CMOS Transistor Notation

- A CMOS transistor has 3 terminals, called the *gate*, *source*, and *drain*

- $V_{AB}$ is the voltage between nodes A and B in a circuit

- Positive power supply (power source)
  - In TTL, written $V_{CC}$ (usually written $V_{CC}$)
  - In NMOS and CMOS, written $V_{DD}$ (also $V_{DD}$)

- Negative power supply (power sink)
  - In TTL, written GND (“ground”)
  - In NMOS and CMOS, sometimes written $V_{SS}$ (also $V_{SS}$)

- CMOS uses positive logic: $V_{DD}$ is logic “1”, $V_{SS}$ is logic “0”
CMOS Transistors as Switches

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997
CMOS NAND and NOR Gates

- Use two transistors to make a CMOS inverter (as shown on previous slide)

- Use four transistors to make a CMOS 2-input NAND gate
  - Rule of thumb: 1 gate = 4 transistors

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997
IC Fabrication Technologies
(Implementing an Inverter)

Figure from *Modern VLSI Design*, Wolf, Prentice Hall, 1994

- **Bipolar** — transistor & resistor (**fastest**)
- **NMOS** — n-channel depletion mode transistor (top) & n-channel enhancement mode transistor (bottom)
- **CMOS** — p-channel (**lowest power**) enhancement mode transistor (top) & n-channel enhancement mode transistor (bottom)
Cross-Section of an N-Channel Enhancement Mode MOS Transistor

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997

- Base is silicon *substrate* (*bulk*, *well*, *tub*) that’s been *doped* with p-type impurities (full of positively-charged *holes*)
  - Two *diffusion* areas heavily doped with n-type impurities (full of negatively-charged *electrons*) form the *source* and *drain*
  - Transistor action takes place at the *channel*, connecting the source and drain

- A very thin layer of silicon dioxide (*SiO_2*), called the *gate oxide*, insulates the *gate*, made of polysilicon, from the channel
Operation of an N-Channel Enhancement Mode MOS Transistor

- Works as a switch — gate-to-source voltage regulates the amount of current that can flow between drain and source

- When \( V_{GS} = 0 \), the p-type channel is full of holes, and the n-type source and drain contain electrons
  - The p-n junctions at source and drain form diodes in opposite directions, so no current flows between source and drain

- As \( V_{GS} \) rises above 0, the few n-type impurities that are present in the p-type channel start to attract electrons
  - The electrons migrate toward the (positively charged) gate, but are stopped by the gate oxide, so collect at the top of the channel

- When \( V_{GS} \) rises to the threshold voltage (\( V_t \)), enough electrons have collected to form an n-channel inversion layer, which allows electrons to flow from source to drain (current flows from drain to source)
Operation of an N-Channel Enhancement Mode MOS Transistor (cont.)

- N-channel vs. p-channel
  - N-channel: $V_{GS}$ and $V_{DS}$ both positive, gate and source are n-type (electrons), substrate & channel is p-type (holes), when $V_{GS} >> V_t$ electrons accumulate in channel and flow from source to drain, current flows from drain to source
  - P-channel: $V_{GS}$ and $V_{DS}$ both negative, gate and source are p-type (holes), substrate & channel is n-type (electrons), when $V_{GS} >> V_t$ holes accumulate in channel and flow from source to drain, current flows from source to drain

- Current proportional to W/L of transistor
  - Length (L) = parallel to current flow
  - As W increases, more current can flow
  - As L increases, less current flows
1. Start with silicon (Si), refined from quartzite, dope it with p- or n-type impurities, and melt it at 1500°C

2. & 3. Draw out a single crystal (6" or 8"), saw it into thin (600 µm) wafers, polish one side, and grind down an edge or two

4. Batch of wafers (a wafer lot) is placed on a boat and put in a furnace to grow a layer (1000 Å) of silicon dioxide (SiO₂) (called the oxide)
IC Fabrication
(cont.)

