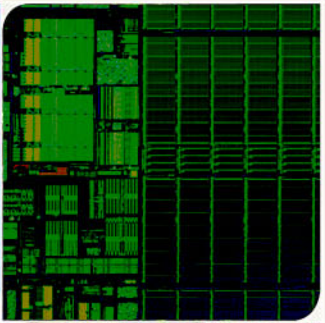
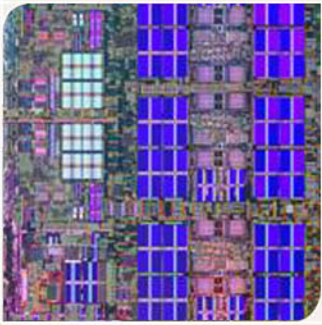
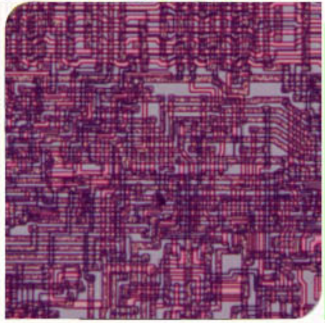
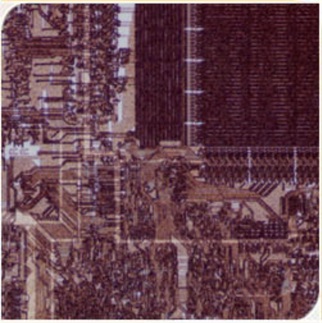





CS4290/CS6290

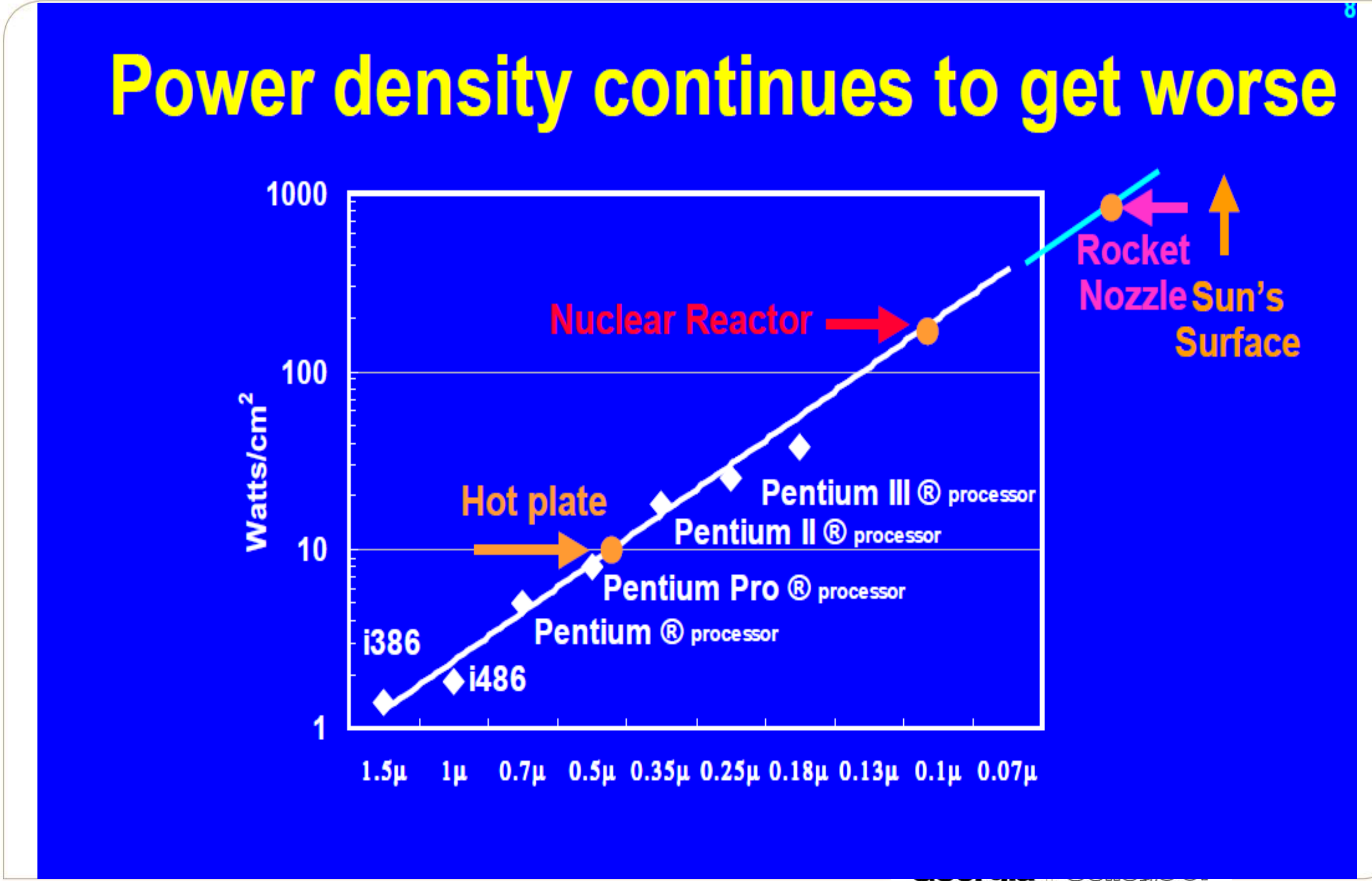
Fall 2011

Prof. Hyesoon Kim



Georgia Tech College of Computing

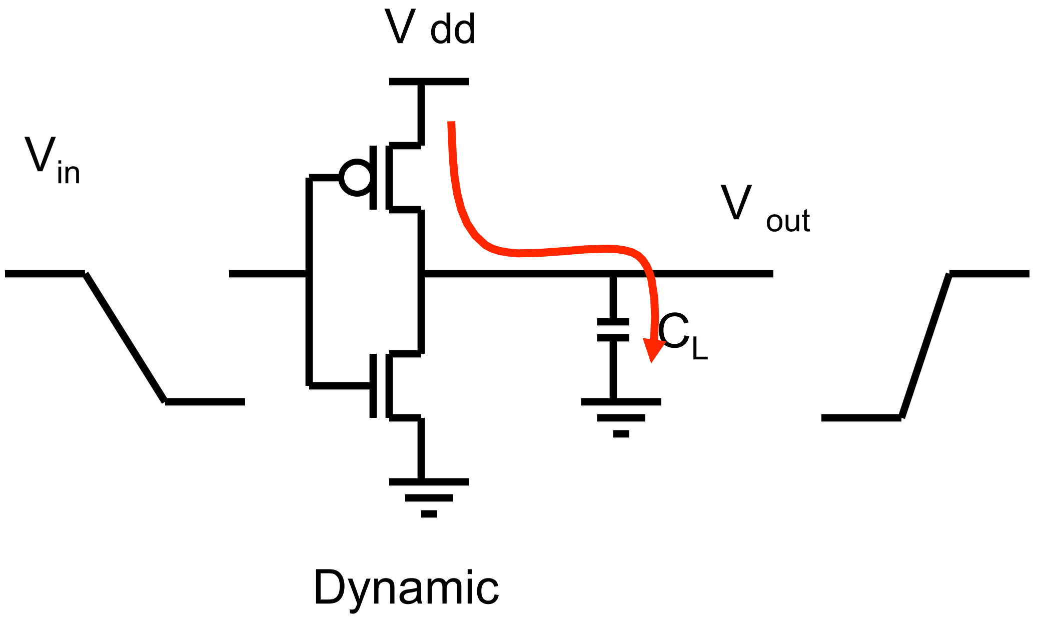
Why Power?

