CMSC 611

Evaluating Cost
Integrated Circuits: Fueling Innovation

• Chips begins with silicon, found in sand
• Silicon does not conduct electricity well and thus called semiconductor
• A special chemical process can transform tiny areas of silicon to either:
  – Excellent conductors of electricity (like copper)
  – Excellent insulator from electricity (like glass)
  – Areas that can conduct or insulate (a switch)
• A transistor is simply an on/off switch controlled by electricity
• Integrated circuits combines dozens of hundreds of transistors in a chip
**Integrated Circuits: Fueling Innovation**

- Technology innovations over time

<table>
<thead>
<tr>
<th>Year</th>
<th>Technology used in computers</th>
<th>Relative performance/unit cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>Vacuum tube</td>
<td>1</td>
</tr>
<tr>
<td>1965</td>
<td>Transistor</td>
<td>35</td>
</tr>
<tr>
<td>1975</td>
<td>Integrated circuits</td>
<td>900</td>
</tr>
<tr>
<td>1995</td>
<td>Very large-scale integrated circuit</td>
<td>2,400,000</td>
</tr>
</tbody>
</table>

Advances of the IC technology affect H/W and S/W design philosophy.
• Silicon ingots:
  – 6-12 inches in diameter and about 12-24 inches long
• Impurities in the wafer can lead to defective devices and reduces the yield
**Integrated Circuits Costs**

\[
\text{Dies\_per\_Wafer} = \frac{\pi \times (\text{Wafer\_diameter}/2)^2}{\text{Die\_Area}} - \frac{\pi \times \text{Wafer\_Diameter}}{\sqrt{2} \times \text{Die\_Area}}
\]

\[
\text{Die\_Yield} = \text{Wafer\_Yield} \times \left[ 1 + \frac{\text{Defects\_per\_Unit\_Area} \times \text{Die\_Area}}{\alpha} \right]^{-\alpha}
\]

\[
\text{Die\_Cost} = \frac{\text{Wafer\_Cost}}{\text{Dies\_per\_Wafer} \times \text{Die\_Yield}}
\]

\[
\text{IC\_Cost} = \frac{\text{Die\_Cost} + \text{Testing\_Cost} + \text{Packing\_Cost}}{\text{Final\_Test\_Yield}}
\]

*Die cost roughly goes with die area$^4$*
What Affects Cost?

1. Learning curve:
   - The more experience in manufacturing a component, the better the yield
   - In general, a chip, board or system with twice the yield will have half the cost.
   - The learning curve is different for different components, complicating design decisions

2. Volume
   - Larger volume increases rate of learning curve
   - Doubling the volume typically reduce cost by 10%

3. Commodities
   - Are essentially identical products sold by multiple vendors in large volumes
   - Foil the competition and drive the efficiency higher and thus the cost down
## Real World Examples

<table>
<thead>
<tr>
<th>Chip</th>
<th>Layers</th>
<th>Wafer cost</th>
<th>Defect/cm²</th>
<th>Area (mm²)</th>
<th>Dies/Wafer</th>
<th>Yield</th>
<th>Die Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>386DX</td>
<td>2</td>
<td>$900</td>
<td>1.0</td>
<td>43</td>
<td>360</td>
<td>71%</td>
<td>$4</td>
</tr>
<tr>
<td>486DX2</td>
<td>3</td>
<td>$1200</td>
<td>1.0</td>
<td>81</td>
<td>181</td>
<td>54%</td>
<td>$12</td>
</tr>
<tr>
<td>PowerPC 601</td>
<td>4</td>
<td>$1700</td>
<td>1.3</td>
<td>121</td>
<td>115</td>
<td>28%</td>
<td>$53</td>
</tr>
<tr>
<td>HP PA 7100</td>
<td>3</td>
<td>$1300</td>
<td>1.0</td>
<td>196</td>
<td>66</td>
<td>27%</td>
<td>$73</td>
</tr>
<tr>
<td>DEC Alpha</td>
<td>3</td>
<td>$1500</td>
<td>1.2</td>
<td>234</td>
<td>53</td>
<td>19%</td>
<td>$149</td>
</tr>
<tr>
<td>SuperSPARC</td>
<td>3</td>
<td>$1700</td>
<td>1.6</td>
<td>256</td>
<td>48</td>
<td>13%</td>
<td>$272</td>
</tr>
<tr>
<td>Pentium</td>
<td>3</td>
<td>$1500</td>
<td>1.5</td>
<td>296</td>
<td>40</td>
<td>9%</td>
<td>$417</td>
</tr>
</tbody>
</table>

