2. CMOS Fabrication, Layout, Design Rules

nMOS Transistor

- Four terminals: gate, source, drain, body
- Gate – oxide – body stack looks like a capacitor
  - Gate and body are conductors
  - SiO$_2$ (oxide) is a very good insulator
  - Called metal – oxide – semiconductor (MOS) capacitor
  - Even though gate is not made of metal

nMOS Operation

- Body is commonly tied to ground (0 V)
- When the gate is at a low voltage:
  - P-type body is at low voltage
  - Source-body and drain-body diodes are OFF
  - No current flows, transistor is OFF

nMOS Operation Cont.

- When the gate is at a high voltage:
  - Positive charge on gate of MOS capacitor
  - Negative charge attracted to body
  - Inverts a channel under gate to n-type
  - Now current can flow through n-type silicon from source through channel to drain, transistor is ON
pMOS Transistor

- Similar, but doping and voltages reversed
  - Body tied to high voltage ($V_{DD}$)
  - Gate low: transistor ON
  - Gate high: transistor OFF
  - Bubble indicates inverted behavior

Power Supply Voltage

- $GND = 0$ V
- In 1980’s, $V_{DD} = 5V$
- $V_{DD}$ has decreased in modern processes
  - High $V_{DD}$ would damage modern tiny transistors
  - Lower $V_{DD}$ saves power
- $V_{DD} = 3.3, 2.5, 1.8, 1.5, 1.2, 1.0, ...$
Transistors as Switches

- We can view MOS transistors as electrically controlled switches
- Voltage at gate controls path from source to drain

\[ g = 0 \quad \text{OFF} \]
\[ g = 1 \quad \text{ON} \]

\[ \begin{array}{c|c}
\text{nMOS} & \text{d} \\
g & s \\
\text{pMOS} & \text{d} \\
g & s \\
\end{array} \]

CMOS Inverter

\[ \begin{array}{c|c}
A & Y \\
0 & \\
1 & \text{V}_{DD} \\
\text{A} & \text{Y} \\
\text{GND} & \text{ON} \\
\text{V}_{DD} & \text{OFF} \\
\end{array} \]
2. CMOS Fabrication, Layout Rules

CMOS NAND Gate

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Y</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>

- A = 1
- B = 1
- Y = 0

ON OFF

CMOS NOR Gate

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Y</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>

- A
- B
- Y

3-input NAND Gate
- Y pulls low if ALL inputs are 1
- Y pulls high if ANY input is 0

3-input NAND Gate
- Y pulls low if ALL inputs are 1
- Y pulls high if ANY input is 0
Characteristics of CMOS Gates

- In general, when the circuit is stable
  - There is a path from one supply (V_{SS} or V_{DD}) to the output (low static power dissipation)
  - There is NEVER a path from one supply to another
- There is a momentary drain of current when a gate switches from one state to the other
  - Dynamic power dissipation
- If a node has no path to power or ground, the previous value retained due to the capacitance of the node

Complementary Switch (Transmission Gate)

- Combine n- and p-channel switches in parallel to get a switch which passes both “1” and “0” well

Multiplexer

- Two-input MUX using only switches

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>S</th>
<th>S</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>X</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

X: don't care

Schematic Vs. Physical Layout

- In schematic layout, lines drawn between device terminals represent connections
  - Any non-planar situation is dealt with by crossing lines
  - Provides more information than logic level (sizes of transistors, etc.)
- Physical layout captures interaction between layers
  - Includes diffusion, polysilicon, metal (many layers of metal), vias (contacts)
Stick Diagram

- Intermediate representation between the schematic level and the mask level
- Gives topological information (identifies different layers and their relationship)
  - Assumes that wires have no width

N-Channel Switch

P-Channel Switch

Basic Layers in CMOS

When two layers of the same material (i.e., on the same layer) touch or cross, they are connected and belong to the same electrical node

When Polysilicon crosses Diffusion (N or P), an N or P transistor is formed

There is no diffusion underneath the poly, but the diffusion must be drawn connecting the source and the drain

The self-aligned gate is automatically formed during fabrication

When a Metal line needs to be connected to a metal line on another layer, or to one of the other three conductors, a contact cut (via) is required

CMOS Fabrication

- CMOS transistors are fabricated on silicon wafers
- Lithography process has been the mainstream chip manufacturing process
  - Similar to printing press
  - See Chris Mack's page for a nice litho tutorial
- On each step, different materials are deposited or etched
- Easiest to understand by viewing both top and cross-section of wafer in a simplified manufacturing process

Inverter Cross-section

- Typically use p-type substrate for nMOS transistors
- Requires n-well for body of pMOS transistors
2. CMOS Fabrication, Layout Rules

Well and Substrate Taps

- Substrate must be tied to GND, n-well to $V_{DD}$
- Metal to lightly-doped semiconductor forms poor connection called Shottky Diode
- Use heavily doped well and substrate contacts / taps

Inverter Mask Set

- Transistors and wires are defined by *masks*
- Cross-section taken along dashed line

