Goals of Technology Scaling

- Make things cheaper:
  - Want to sell more functions (transistors) per chip for the same money
  - Or build same products cheaper
  - Price of a transistor has to be reduced
- But also want to be faster, smaller, lower power…

Technology Scaling

  - Double transistor density
  - Reduce gate delay by 30% (increase operating frequency by 43%)
  - Reduce energy per transition by 65% (50% power savings @ 43% increase in frequency)
- Die size used to increase by 14% per generation (not any more)
- Technology generation spans 2-3 years

Technology Scaling (1)

CMOS Transistor Scaling

Announcements

- Homework #7 due today
  - Project #1 out today, due next Thurs.
- Midterm 2: Thurs. Nov. 6th, 6:30-8:00pm, room TBA
  - Review session likely on Tues. Nov. 4th
- Elad out of town Thurs. Oct. 30th
  - No lecture – instead will hold lecture on Thurs. Nov. 6th during normal time
Technology Scaling (2)

Number of components per chip

Technology Scaling (3)

Propagation Delay

Technology Scaling (4)

Full Scaling (Dennard, Long-Channel)

- W, L, tox: 1/S
- VDD, VT: 1/S
- Area: WL
- Cox: 1/tox
- CL: CoxWL
- I0: Cox(W/L)(VDD-VT)²
- R everlasting: VDD/IDSAT

Technology Scaling Models

- Full Scaling (Constant Electrical Field)
  ideal model — dimensions and voltage scale together by the same factor S

- Fixed Voltage Scaling
  most common model until 1990’s
  only dimensions scale, voltages remain constant

- General Scaling
  most realistic for today’s situation —
  voltages and dimensions scale with different factors

Scaling
**Full Scaling (Dennard, Long-Channel)**

- $W, L, t_{ox}: 1/S$
- $V_{DD}, V_T: 1/S$
- $t_p: R_{eq}C_L$
- $P_{avg}: C_L V_{DD}^2/\eta_p$
- $P_{avg}/A: C_{ox} V_{DD}^2/\eta_p$

**Full Scaling (Dennard, Short-Channel)**

- $W, L, t_{ox}: 1/S$
- $V_{DD}, V_T: 1/S$
- $t_p: R_{eq}C_L$
- $P_{avg}: C_L V_{DD}^2/\eta_p$
- $P_{avg}/A: C_{ox} V_{DD}^2/\eta_p$

**Scaling Relationships for Long Channel Devices**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Relation</th>
<th>Full Scaling</th>
<th>General Scaling</th>
<th>Fixed Voltage Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W, L, t_{ox}$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td></td>
</tr>
<tr>
<td>$V_{DD}, V_T$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td></td>
</tr>
<tr>
<td>$N_{eq}$</td>
<td>$V_{DD}^{m/p}$</td>
<td>$S$</td>
<td>$S^2$</td>
<td>$S^2$</td>
</tr>
<tr>
<td>Area/Device</td>
<td>$W/L$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
<tr>
<td>$C_{eq}$</td>
<td>$1/t_{ox}$</td>
<td>$S$</td>
<td>$S$</td>
<td>$S$</td>
</tr>
<tr>
<td>$C_L$</td>
<td>$C_{eq}WL$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
</tr>
<tr>
<td>$I_{D}$</td>
<td>$I_{D} = \frac{C_{eq}V_{DD}^2}{2}(W/L)&lt;/br&gt;$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
<tr>
<td>$P_{avg}$</td>
<td>$C_{eq}V_{DD}^2/\eta_p$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
<tr>
<td>$P_{avg}/A$</td>
<td>$C_{ox}V_{DD}^2/\eta_p$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
</tbody>
</table>

**Transistor Scaling (Velocity-Saturated Devices)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Relation</th>
<th>Full Scaling</th>
<th>General Scaling</th>
<th>Fixed Voltage Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W, L$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
</tr>
<tr>
<td>$I_{D}$</td>
<td>$I_{D} = \frac{C_{eq}V_{DD}^2}{2}(W/L)&lt;/br&gt;$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
<tr>
<td>$C_{eq}$</td>
<td>$C_{eq}WL$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
</tr>
<tr>
<td>$C_L$</td>
<td>$C_{eq}WL$</td>
<td>$1/S$</td>
<td>$1/S$</td>
<td>$1/S$</td>
</tr>
<tr>
<td>$I_{D_{sat}}$</td>
<td>$I_{D_{sat}} = \frac{V_{DD}^2}{2}(W/L)&lt;/br&gt;$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
<tr>
<td>$P_{avg}$</td>
<td>$C_{eq}V_{DD}^2/\eta_p$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
<tr>
<td>$P_{avg}/A$</td>
<td>$C_{ox}V_{DD}^2/\eta_p$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
<td>$1/S^2$</td>
</tr>
</tbody>
</table>

**Interconnect Scaling**

- $W, L, t_{ox}: 1/S$
- $V_{DD}, V_T: 1/S$
**Resistance Scaling (local)**

Scale W, H, and L:

\[ R_w = \frac{\rho}{L(WH)} \]

\[ R_w \propto \frac{1}{S^2} \]

\[ (R \propto S) \]

**Resistance Scaling (global)**

Scale W, H, constant L:

\[ R_w = \frac{\rho}{L(WH)} \]

\[ R_w \propto \frac{1}{S^2} \]

**Scenario 2: Intralayer Capacitance**

- \( C_{pp,side} \propto \frac{LT}{D} \) \( \propto \) const.
- Aspect ratio limited – eventually have to scale T
- Different metal layers have different T

**Wire Scaling (Scenario 2)**

- \( C_{pp}' \propto \frac{1}{S} \)
- \( C_{fringe}' \propto \) const.

**Wire Scaling (Scenario 1)**

- \( C_{pp} \propto \frac{WL}{H} \)
- \( C_{fringe} \propto L \)
- \( R_w \propto \frac{L(WT)}{R_w C_w} \)

Bad news: gates speed up by S…

**Wire Scaling (Scenario 2, Global)**

- \( C_{pp}' \propto \frac{1}{S} \)
- \( C_{fringe}' \propto \) const.

Very bad: wire delay \( S^2 \) worse than gates
Modern Interconnect

- 90nm process

Next Lecture

- Pass Transistor Logic