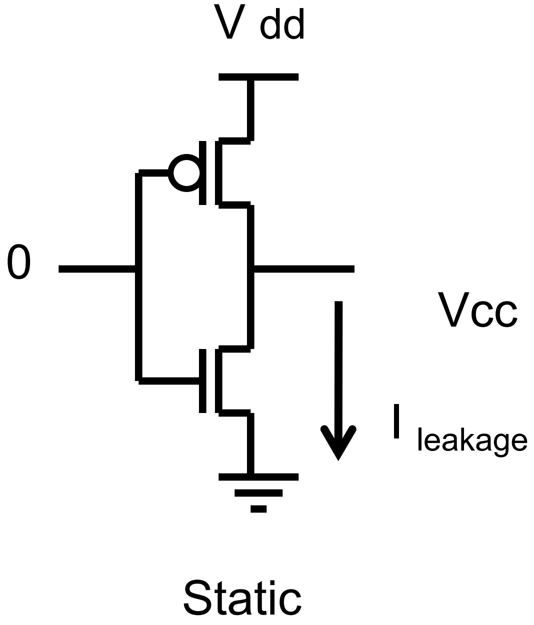


Power density continues to get worse

Fred Pollack, Intel, Micro-32 Keynote

Power Dissipation in CMOS





$$P_{tot} = P_{dyn} + P_{sta} = C_L V_{dd}^2 f + V_{dd} I_{leak}$$

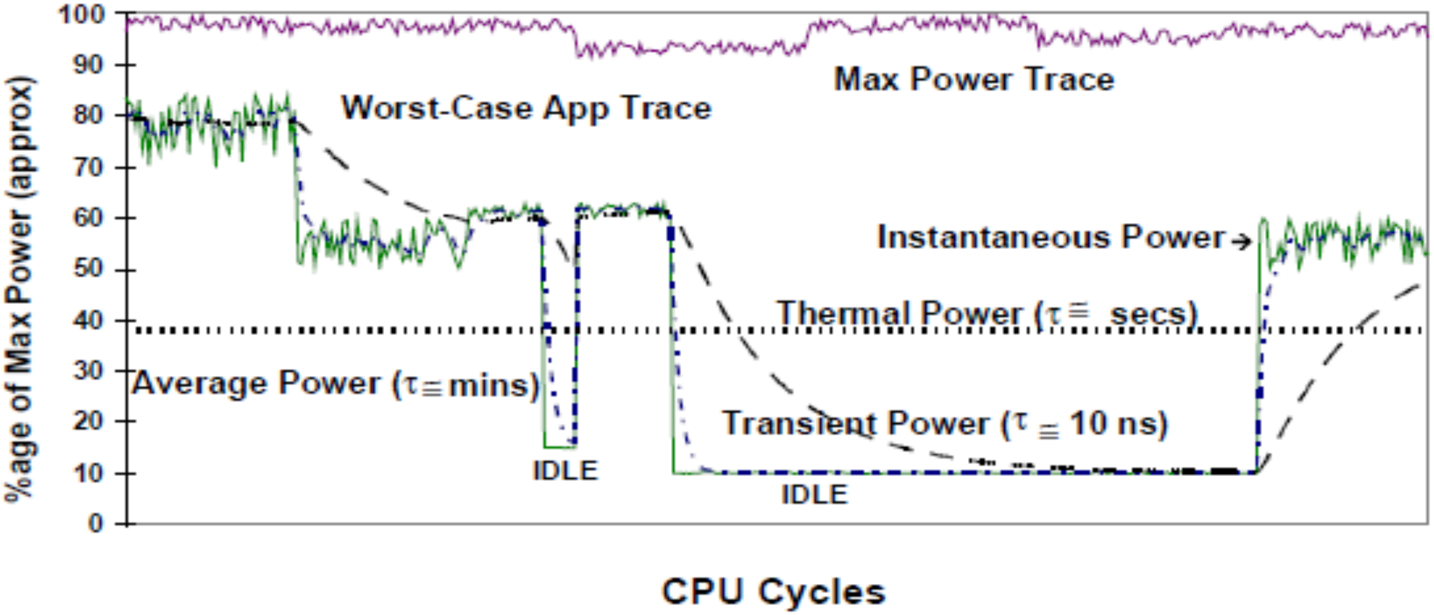
Power Basics

- Power vs. Energy
- Dynamic power vs. Static power
  - Dynamic: “switching” power
  - Static: “leakage” power
  - Dynamic power dominates, but static power is increasingly important

Why do we care?

- 1) Increase the CPU cost
  - Thermal cost: keeping the devices below the special temperature
- 2) the cost of power delivery

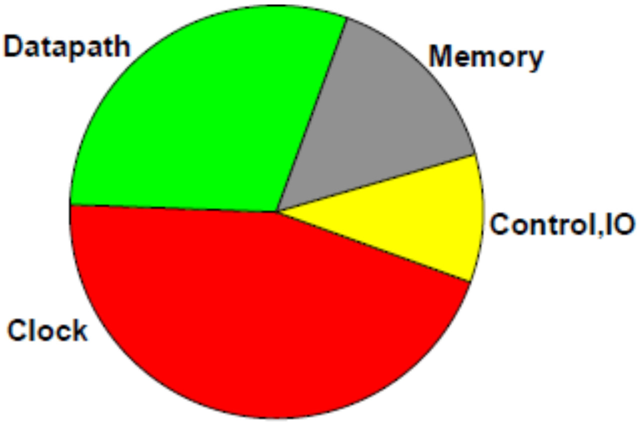
Power power power



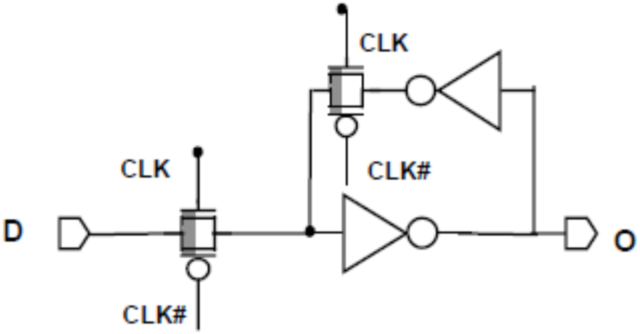
- Max Power
- Worst-case application
- Average power
- Thermal power: running average of worst-case app for several seconds : used to decide cooling option
- Transient power (power delivery) , standby power (battery life),

Reducing Power in High-performance Microprocessors, Tiwari et al. ‘98

Where does Power Go?



