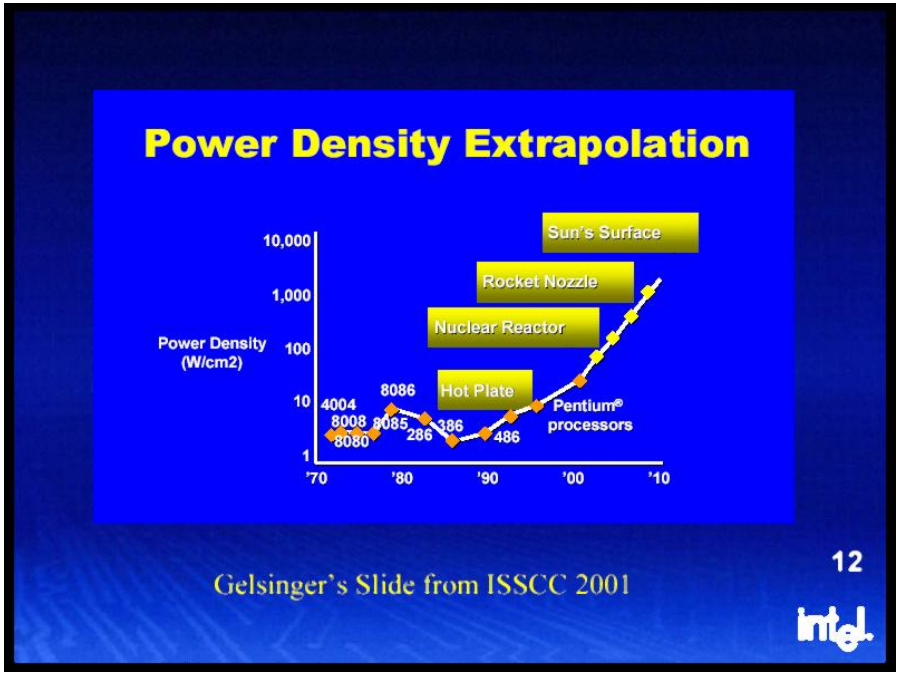
EECS 4340
Unit 6: Power and Energy



Prof. Simha Sethumadhavan

Reference Book: Low Power Design Essentials, J. Rabaey, 2009

Topics

- Power/Energy Basics
- Reducing Standby Power
- Reducing Dynamic Power
- Low Power Design Methodology
- Power Delivery Architecture

Importance of Power and Energy

- Power
  - Power is dissipated as heat; hot devices need cooling
  - Air cooling is $5, vs. liquid nitrogen gamer cooling is $200
  - At the data center level, cooling is ~ 30% op costs

- Energy
  - Battery life and energy bills matter
  - Ultimately energy/op determines what is computable
  - Can you simulate the brain today?

Measuring What Matters

- Performance metric: Time delay to perform a task (s)
- Efficiency metric: Effort to perform a task (Joule)
- Power metric: Energy/Time (Watt)
- Power*Delay: Efficiency to perform a task (Joule)
  - Often used to characterize efficiency at a technology node
- Energy*Delay: Power * Delay * Delay (Joule s)
  - Combined performance and energy metric
  - Figure of merit of design style

Where is Power Dissipated in CMOS?

- Dynamic (Active) Power
  - Charging and discharging capacitors
  - Temporary glitches (dynamic hazards)
  - Short-circuit (pull-up and pull-down ON during transition)

