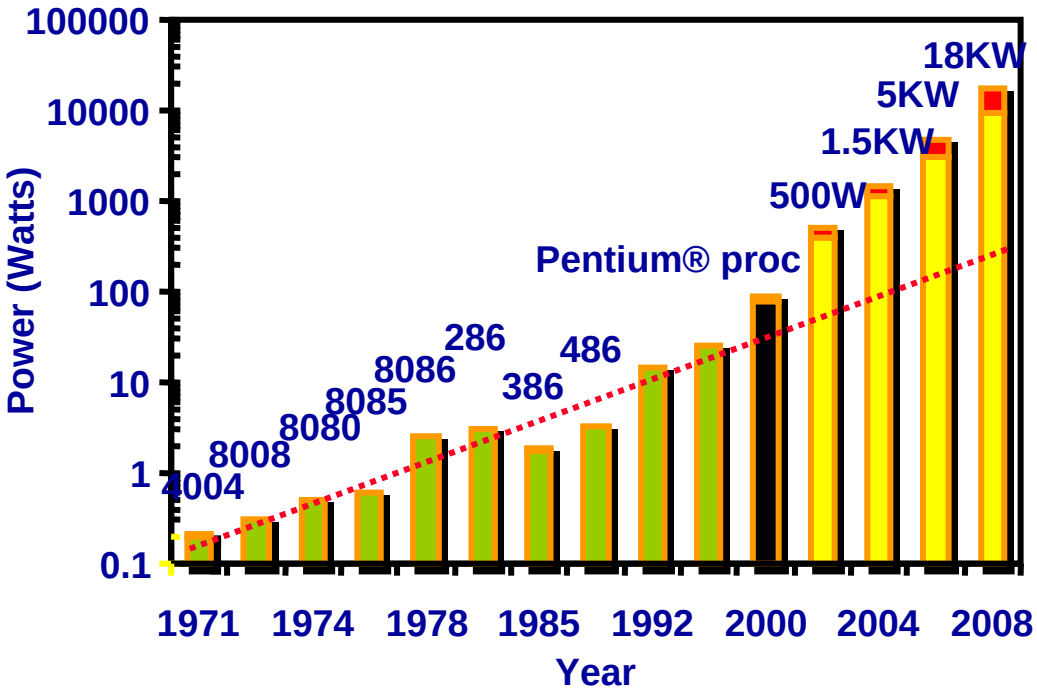
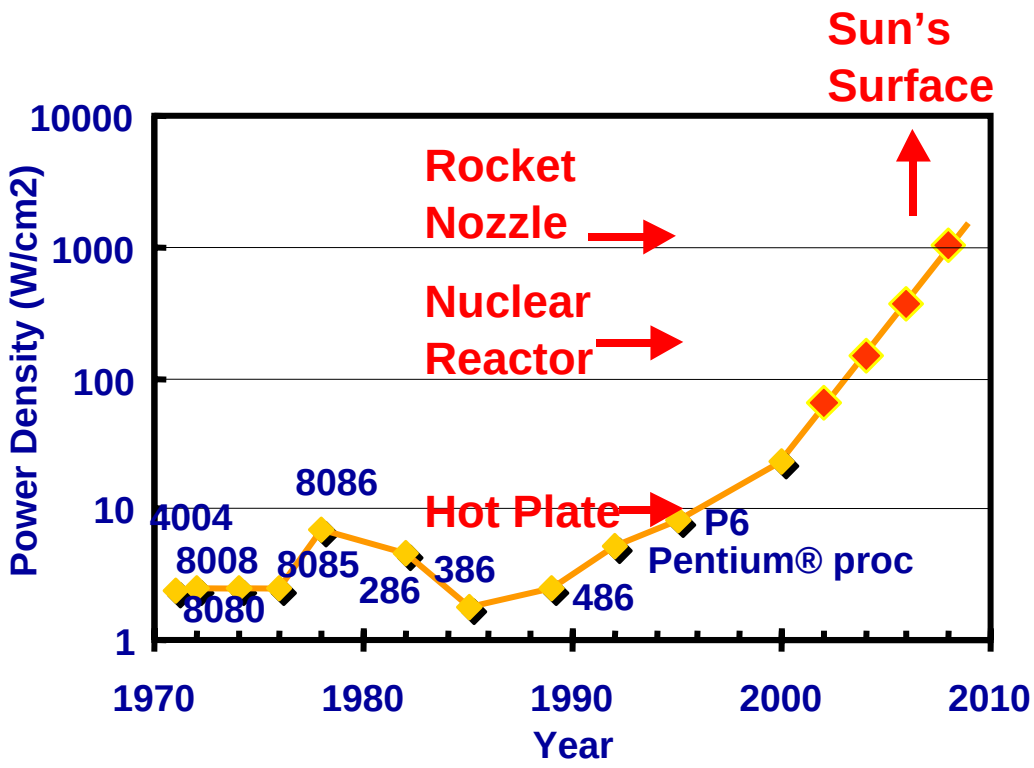
L16: Power Dissipation in Digital Systems

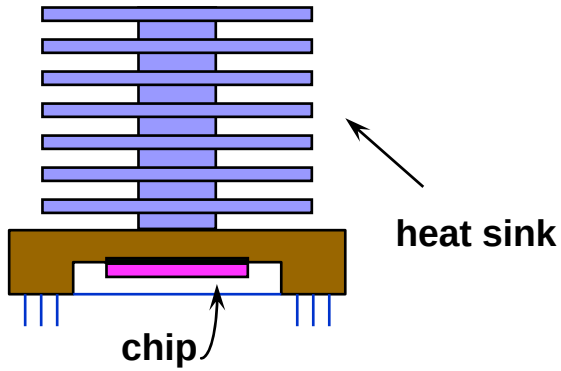
Problem #1: Power Dissipation/Heat



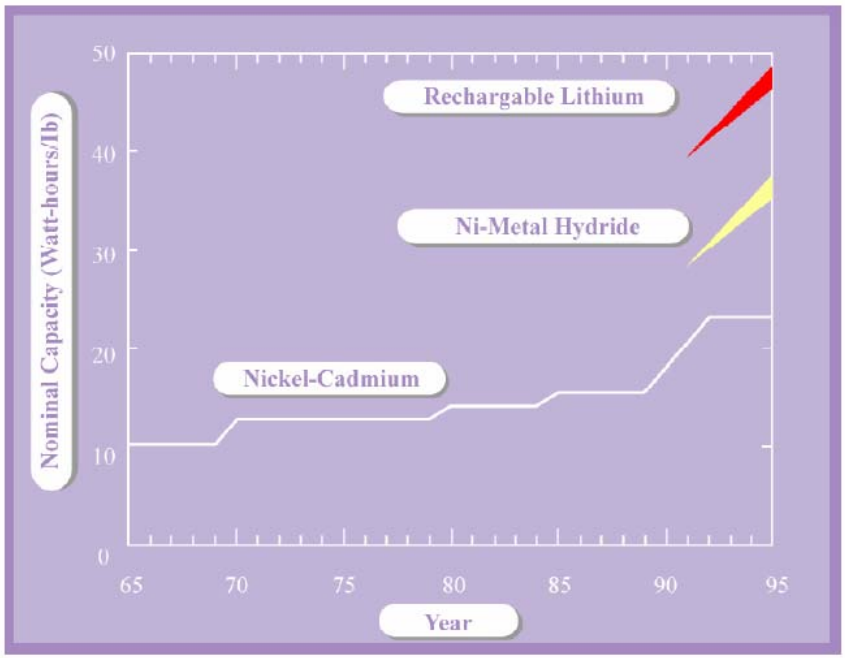


Courtesy Intel (S. Borkar)

How do you cool these chips??



Problem #2: Energy Consumption



(Image by MIT OCW. Adapted from Jon Eager, Gates Inc. , S. Watanabe, Sony Inc.)

Image by MIT OCW.

The Energy Problem

7.5 cm³ AA battery

Alkaline: ~10,000J

What can One Joule of energy do?

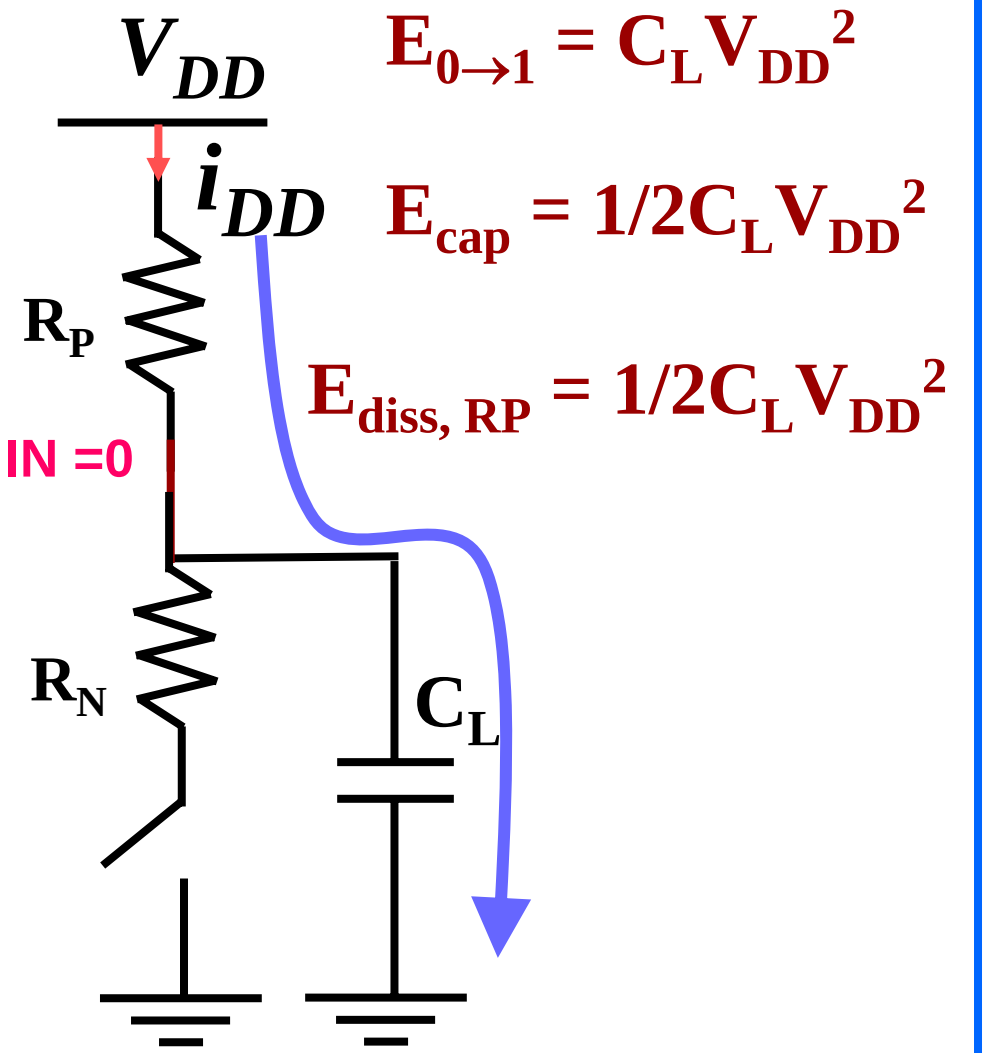
- **Mow your lawn for 1 ms**
- **Operate a processor for ~ 7s**
- **Send a 1 Megabyte file over 802.11b**