Clock power consumption is high
Even when AF is 0.5



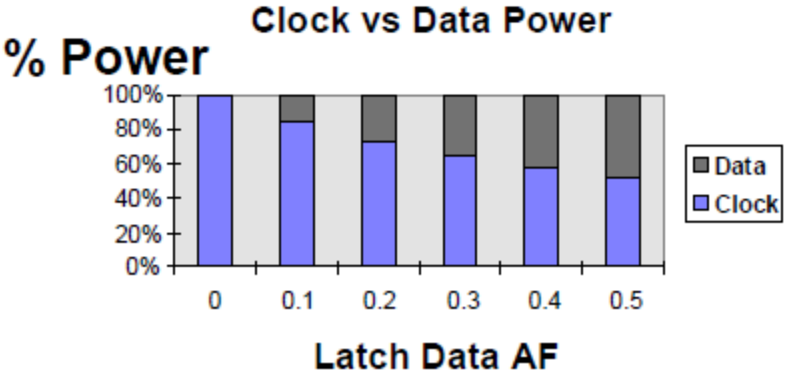


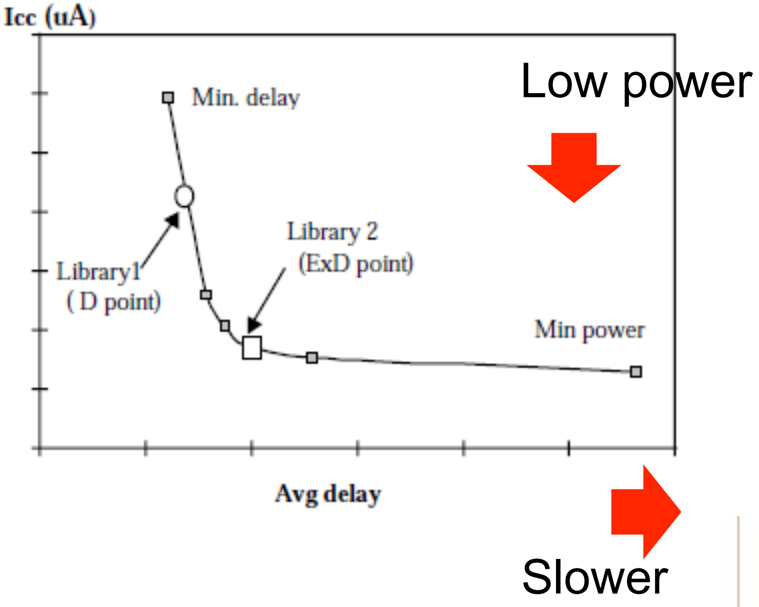
Figure 5: Power consumption in latches

AF (activity factor)

Reducing Power in High-performance Microprocessors, Tiwari et al. ‘98

Power Reduction Techniques

- Voltage scaling
- Clock gating
- Utilize circuit design techniques
- Low power logic synthesis
  - Non-critical path→ low power circuit (slow but so what?)
- Specific circuit technology
  - Reduce AF (domino circuit)



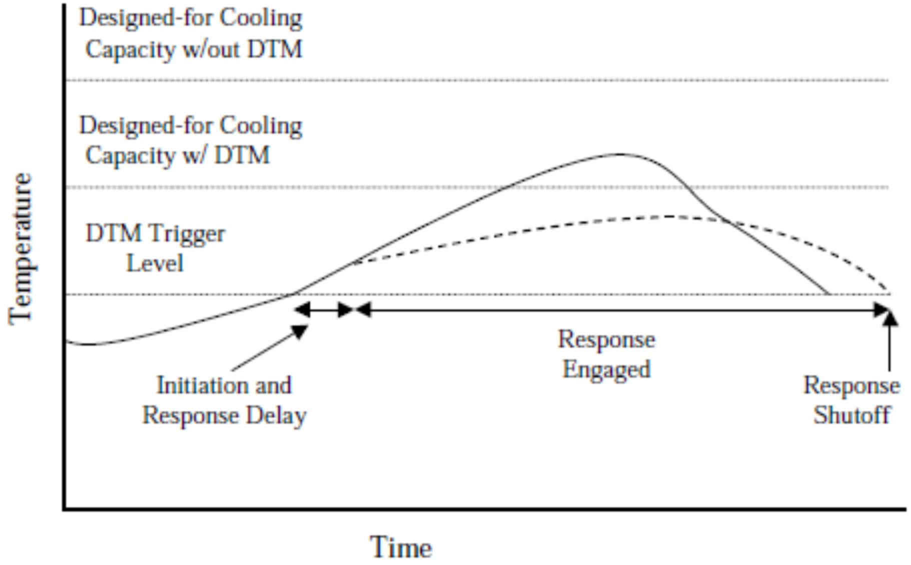
Dynamic Thermal Management for High-Performance Microprocessors, Brooks and Martonosi (01)

DVS (Dynamic Voltage Scaling)

- O/S controls the processor speed: find the minimum voltage required for the desired speed.
- DVFS (Dynamic Voltage Frequency Scaling): Intel's CPU throttling technology, SpeedStep

DTM (Dynamic Thermal Management) Techniques

- DTM: software and hardware techniques at run-time to control a chip's operating temperature
- Thermal package is designed for normal operating conditions rather than worst case
- Key goals:
  - To provide inexpensive hardware/software responses
  - Reduce power
  - Reduce impacts on performance as little as possible


Designed-for Cooling Capacity w/out DTM
Designed-for Cooling Capacity w/ DTM
DTM Trigger Level
Temperature
Initiation and Response Delay
Response Engaged
Response Shutoff
Time

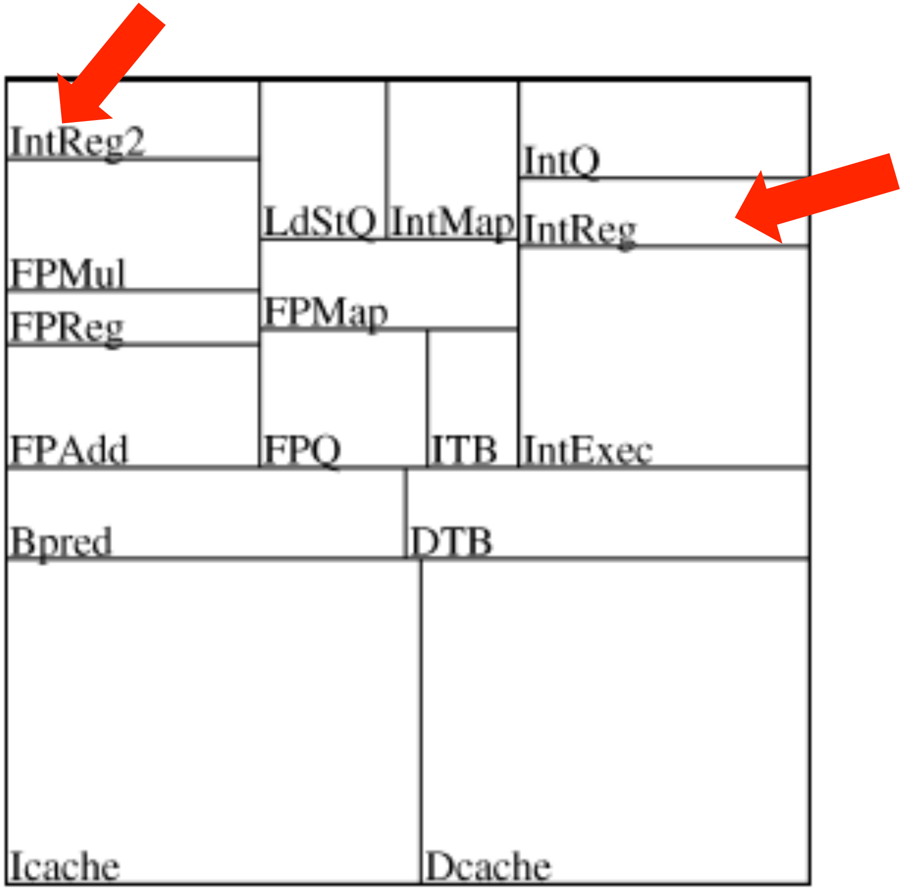
Dynamic Thermal Management for High-Performance Microprocessors, Brooks and Martonosi (01)

DTM

- Trigger mechanisms
  - Temperature sensors, on-chip activity counters
- DTM Response mechanisms
  - Clock frequency scaling
  - Voltage and frequency scaling
  - Decode throttling (PowerPC G3)
  - Speculation control
  - I-cache toggling (disabling instruction cache)
  - Migration computation

Dynamic Thermal Management for High-Performance Microprocessors,
Brooks and Martonosi (01), Temperature-Aware Microarchitecture:
Modeling and Implementation, skadron et al. '04