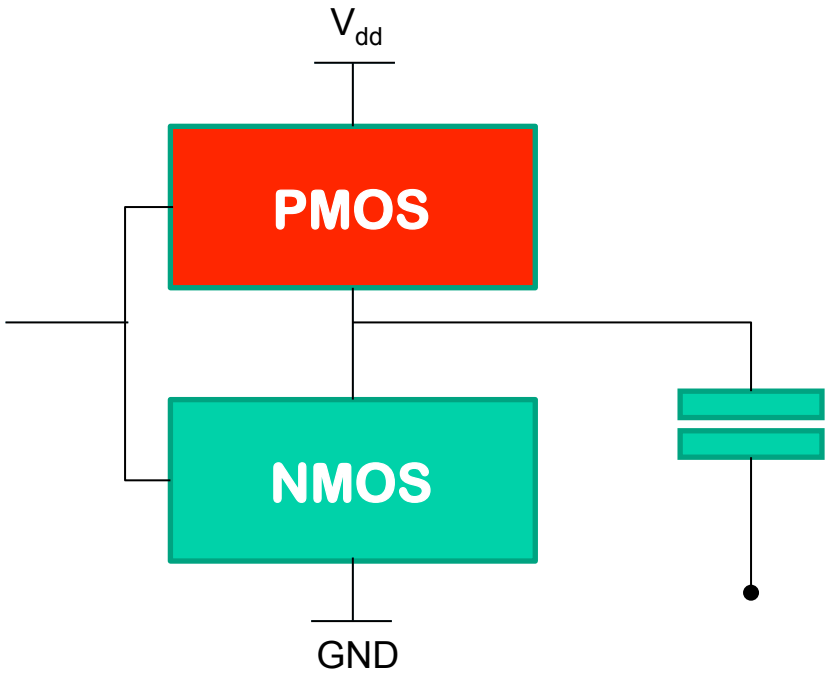
- Static (Leakage) Power
  - Transistors are imperfect switches
  - Drain leakage
  - Junction leakage (gate induced drain leakage)
  - Gate leakage (tunneling currents through gate oxide)

Active Power and Heat (1)

- To charge a capacitance C by applying a voltage V, an amount of energy equal to CV^2 is taken from the supply.
- Half of the energy is stored in the capacitor, and the other half is dissipated as heat in the resistance of charging the network.
- During discharge the stored energy is turned into heat as well.

Active Power and Heat (2)

- One half of the power from the supply is consumed in the pull-up network and one half is stored in C_{load}
- Charge in C_{load} is dumped during the 1->0 transition



Active Power Formula

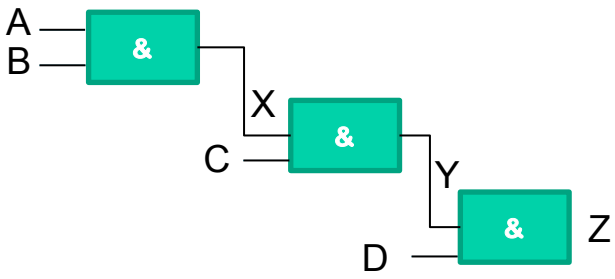
- Power = Energy/transition * Transition Rate

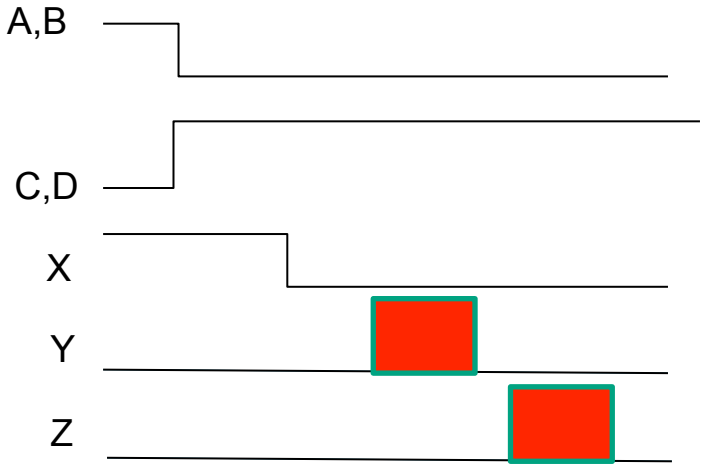
$$C_{load} * (V_{dd})^2 * f_{0->1}$$

- Power dissipation is data dependent – depends on the probability of switching from 0 -> 1 (activity factor): Hard to estimate, varies with gate
  - 2-input NAND/NOR = 3/16
  - 2-input XOR gate = 1/4

- Switched capacitance
  - $C_{switched} = P_{0->1} * C_{load}$
  - Also known as switching activity load

Glitching in Static CMOS





- Uneven arrival times of input signals due to unbalanced delay paths
- Intermediate charging and discharging of capacitors

- Fix: Build balanced paths

Short-circuit Currents

- For short periods of time Pull-up and Pull-down circuits are simultaneously on
  - Transition from 0->1 has a slope in real life
- Causes current to flow from V_{dd} to GND
- Proportional to switching activity