- IC fabrication process uses a series of masking steps to create the layers that form the transistors etc. on the chip

5. A thin layer (1 μm) of liquid photoresist (*resist*) is spun onto each wafer, and it is baked at 100°C to harden the resist

6. The wafer is partially exposed to ultraviolet light through a *mask*, which polymerizes the exposed areas; the polymerized part is then removed using an organic solvent

7. The exposed oxide is *etched* away, making the oxide match the mask

8. The exposed silicon substrate is doped with appropriate ions by an *ion implanter*

9. & 10. Resist and oxide are removed
Fabricating a CMOS Transistor

Continuing the fabrication process:

- Polycrystalline silicon (poly) is deposited using chemical vapor deposition (CVD) to deposit dopants using a gaseous source in a furnace.
  - Poly wires (e.g., transistor gates) are deposited before diffusion to create self-aligned transistors — this avoids small gaps that might otherwise occur if the order is reversed.

- Metal layers are deposited in a similar manner, called sputtering.

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997.
Fabricating a CMOS Transistor (cont.)

Figure from Principles of CMOS VLSI Design, Weste, Addison-Wesley, 1993
Wires and Vias

- “Wires” can be fabricated using diffusion, polysilicon, or metal
  - Must be insulated from each other using silicon dioxide; can be built up as layers
  - Diffusion used only for within a cell
  - Poly used between adjacent cells
  - Metal used for longer connections
    - Chip may have 5-6 layers of metal
  - Cuts in the silicon dioxide between layers are called *vias*

- Metal layer 1 is used for VDD and VSS; other layers are used for interconnection
  - Under high currents, electron collisions with metal grains can cause the metal to move (*metal migration*), so in large designs, sizing the power supply lines is critical to keep chip from failing
CMOS Inverter in a n-well Process

Figure from *Principles of CMOS VLSI Design*, Weste, Addison-Wesley, 1993
An n-channel (p-channel) transistor’s substrate must be tied to VSS (VDD)

- The special vias that provide these connections are called tub ties
  - Need one every 1-2 transistors (SCMOS)
- Tie area in substrate is heavily doped to provide a low-resistance connection
Design Rules & Fabrication Errors

- Common fabrication errors:
  - Wire too wide — may *short* (contact) an adjacent wire
  - Wire too narrow — may break under load, and become *open*

- Solution — impose *design rules* to specify what’s legal and illegal
  - Wires — specify minimum width and minimum spacing between wires
  - Poly — must extend beyond channel to ensure that there is no short between source and drain
  - Diffusion — must extend enough to have room for a contact to that region
  - Via — must be smaller than what it’s contacting, what it’s contacting must extend back under SiO₂
Fabrication processes are constantly being improved

- (Gordon) Moore’s Law (version 2) says that the number of transistors on a chip is doubling every 18 months

We take advantage of these improvements by designing according to *scaleable design rules*

- Specified in terms of $\lambda$, the minimum feature size possible in that process
- In MOSIS SCMOS rules, minimum channel width (poly) is $2\lambda$, and minimum wire width is $2\lambda$

MOSIS = MOS Implementation Service, located at the Information Sciences Institute at the University of Southern California (USC), does small-volume fabrication for universities (partially NSF-funded) and commercially (www.isi.edu)
SCMOS Design Rules (1997) (cont.)

- MOSIS (rev. 7), dimensions in $\lambda$

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997
SCMOS Design Rules (1997)
(cont.)

**MOSIS (rev. 7), dimensions in $\lambda$**

<table>
<thead>
<tr>
<th>Layer</th>
<th>Rule</th>
<th>Explanation</th>
<th>Value / $\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>well (CWN, CWP)</td>
<td>1.1</td>
<td>minimum width</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>minimum space (different potential, a hot well)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>minimum space (same potential)</td>
<td>0 or 6</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>minimum space (different well type)</td>
<td>0</td>
</tr>
<tr>
<td>active (CAA)</td>
<td>2.1/2.2</td>
<td>minimum width/space</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>source/drain active to well edge space</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>substrate/well contact active to well edge space</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>minimum space between active (different implant type)</td>
<td>0 or 4</td>
</tr>
<tr>
<td>poly (CPG)</td>
<td>3.1/3.2</td>
<td>minimum width/space</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3.3</td>
<td>minimum gate extension of active</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3.4</td>
<td>minimum active extension of poly</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>minimum field poly to active space</td>
<td>1</td>
</tr>
<tr>
<td>select (CSN, CSP)</td>
<td>4.1</td>
<td>minimum select spacing to channel of transistor $1$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4.2</td>
<td>minimum select overlap of active</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>minimum select overlap of contact</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>minimum select width and spacing $2$</td>
<td>2</td>
</tr>
<tr>
<td>poly contact (CCP)</td>
<td>5.1.a</td>
<td>exact contact size</td>
<td>$2 \times 2$</td>
</tr>
<tr>
<td></td>
<td>5.2.a</td>
<td>minimum poly overlap</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>5.3.a</td>
<td>minimum contact spacing</td>
<td>2</td>
</tr>
<tr>
<td>active contact (CCA)</td>
<td>6.1.a</td>
<td>exact contact size</td>
<td>$2 \times 2$</td>
</tr>
<tr>
<td></td>
<td>6.2.a</td>
<td>minimum active overlap</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>6.3.a</td>
<td>minimum contact spacing</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6.4.a</td>
<td>minimum space to gate of transistor</td>
<td>2</td>
</tr>
</tbody>
</table>