No Moore's law for batteries…

Today: Understand where power goes and ways to manage it

Dynamic Energy Dissipation

Charging

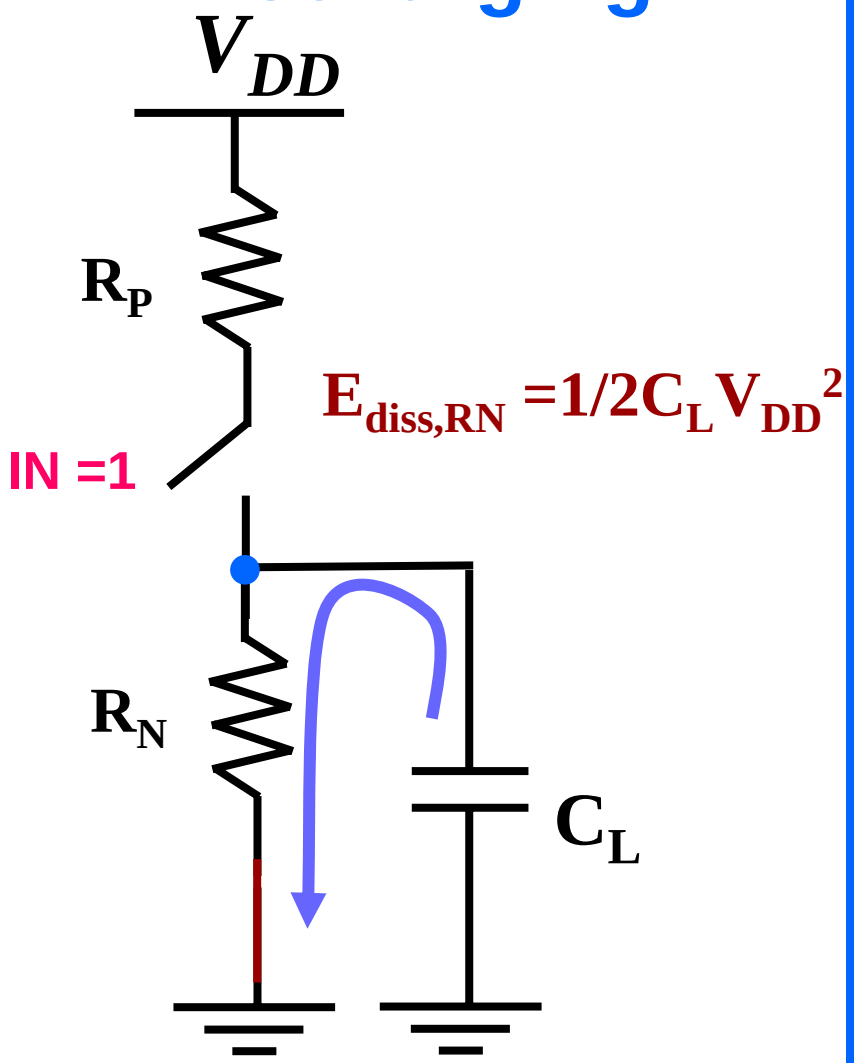


$E_{0\rightarrow1} = C_L V_{DD}^2$

$E_{cap} = 1/2 C_L V_{DD}^2$

$E_{diss,\ RP} = 1/2 C_L V_{DD}^2$

Discharging

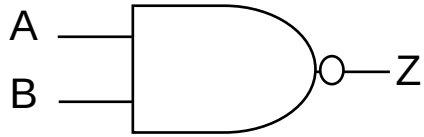


$E_{diss,RN} = 1/2 C_L V_{DD}^2$

$$P = C_L\ V_{DD}^2\ f_{clk}$$

The Transition Activity Factor $\alpha_{0->1}$

Current Input	Next Input	Output Transition
00	00	1 –> 1
00	01	1 –> 1
00	10	1 –> 1
00	11	1 –> 0
01	00	1 –> 1
01	01	1 –> 1
01	10	1 –> 1
01	11	1 –> 0
10	00	1 –> 1
10	01	1 –> 1
10	10	1 –> 1
10	11	1 –> 0
11	00	0 –> 1
11	01	0 –> 1
11	10	0 –> 1
11	11	0 –> 0



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

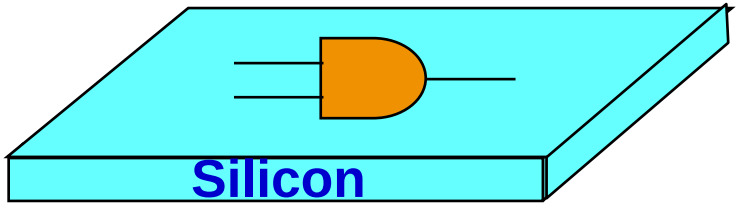
What is the average power dissipation?

$$\alpha_{0->1} = 3/16$$

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

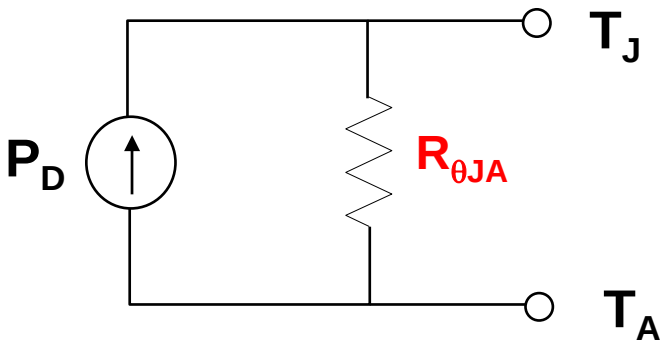
Junction (Silicon) Temperature

Simple Scenario



$$T_j - T_a = R_{\theta JA} P_D$$

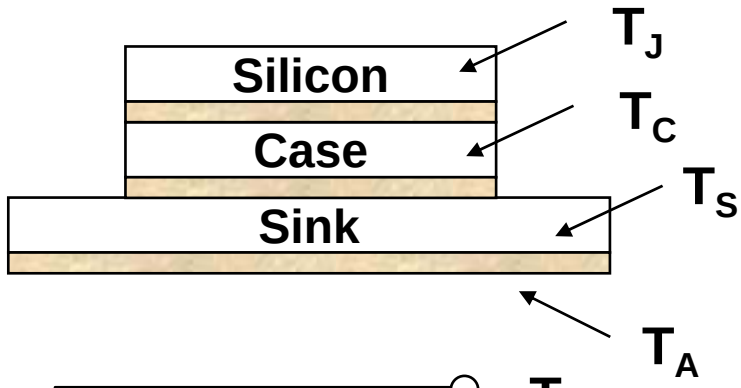
$R_{\theta JA}$ is the thermal resistance between silicon and Ambient

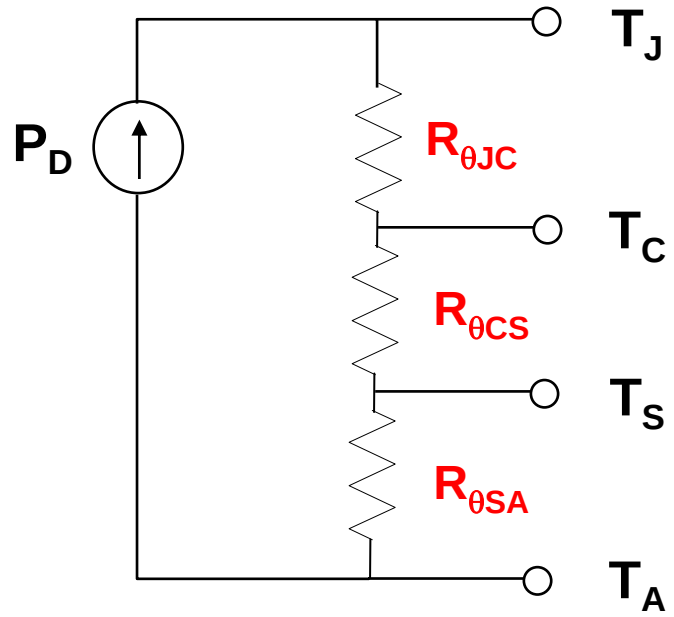


$$T_j = T_a + R_{\theta JA} P_D$$

Make this as low as possible

Realistic Scenario





$$R_{\theta CA} = R_{\theta CS} + R_{\theta SA}$$

is minimized by facilitating heat transfer
(bolt case to extended metal surface – heat sink)

Intel Pentium 4 Thermal Guidelines

- **Pentium 4 @ 3.06 GHz dissipates 81.8W!**
- **Maximum T_C = 69 °C**
- **R_{CA} < 0.23 °C/W for 50 C ambient**
- **Typical chips dissipate 0.5-1W (cheap packages without forced air cooling)**