Static (Leakage) Power

- Drain Leakage
  - Main cause of concern
  - Diffusion current in sub-threshold region
  - Threshold voltage needs to be lowered with supply voltage is lowered
  - $P_{leak} = V_{dd} * I_{leak}$
- Junction Leakage
  - Gate-Induced drain leakage
- Gate Leakage
  - Tunneling currents through thin oxide

Summary: Power Dissipation Sources

- Switching activity * (load + short circuit) capacitance * voltage swing * voltage*frequency + (leakage current * voltage)

Topics

- ~~Power/Energy Basics~~
- Reducing Standby Power
- Reducing Dynamic Power
- Low Power Design Methodology
- Modern Microprocessor Power Supply

Standby Power

- Power consumed with no computational activity
  - Many apps are "bursty" e.g., cell phone usage
  - Power dissipation in standby should be absolutely minimum
- Reducing dynamic power in standby: Clock gating
  - Turn off clocks to idle modules
- Reducing static power in standby: Power gating

Clock Gating

- Motivation: Clocks consume a significant fraction of the total power (~ 30% for microprocessors)
  - Even if the circuit is idle the flip-flops load the clock which consumes power
- Clock gating
  1. Turn off clocks to idle modules
  2. Ensure inputs to idle logic are stable

- Implementation: AND the clock signal with an Enable

RECOMMENTATION: Include the enable signal, but do not manually AND the clock and enable signals. (Clocking becomes difficult). Let the tools handle this.

Advanced Standby (sleep) Modes

- Gate the clock to idle module
  - Some leaves in the clock tree are inactive
- Disable the clock distribution network
  - All leaves and wires are inactive but the root (clk generator) is still active
- Turn off the clock driver and the phase locked loops that generate the clock signal
- Turn off the clock completely
  - Only the wakeup circuit is active

Leakage Challenge

- With clock gating, leakage power becomes the dominant standby power source

  ⇒ Leakage should be minimized

- Challenge: how to disable a unit most effectively given that the transistors are no ideal switches.

Power Gating & Supply Voltage Shutoff

- Power gating
  - Disconnect modules from the supply rail(s) during standby
  - How to do this when we do not have a perfect switch?
  - Use a different V_t transistor for putting the circuits to sleep
    - Called MT-CMOS (multi-threshold)
  - Can impact circuit operation time (charging and discharging)
  - Also need to preserve state: usually only logic is gated

- But, ideally want to turn off the voltage to zero
  - Supply voltage should be ramped down to zero
  - Need a controllable voltage regulator
  - Challenge: integrating voltage regulators on die
  - Con: Slow activation time after power-down
  - Possibly huge gains

Topics

- ~~Power/Energy Basics~~
- ~~Reducing Standby Power~~
- Reducing Dynamic Power
- Low Power Design Methodology
- Power Delivery Architecture

Runtime Optimizations for Power

- Power dissipation is a strong function of activity
- Activity varies over time (e.g., web page browsing)
- Choosing one operational point is sub-optimal

- Changing the operational point:What can we change?
  - V_{dd}: Supply voltage
  - V_{th}: Threshold voltage (not considered in this class)
  - f: Frequency

