

Extreme Scale Computing: Building Computer Architectures for Energy Efficiency

Josep Torrellas

Department of Computer Science
University of Illinois at Urbana-Champaign
<http://iacoma.cs.uiuc.edu>



The State of The Art



Blue Waters

- ~1 PF sustained
- >300,000 cores
- >1 PB of memory
- >10 PB of disk storage
- ~500 PB of archival storage
- >100 Gbps connectivity

10-20 MW



Blue Waters Building Block

- 32 IH server nodes
- 32 TB memory
- 256 TF (peak)
- 4 Storage systems
- 10 Tape drive connections

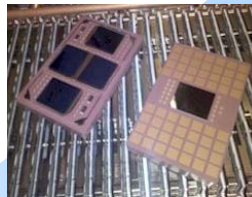


IH Server Node

- 8 MCM's (256 cores)
- 1 TB memory
- 8 TF (peak)

Fully water cooled

800 W



Multi-chip Module

- 4 Power7 chips
- 128 GB memory
- 512 GB/s memory bandwidth
- 1 TF (peak)



Power7 Chip

- 8 cores, 32 threads
- L1, L2, L3 cache (32 MB)
- Up to 256 GF (peak)
- 45 nm technology

Router

- 1,128 GB/s bandwidth

10MW = slightly over 10K American homes
Nuclear Power Station in Clinton, IL=1,043MW

Wanted: Computing at Affordable Power

- Current peta-scale machines:
 - Footprint of $1/10^{\text{th}}$ of a US football field
 - Consume over 10MW
 - 1MW == \$1M per year electricity
- **Extreme Scale** computing: 100-1000x more capable for the same power consumption and physical footprint
 - Exascale (10^{18} ops/cycle) datacenter: 20MW
 - Petascale (10^{15} ops/cycle) departmental server: 20KW
 - Terascale (10^{12} ops/cycle) portable device: 20W

Energy-Efficiency Gap

- **Goal:**
 - 20W Tera-Op chip
 - 20 pJoules/operation
- **Some numbers:**
 - Intel CoreDuo mobile microprocessor 2006: ($>10K$ pJ/instr)
 - IBM Power7 released 2010: MCM 800W for 1TFlop
 - Problem is harder than it looks:
 - Machines spend much of the energy transferring data
 - Minimizing E in data transfer, not ALU op is the challenge

Types of Power

- **Dynamic power:**

- Related to switching activity of logic
- Prop. to square of V_{dd} (cube)

$$P_{dyn} \propto CV_{dd}^2 f$$

- About 70% of all power

- **Static (leakage) power:**

- Leakage of a transistor even if it does nothing
- Exponential to T (also function of V_{dd})

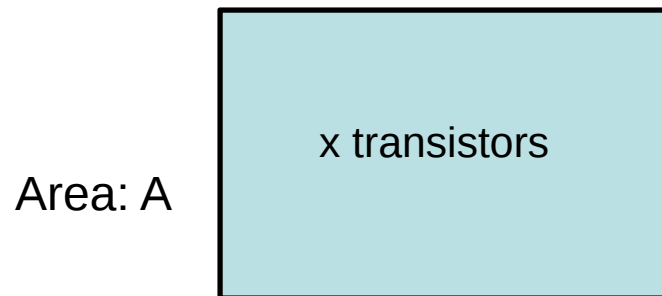
$$P_{sta} \propto V_{dd} T^2 e^{-qV_t/kT}$$

- About 30% of all power

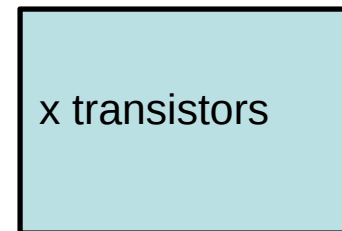
Why Are Energy & Power an Issue?

- **Ideal Scaling** (or **Dennard Scaling**): Every semicond. generation:
 - Dimension: 0.7
 - Area of transistor: $0.7 \times 0.7 = 0.49$
 - Supply Voltage (V_{dd}), C: 0.7
 - Frequency: $1/0.7 = 1.4$

$$P_{dyn} \propto CV_{dd}^2 f$$



Power density: $CV_{dd}^2 f/A$



Area: $0.7^2 A$

Power density: $0.7C \cdot 0.7^2 V_{dd}^2 \cdot 1.4f / 0.7^2 A$
 $= CV_{dd}^2 f/A$

Constant power density

Why Are Energy & Power an Issue?

- **Real Scaling:** V_{dd} does not decrease much.
 - If too close to threshold voltage (V_{th}) $\rightarrow$ slow transistor
 - Switching speed is prop to $(V_{dd} - V_{th})$

$$T_g \propto \frac{V_{dd} L_{eff}}{\mu (V_{dd} - V_t)^\alpha}$$

- Dynamic power density increases with smaller tech
- Additionally: There is the static power

Power density increases rapidly

What To Do?

- Evolutionary approaches
- Design computers for E & P efficiency **from the ground up**



Extreme Scale Computing

Evolutionary Approaches

- Design circuits for E & P efficiency rather than speed
 - Low-swing on-chip interconnection network circuits
 - New memory layouts and bank organizations that minimize the capacitance switched per access
- Simplify the processor, shallow pipeline, less speculation
- Augment processing nodes with accelerators

Not enough

Designing Computers for E & P from the Ground Up

- New technologies:
 - Low supply voltage (V_{dd}) operation
 - Resistive memory
 - 3D die stacking
 - Efficient on-chip voltage conversion
 - Photonic interconnects
- New architectural designs:
 - Efficient support for high concurrency
 - Data transfer minimization

NTC Operation

- Advantages:
 - Reduces energy of an operation by 8-10x
 - Increases the delay by 10x
 - Hence: potentially reduces **power consumption by 80-100x**
- Drawbacks:
 - Lower speed (1/10)
 - Induces a 5x increase in gate delay variation
 - Potentially increases faults several orders of magnitude