L16: Power Dissipation in Digital Systems
Problem #1: Power Dissipation/Heat

How do you cool these chips??

Courtesy Intel (S. Borkar)
Problem #2: Energy Consumption

No Moore’s law for batteries…
Today: Understand where power goes and ways to manage it

(from Jon Eager, Gates Inc., S. Watanabe, Sony Inc.)

Nominal Capacity (Watt-hours/lb)

Year

65 70 75 80 85 90 95

Rechargeable Lithium
Ni-Metal Hydride
Nickel-Cadmium

Battery (40+ lbs)
Dynamic Energy Dissipation

### Charging

\[ E_{0\rightarrow1} = C_L V_{DD}^2 \]
\[ E_{\text{cap}} = \frac{1}{2} C_L V_{DD}^2 \]
\[ E_{\text{diss, } R_P} = \frac{1}{2} C_L V_{DD}^2 \]

### Discharging

\[ E_{\text{diss, } R_N} = \frac{1}{2} C_L V_{DD}^2 \]

\[ P = C_L V_{DD}^2 f_{clk} \]
Transition Activity Factor $\alpha_{0\rightarrow1}$

<table>
<thead>
<tr>
<th>Current Input</th>
<th>Next Input</th>
<th>Output Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>00</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>00</td>
<td>01</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>00</td>
<td>10</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>00</td>
<td>11</td>
<td>1 $\rightarrow$ 0</td>
</tr>
<tr>
<td>01</td>
<td>00</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>01</td>
<td>01</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>01</td>
<td>10</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>01</td>
<td>11</td>
<td>1 $\rightarrow$ 0</td>
</tr>
<tr>
<td>10</td>
<td>00</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>10</td>
<td>01</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>1 $\rightarrow$ 1</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>1 $\rightarrow$ 0</td>
</tr>
<tr>
<td>11</td>
<td>00</td>
<td>0 $\rightarrow$ 1</td>
</tr>
<tr>
<td>11</td>
<td>01</td>
<td>0 $\rightarrow$ 1</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>0 $\rightarrow$ 1</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>0 $\rightarrow$ 0</td>
</tr>
</tbody>
</table>

Assume inputs $(A,B)$ arrive at $f$ and are uniformly distributed.

What is the average power dissipation?

$\alpha_{0\rightarrow1} = \frac{3}{16}$

$P = \alpha_{0\rightarrow1} \cdot C_L \cdot V_{DD}^2 \cdot f$
### Simple Scenario

- **Temperature Difference:** \( T_j - T_a = R_{\theta JA} P_D \)
- **Thermal Resistance:** \( R_{\theta JA} \) is the thermal resistance between silicon and Ambient.
- **Equation:** \( T_j = T_a + R_{\theta JA} P_D \)
- **Objective:** Make this as low as possible.

### Realistic Scenario

- **Temperature Levels:**
  - \( T_j \)
  - \( T_a \)
  - \( T_C \)
  - \( T_S \)
  - \( T_A \)
- **Resistance Components:**
  - \( R_{\theta JA} \)
  - \( R_{\theta JC} \)
  - \( R_{\theta CS} \)
  - \( R_{\theta SA} \)
- **Equation:** \( R_{\theta CA} = R_{\theta CS} + R_{\theta SA} \)
- **Objective:** Minimized by facilitating heat transfer (bolt case to extended metal surface – heat sink).
Large Thermal Gradients

Temp \( ^\circ C \)

Cache

Execution core

Integer & FP ALUs

120\(^\circ C\)

70\(^\circ C\)

Courtesy of Intel (Ram Krishnamurthy)
Pentium 4 @ 3.06 GHz dissipates 81.8W!

- Maximum $T_C = 69 \, ^\circ C$
- $R_{CA} < 0.23 \, ^\circ C/W$ for 50 C ambient
- Typical chips dissipate 0.5-1W (cheap packages without forced air cooling)

<table>
<thead>
<tr>
<th>Processor and Core Frequency</th>
<th>Thermal Design Power $^1,2$ (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processors with $V_D=1.500V$</td>
<td></td>
</tr>
<tr>
<td>2 GHz</td>
<td>52.4</td>
</tr>
<tr>
<td>2.20 GHz</td>
<td>55.1</td>
</tr>
<tr>
<td>2.26 GHz</td>
<td>56.0</td>
</tr>
<tr>
<td>2.40 GHz</td>
<td>57.8</td>
</tr>
<tr>
<td>2.50 GHz</td>
<td>58.3</td>
</tr>
<tr>
<td>2.53 GHz</td>
<td>58.3</td>
</tr>
<tr>
<td>Processors with $V_D=1.525V$</td>
<td></td>
</tr>
<tr>
<td>2 GHz</td>
<td>54.3</td>
</tr>
<tr>
<td>2.20 GHz</td>
<td>57.1</td>
</tr>
<tr>
<td>2.26 GHz</td>
<td>58.0</td>
</tr>
<tr>
<td>2.40 GHz</td>
<td>59.8</td>
</tr>
<tr>
<td>2.50 GHz</td>
<td>61.0</td>
</tr>
<tr>
<td>2.53 GHz</td>
<td>61.5</td>
</tr>
<tr>
<td>2.60 GHz</td>
<td>62.6</td>
</tr>
<tr>
<td>2.66 GHz</td>
<td>66.1</td>
</tr>
<tr>
<td>2.80 GHz</td>
<td>68.4</td>
</tr>
<tr>
<td>Processors with multiple VIDs</td>
<td></td>
</tr>
<tr>
<td>2 GHz</td>
<td>54.3</td>
</tr>
<tr>
<td>2.20 GHz</td>
<td>57.1</td>
</tr>
<tr>
<td>2.26 GHz</td>
<td>58.0</td>
</tr>
<tr>
<td>2.40 GHz</td>
<td>59.8</td>
</tr>
<tr>
<td>2.50 GHz</td>
<td>61.0</td>
</tr>
<tr>
<td>2.53 GHz</td>
<td>61.5</td>
</tr>
<tr>
<td>2.60 GHz</td>
<td>62.6</td>
</tr>
<tr>
<td>2.66 GHz</td>
<td>66.1</td>
</tr>
<tr>
<td>2.80 GHz</td>
<td>68.4</td>
</tr>
<tr>
<td>3.06 GHz</td>
<td>81.8</td>
</tr>
</tbody>
</table>
Power Reduction Strategies

