Research Challenges on Temperature-Aware Computer Systems

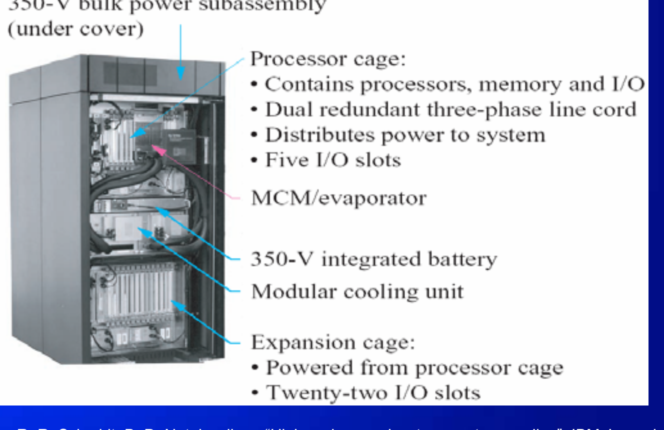
Antonio González

Intel-UPC Barcelona Research Center

2nd. Workshop on Temperature-Aware Computer Systems, Madison, WI, June 5, 2005

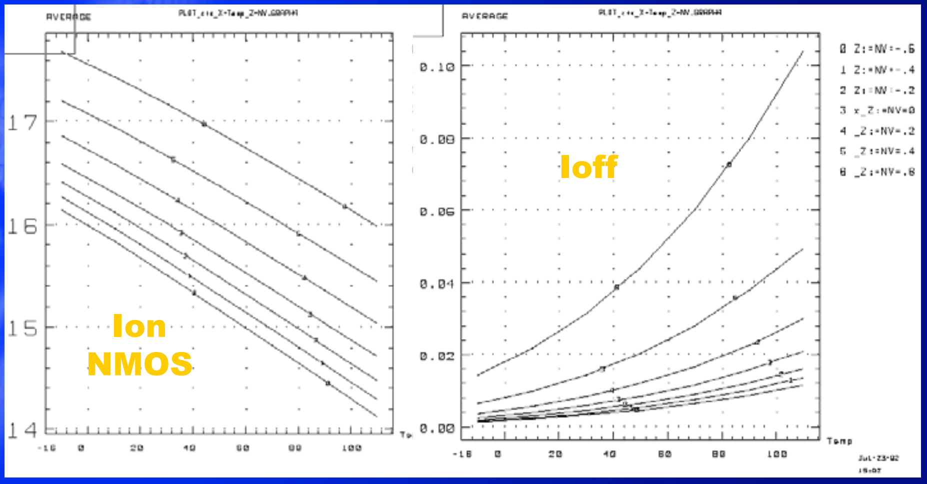
Temperature Affects Cooling Cost

- IBM S/390:



Source: R. R. Schmidt, B. D. Notohardjono "High-end server low temperature cooling", IBM Journal of R&D

Temperature Affects Performance and Power



Source: K. Skadron et al., Tutorial ISCA 2004

Temperature Affects Reliability

The Arrhenius Equation: $MTF = A * exp^{(E_a / K*T)}$

MTF: mean time to failure at T
A: empirical constant
E_a: activation energy
K: Boltzmann's constant
T: absolute temperature

Failure mechanisms

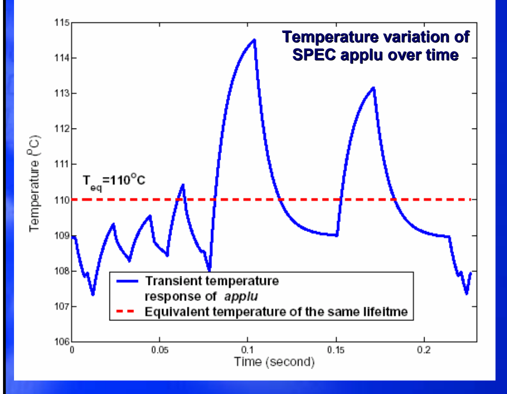
- Die metalization (Corrosion, Electromigration, Contact spiking)
- Oxide (charge trapping, gate oxide breakdown, hot electrons)
- Device (ionic contamination, second breakdown, surface-charge)
- Die attach (fracture, thermal breakdown, adhesion fatigue)
- Interconnect (wirebond failure, flip-chip joint failure)
- Package (cracking, whisker and dendritic growth, lid seal failure)