- **Component Costs**: raw material cost for the system’s building blocks
- **Direct Costs** (add 25% to 40%) recurring costs: labor, purchasing, scrap, warranty
- **Gross Margin** (add 82% to 186%) nonrecurring costs: R&D, marketing, sales, equipment maintenance, rental, financing cost, pretax profits, taxes
- **Average Discount** to get List Price (add 33% to 66%): volume discounts and/or retailer markup
### Chip Prices (August 1993) for a volume of 10,000 units

<table>
<thead>
<tr>
<th>Chip</th>
<th>Area (mm²)</th>
<th>Total Cost</th>
<th>Price</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>386DX</td>
<td>43</td>
<td>$9</td>
<td>$31</td>
<td></td>
</tr>
<tr>
<td>486DX2</td>
<td>81</td>
<td>$35</td>
<td>$245</td>
<td>No Competition</td>
</tr>
<tr>
<td>PowerPC 601</td>
<td>121</td>
<td>$77</td>
<td>$280</td>
<td></td>
</tr>
<tr>
<td>DEC Alpha</td>
<td>234</td>
<td>$202</td>
<td>$1231</td>
<td>Recoup R&amp;D?</td>
</tr>
<tr>
<td>Pentium</td>
<td>296</td>
<td>$473</td>
<td>$965</td>
<td></td>
</tr>
</tbody>
</table>
Defining Performance

- Performance means different things to different people, therefore its assessment is subtle.

Analogy from the airlines industry:

- How to measure performance for an airplane?
  - Cruising speed (How fast it gets to the destination)
  - Flight range (How far it can reach)
  - Passenger capacity (How many passengers it can carry)

<table>
<thead>
<tr>
<th>Airplane</th>
<th>Passenger capacity</th>
<th>Cruising range (miles)</th>
<th>Cruising speed (m.p.h)</th>
<th>Passenger throughput (Passenger x m.p.h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boeing 777</td>
<td>375</td>
<td>4630</td>
<td>610</td>
<td>228,750</td>
</tr>
<tr>
<td>Boeing 747</td>
<td>470</td>
<td>4150</td>
<td>610</td>
<td>286,700</td>
</tr>
<tr>
<td>BAC/Sud Concorde</td>
<td>132</td>
<td>4000</td>
<td>1350</td>
<td>178,200</td>
</tr>
<tr>
<td>Douglas DC-8-50</td>
<td>146</td>
<td>8720</td>
<td>544</td>
<td>79,424</td>
</tr>
</tbody>
</table>

Criteria of performance evaluation differs among users and designers.
Performance Metrics

• Response (execution) time:
  – The time between the start and the completion of a task
  – Measures user perception of the system speed
  – Common in reactive and time critical systems, single-user computer, etc.

• Throughput:
  – The total number of tasks done in a given time
  – Most relevant to batch processing (billing, credit card processing)
  – Mainly used for input/output systems (disk access, printer, etc.)
Maximizing performance means minimizing response (execution) time.

\[
\text{Performance} = \frac{1}{\text{Execution time}}
\]
Response-time Metric

\[
\text{Performance} = \frac{1}{\text{Execution time}}
\]

- Performance of Processor P1 is better than P2 if
  - For a given work load L
  - P1 takes less time to execute L than P2

\( \text{Performance (P}_1\text{)} > \text{Performance (P}_2\text{)} \) w.r.t L

\( \Rightarrow \) Execution time \( (P_1, L) < \) Execution time \( (P_2, L) \)
Response-time Metric

\[
\text{Performance} = \frac{1}{\text{Execution time}}
\]

- Relative performance captures the performance ratio
  - For the same work load

\[
\frac{\text{CPU Performance (P}_2\text{)}}{\text{CPU Performance (P}_1\text{)}} = \frac{\text{Total execution time (P}_1\text{)}}{\text{Total execution time (P}_2\text{)}}
\]