole energy levels:

Donor side

- Conduction Band
- Valence Band

Extra electron energy levels:

Diode

- P-type silicon
- N-type silicon

p-n junction
The P/N Junction

DEPLETION REGION

If not touching, nothing happens
The P/N Junction

DEPLETION REGION

With a connection, electrons from n-region in conduction band diffuse across junction and combine with holes in p-region

(why doesn’t this continue indefinitely?)
The P/N Junction

DEPLETION REGION

Ions are formed on both sides of junction (negative ion from filled hole; positive ion from removed electron). This forms a space charge that impedes further electron flow.
The P/N Junction

DEPLETION REGION

Parasitic capacitance:

$$C_j = \frac{C_j^0}{V_d^m} \left[ 1 - \frac{V_d}{\phi_0} \right]$$

Built-in junction potential:

$$\phi_0 = \phi_T \cdot \ln \left( \frac{N_A N_D}{n_i^2} \right)$$
The P/N Junction

BIAS EFFECT on DEPLETION REGION

Equilibrium

- Upward = increased electron energy (must supply energy to make electron go up or hole to go down)
- Drift-diffusion equilibrium (current is flowing)
The P/N Junction

BIAS EFFECT on DEPLETION REGION

Forward Bias

P-side is made more positive relative to N-side, making it “downhill” to move an electron across the junction. Electron on N-side can fill a vacancy (“hole”) on P-side & move from hole to hole to the left to positive terminal (hole “moves” right).
The P/N Junction

BIAS EFFECT on DEPLETION REGION

Reverse Bias

- +

P-side is made more negative relative to N-side, making it “uphill” to move an electron across the junction. Applied voltage impedes the flow of N-region electrons across the p/n junction. Initial transient electron flow is left to right; it stops when potential (widening depletion region) equals the applied voltage.

P/N Junction

Valence Band

Conduction Band

Extra hole energy levels

Valence Band

Conduction Band

Extra electron energy levels

depletion region increases in size until new potential = applied bias
MOS Transistors

MOS Transistor, reverse-biased:

- p-doped semiconductor substrate
- PN Junction
- N-Doped Region
  - [mobile electrons]
- P-Doped Region
  - [mobile holes]
MOS Transistors

MOS Transistor, reverse-biased:

- P-Doped Region [acceptor holes]
- N-Doped Regions [donor electrons]

VDD

VDD

p-doped semiconductor substrate

VSS
MOS Transistors

MOS Transistor, reverse-biased:

- p-doped semiconductor substrate
- VDD
- VSS
MOS Transistors

NMOS Transistor with gate:

- NMOS Transistor with gate:
  - p-doped semiconductor substrate
  - Insulator (gate oxide)
  - Conductor

0

+ +

0

n n

p-doped semiconductor substrate
MOS Transistors

NMOS Transistor with bias voltages:

- p-doped semiconductor substrate
- Insulator (gate oxide)
- Conductor

0 +

0 0
MOS Transistors

NMOS Transistor with bias voltages:

- **NMOS Transistor with bias voltages:**
  - **p-doped semiconductor substrate**
  - **Insulator (gate oxide)**
  - **Gate (conductor)**
  - **CURRENT**

[Diagram of MOS Transistor]
MOS Transistors

NMOS Transistor, two views:

**TOP VIEW**

**SIDE VIEW**

- Length
- Width

- p-doped semiconductor substrate

- Gate oxide

NMOS Transistor, two views:

- n-doped semiconductor channel
- n-doped source and drain
- Gate oxide
- Gate
NMOS Transistor with bias voltages:

- NMOS Transistor with bias voltages:
- Source: 0 V
- Gate: 0 V
- Drain: V > 0

Electron Flow
MOS Transistors

PMOS Transistor with bias voltages:

- **PMOS Transistor with Bias Voltages:**
  - Drain: 0
  - Gate: VDD
  - Source: V > 0

- **Operation:**
  - **VDD:** Drain to Source
  - **V < VDD:** Source to Drain

- **Electron Flow:**
  - From Source to Drain

- **Diagram:**
  - MOS Transistor with n-doped semiconductor substrate
  - Channel p

- **Bruce Jacob**
  - University of Maryland
  - ECE Dept.
MOS Transistors

MOS Transistors:

NMOS:
- p-doped semiconductor substrate
- n-channel
- Source
- Drain
- Gate

PMOS:
- n-doped semiconductor substrate
- p-channel
- Source
- Drain
- Gate

Source        Drain
Gate
Substrate

VSS

VDD
0.25 µm transistor (Bell Labs)

Poly+silicide = “polycide gate” (lower R)
MOS Behavior

Depletion Regions

p (bulk)
MOS Behavior

\[ V_S = 0V \quad V_G = 0.5V \quad V_D = 0V \]

Depletion Regions

Charge Density

x (depth)

Depletion layer
MOS Behavior

V_{Source} \quad V_{Gate} \quad V_{Drain}

Inversion Layer forms when \( V_{GS} > V_T \)

Assume \( V_T = 0.75V \) (threshold voltage)

\( V_S = 0V \quad V_G = 1V \quad V_D = 0V \)

Charge Density

Gate

Oxide

Inversion layer

Substrate (p-type)

Depletion layer

x (depth)
MOS Behavior: linear region

Assume $V_T = 0.75V$ (threshold voltage)

$V_S = 0V$  
$V_G = 1V$  
$V_D = 0.001V$

$$I_{DS} = \mu_n \frac{\varepsilon_{OX}}{t_{OX}} \left( \frac{W}{L} \right) (V_{GS} - V_T) V_{DS}$$

True when $V_{GS} > V_T$ & $V_{DS} \ll V_{GS} - V_T$
MOS Behavior: ‘linear’ region

Assume $V_T = 0.75V$ (threshold voltage)

- $V_S = 0V$
- $V_G = 1V$
- $V_D = 0.15V$

\[
I_{DS} = \mu_n \frac{\varepsilon_{\text{ox}} W}{L} \left[(V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2\right]
\]

True when $V_{GS} > V_T$ & $V_{DS} \leq V_{GS} - V_T$
MOS Behavior: saturation

\[ V_S = 0V \quad V_G = 1V \quad V_D = 0.25V \]

\[ I_{DS} = \frac{1}{2} \left( \mu_n \frac{\varepsilon_{ox}}{t_{ox}} \frac{W}{L} \right) (V_{GS} - V_T)^2 \]

True when \( V_{GS} > V_T \) & \( V_{DS} = V_{GS} - V_T \)
MOS Behavior: modulation

\[ V_S = 0V \quad V_G = 1V \quad V_D = 0.35V \]

\[ I_{DS} = \frac{1}{2} \left( \mu_n \frac{\varepsilon_{ox}}{t_{ox}} \left( \frac{W}{L} \right) \right) \left( V_{GS} - V_T \right)^2 \left( 1 + \lambda V_{DS} \right) \]

Inversion Layer does not exist here
\[ V_{DS} \geq V_{GS} - V_T \]

Assume \( V_T = 0.75V \) (threshold voltage)

True when \( V_{GS} > V_T \) & \( V_{DS} \geq V_{GS} - V_T \)
Example of Drain Current

Values for generic 0.5 µm process:

\[
k' (\text{transconductance}) = \frac{\varepsilon_{\text{ox}}}{\mu_n t_{\text{ox}}} V_T
\]

<table>
<thead>
<tr>
<th>Type</th>
<th>(k'_n)</th>
<th>(V_T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-type</td>
<td>73 µA/V^2</td>
<td>0.7V</td>
</tr>
<tr>
<td>p-type</td>
<td>21 µA/V^2</td>
<td>-0.8V</td>
</tr>
</tbody>
</table>

Assume \(W/L = 3/2\), \(V_{GS} = 2V\), find \(I_{DS}\) for NMOS device at saturation point:

\[
I_{DS} = \frac{1}{2} \left( k' \frac{W}{L} \right) (V_{GS} - V_T)^2
\]

\[
I_{DS} = \frac{1}{2} \left( 73 \frac{\mu A}{V^2} \right) \left( \frac{3}{2} \right) (2V - 0.7V)^2 = 93 \mu A
\]
NMOS I-V Plot

NMOS transistor, 0.25um, $L_d = 0.25\mu$m, $W/L = 1.5$, $V_{DD} = 2.5V$, $V_T = 0.4V$
PMOS I-V Plot