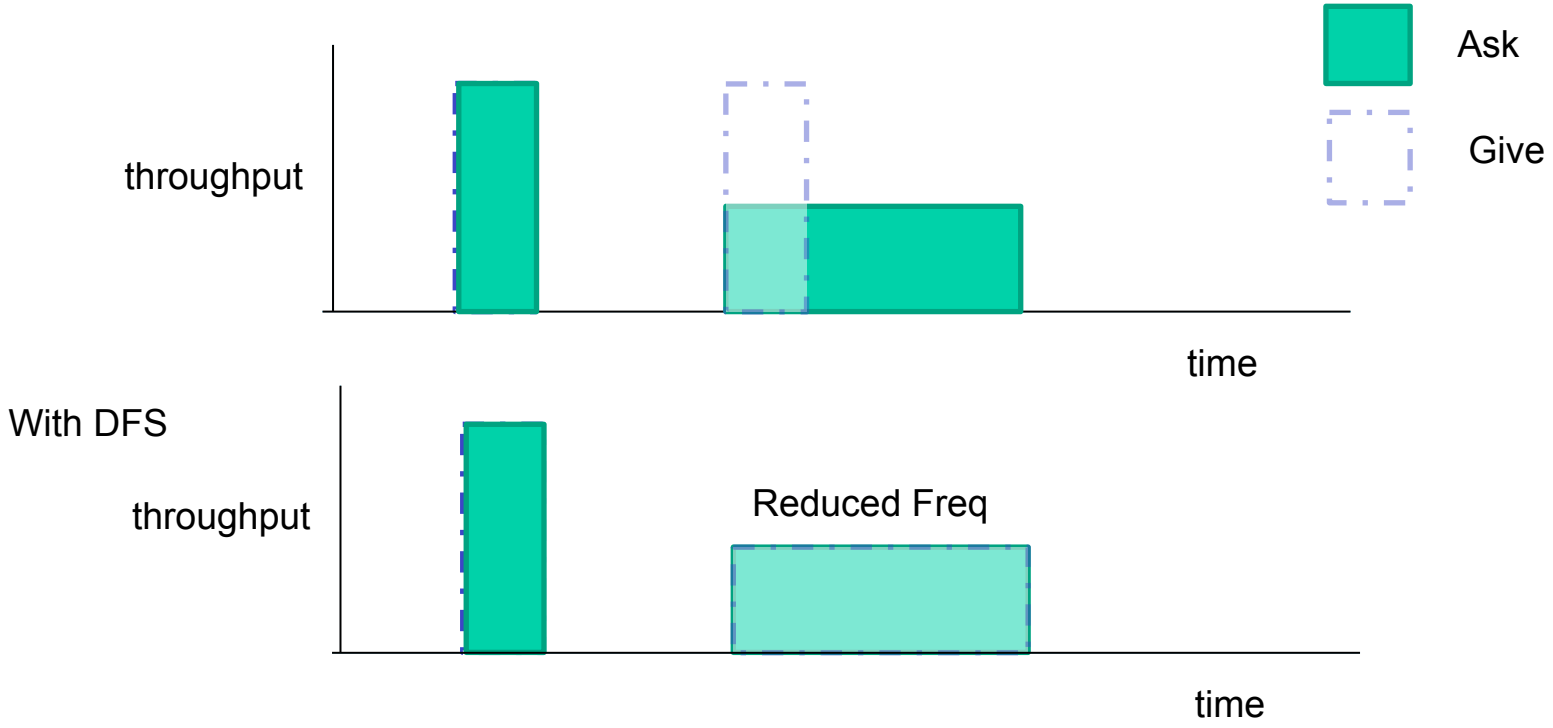
Two techniques:

- Dynamic Frequency Scaling
- Dynamic Voltage and Frequency Scaling

Dynamic Frequency Scaling

- **Dial down frequency for slow tasks**
- **Only reduces power – leaves energy/op unchanged**

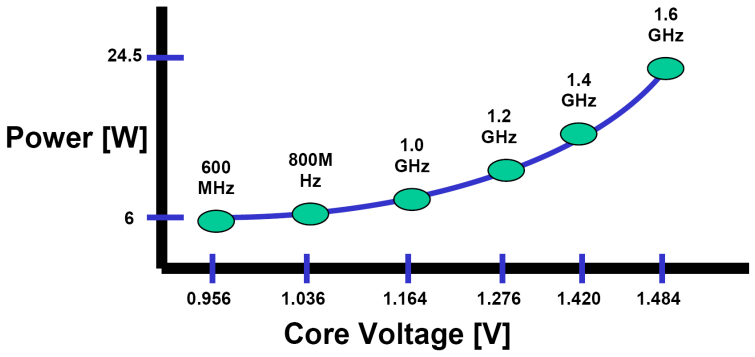
  ⇒ **Battery life does not change**



Dynamic Voltage & Frequency Scaling

- Vary V_{dd} and F to based on throughput "ask"
- Minimizes energy and power
- Relationship between V_{dd} and F
  - $F = (v - v_t)/(1 - v_t)^{\alpha} * (1/v)$
  - For alpha = 2 and $V_{dd} >> V_{th}$; Frequency and Voltage are linear
  - Cubic reductions when speed is reduced!
- Most Microprocessors switch between discrete preselected voltage/frequencies

Frequency	Voltage	P-State
1.6 GHz	1.484 V	P0
1.4 GHz	1.420 V	P1
1.2 GHz	1.276 V	P2
1.0 GHz	1.164 V	P3
800 MHz	1.036 V	P4
600 MHz	0.956 V	P5