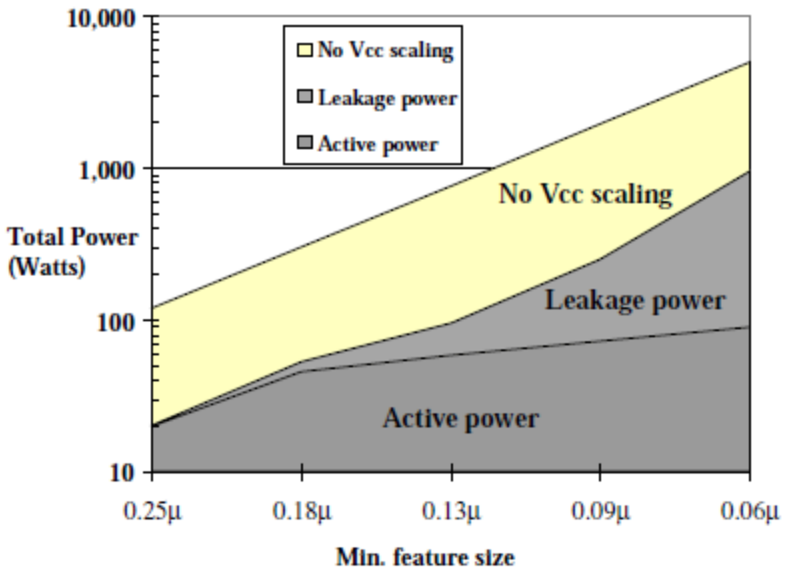
DTM: Migration Computation

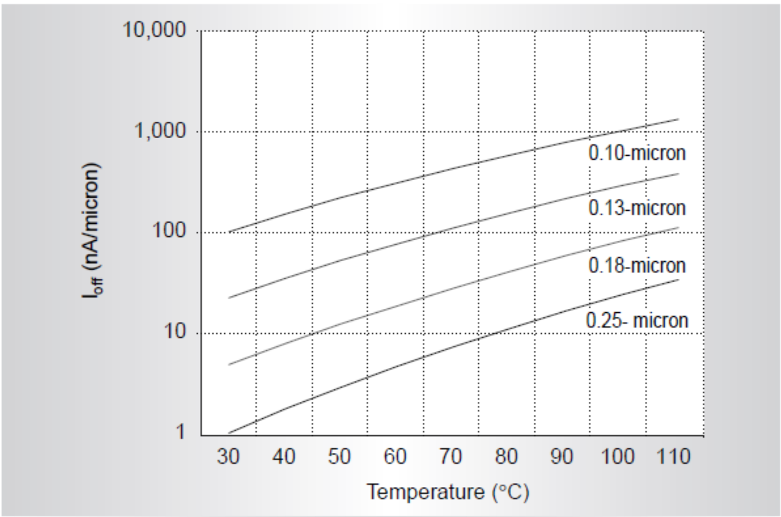


- Spare unit is located in cold area of chip
- Primary unit reaches 81.6C, issue is stalled, instructions ready to write back is allowed.
- All instructions use second register file.
- When the primary register file reaches 81.5C the process is reversed

Temperature-Aware Microarchitecture: Modeling and Implementation, Skadron et al. '04

Leakage Power Trend





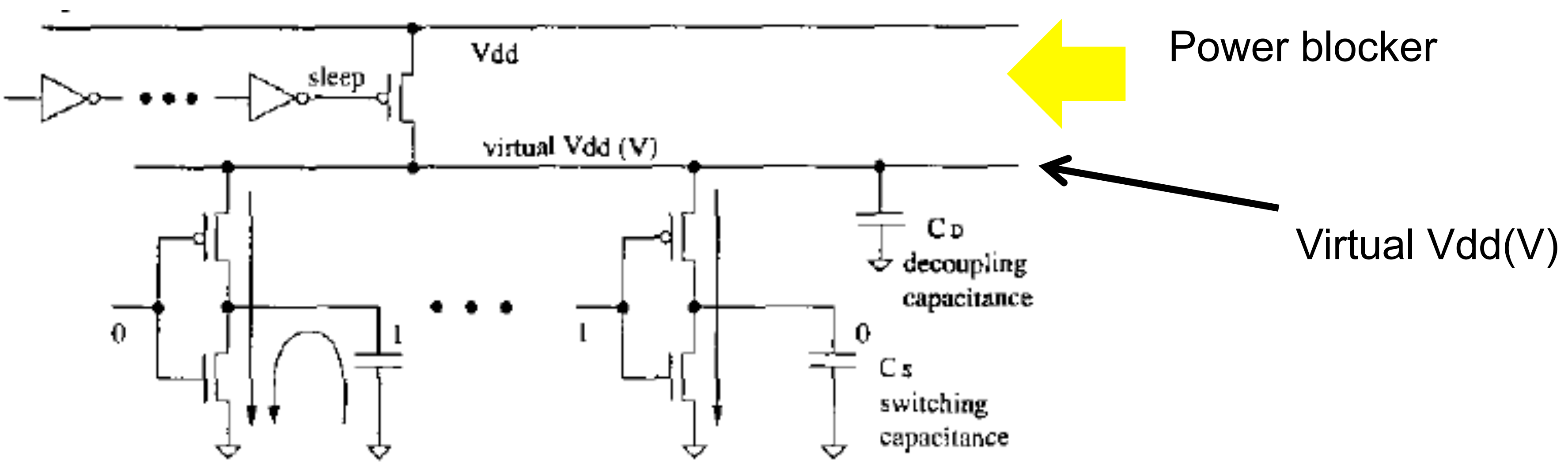
- Technology scales, leakage power consumption is increased
- Leakage power/current increases as temperature increases

DESIGN CHALLENGES OF TECHNOLOGY SCALING, Borkar'99

Leakage Power Control Techniques

- Body-bias control
- Dual-threshold domino circuits
- Input vector control (by inputting all 0's for a NAND gate)
- Power gating

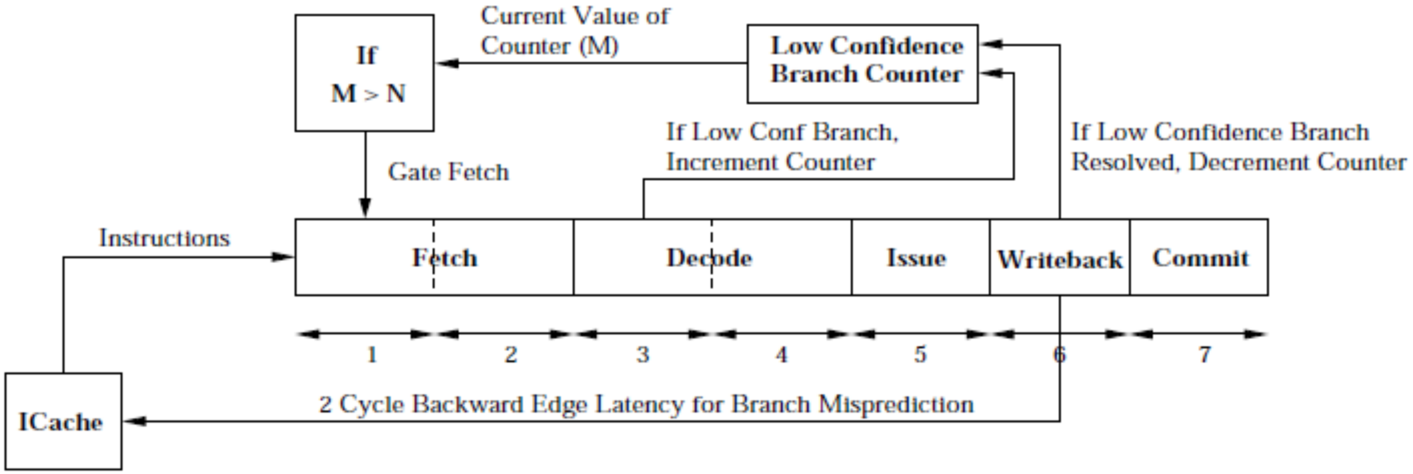
Power Gating



- Sleep signal to turn off the supply voltage
- Save both dynamic power and leakage power

Microarchitectural Techniques for Power Gating of Execution Units, Hu et al.