All polarities of all voltages and currents are reversed

PMOS transistor, 0.25μm, \( L_d = 0.25\mu m \), W/L = 1.5, \( V_{DD} = 2.5V \), \( V_T = -0.4V \)

Note y-axis scale (because \( W/L_p = W/L_n \))

*(drive current)*: \( I_D \) when \( V_{GS} = V_{DS} = V_{DD} \)
Review: RC Circuits

\[ v_{\text{out}}(t) = (1 - e^{-t/\tau})V \]

\[ \tau = RC \]

**RC time-constant**: dictates how rapidly the output voltage reacts to the voltage rise on input (step function).

**Larger RC, slower response**
Capacitances

Yes, there are others ...

Result: parasitic capacitances hinder switching speeds
Body Effect

- Suppose source and body are not in equilibrium: reverse bias increases size of depletion region around that diode (and changes its parasitic capacitance)
- Called “body effect” ... it changes the threshold voltage for that device

\[
\Delta V_t = \sqrt{\frac{2q\varepsilon_{si}N_A}{C_{ox}}} (\sqrt{\phi_S} + V_{SB} - \sqrt{\phi_S})
\]

But can it happen?
Body Effect

NAND gate

- If \#B propagates signal in non-zero time, the effective source voltage for \#A can go positive (higher than ground)
- Perspective: Things start to get interesting when you start connecting these things together ...
CMOS Inverter Layout I

- **N-regions** for source, drain
- **P-regions** for PMOS device
- **Gate** (poly)

**Diagram Details:**
- Input
- GND
- **NMOS**
- **PMOS**
- Output
- VDD
- N-Well
- Cut line
CMOS Inverter Layout II

Another view (note: wells/tubs not shown)
CMOS Inverter: Analysis

- Gate response time is determined by the time to charge $C_L$ through $R_p$ or discharge $C_L$ through $R_n$
CMOS Inverter: Transfer Plot

- NMOS off
- PMOS res
- NMOS sat
- PMOS res
- NMOS res
- PMOS sat
- NMOS res
- PMOS off

Graph showing the transfer characteristics of a CMOS inverter with key points indicating different stages of operation.
CMOS Inverter: Current

- NMOS off
- PMOS res
- NMOS sat
- PMOS res
- NMOS res
- PMOS sat
- NMOS res
- PMOS off

$V_{out}$ (V)

$V_{in}$ (V)
Properties of CMOS

- Full rail-to-rail swing -> high noise margins
- Logic levels not dependent upon the relative device sizes -> transistors can be minimum size -> ratioless
- Always a path to Vdd or GND in steady state -> low output impedance (output resistance in kΩ range) -> large fan-out (albeit with degraded performance)
- Extremely high input resistance (gate of MOS transistor is near perfect insulator) -> nearly zero steady-state input current
- No direct path steady-state between power and ground -> no static power dissipation
- Propagation delay function of load capacitance and resistance of transistors
Capacitive Load, etc.

**Fan-out**: number of gates connected to the output of the driving date
- Gates with large fan-out are slower

**Fan-in**: the number of inputs to the gate
- Gates with large fan-in are bigger and slower
Aside: is capacitance all bad?

Slows down output ...

Bigger capacitor, more charge to change voltage => SLOWER

... but stabilizes power supply

Bigger capacitor, more charge to change voltage => more stable power-supply voltage levels

Capacitors are *de facto* frequency filters ... can be a good thing ("bypass caps")
The electrical characteristics of transistors determine the switching speed of a circuit

- Need to select the aspect ratios \((W/L)_n\) and \((W/L)_p\) of every FET in the circuit

Define *Unit Transistor* \((R_1, C_1)\)

- \(L/W_{\text{min}}\) -> highest resistance
- \(R_2 = R_1 \div 2\) and \(C_2 = 2 \cdot C_1\)
- Separate nFET and pFET unit transistors
Long MOSFETs

- **poly**
- **metal**
- **active**
- **n-well**
- **via**

Short

Long

Really Long
Wide MOSFETs

---

**Wide MOSFETs**

- **poly**
- **metal**
- **active**
- **n-well**
- **via**

**Narrow**

- **Wide**

**Really Wide**
Transistor Sizing II

Resistance of MOSFET:

\[ R_n = \frac{1}{\mu_n C_{ox} (V_{GS} - V_{Tn})} \left( \frac{L}{W} \right) \]