Most of the above increase with T (Arrhenius)
Notable exception: hot electrons are worse at low temperatures

Source: K. Skadron et al., Tutorial ISCA 2004

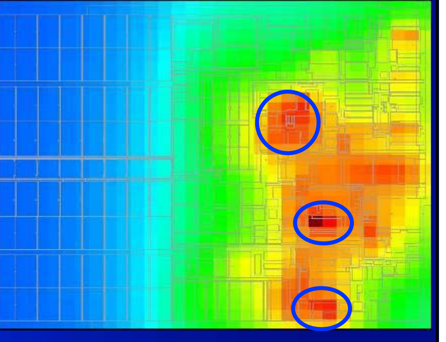
Temperature Variation

Temporal



Source: "A Quick Thermal Tutorial"
Kevin Skadron, Mircea Stan, U. of Virginia 2005

Spatial

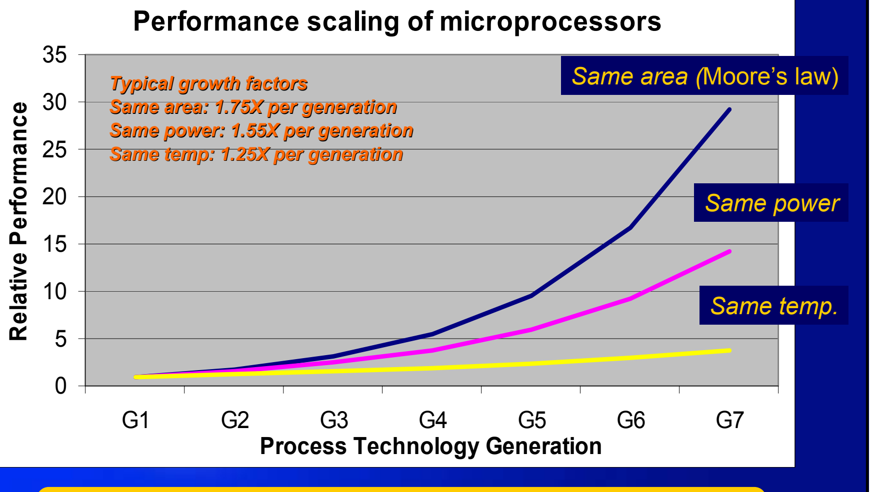


Thermal Map Pentium M (simulated)

(lowest) blue, green, yellow, orange, red (highest)

Source: Lev Finkelstein, Intel 2005

The Thermal Wall



In a thermally limited environment
Evolutionary Uarch will diminish its performance return

Source: Ronny Ronen

Basic Concepts

- Temperature is a function of power density
- Reducing temperature implies
  - Increasing area
    - Increases wire delays → Big impact on performace
  - Reducing power (slower transistors, simpler blocks)
    - May impact performance if not done carefully

Critical Areas of Research

- Modeling
  - Heat transfer
  - Thermal sensor's response
- Floorplan
  - Tradeoff between wire delays and peak temperature
- Microarchitecture techniques
  - Throttling
  - Clustering
    - Thermal steering
    - Cluster hopping
  - DVS/DFS
    - GALS
    - At the core granularity (multi-core)
  - Adaptive microarchitectures
- More effective cooling solutions
  - Constrained by weight, noise and power

Will Put Off the Thermal Wall but…

- A breakthrough is needed
  - New materials (e.g. High-K gate dielectric)
  - New devices (e.g. Tri-gate transistors)
  - New technology (e.g. carbon nanotubes, III-V transistors)
  - New circuit design techniques (e.g. asynchronous)
  - New microarchitectures (e.g. many simple cores)
  - …

The Challenge:
Reduce Energy While Increasing Performance