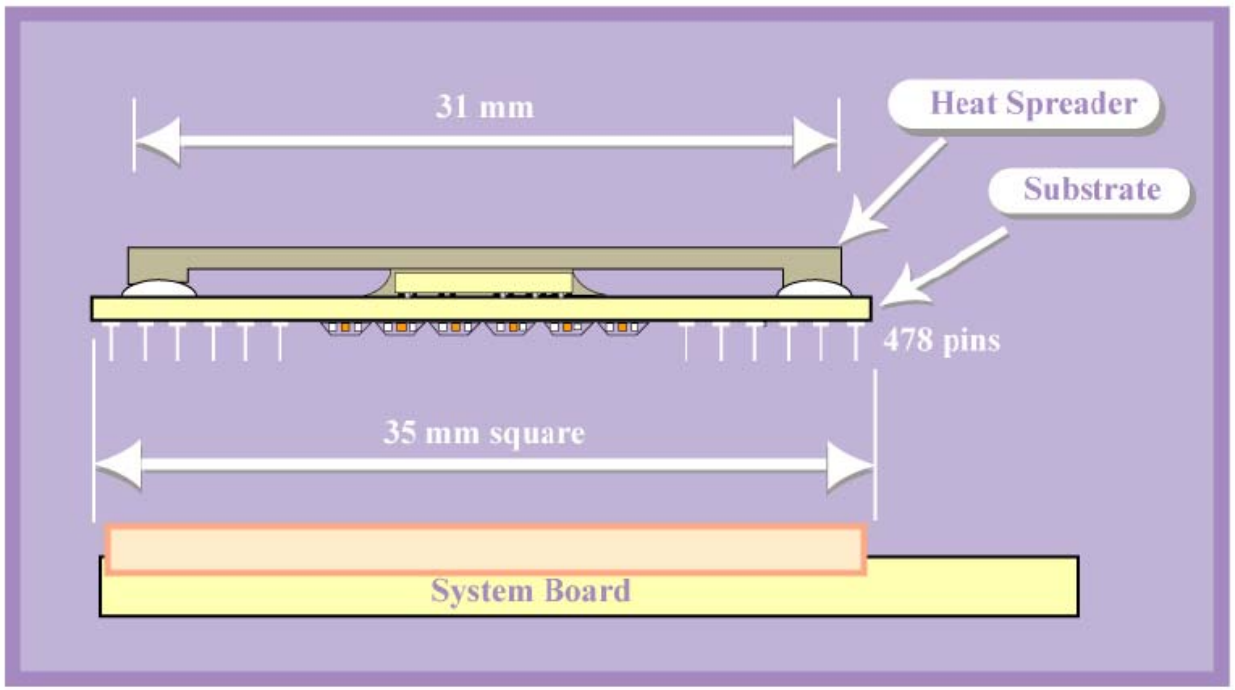


Image by MIT OpenCourseWare.

Processors and Core Frequency	Thermal Design Power (W)
Processors with VID = 1.500V	
2 GHz	52.4
2.20 GHz	55.1
2.26 GHz	56.0
2.40 GHz	57.8
2.50 GHz	59.3
2.53 GHz	59.3
Processors with VID = 1.525V	
2 GHz	54.3
2.20 GHz	57.1
2.26 GHz	58.0
2.40 GHz	59.8
2.50 GHz	61.0
2.53 GHz	61.5
2.60 GHz	62.6
2.66 GHz	66.1
2.80 GHz	68.4
Processors with Multiple VIDs	
2 GHz	54.3
2.20 GHz	57.1
2.26 GHz	58.0
2.40 GHz	59.8
2.50 GHz	61.0
2.53 GHz	61.5
2.60 GHz	62.6
2.66 GHz	66.1
2.80 GHz	68.4
3.06 GHz	81.8

Image by MIT OpenCourseWare. Adapted from Intel Pentium 4 documentation.

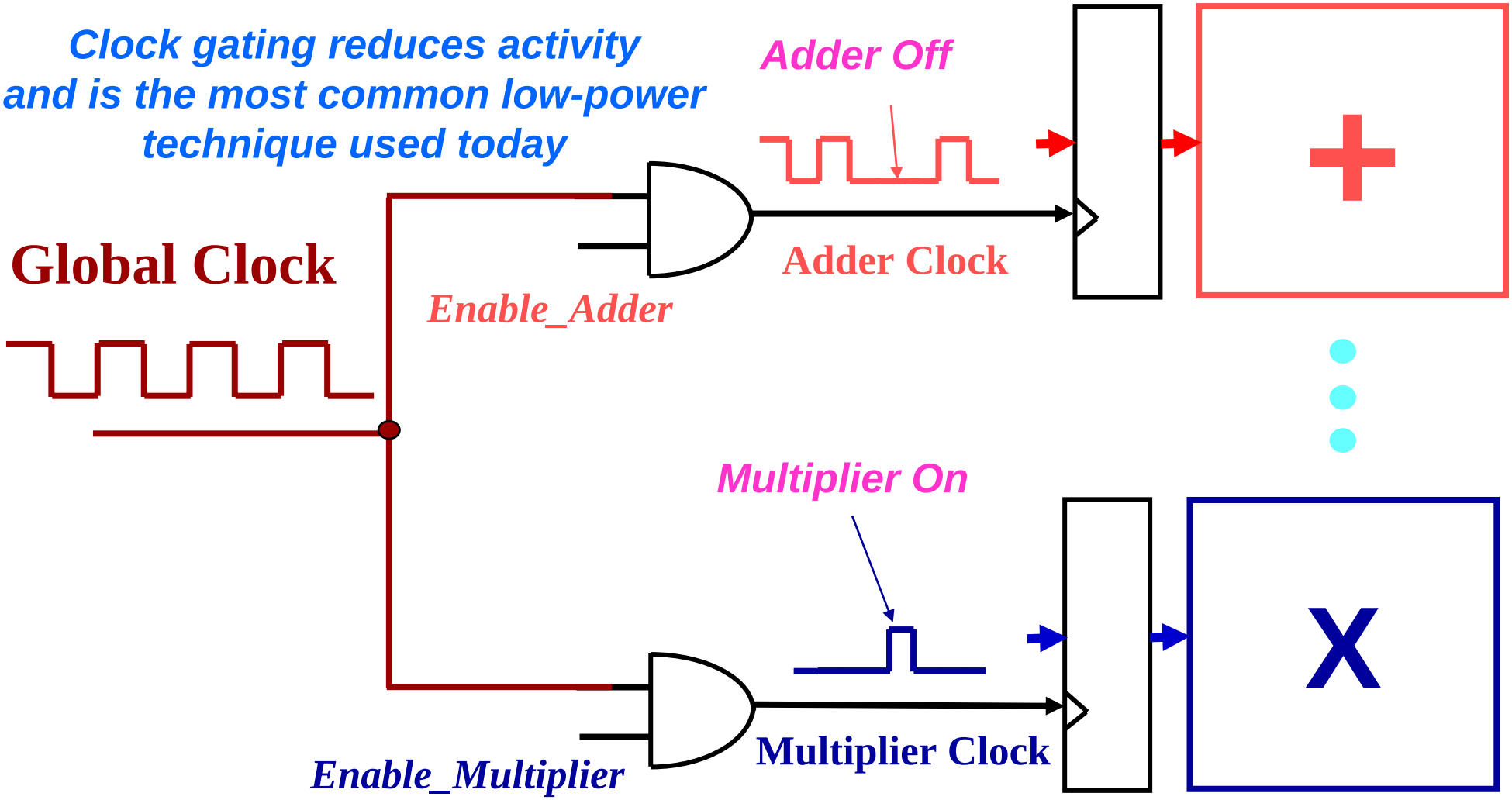
Power Reduction Strategies

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

- 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

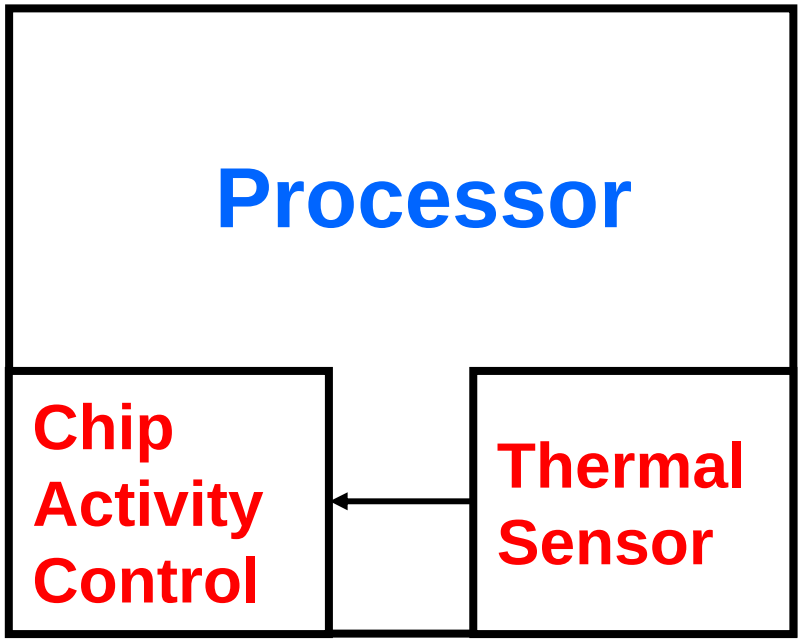
Optimize at all levels of design hierarchy

Clock Gating is a Good Idea!



Clock Gating Reduces Energy, does it reduce Power?