Pentium – M Data (from 04)

Challenges

- Estimating workloads (architectural prediction)
- Multiple voltage domain generation and supply
  - Voltage regulator efficiency, integration on-die
  - Reliable distribution
  - Interface circuits between voltage domains
  - Switching between power modes
  - How to convert a “request for perf” to Voltage/freq?
- Validation: Need to verify at every voltage point?
  - Functionality, Timing, Integrity?
  - Luckily does not matter for Static CMOS; think inverters
  - No non-linearities
  - RECOMMENDATION: Use Static CMOS

Topics

- ~~Power/Energy Basics~~
- ~~Reducing Standby Power~~
- ~~Reducing Dynamic Power~~
- Low Power Design Methodology
- Power Delivery Architecture

Low Power Methodology

- Motivation minimize power, time and effort

Design

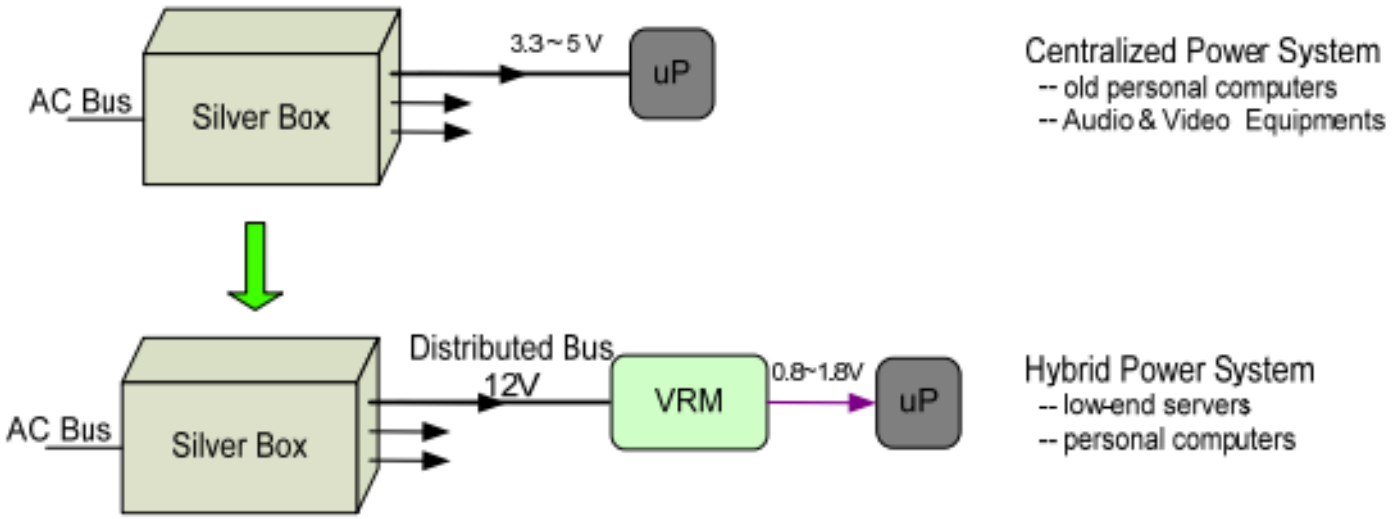
1. Explore architecture and algorithms for pwr. Eff.
2. Map functions to s/w or h/w blocks for pwr. Eff.
3. Choose voltages and frequencies
4. Evaluate power consumption for different ops
5. Generate budgets for power/perf/area
6. Implement RTL
7. Validate power reported against power budget
8. Iterate until convergence