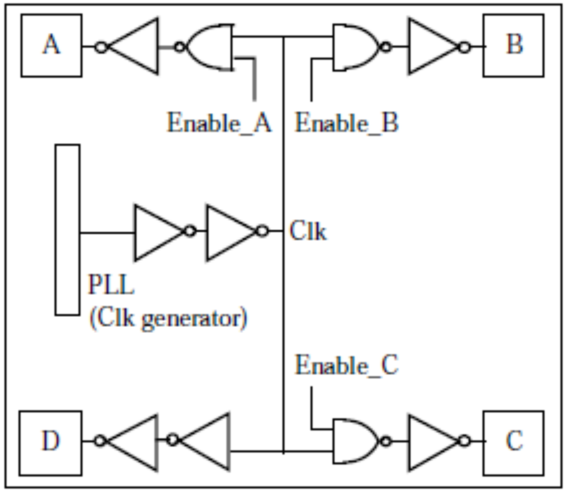
Pipeline Gating (Manne et al. '98)



- Determine when a branch is more likely mispredicted and gate a pipeline
- Use confidence estimator
- Other metrics (number of instructions)
- JVM

Clock Gating

- Adds additional logic to a circuit to prune the clock tree
- Reduce dynamic power consumption
- Power up delay (timing problem)
- Variations in current



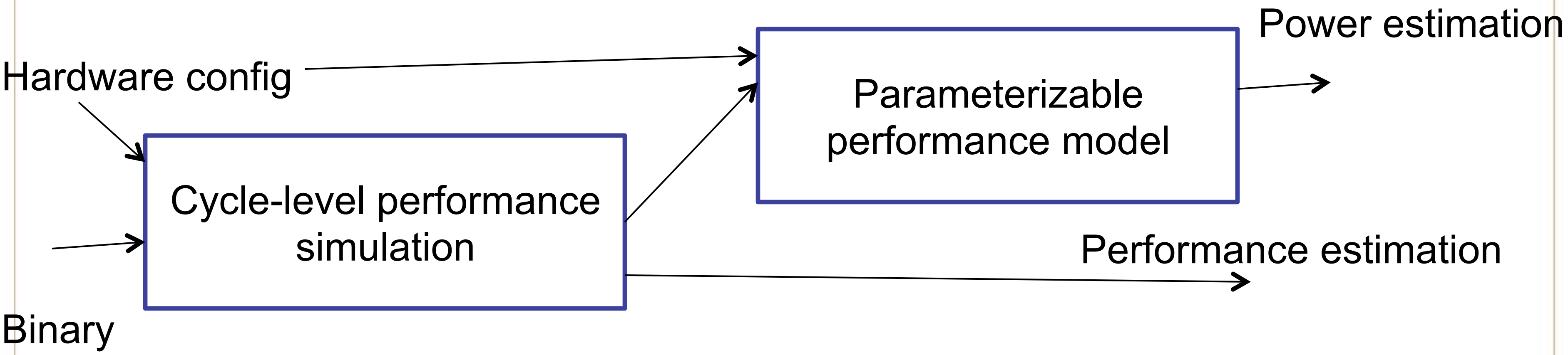
Benefits of Power Gating and Clock Gating

- Clock Gating → reduce dynamic power consumption but not the static power consumption
- Power gating → eliminate both dynamic and static power consumption

Architecture Power Simulators

- SimpleScalar (Performance simulator)
- Wattch : Dynamic Power Simulator
- HotLeakage: Leakage current simulator
- HotPower: thermal spot
- MacPat:

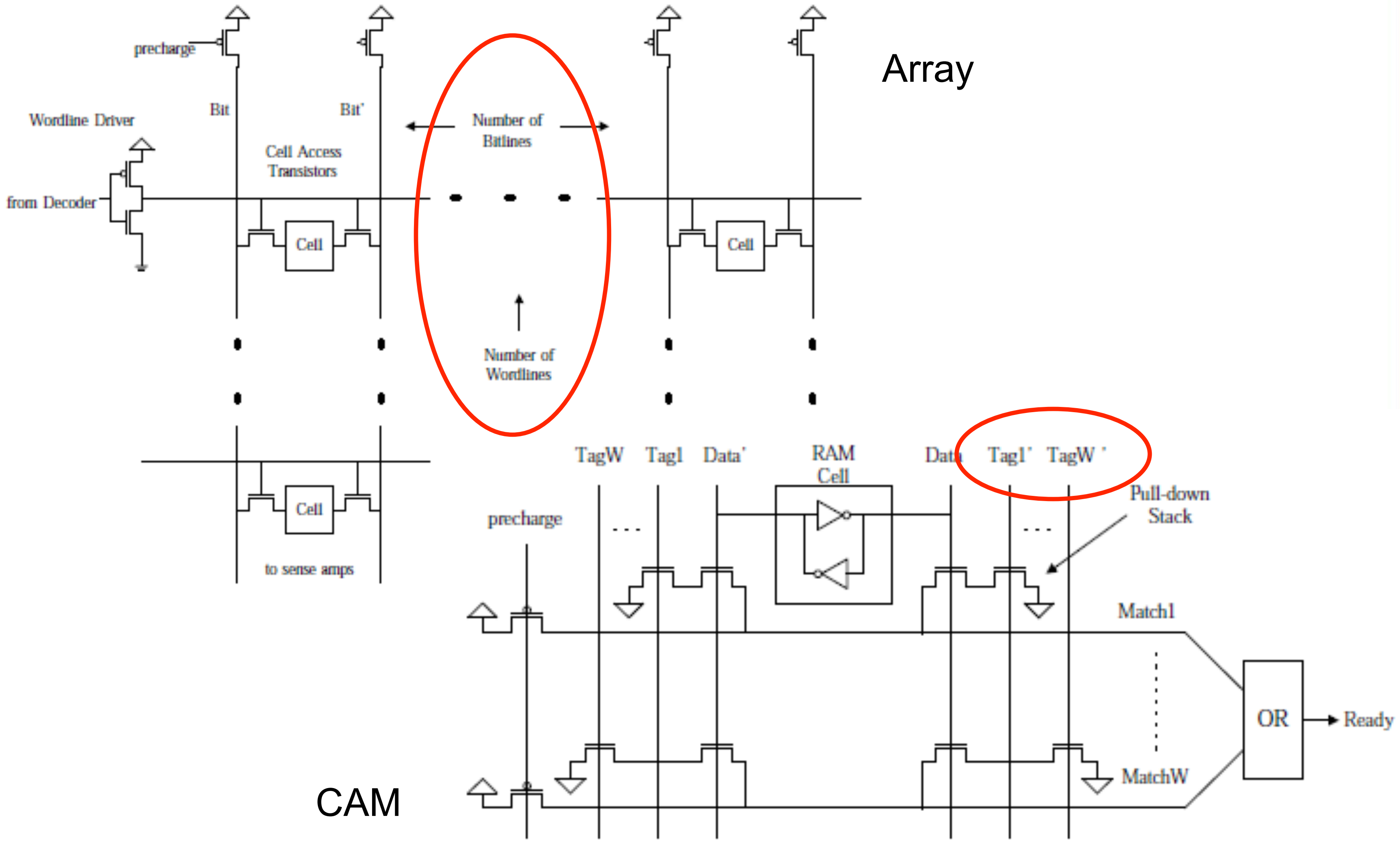
Wattch (Brooks et al. ‘00)