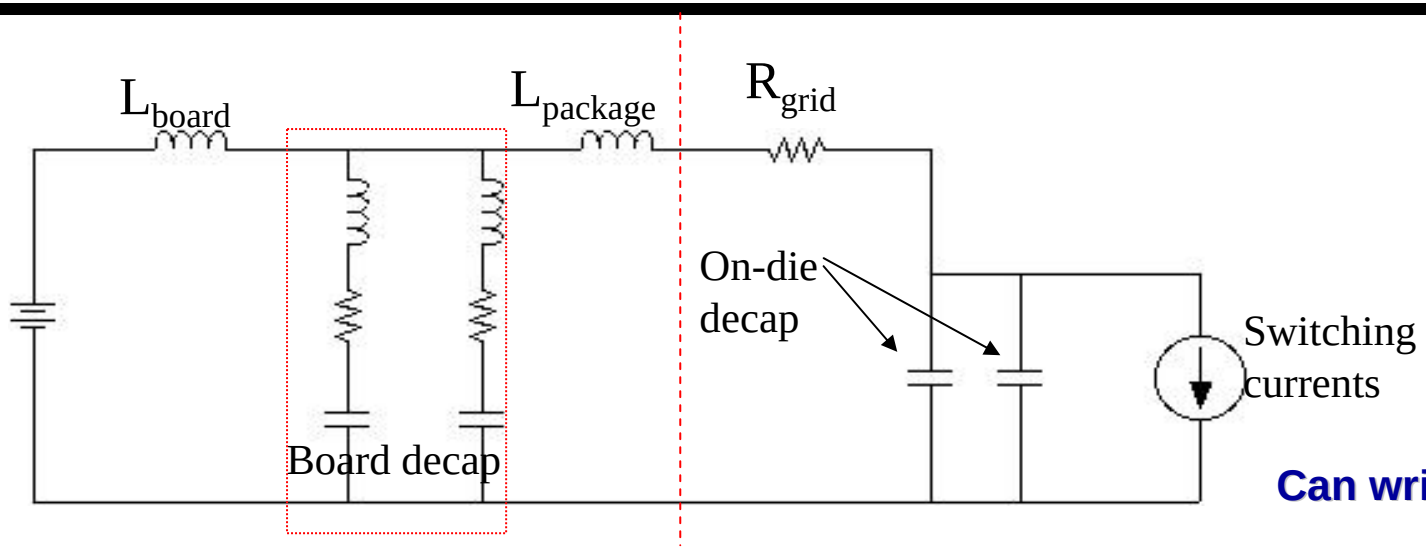
Does your GHz Processor run at a GHz?



- 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 (say 125 °C), then the activity is reduced (e.g., reduce clock rate or use clock gating)

Use of Thermal Feedback

Power Supply Resonance



Can write a Virus to Activate Power Supply Resonance!

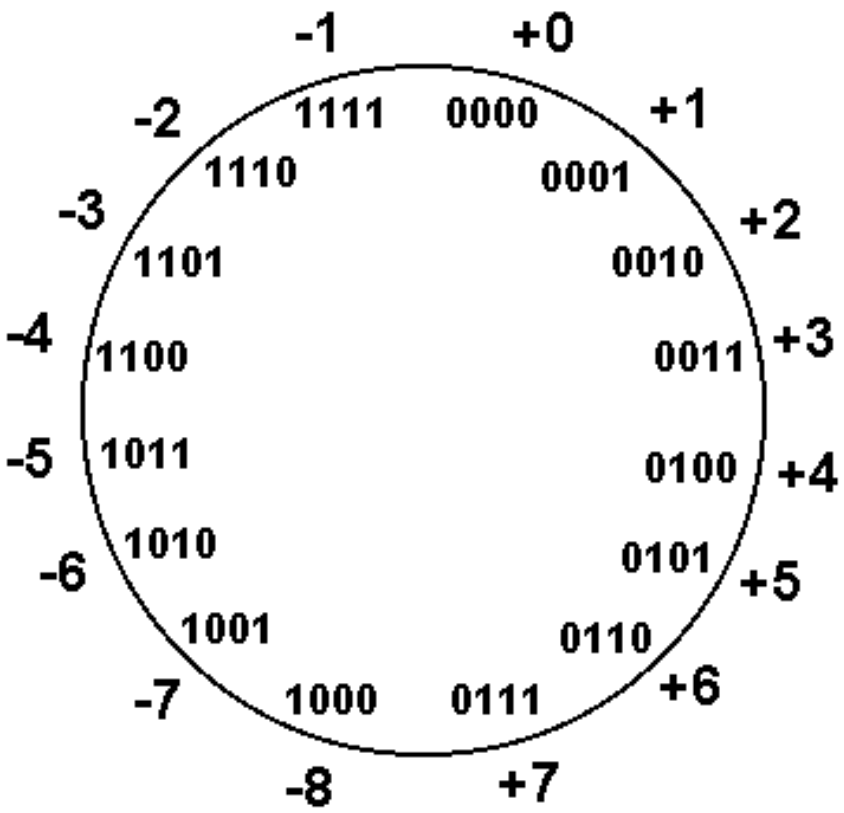
Image removed due to copyright restrictions.

Image removed due to copyright restrictions.

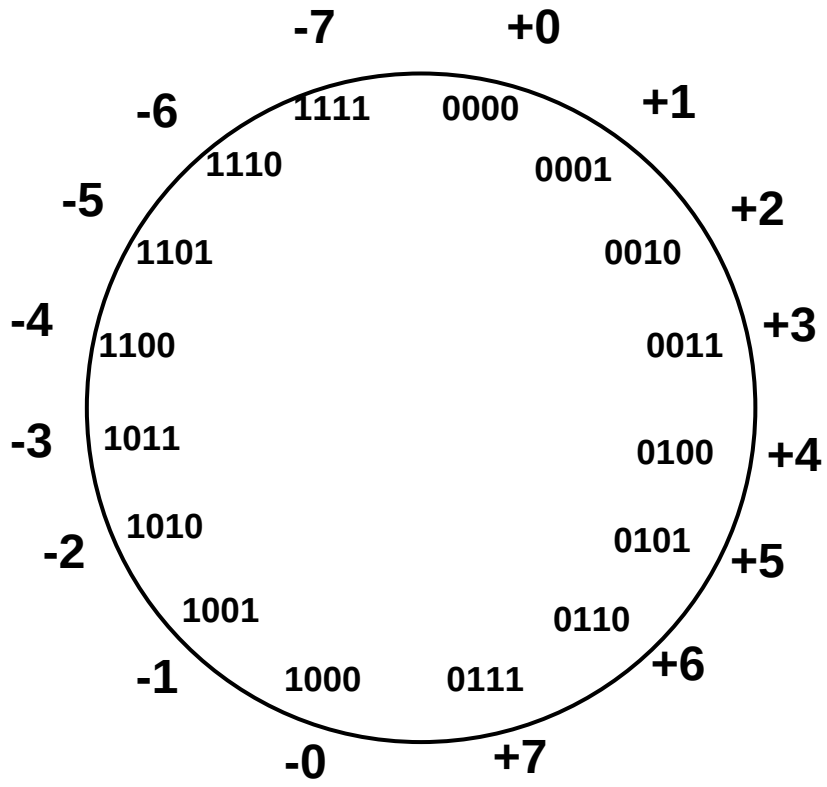
Image removed due to copyright restrictions.

Number Representation: Two's Complement vs. Sign Magnitude

Two's complement



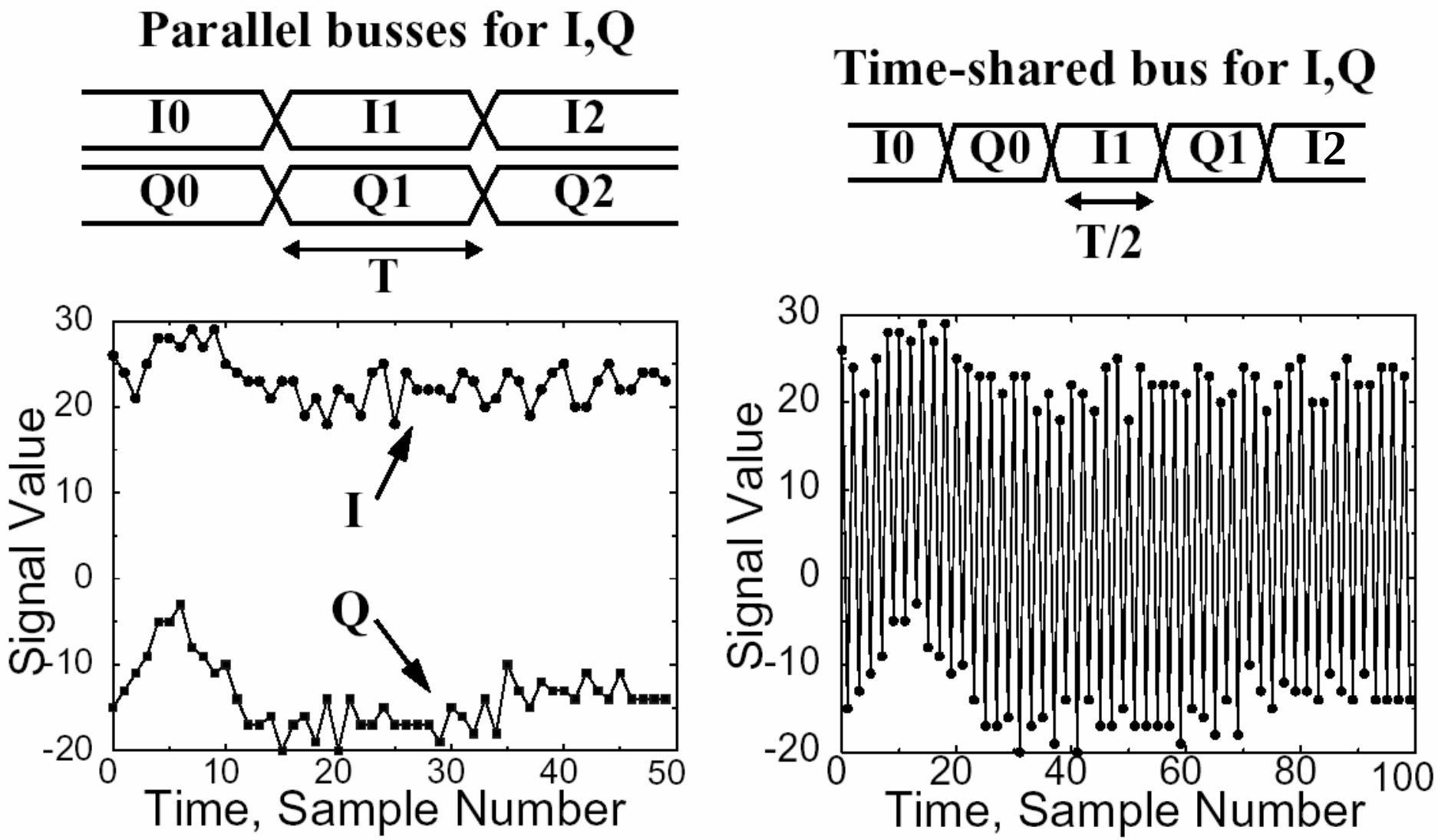
Sign-Magnitude



Consider a 16 bit bus where inputs toggles between +1 and –1 (i.e., a small noise input)