Topics

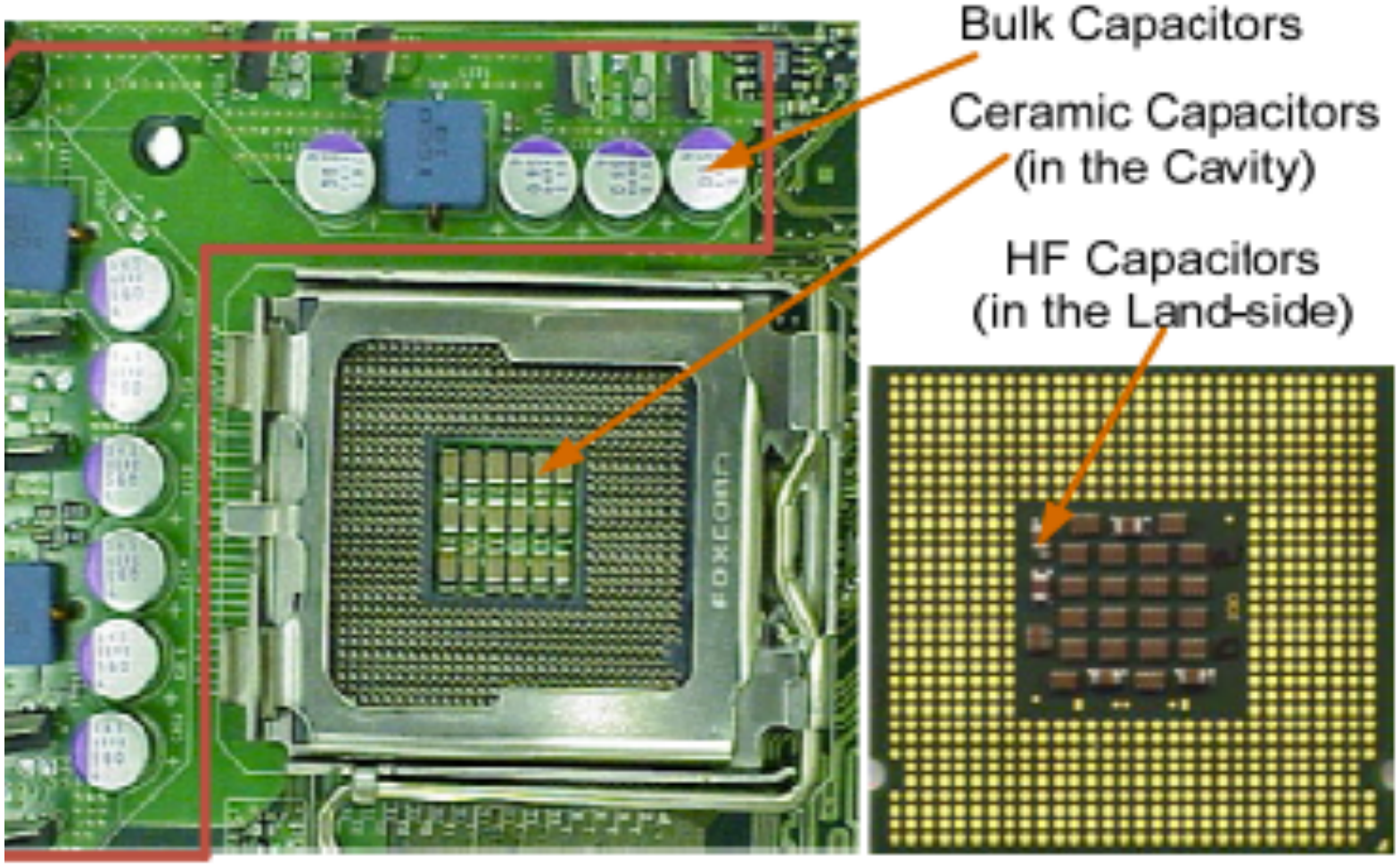
- ~~Power/Energy Basics~~
- ~~Reducing Standby Power~~
- ~~Reducing Dynamic Power~~
- ~~Low Power Design Methodology~~
- Power Delivery Architecture