- Increasing \( W \) decreases the resistance; allows more current to flow

Oxide capacitance \( C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \text{[F/cm}^2\text{]} \)

Gate capacitance \( C_G = C_{ox} WL \text{[F]} \)

Transconductance \( \beta_n = \mu_n C_{ox} \left( \frac{W}{L} \right) = k'_n \left( \frac{W}{L} \right) \)

(units [A/V^2])
Transistor Sizing II

nFET vs. pFET

\[
R_n = \frac{1}{\beta_n (V_{DD} - V_{Tn})} \quad \beta_n = \mu_n C_{ox} \left(\frac{W}{L}\right)_n
\]

\[
R_p = \frac{1}{\beta_p (V_{DD} - |V_{Tp}|)} \quad \beta_p = \mu_p C_{ox} \left(\frac{W}{L}\right)_p
\]

\[
\frac{\mu_n}{\mu_p} = r \quad \text{Typically (2 .. 3)}
\]

(\(\mu\) is the carrier mobility through device)
Inverter Switching Point

Where $V_{in} = V_{out}$

- NMOS off
- PMOS res
- NMOS sat
- PMOS res
- NMOS res
- PMOS sat
- NMOS res
- PMOS off
Inverter Switching Point

At all points $I_{DSn} = I_{DSP}$ (drain currents)

At switching point, $V_{in} = V_{out} = V_{sp}$

\[
\frac{\beta_n}{2} (V_{SP} - V_{Tn})^2 = \frac{\beta_p}{2} (V_{DD} - V_{SP} - V_{Tp})^2
\]

\[
V_{SP} = \frac{\sqrt{\frac{\beta_n}{\beta_p}} \cdot V_{Tn} + (V_{DD} - V_{Tp})}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}
\]

For $V_{sp} = V_{dd}/2$, assuming $V_{Tn} = V_{Tp}$, $\beta_n = \beta_p \Rightarrow W_p \approx 2–3W_n$

(equal drive currents, equal $R_{eff}: R_n = R_p$)
The Result ($W_p = 2W_n$, .25µm)

- **metal1**
- **metal2**
- **pdiff**
- **metal1-diff via**
- **metal1-poly via**
- **polysilicon**
- **V_{DD}**
- **PMOS** ($4/.24 = 16/1$)
- **NMOS** ($2/.24 = 8/1$)
- **ndiff**
- **metal2-metal1 via**

**GND**
The Result II \((W_p = 3W_n)\)

PMOS devices 3x larger than NMOS devices
Delay Definitions

Propagation Delay
\[ t_p = \frac{t_{PLH} + t_{PHL}}{2} \]
Inverter Switching Delay

If \((W/L)_p = r(W/L)_n\) then \(\beta_n = \beta_p\)
(and \(R_n = R_p\))

... symmetric inverter

Make pFET bigger (wider) by factor of \(r\)
Inverter Switching Delay

\[ t_{pLH} = \ln(2) R_p C_L = 0.69 R_p C_L \]

\[ t_{pHL} = \ln(2) R_n C_L = 0.69 R_n C_L \]

\[ t_p = \frac{t_{pHL} + t_{pLH}}{2} = 0.69 C_L (R_n + R_p)/2 \]

(note: the \( \ln(2)R_c \) term comes from first-order analysis of simple RC circuit’s response to step input ... time for output to reach 50% value)
Inverter Pair

PMOS
1.125/0.25

NMOS
0.375/0.25

Polysilicon

V_{DD}

In

Out

GND

Metal1
Inverter Pair

Output Waveform

2.5V

Vh
Vth

Vt
Vl

Vout

UNIVERSITY OF MARYLAND
Dynamic Power Dissipation

\[ I_{\text{avg}} = \frac{Q_{\text{Ctot}}}{T} = \frac{V_{\text{DD}} \cdot C_{\text{tot}}}{T} \]

\[ P_{\text{avg}} = V_{\text{DD}} \cdot I_{\text{avg}} = \frac{C_{\text{tot}} \cdot V_{\text{DD}}^2}{T} = C_{\text{tot}} \cdot V_{\text{DD}}^2 \cdot f_{\text{CLK}} \]