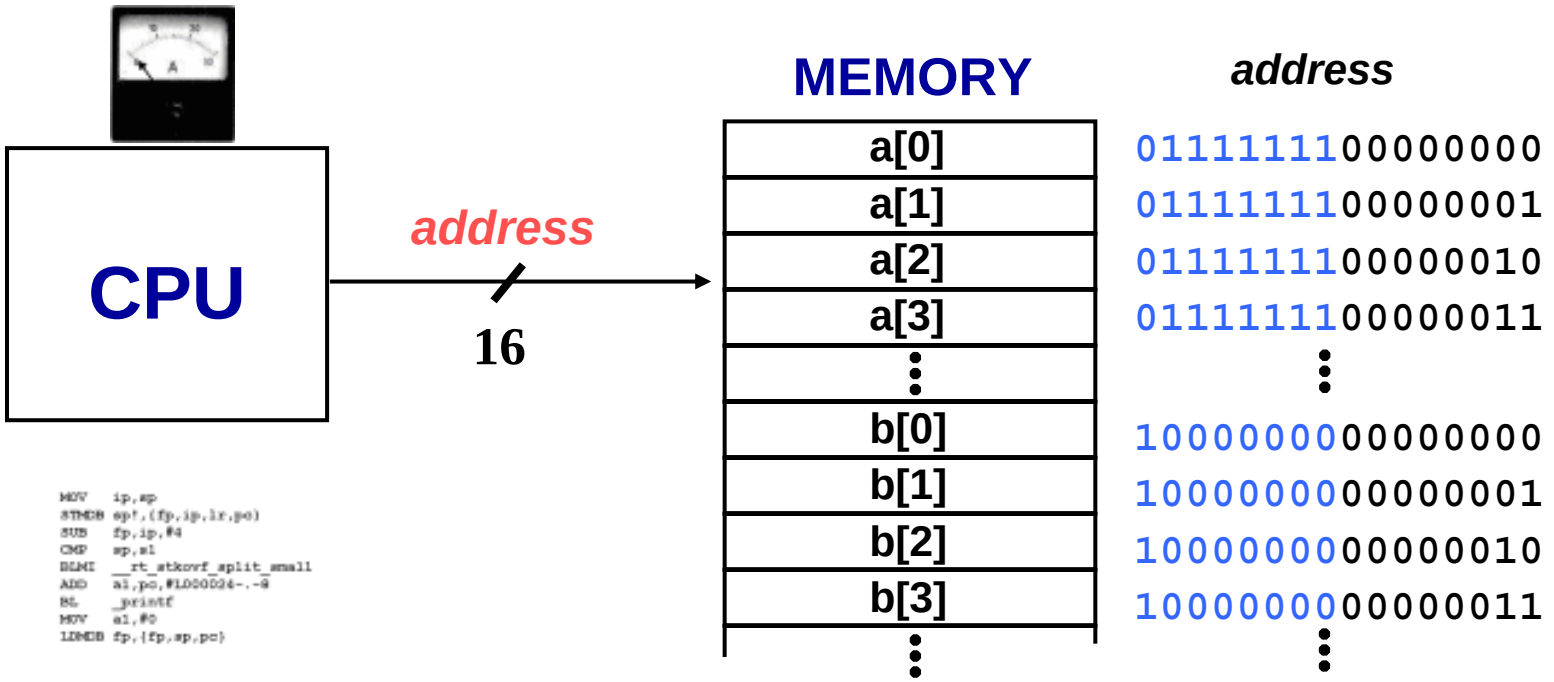
Which representation is more energy efficient?

Time Sharing is a Bad Idea



Time Sharing Increases Switching Activity

Not just a 6-1 Issue: "Cool" Software ???




```
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

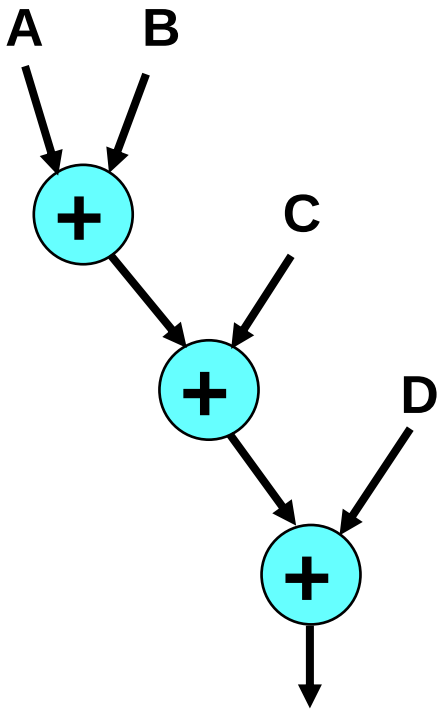
```
float a [256], b[256];
float pi= 3.14;

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

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

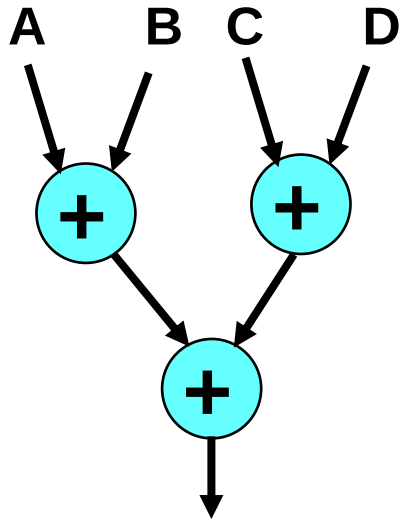
Glitching Transitions

Chain Topology



(((A+B) + C)+D)

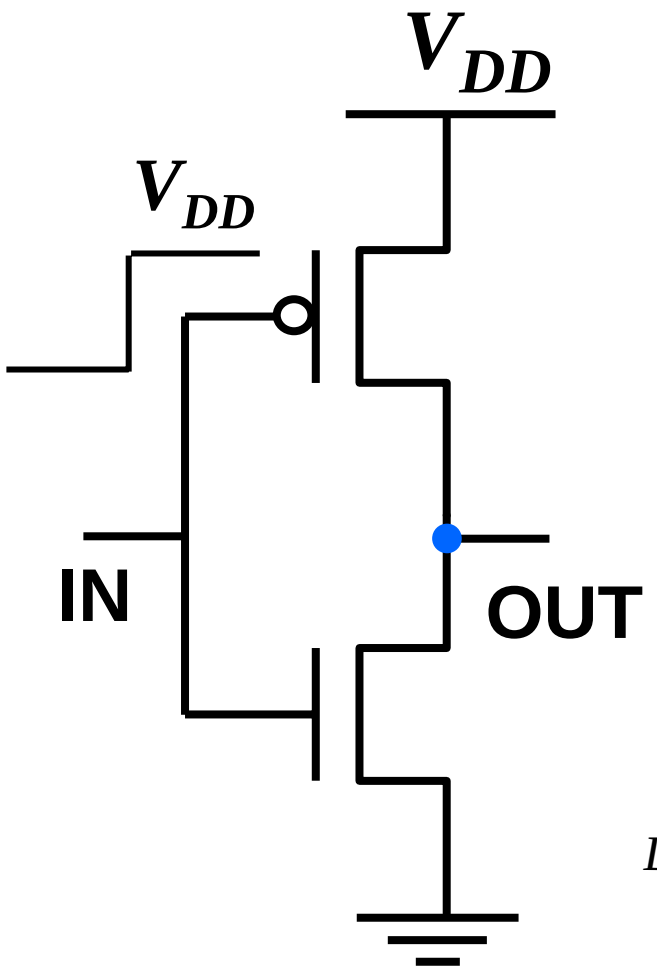
Tree Topology

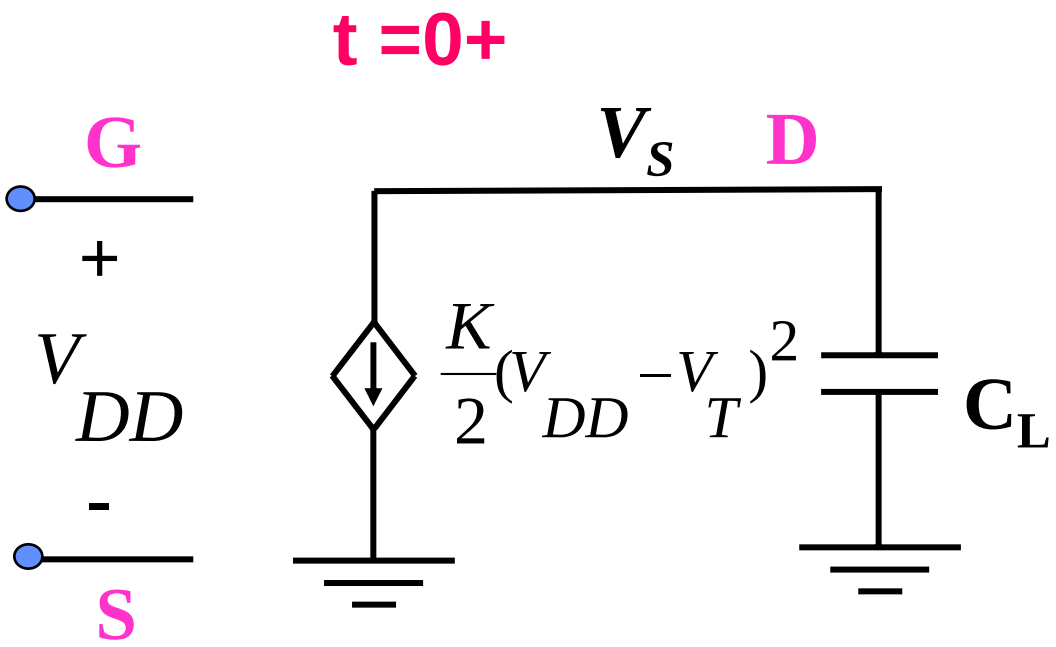


(A+B) + (C+D)

- Balancing paths reduces glitching transitions
- 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?