Ref: HIGH PERFORMANCE DIGITAL CONTROL TECHNIQUES FOR POWERING MICROPROCESSORS by SHANGZHI PAN

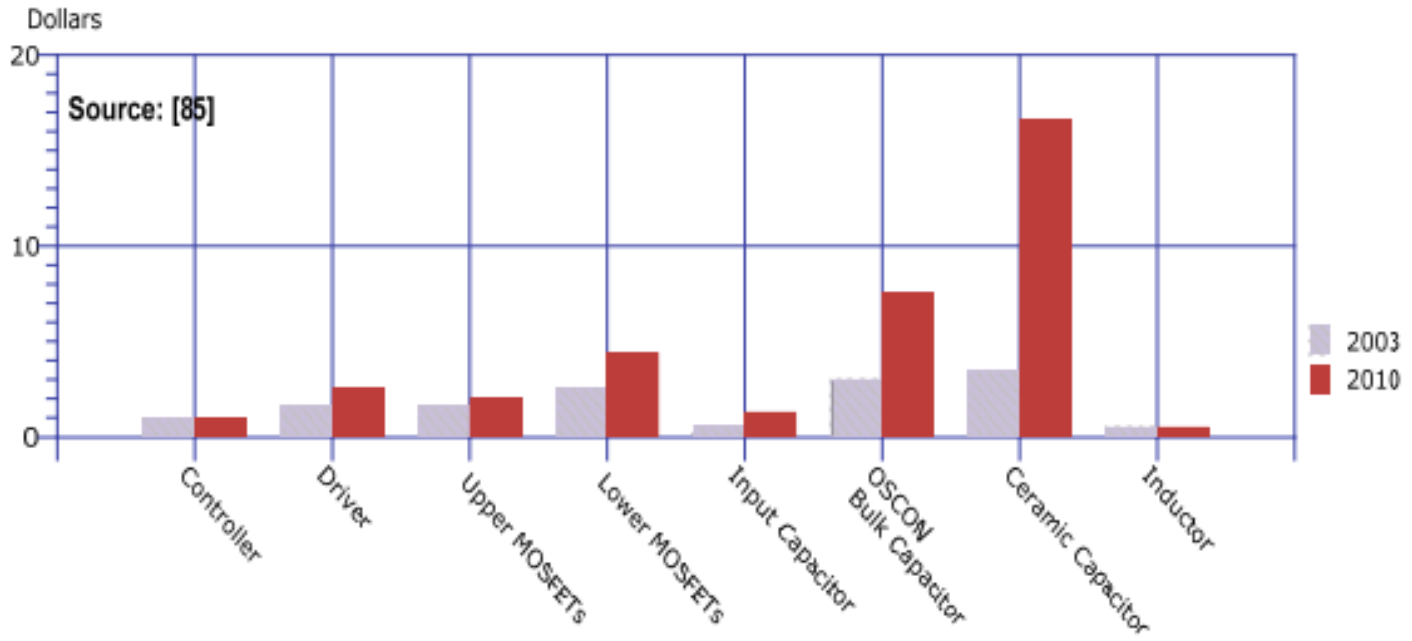
Power Delivery Architecture



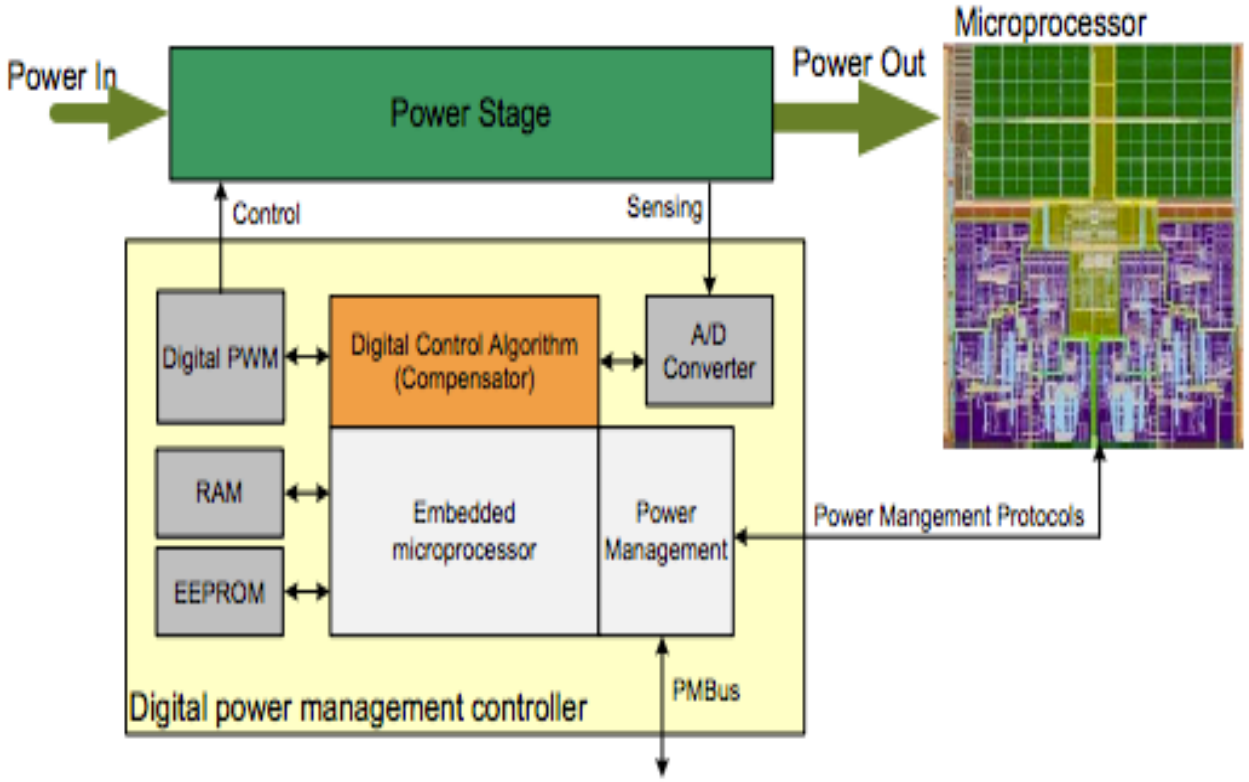