Power estimation
Hardware config
Parameterizable performance model
Cycle-level performance simulation
Performance estimation
Binary

$Pd = CvV_{dd}^2af$ C: load capacitance, Vdd: supply voltage, f: clock frequency
A: switching activity

- Array structures
- Fully associative content-addressable memories (CAM)
- Combinational logic and wires
- Clocking

Array structure vs. CAM



Array

CAM

Wattch (Brooks et al. '00)

Hardware Structures in Wattch

Hardware structures	Model Type
Instruction cache	
Wakeup logic	
Issue selection logic	
Instruction window	
Branch predictor	
Register file	
TLB	
Load/Store Queue	
Data Cache	
Integer functional units	
FP functional units	
Global clock	

Wattch (Brooks et al. ‘00)

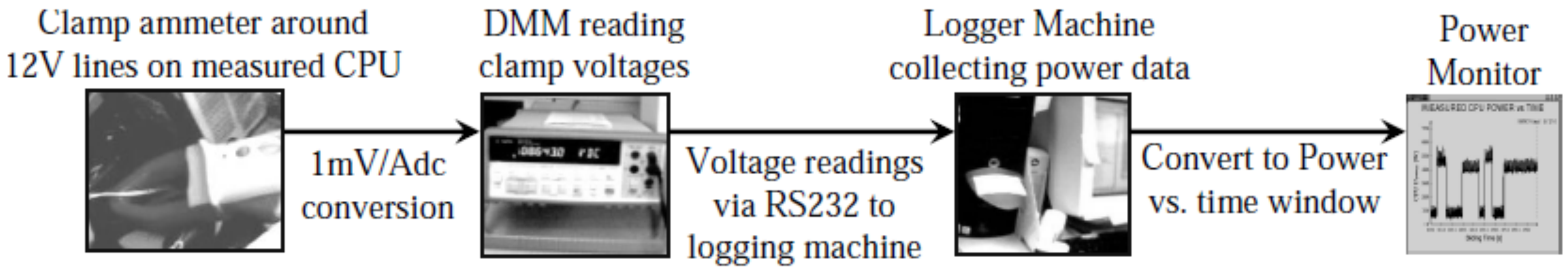
Metrics

- Power
- Energy
- EDP (Energy Delay Product)
- EDDP (Energy Delay^2 Product) : more emphasis on performance
- EPI (Energy per instructions)

Review: Performance vs. Power

- Cooling capacity also decides the maximum power
- Back-of-the-Envelope calculation:
  - 3.8 GHz CPU at 100W
  - Dual-core: 50W per CPU
  - $P \propto V^3$: $V_{orig}^3/V_{CMP}^3 = 100W/50W$ → $V_{CMP} = 0.8\ V_{orig}$
  - $f \propto V$: $f_{CMP} = 3.0GHz$

Runtime Power Monitoring

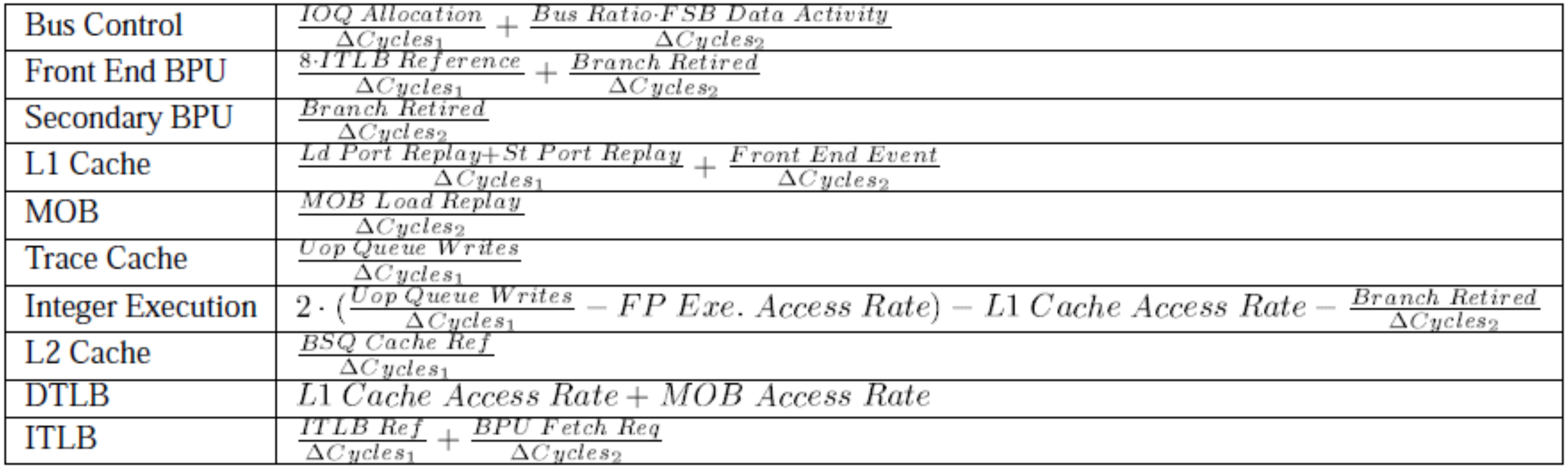

Clamp ammeter around 12V lines on measured CPU
1mV/Adc conversion
DMM reading clamp voltages
Voltage readings via RS232 to logging machine
Logger Machine collecting power data
Convert to Power vs. time window
Power Monitor

- Measuring current
- Real time power monitoring
- Power(Ci) = AccessRate(Ci)* architecturalScaling(Ci) * MaxPower(Ci) + NongatecClockPower(Ci)
- MaxPower→ proportional to area
- Accessrate → dynamic events
- MaxPower, architecturalScaling, NongatedClockPower: foundout from empirical data

Isci and Martonosoni'03

Examples of Processor Components

- Access Rate calculation
- Use hardware performance counters to get events



Bus Control	$\frac{IOQ\ Allocation}{\Delta Cycles_1} + \frac{Bus\ Ratio \cdot FSB\ Data\ Activity}{\Delta Cycles_2}$
Front End BPU	$\frac{8 \cdot ITLB\ Reference}{\Delta Cycles_1} + \frac{Branch\ Retired}{\Delta Cycles_2}$
Secondary BPU	$\frac{Branch\ Retired}{\Delta Cycles_2}$
L1 Cache	$\frac{Ld\ Port\ Replay + St\ Port\ Replay}{\Delta Cycles_1} + \frac{Front\ End\ Event}{\Delta Cycles_2}$
MOB	$\frac{MOB\ Load\ Replay}{\Delta Cycles_2}$
Trace Cache	$\frac{Uop\ Queue\ Writes}{\Delta Cycles_1}$
Integer Execution	$2 \cdot (\frac{Uop\ Queue\ Writes}{\Delta Cycles_1} - FP\ Exe.\ Access\ Rate) - L1\ Cache\ Access\ Rate - \frac{Branch\ Retired}{\Delta Cycles_2}$
L2 Cache	$\frac{BSQ\ Cache\ Ref}{\Delta Cycles_1}$
DTLB	$L1\ Cache\ Access\ Rate + MOB\ Access\ Rate$
ITLB	$\frac{ITLB\ Ref}{\Delta Cycles_1} + \frac{BPU\ Fetch\ Req}{\Delta Cycles_2}$

- Use train benchmarks to stress particular units
- Pentium 4 based design
- Different architectures have different units

