This week
• Lab 4 this week
• Homework #4 due Thursday 9/29, 5pm

Next week
• No new homework due next week
• No labs next week
• Midterm 1 is next Thursday, Oct 6, 6:30-8pm (TBD)

Hw-2 stats
• 94% had > 80% (32/40)
• 85% had > 90% (36/40)
Midterm #1

- Thursday, October 6, 6:30-8:00pm, (TBD)
- Material up to (including) Lecture-9 (Scaling)
  - Ch-1, Ch-2, Ch-3, Ch-5
- Open book, open notes
- Past Midterms on the web
- Review: Tuesday, Oct 4, 6:30-8:30pm (TBD)

Interesting Presentations...

- Wednesday, Sep-28, 4pm, 306 Soda Hall
  “Multi-Core Systems”
  by Michael Rosenfield
  Director, VLSI Systems, IBM Research
- IC Seminar, Mondays, 4pm, 521 Cory
Last Lecture

- Last lecture
  - Buffer sizing
  - Power dissipation

- Today’s lecture
  - CMOS scaling

Impact of Technology Scaling
Goals of Technology Scaling

- **Make things cheaper:**
  - Want to sell more functions (transistors) per chip for the same money
  - Build same products cheaper, sell the same part for less money
  - Price of a transistor has to be reduced

- But also want to be faster, smaller, lower power

Technology Scaling

- Technology generation spans 2-3 years

- Benefits of scaling the dimensions by 30%:
  - Reduce gate delay by 30% (increase operating frequency by 43%)
  - Double transistor density
  - Reduce energy per transition by 65% (50% power savings @ 43% increase in frequency)

- Die size used to increase by 14% per generation
### Technology Generations

#### Table 2. Time overlap of semiconductor technology generations.

<table>
<thead>
<tr>
<th>Year</th>
<th>2001</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2010</th>
<th>2013</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM ½ pitch [nm]</td>
<td>130</td>
<td>100</td>
<td>80</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>MPU transistors/chip</td>
<td>97M</td>
<td>153M</td>
<td>243M</td>
<td>386M</td>
<td>773M</td>
<td>1.55G</td>
<td>3.09G</td>
</tr>
<tr>
<td>Wiring levels</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>High-perf. phys. gate [nm]</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>25</td>
<td>18</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>High-perf. VDD [V]</td>
<td>1.2</td>
<td>1.0</td>
<td>0.9</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Local clock [GHz]</td>
<td>1.7</td>
<td>3.1</td>
<td>5.2</td>
<td>6.7</td>
<td>11.5</td>
<td>19.3</td>
<td>28.8</td>
</tr>
<tr>
<td>High-perf. power [W]</td>
<td>130</td>
<td>150</td>
<td>170</td>
<td>190</td>
<td>218</td>
<td>251</td>
<td>288</td>
</tr>
<tr>
<td>Low-power phys. gate [nm]</td>
<td>90</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>22</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Low-power VDD [V]</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>Low-power power [W]</td>
<td>2.4</td>
<td>2.8</td>
<td>3.2</td>
<td>3.5</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

ITRS Technology Roadmap Acceleration Continues

Technology Scaling (1)

Minimum Feature Size
Technology Scaling (2)

Number of components per chip

Technology Scaling (3)

$\tau_p$ decreases by 30%/year
$R$ increases by 43%

Propagation Delay
Technology Scaling (4)

ISSCC data

(a) Power dissipation vs. year.

(b) Power density vs. scaling factor.

From Kuroda

Technology Scaling Models

- **Full Scaling (Constant Electric Field)**
  - Ideal model
  - Dimensions and voltages scale by the same factor $S$

- **Fixed Voltage Scaling**
  - Most common model until recently
  - Only dimensions scale, voltages remain constant

- **General Scaling**
  - Most realistic for today situation
  - Voltages and dimensions scale with different factors
### Scaling (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/U</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>N SUB</td>
<td>V/W dep 2</td>
<td>S</td>
<td>S^2/U</td>
<td>S^2</td>
</tr>
<tr>
<td>Area/Device</td>
<td>WL</td>
<td>1/S^2</td>
<td>1/S^2</td>
<td>1/S^2</td>
</tr>
<tr>
<td>C ox</td>
<td>1/t ox</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>C L</td>
<td>C ox * WL</td>
<td>1/S</td>
<td>1/S</td>
<td>1/S</td>
</tr>
<tr>
<td>k_m, k p</td>
<td>C ox * WL</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>I av</td>
<td>k_n, p * V^2</td>
<td>1/S</td>
<td>S/U^2</td>
<td>S</td>
</tr>
<tr>
<td>t_p (intrinsic)</td>
<td>C L * V / I av</td>
<td>1/S</td>
<td>U/S^2</td>
<td>1/S^2</td>
</tr>
<tr>
<td>P av</td>
<td>C_L * V^2 / t_p</td>
<td>1/S^2</td>
<td>S/U^3</td>
<td>S</td>
</tr>
<tr>
<td>PDP</td>
<td>C_L * V^2</td>
<td>1/S^3</td>
<td>1/SU^2</td>
<td>1/S</td>
</tr>
</tbody>
</table>