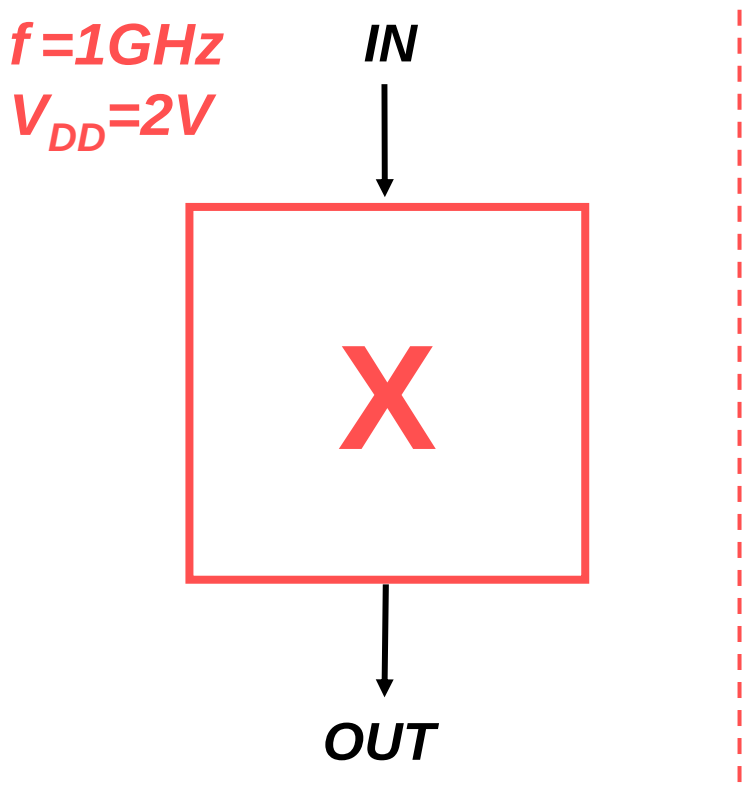




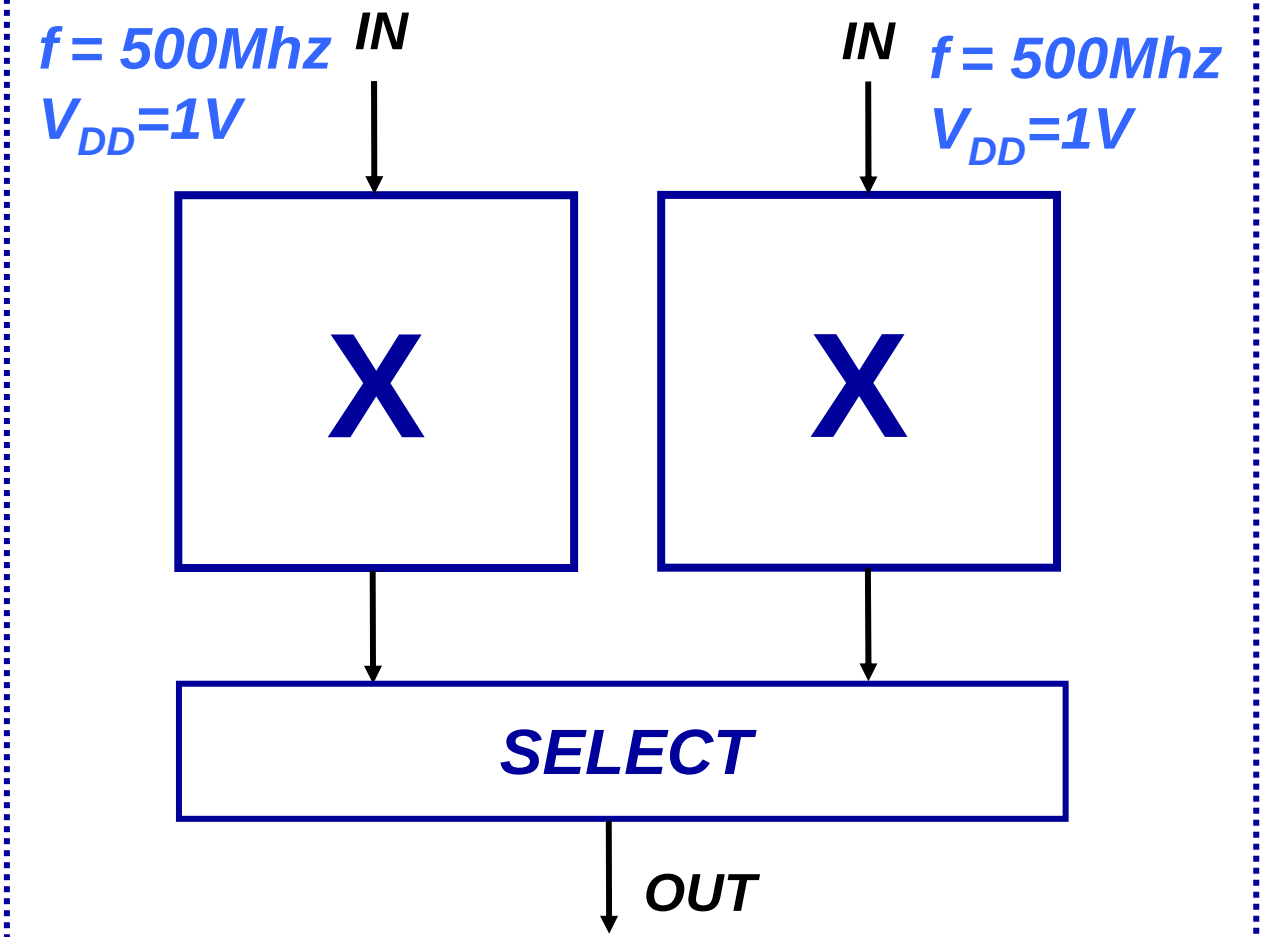
$$Delay = \frac{C_L \cdot \Delta V}{i_D} = \frac{C_L \cdot \frac{V_{DD}}{2}}{\frac{k}{2}(V_{DD}-V_T)^2} \propto \frac{V_{DD}}{(V_{DD}-V_T)^2} \approx \frac{1}{V_{DD}}$$

V_{DD} from 2V to 1V, energy ↓ by x4, delay ↑ x2

Transistors Are Free…
(What do you do with a Billion Transistors?)



$P_{serial} = C_{mult}\ 2^2\ f$



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

Trade Area for Low Power

Algorithmic Workload

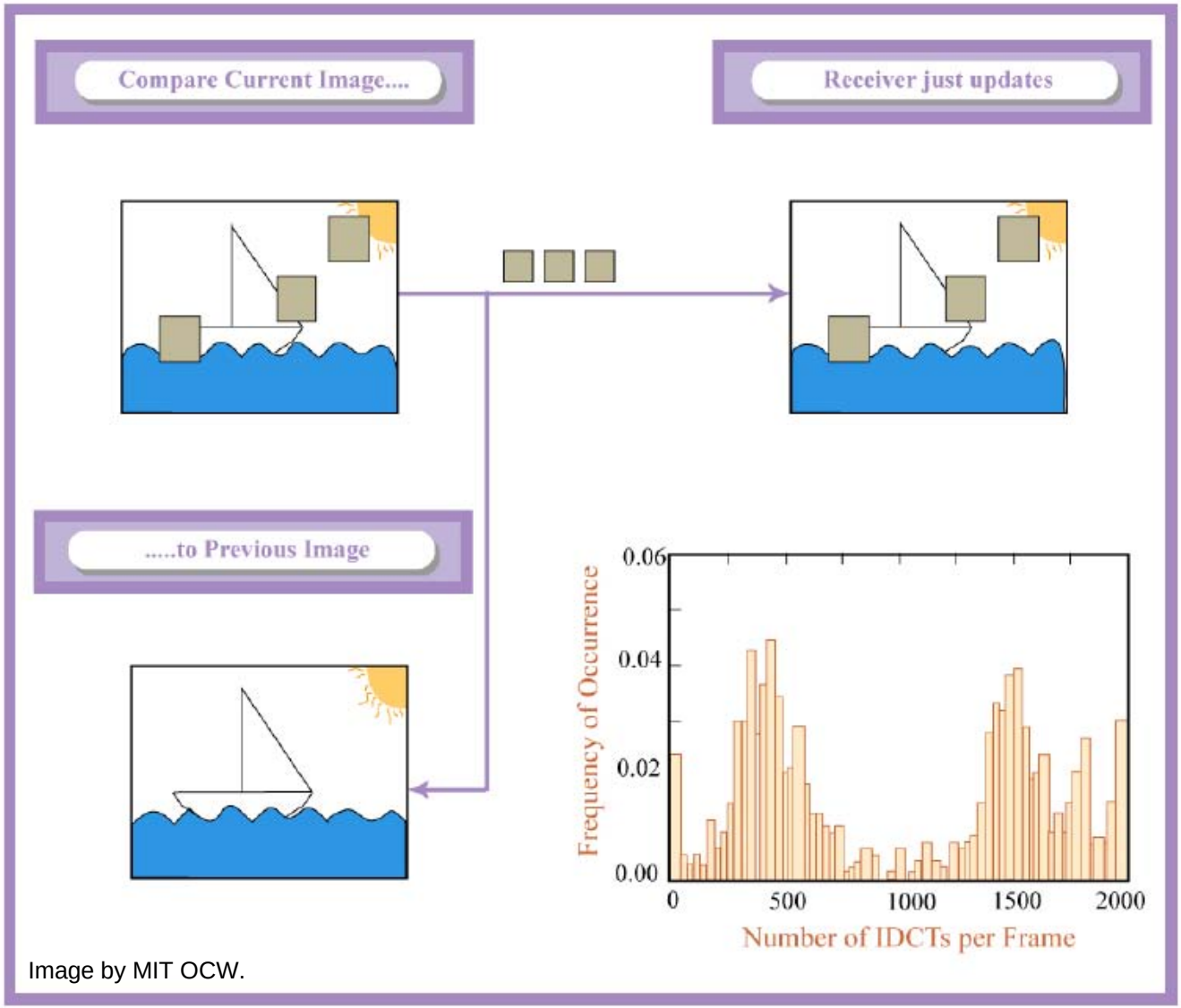


Image by MIT OCW.