\[ P = \alpha_{0->1} C_L V_{DD}^2 f \]

- Reduce Transition Activity or Switching Events
- Reduce Capacitance (e.g., keep wires short)
- Reduce Power Supply Voltage
- Frequency is typically fixed by the application, though this can be adjusted to control power

Optimize at all levels of design hierarchy
Clock Gating is a Good Idea!

Clock gating reduces activity and is the most common low-power technique used today.

Clock Gating Reduces Energy, does it reduce Power?

Global Clock

Enable_Adder

Adder Clock

Adder Off

Enable_Multiplier

Multiplier Clock

Multiplier On

Clock Gating Reduces Energy, does it reduce Power?
Use of Thermal Feedback

- Note that there is a difference between average and peak power
- On-chip thermal sensor (diode based), measures the silicon temperature
- If the silicon junction gets too hot, then the activity is reduced (e.g., reduce clock rate)
Power Supply Parasitics

![Circuit Diagram](image)

- $L_{\text{board}}$
- $L_{\text{package}}$
- $R_{\text{grid}}$
- On-die decap
- Board decap
- Switching currents

**200Mhz Design**

*Courtesy of Motorola (David Blaauw)*
Power Supply Resonance!

- Graph 1: Voltage vs. Time
- Graph 2: Current vs. Time
Number Representation: Two’s Complement vs. Sign Magnitude

Two’s complement

Consider a 16 bit bus where inputs toggles between +1 and −1 (i.e., a small noise input)
Which representation is more energy efficient?

Sign-Magnitude
Bus Coding to Reduce Activity

Majority Function

Extra bit to indicated if the bus is inverted

Input

Data Bus

Output

[Stan94]
Time Sharing Increases Switching Activity
float a[256], b[256];
float pi = 3.14;

for (i = 0; i < 255; i++) {
    a[i] = sin(pi * i / 256);
    b[i] = cos(pi * i / 256);
}

512(8) + 2 + 4 + 8 + 16 + 32 + 64 + 128 + 256 = 4607 bit transitions

2(8) + 2(2 + 4 + 8 + 16 + 32 + 64 + 128 + 256) = 1030 transitions
Pre-Computation

from [Alidina94]
(1994 International Workshop on Low-power Design)
Glitching Transitions

Chain Topology

Tree Topology

Balancing paths reduces glitching transitions

For 4 inputs, 50% less transitions using a tree approach

Structures such as multipliers have lot of glitching transitions

Keeping logic depths short (e.g., pipelining) reduces glitching
Reduce Supply Voltage: But is it Free?

\[ V_{DD} \text{ from } 2V \text{ to } 1V, \text{ energy } \downarrow \text{ by } x4, \text{ delay } \uparrow \times 2 \]
Voltage Scaling Using Parallelism

\[ P_{\text{serial}} = C_{\text{mult}} 2^2 f \]

\[ P_{\text{parallel}} = (2C_{\text{mult}} 1^2 f / 2) = P_{\text{serial}} / 4 \]

Trade Area for Low Power
Algorithmic Workload

Compare Current Image...
...to Previous Image

Receiver just updates

Exploit Time Varying Algorithmic Workload
To Vary the Power Supply Voltage
Fixed Power Supply

\[ E_{\text{FIXED}} = \frac{1}{2} C V_{DD}^2 \]

Variable Power Supply

\[ E_{\text{VARIABLE}} = \frac{1}{2} C \left( \frac{V_{DD}}{2} \right)^2 = \frac{E_{\text{FIXED}}}{4} \]

[Graph showing normalized energy vs. normalized workload with fixed and variable supply]

\[ [\text{Gutnik97}] \]
Digitally adjustable DC-DC converter powers SA-1110 core

\[ \text{Controller} \rightarrow 3.6V \rightarrow \text{V}_{\text{out}} \rightarrow \text{SA-1110} \rightarrow \text{\(\mu\text{OS}\)} \]

\(\mu\text{OS}\) selects appropriate clock frequency based on workload and latency constraints
- User adjusts number of filter taps
- Frequency/Voltage adjusted appropriately (via eCOS based µOS)
Energy Scavenging

MEMS Generator

Vibration-to-Electric Conversion

~ 10\(\mu\)W

Jose Mur Miranda/
Jeff Lang

Power Harvesting Shoes

Joe Paradiso
(Media Lab)

After 3-6 steps, it provides 3 mA for 0.5 sec

~10mW