Table from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997
Economics of ASICS

- For a given design, which type of ASIC is the most cost-effective?
  - (full-custom) ASIC?
  - MGA (mask-programmable gate array)?
  - CBIC (cell-based integrated circuit = standard-cell-based ASIC)?

- Answer: consider the ASIC as a product, and examine the fixed costs and variable costs
  - total product cost = fixed product cost + variable product cost
  - Fixed product cost is independent of sales volume
    - Fixed product costs amortized per product sold decrease as sales volume increases
  - Variable product cost includes assembly costs and manufacturing costs
Example of ASIC Economics  
(Warning — 1997 Numbers!)

- Sample costs:
  - CBIC: fixed cost $146,000; part cost $8
  - MGA: fixed cost $86,000; part cost $10
  - FPGA: fixed cost $21,800; part cost $39

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997

- Break-even points:
  - FPGA to MGA is around 2,000 parts
  - FPGA to CBIC is around 4,000 parts
  - MGA to CBIC is around 20,000 parts
ASIC Fixed Costs (1997 Numbers!)

- Design: estimate of designer productivity
- Production test: make sure the IC works
- Non-recurring engineering (NRE): work done by ASIC vendor — developing mask, production tests, prototypes, etc.

Figure from *Application-Specific Integrated Circuits*, Smith, Addison-Wesley, 1997
ASIC Variable Costs (1997 Numbers!)

- Wafer size: 6” & 8” common, 12” soon
- 10k gates = small design, 100k = large
- Gate utilization: space used for gates, not used for interconnect
- Defect density is measure of fabrication quality (defect on a die is usually fatal)
- Yield is percentage of usable dies

<table>
<thead>
<tr>
<th></th>
<th>FPGA</th>
<th>MGA</th>
<th>CBIC</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer size</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6 inches</td>
</tr>
<tr>
<td>Wafer cost</td>
<td>1,400</td>
<td>1,300</td>
<td>1,500</td>
<td>$</td>
</tr>
<tr>
<td>Design</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>gates</td>
</tr>
<tr>
<td>Density</td>
<td>10,000</td>
<td>20,000</td>
<td>25,000</td>
<td>gates/sq.cm</td>
</tr>
<tr>
<td>Utilization</td>
<td>80</td>
<td>85</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Die size</td>
<td>1.67</td>
<td>0.59</td>
<td>0.40</td>
<td>sq.cm</td>
</tr>
<tr>
<td>Die/Wafer</td>
<td>88</td>
<td>248</td>
<td>355</td>
<td></td>
</tr>
<tr>
<td>Defect density</td>
<td>1.10</td>
<td>0.90</td>
<td>1.00</td>
<td>defects/sq.cm</td>
</tr>
<tr>
<td>Yield</td>
<td>65</td>
<td>72</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>Die cost</td>
<td>25</td>
<td>7</td>
<td>5</td>
<td>$</td>
</tr>
<tr>
<td>Profit margin</td>
<td>80</td>
<td>45</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>Price/gate</td>
<td>0.39</td>
<td>0.10</td>
<td>0.08</td>
<td>cents</td>
</tr>
</tbody>
</table>

Part cost $39 $10 $8

Figure from Application-Specific Integrated Circuits, Smith, Addison-Wesley, 1997