Exploit Time Varying Algorithmic Workload To Vary the Power Supply Voltage

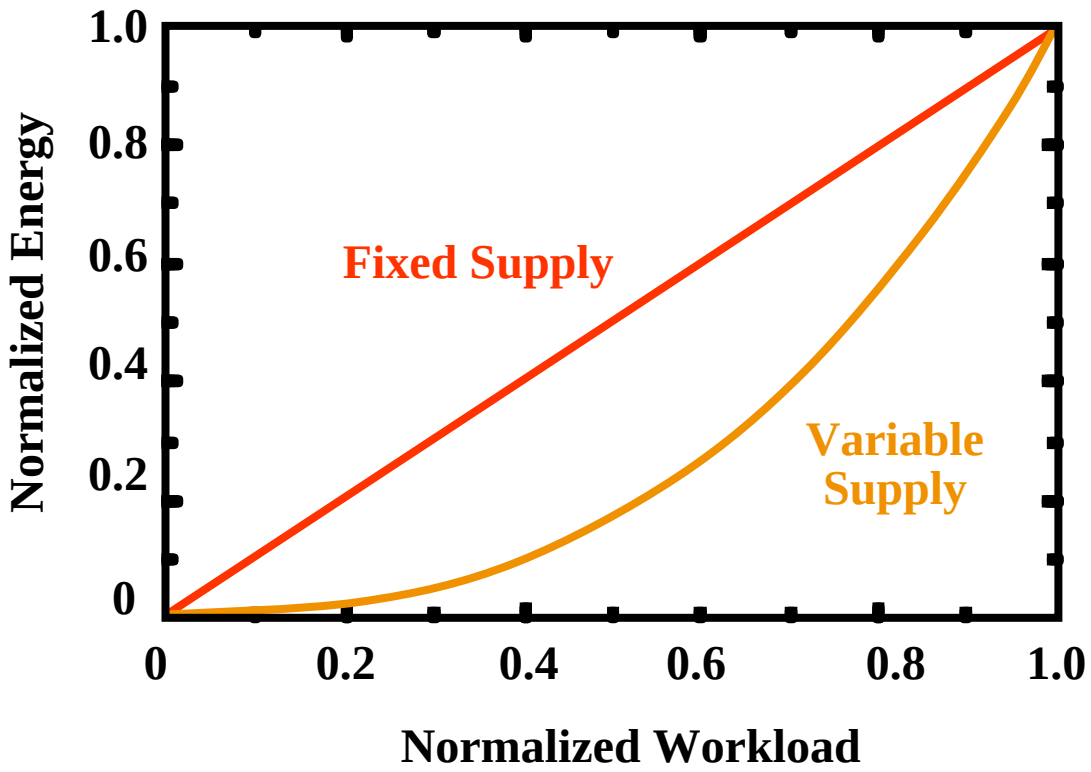
Dynamic Voltage Scaling (DVS)

Fixed Power Supply

$E_{FIXED} = \frac{1}{2} C V_{DD}^2$

Variable Power Supply

$E_{VARIABLE} = \frac{1}{2} C (V_{DD}/2)^2 = E_{FIXED} / 4$



[Gutnik97]

DVS on a Processor

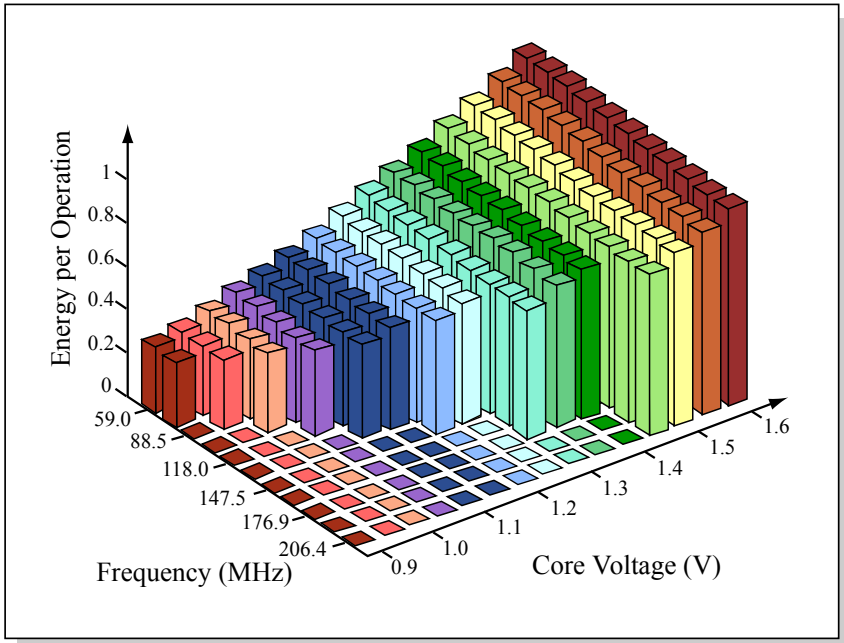
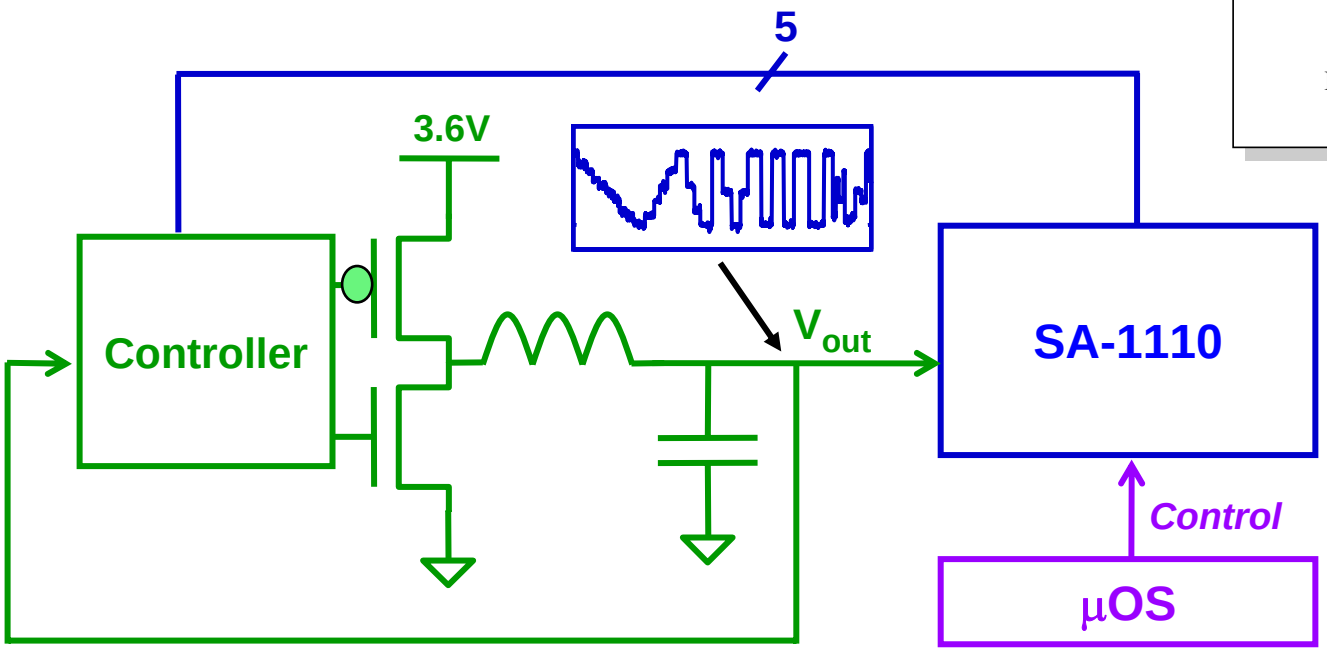


Figure by MIT OpenCourseWare. Adapted from R. Min, T. Furrer, and A. P. Chandrakasan. "Dynamic Voltage Scaling Techniques for Distributed Microsensor Networks." Workshop on VLSI (April 2000): 43-46.

Digitally adjustable DC-DC converter powers SA-1110 core



µOS selects appropriate clock frequency based on workload and latency constraints

Energy Efficiency of Software

Processor (StrongARM-1100)

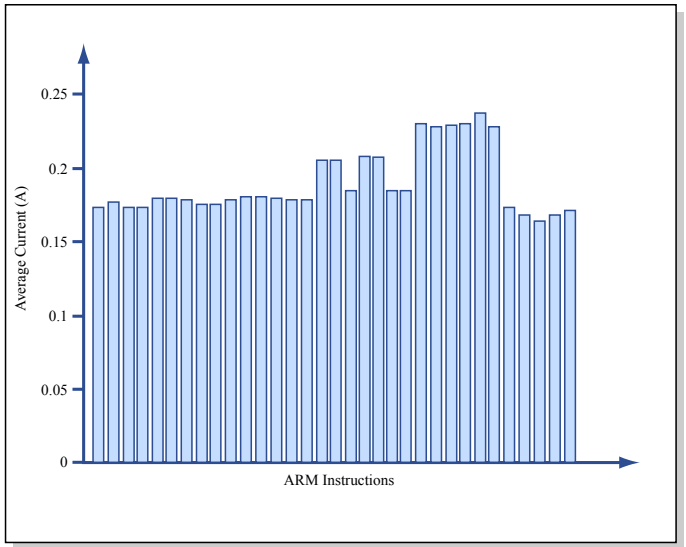
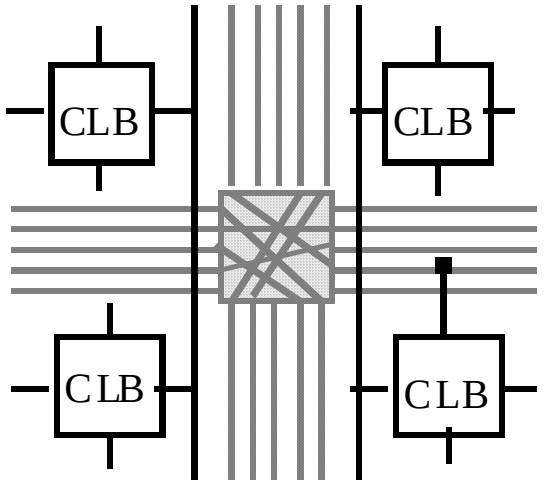


Figure by MIT OpenCourseWare. Adapted from A. Sinha, DAC.