Decoupling Caps



Motherboard Cost Breakdown

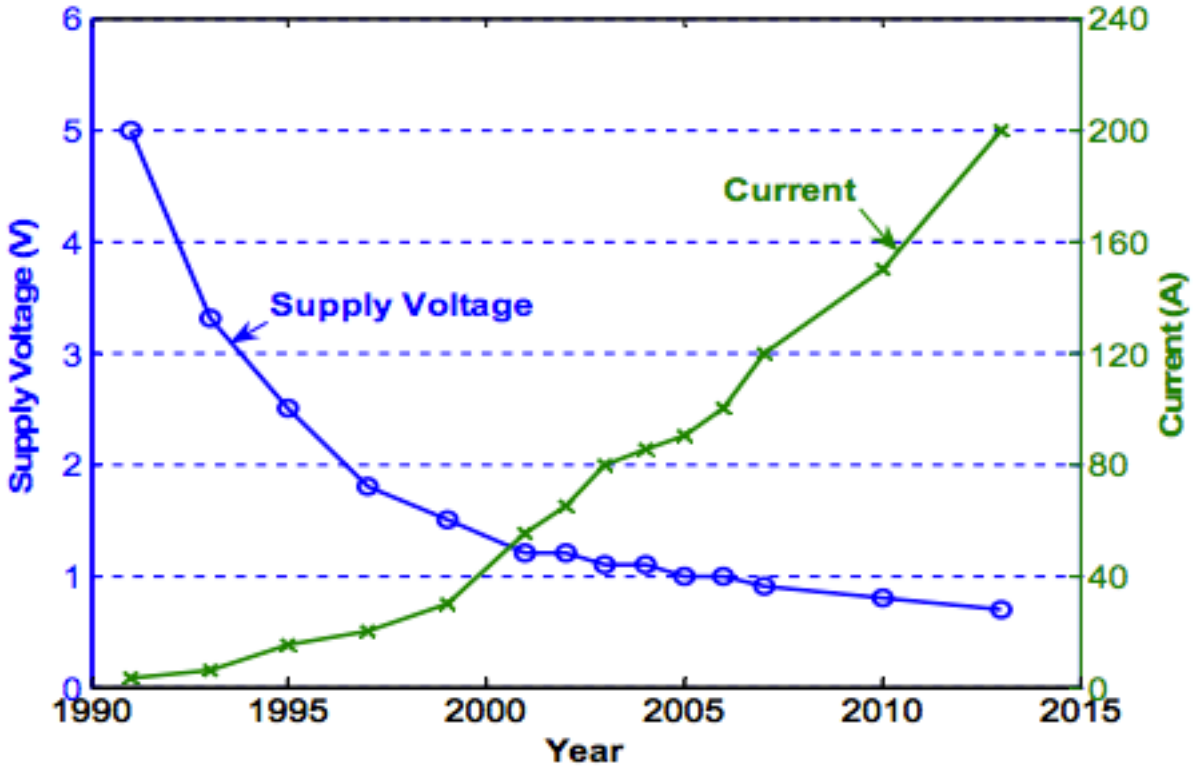


Typical Power Mgmt. Controller



Outlook

- **Scaling Trends**



Possibilities: Analog and Digital Accelerators