Isci and Martonosoni'03

Max Power Calculation

$$Runtime\_power\_component = AccessRate \times \boxed{MaxPower}$$

- **Allowable** maximum power consumption per arch. unit

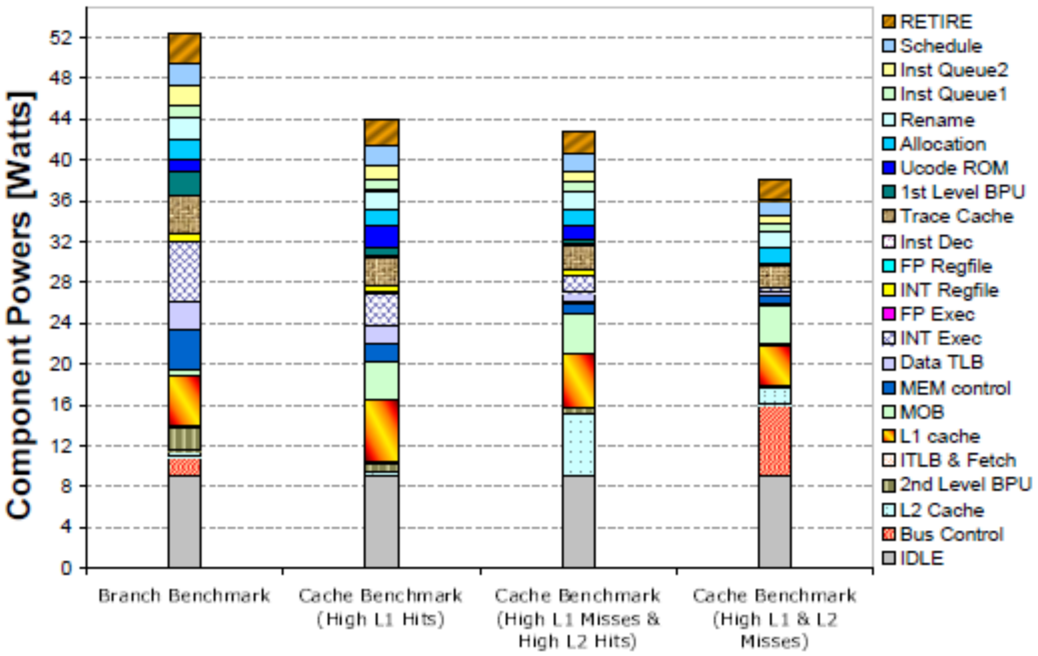
Training benchmarks

Bus control
Branch predictor
L1 cache
MOB (MSHR)
Trace cache
Integer Unit
L2 cache

$$\underbrace{\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ . \\ . \\ . \\ P_n \end{bmatrix}}_{\textbf{Measured Power}} = \underbrace{\begin{bmatrix} AR_{1A} & AR_{1B} & \ldots & AR_{1Z} \\ AR_{2A} & AR_{2B} & \ldots & AR_{2Z} \\ AR_{3A} & AR_{3B} & \ldots & AR_{3Z} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ AR_{nA} & AR_{nB} & \ldots & AR_{nZ} \end{bmatrix}}_{\textbf{AccessRate}} \mathrm{X} \underbrace{\begin{bmatrix} M_A \\ M_B \\ M_C \\ . \\ . \\ . \\ M_Z \end{bmatrix}}_{\textbf{Maximum power per arch. unit}}$$

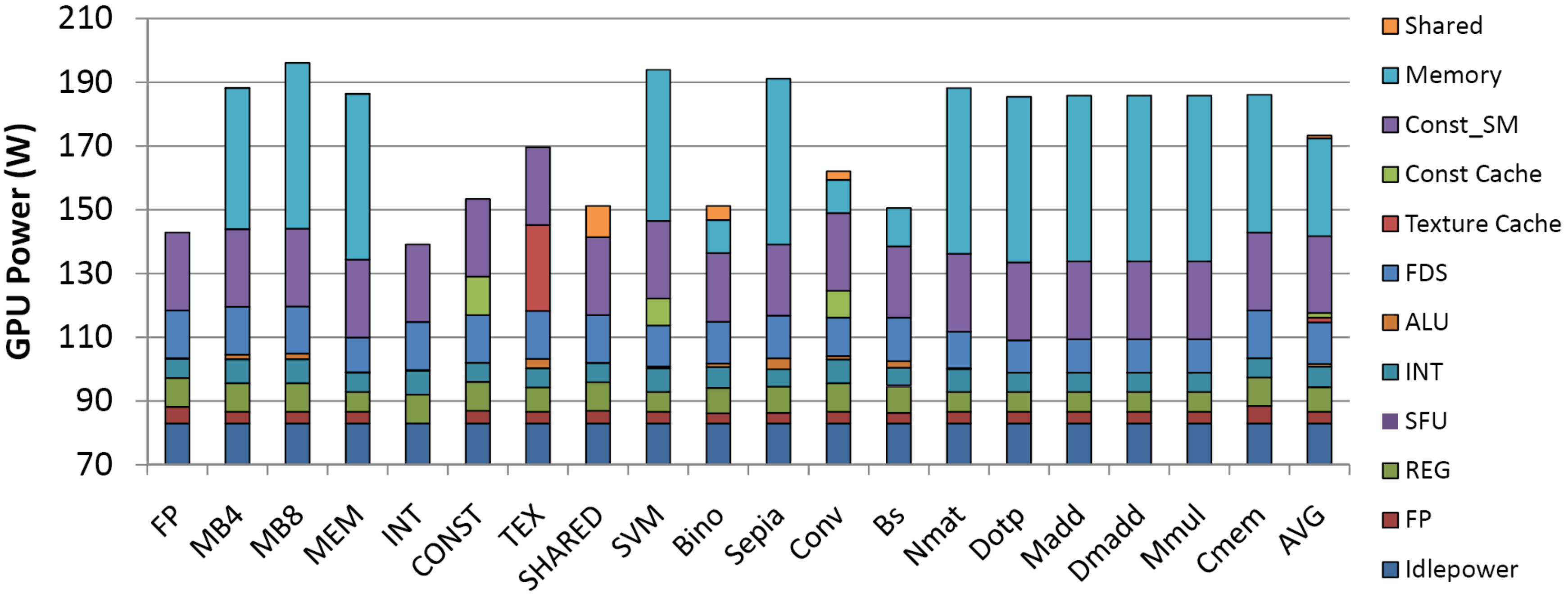
Solve for the set of ***MaxPower*** values (M_A, M_B, … , M_Z)

Power Breakdowns



Isci and Martonosoni'03

GPU Power Breakdown



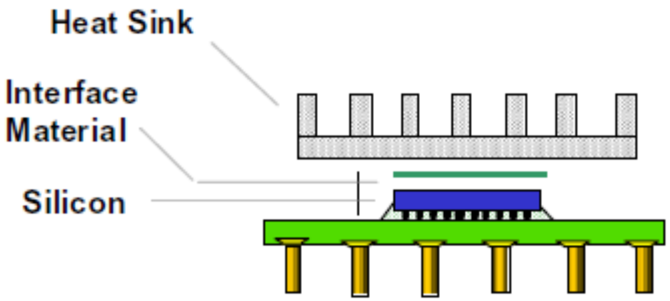
Hong and Kim '10

Georgia Tech College of Computing

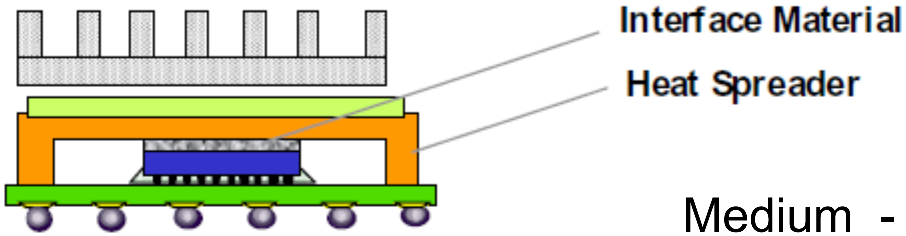
Hardware Performance Counters

- Built in counters inside hardware
- Example counters
  – Branch misprediction, cache misses, retired instructions, pipeline bubbles, DRAM traffics
  – 10s (even 100s) of even counters, but typically only few can be read simultaneously
- Software
  – Typically windows/Linux (Linux requires kernel recompilation)
  – PAPI, PerfMon, Vtune, etc.