Detailed Mask Views

- Six masks
  - n-well
  - Polysilicon
  - n+ diffusion
  - p+ diffusion
  - Contact
  - Metal

Fabrication Steps

- Start with blank wafer
- Build inverter from the bottom up
- First step will be to form the n-well
  - Cover wafer with protective layer of SiO$_2$ (oxide)
  - Remove layer where n-well should be built
  - Implant or diffuse n dopants into exposed wafer
  - Strip off SiO$_2$
Oxidation

- Grow SiO\textsubscript{2} on top of Si wafer
  - 900 – 1200 C with H\textsubscript{2}O or O\textsubscript{2} in oxidation furnace

Photoresist

- Spin on photoresist
  - Photoresist is a light-sensitive organic polymer
  - Softens where exposed to light

Lithography

- Expose photoresist through n-well mask
- Strip off exposed photoresist

Etch

- Etch oxide with hydrofluoric acid (HF)
  - Seeps through skin and eats bone; nasty stuff!!!
- Only attacks oxide where resist has been exposed
Strip Photoresist

- Strip off remaining photoresist
  - Use mixture of acids called piranah etch
- Necessary so resist doesn’t melt in next step

n-Well

- n-well formed with diffusion or ion implant
- Diffusion
  - Place wafer in furnace with arsenic gas
  - Heat until As atoms diffuse into exposed Si
- Ion Implantation
  - Blast wafer with beam of As ions
  - Ions blocked by SiO$_2$, only enter exposed Si

Strip Oxide

- Strip off the remaining oxide using HF
- Back to bare wafer with n-well
- Subsequent steps involve similar series of steps

Polysilicon

- Deposit very thin layer of gate oxide
  - < 20 Å (6-7 atomic layers)
- Chemical Vapor Deposition (CVD) of Si layer
  - Place wafer in furnace with Silane gas (SiH$_4$)
  - Forms many small crystals called polysilicon
  - Heavily doped to be good conductor

Trend towards metal gates and rare earth (Hf, etc.) oxides in nanometer-scale processes
2. CMOS Fabrication, Layout Rules

Polysilicon Patterning

- Use same lithography process to pattern polysilicon

Self-Aligned Process

- Use oxide and masking to expose where n+ dopants should be diffused or implanted
- N-diffusion forms nMOS source, drain, and n-well contact

N-diffusion

- Pattern oxide and form n+ regions
- Self-aligned process - gate blocks diffusion
- Polysilicon is better than metal for self-aligned gates because it doesn't melt during later processing

N-diffusion, Cont’d

- Historically dopants were diffused
- Usually ion implantation today
- But regions are still called diffusion
2. CMOS Fabrication, Layout Rules

N-diffusion, Cont’d

- Strip off oxide to complete patterning step

P-Diffusion

- Similar set of steps form p+ diffusion regions for pMOS source and drain and substrate contact

Contacts

- Now we need to wire together the devices
- Cover chip with thick field oxide
- Etch oxide where contact cuts are needed

Metallization

- Sputter on aluminum over whole wafer
- Pattern to remove excess metal, leaving wires
Layout

- Describes actual layers and geometry on the silicon substrate to implement a function
- Need to define transistors, interconnection
  - Transistor widths (for performance)
  - Spacing, interconnect widths, to reduce defects, satisfy power requirements
  - Contacts (between poly or active and metal), and vias (between metal layers)
  - Wells and their contacts (to power or ground)
- Layout of lower-level cells constrained by higher-level requirements: “floorplanning”
  - "design iteration"

Layout, Cont’d

- Chips are specified with set of masks
- Minimum dimensions of masks determine transistor size (and hence speed, cost, and power)
- Feature size $f =$ distance between source and drain
  - Set by minimum width of polysilicon
- Feature size improves 30% every 3 years or so
- Normalize for feature size when describing design rules
- Express rules in terms of $\lambda = f/2$
  - E.g. $\lambda = 0.3 \mu m$ in 0.6 $\mu m$ process

CMOS Inverter Layout

Another CMOS Inverter Layout

Note: the N- and P- wells are not shown here
2. CMOS Fabrication, Layout Rules

**Simplified Layout of NAND Gate**

- sub Diffusion
- p-type Transistor
- n-type Transistor
- n-tub
- poly
- metal-1
- metal-1-p-diff via
- metal 1

**“Stick” Diagram for NAND Gate**

- Identifies actual layers, can be annotated with transistor sizes

**Simplified Design Rules**

- Conservative rules to get you started

**Inverter Layout**

- Transistor dimensions specified as Width / Length
  - Minimum size $4\lambda / 2\lambda$, sometimes called 1 unit
  - In $f = 0.6 \mu m$ process, this is 1.2 $\mu m$ wide, 0.6 $\mu m$ long
The MOSIS Scalable CMOS Rules

- λ-based rules
- Designs using these rules are fabricated by a variety of companies
- Multiple designs are put on a single die
  - Each chip wired to a particular design
- Support for submicron digital CMOS, analog (buried poly layer for capacitor), micromachines, etc.