Table 3.1: Scaling Relationships for Long Channel Devices

### Scaling (Short Channel Devices)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Relation</th>
<th>Full Scaling</th>
<th>General Sc.</th>
<th>Fixed V Sc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>W, L, t ox</td>
<td>1/S</td>
<td>1/S</td>
<td>1/S</td>
<td>1/S</td>
</tr>
<tr>
<td>V DD, V T</td>
<td>1/S</td>
<td>1/U</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Area/Device</td>
<td>WL</td>
<td>1/S^2</td>
<td>1/S^2</td>
<td>1/S^2</td>
</tr>
<tr>
<td>C ox</td>
<td>1/t ox</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>C gate</td>
<td>C ox * WL</td>
<td>1/S</td>
<td>1/S</td>
<td>1/S</td>
</tr>
<tr>
<td>k_m, k p</td>
<td>C ox * W/L</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>I sat</td>
<td>C ox * W</td>
<td>1/S</td>
<td>1/U</td>
<td>1</td>
</tr>
<tr>
<td>Cnt Density</td>
<td>I sat / Area</td>
<td>S</td>
<td>S^2/U</td>
<td>S^2</td>
</tr>
<tr>
<td>R on</td>
<td>V / I sat</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Intr. Delay</td>
<td>R on * C gate</td>
<td>1/S</td>
<td>1/S</td>
<td>1/S</td>
</tr>
<tr>
<td>Power</td>
<td>I sat * V</td>
<td>1/S^2</td>
<td>1/U^2</td>
<td>1</td>
</tr>
<tr>
<td>P Density</td>
<td>Power/Area</td>
<td>1</td>
<td>S^2/U^2</td>
<td>S^2</td>
</tr>
</tbody>
</table>

EE141
**μProcessor Scaling**

2X growth in 1.96 years!

S. Borkar, IEEE Micro 1999,

**μProcessor Power**

S. Borkar, IEEE Micro 1999,
μProcessor Performance

![Graph showing performance comparison between different processors](image)

P. Gelsinger, μProcessors for the New Millennium, ISSCC 2001

---

2010 Outlook

- **Performance**
  - 1 TIP (tera instructions/s)
  - 30 GHz clock

- **Size**
  - No of transistors: 2 Billion
  - Die: 40*40 mm

- **Power**
  - 10kW!!
  - Leakage: 1/3 of total Power

P. Gelsinger, μProcessors for the New Millennium, ISSCC 2001
Some Interesting Questions

- What will cause this model to break?
- When will it break?
- Will the model gradually slow down?
  - Power and power density
  - Leakage
  - Process Variation

Wires
The Wire

schematics  physical

transmitters  receivers

Interconnect Impact on Chip
Wire Models

All-inclusive model  Capacitance-only

Impact of Interconnect Parasitics

- Interconnect parasitics
  - reduce reliability
  - affect performance and power consumption

- Classes of parasitics
  - Capacitive
  - Resistive
  - Inductive
Nature of Interconnect

Local Interconnect

Global Interconnect

$S_{Local} = S_{Technology}$

$S_{Global} = S_{Die}$

Source: Intel

Interconnect Capacitance
Capacitance of Wire Interconnect

Capacitance: The Parallel Plate Model

\[ c_{int} = \frac{\varepsilon_{di} WL}{t_{di}} \]

\[ S_{wire} = \frac{S}{S \cdot S_L} = \frac{1}{S_L} \]
Permittivity

<table>
<thead>
<tr>
<th>Material</th>
<th>$\varepsilon_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free space</td>
<td>1</td>
</tr>
<tr>
<td>Aerogels</td>
<td>~1.5</td>
</tr>
<tr>
<td>Polyimides (organic)</td>
<td>3-4</td>
</tr>
<tr>
<td>Silicon dioxide</td>
<td>3.9</td>
</tr>
<tr>
<td>Glass-epoxy (PC board)</td>
<td>5</td>
</tr>
<tr>
<td>Silicon Nitride ($\text{Si}_3\text{N}_4$)</td>
<td>7.5</td>
</tr>
<tr>
<td>Alumina (package)</td>
<td>9.5</td>
</tr>
<tr>
<td>Silicon</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Fringing Capacitance

$$c_{\text{wire}} = c_{pp} + c_{fringe} = \frac{w e_{di}}{t_{di}} + \frac{2\pi e_{di}}{\log(t_{di}/H)}$$

(a)

(b)
Fringing vs. Parallel Plate

Interwire Capacitance
Impact of Interwire Capacitance

Wiring Capacitances (0.25µm)

<table>
<thead>
<tr>
<th></th>
<th>Field</th>
<th>Active</th>
<th>Poly</th>
<th>AI1</th>
<th>AI2</th>
<th>AI3</th>
<th>AI4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI1</td>
<td>30</td>
<td>41</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>47</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI2</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>27</td>
<td>29</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI3</td>
<td>8.9</td>
<td>9.4</td>
<td>10</td>
<td>15</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>27</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI4</td>
<td>6.5</td>
<td>6.8</td>
<td>7</td>
<td>8.9</td>
<td>15</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>27</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>AI5</td>
<td>5.2</td>
<td>5.4</td>
<td>5.4</td>
<td>6.6</td>
<td>9.1</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>19</td>
<td>27</td>
<td>32</td>
</tr>
</tbody>
</table>
Next Lecture

- Wires
  - Resistance
  - Capacitance
  - Inductance