ENEE 302H, Fall 2004
Digital Electronics

P/N Junction, MOS Transistors, CMOS Inverter

Prof. Bruce Jacob
blj@eng.umd.edu

TA: Katie Baynes
ktbaynes@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).
Silicon, Specifically

14 protons in nucleus
4 valence electrons
Silicon, Specifically

Shared electrons of covalent bonds
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 w/ 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.
The P/N Junction

- **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 allow easy excitation of electrons into conduction band
The P/N Junction

Acceptor 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:

\[ C_j = \frac{C_{j0}}{V_d^m} \left[ 1 - \frac{V_d}{\phi_0} \right] \]

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)
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.

+ N-side

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**
- **VSS**

p-doped semiconductor substrate
MOS Transistors

MOS Transistor, reverse-biased:

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

VDD

VDD

VSS
MOS Transistors

NMOS Transistor with bias voltages:

p-doped semiconductor substrate

Gate (conductor)

Insulator (gate oxide)

CURRENT

+ + 0

0
MOS Transistors

NMOS Transistor, two views:

- **TOP VIEW**
  - Length
  - Width

- **SIDE VIEW**
  - Gate
  - Gate oxide
  - n
  - p-doped semiconductor substrate
MOS Transistors

NMOS Transistor with bias voltages:

- **Source**
- **Gate**
- **Drain**

- **Source** and **Drain** connected to 0 Volt.
- **Gate** connected to V > 0 Volt.

- n-channel in a p-doped semiconductor substrate.

Electron Flow.
MOS Transistors

PMOS Transistor with bias voltages:

Drain Source Gate

VDD

0

V>VDD

V< VDD

V>0

Electron Flow
MOS Transistors

MOS Transistors:

NMOS

PMOS

Source

Drain

Gate

p-doped semiconductor substrate

n-channel

VSS

Source

Drain

Gate

n-doped semiconductor substrate

p-channel

VDD

Source

Drain

Gate

Substrate

Substrate
0.25 µm transistor (Bell Labs)

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

V\text{Source}  
\text{Gate}  
V\text{Gate}  
V\text{Drain}  
n+  
n+  
p (bulk)  
Depletion Regions
MOS Behavior

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_S$</td>
<td>Source Voltage</td>
</tr>
<tr>
<td>$V_G$</td>
<td>Gate Voltage</td>
</tr>
<tr>
<td>$V_D$</td>
<td>Drain Voltage</td>
</tr>
</tbody>
</table>

Depletion Regions

Charge Density

Gate

Oxide

Substrate (p-type)

Depletion layer

$x$ (depth)
MOS Behavior

Assume $V_T = 0.75\text{V}$ (threshold voltage)

Regions
Inversion Layer forms when $V_{GS} > V_T$

Charge Density

$V_S = 0\text{V}$  $V_G = 1\text{V}$  $V_D = 0\text{V}$

Depletion layer
Inversion layer

x (depth)
MOS Behavior: linear region

V_s = 0V  \quad V_G = 1V  \quad V_D = 0.001V

\[ I_{DS} = \mu n \frac{\epsilon_{OX}}{t_{OX}} \left( \frac{W}{L} \right) (V_{GS} - V_T) V_{DS} \]

p (bulk)  

Inversion Layer 

existence requires 

\[ V_{GS} - V_T > V(y) \]

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

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

V_s = 0V  \quad V_G = 1V  \quad V_D = 0.15V

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

Assume V_T = 0.75V (threshold voltage)

Inversion Layer existence requires
V_{GS} - V_T > V(y)

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

Assume $V_T = 0.75V$ (threshold voltage)

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

**Graphs:**
- $V(y)$ vs. $V_{DS}$
- $V_{GS} - V(y)$ vs. $y$ (channel)

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

**Graphical Representation**

- **V_S** = 0V
- **V_G** = 1V
- **V_D** = 0.35V

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

- Effective channel length decreases
- Over this range, effective gate potential is not sufficient to create inversion layer

**Equation Details**

- \( I_{DS} \) represents the drain current
- \( \mu_n \) is the electron mobility
- \( \varepsilon_{ox} \) is the oxide permittivity
- \( t_{ox} \) is the oxide thickness
- \( W \) is the width
- \( L \) is the length
- \( V_T \) is the threshold voltage
- \( \lambda \) is the length modulation factor

**Inversion Layer**

- Inversion layer does not exist here

**Assumptions**

- Assume \( V_T = 0.75V \)

**Truth Conditions**

- True when \( V_{GS} > V_T \) and \( V_{DS} \geq V_{GS} - V_T \)
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>Expression</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-type</td>
<td>(k'_n = 73 , \mu A/V^2)</td>
<td>(0.7V)</td>
</tr>
<tr>
<td>p-type</td>
<td>(k'_p = 21 , \mu A/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:

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

\[
I_{DS} = \frac{1}{2} \left( 73 \frac{\mu A}{V^2} \right) \left( \frac{3}{2} \right) (2V - 0.7V)^2 = 93 \mu A
\]
Review: RC Circuits

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

\[ \tau = RC \]

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

Larger RC, slower response
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.

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

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

- Gate response time is determined by the time to charge $C_L$ through $R_p$ or discharge $C_L$ through $R_n$. 

- The diagram shows a CMOS inverter with a load capacitance $C_L$, pull-up resistor $R_p$, and pull-down resistor $R_n$. 

- When $V_{in} = 0$, the output $V_{out}$ is charged through $R_p$. 

- When $V_{in} = V_{DD}$, the output $V_{out}$ is discharged through $R_n$. 

- The gate response time is determined by the time it takes to charge or discharge $C_L$ through $R_p$ or $R_n$. 

- The equation for the gate response time can be derived from the circuit diagram, considering the time constants associated with $R_p$ and $R_n$. 

- The circuit diagram illustrates the operation of the CMOS inverter under different input conditions.
CMOS Inverter

- **NMOS off**
- **PMOS res**
- **NMOS sat**
- **PMOS sat**
- **NMOS res**
- **PMOS off**
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 \emph{de facto} frequency filters ... can be a good thing ("bypass caps")
**Delay Definitions**

**Input Waveform**
- Vin
- 50%

**Output Waveform**
- Vout
- 50%
- 10%
- 90%

**Signal slopes**
- $t_p = \frac{(t_{PHL} + t_{PLH})}{2}$

**Graph**
- Vin → Vout
- Propagation Delay
- $t_p = (t_{PHL} + t_{PLH}) / 2$
Inverter Pair (preview)

Output Waveform

- V\text{out}
- V_h
- V_\text{th}
- V_t\text{l}
- V_l
- 2.5V
CMOS Inverter Layout I

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

**Input**

**NMOS**

**Output**

**PMOS**

**GND**

**VDD**

**Cut line**
CMOS Inverter Layout II

Another view (note: wells/tubs not shown)