Heat Sink



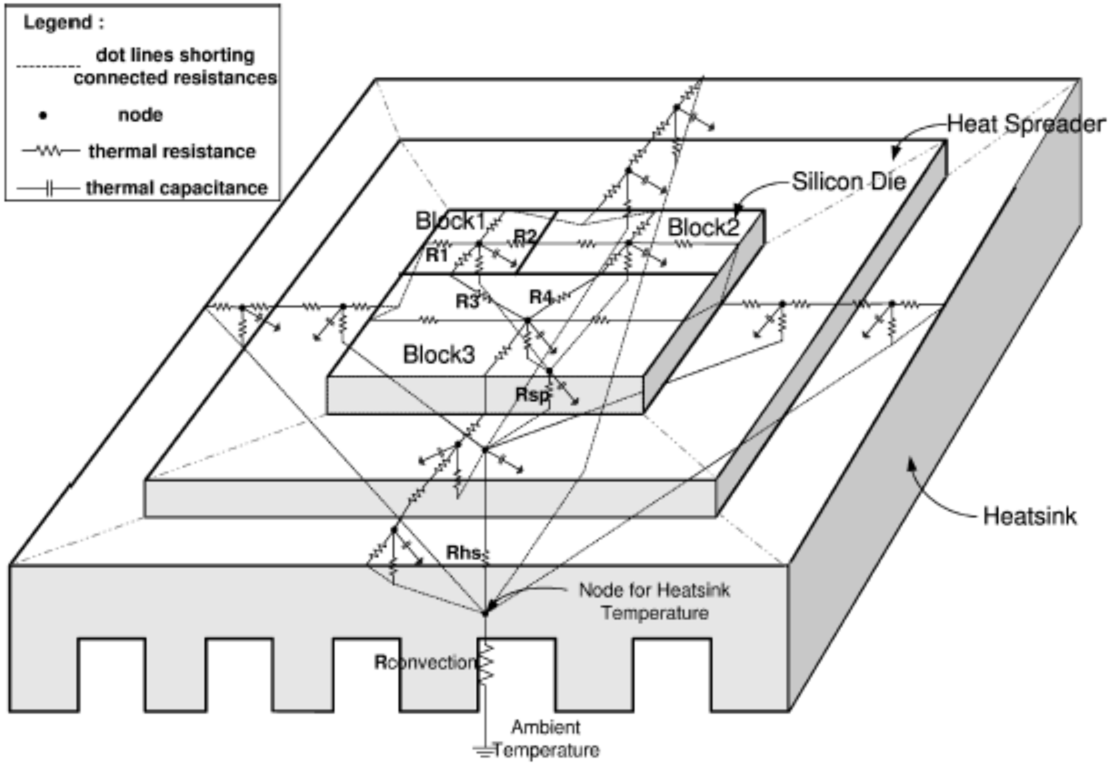
Low power pressure



Medium - high power pressure

http://download.intel.com/technology/itj/q32000/pdf/thermal_perf.pdf

Thermal Model

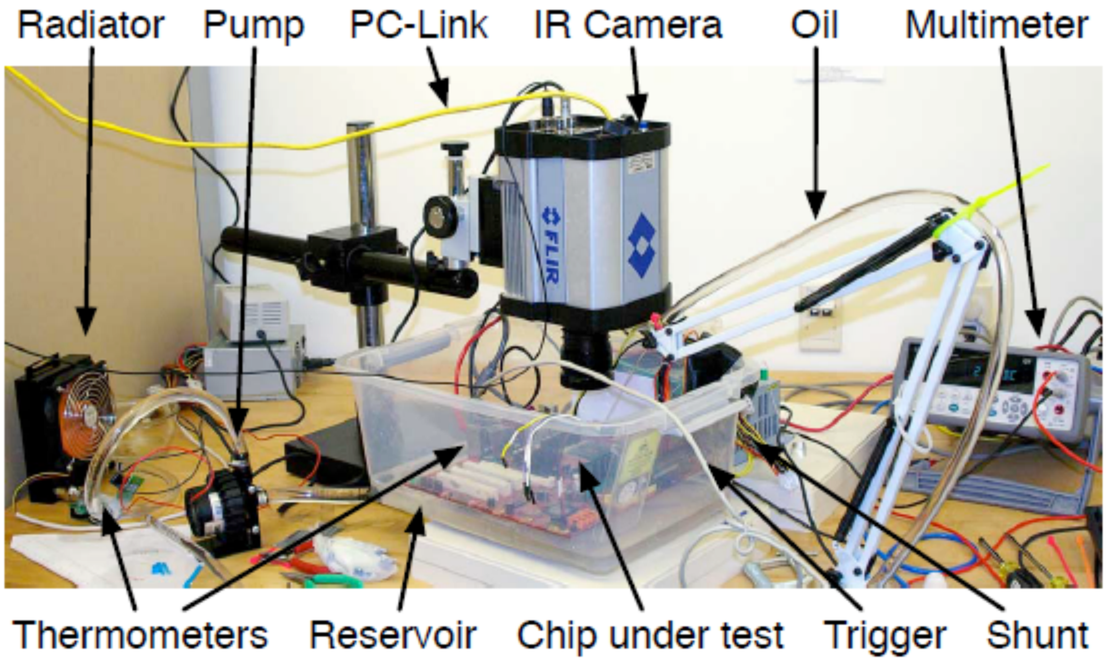


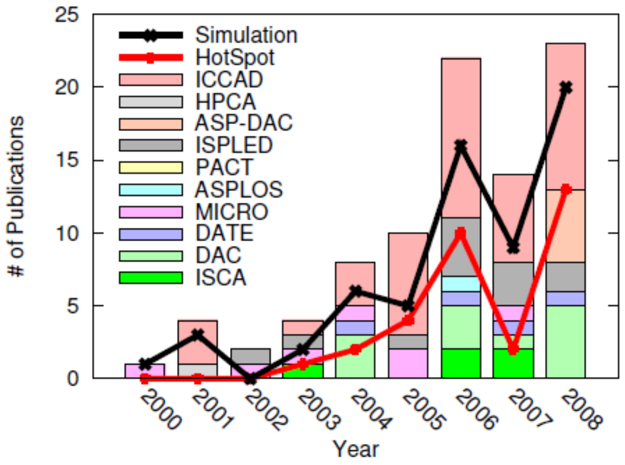
- Thermal behavior is modeled using RC circuit

Temperature-Aware Microarchitecture:
Modeling and Implementation, skadron et al. '04

Temperature Map Measurements

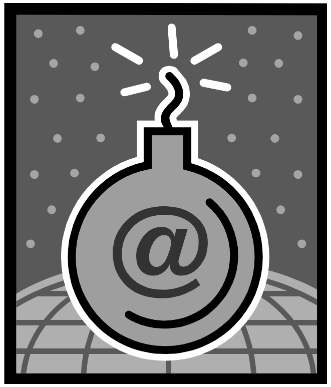
- Use IR camera (Jose Renau, UCSC)





Power Virus

- Maximum power consumption code
- How?
  - – Use data from L1 or L2
  - – Pipelines and queues are maintained full
  - – For longer period (meaningful program



Other Issues

- Power consumption
  - Not only CPUS
  - Memory, I/O devices, other units