FPGA (Xilinx)



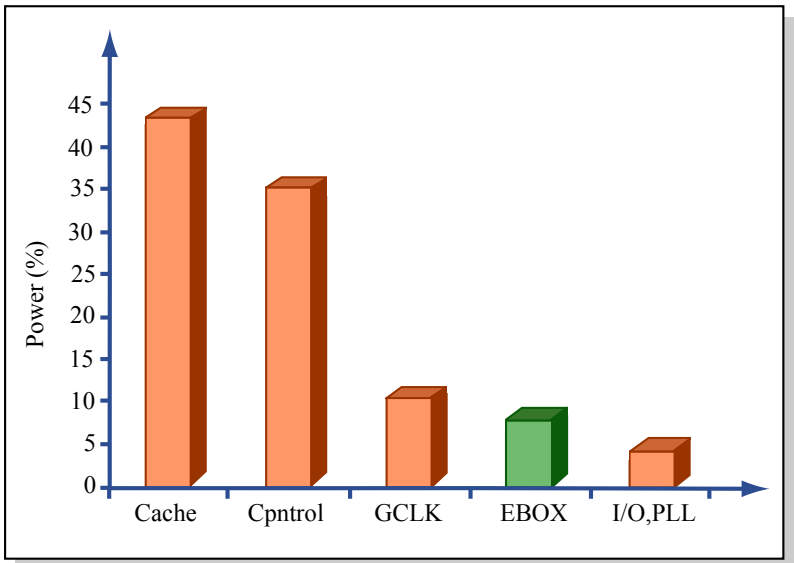


Figure by MIT OpenCourseWare. Adapted from Montanaro 1996, JSSC.

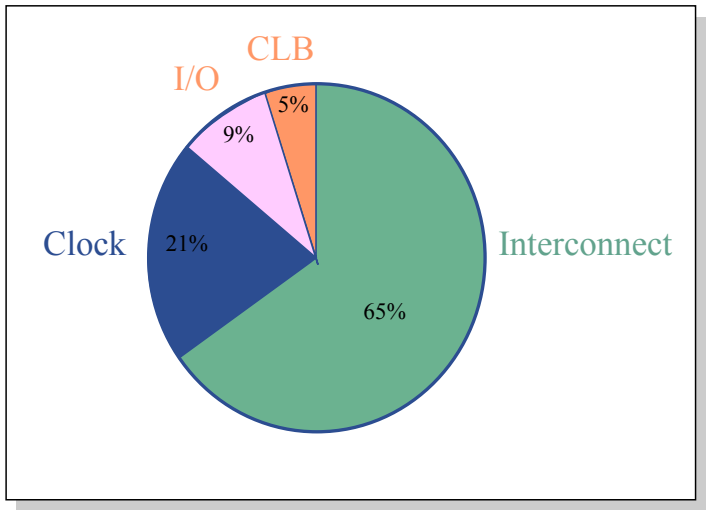
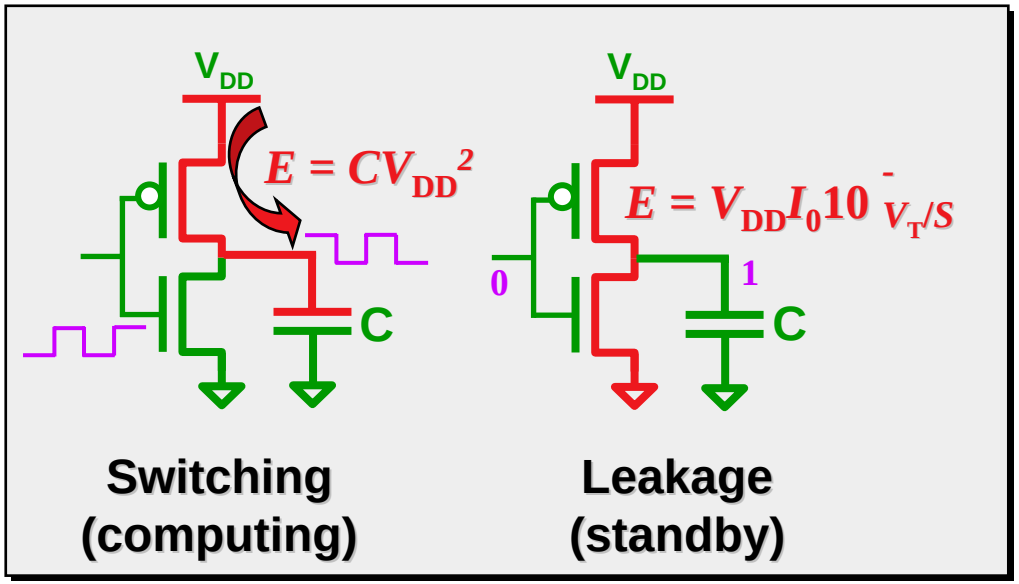
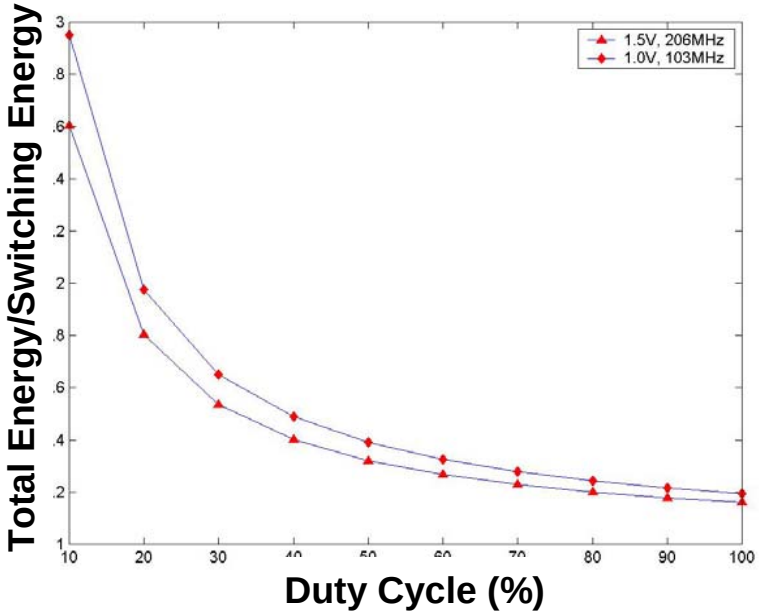


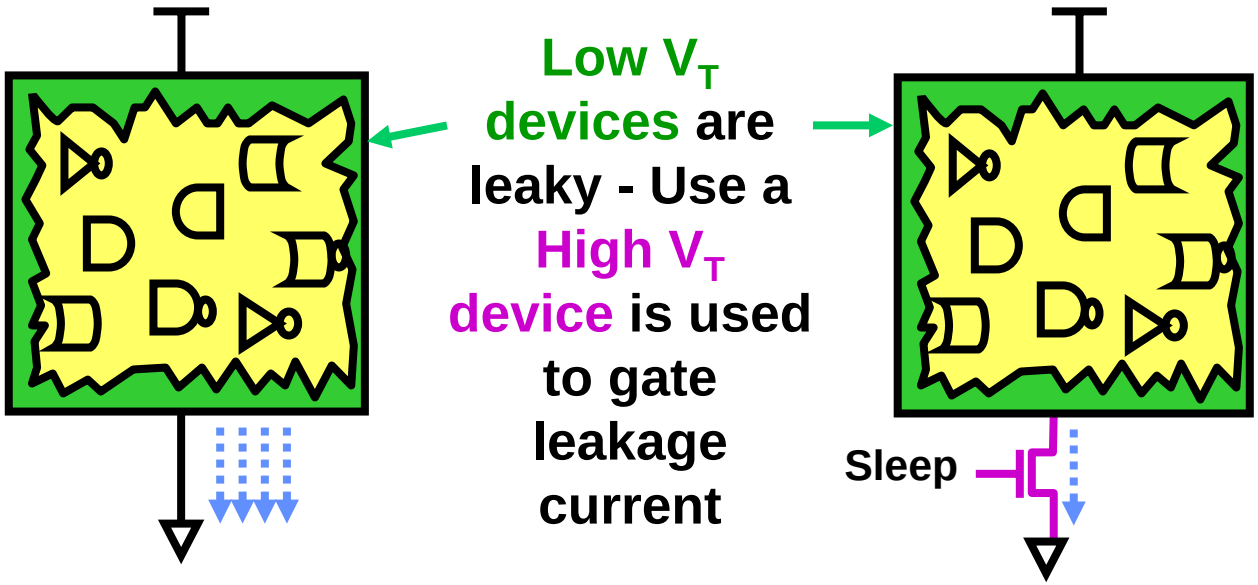
Image by MIT OpenCourseWare. Adapted from Kusse 1998, UCB.

"Software" Energy Dissipation has Large Overhead

Trends: Leakage and Power Gating







Trends: Energy Scavenging

MEMS Generator

Image removed due to copyright restrictions.

Power Harvesting Shoes

Courtesy of Joe Paradiso (MIT Media Lab).
Used with permission.

Vibration-to-Electric Conversion

~ 10μW

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

~10mW