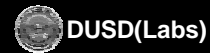
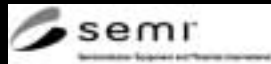


Chip-scale networks: Power and thermal impact



Li-Shiuan Peh

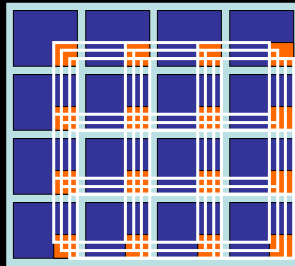
Assistant Professor of Electrical Engineering
Princeton University



Multi-core processors

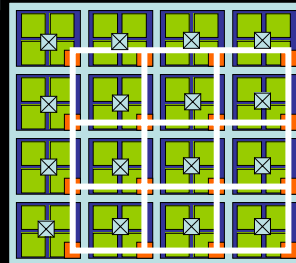
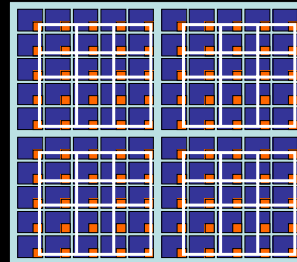
- Intel
 - Itanium: 4 cores (2006), 8, 16
- IBM
 - Power5: 2 cores (2004)
- Sun
 - Niagara: 8 cores (2005)

Future multi-core processors: networks form the backbone



MIT Raw
Four 4x4 networks

Austin TRIPS
Four 5x5 networks

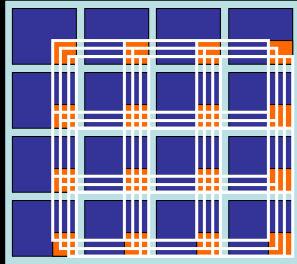


Stanford Smart Memories
A 4x4 network + xbars

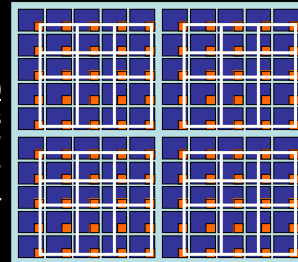
Outline

- Case studies examining power and thermal impact of chip-scale networks
- Systems-level approach
 - Power and thermal models
 - Distributed thermal management of networks
- Foundation for power and thermal-aware multi-core chips

Power impact of chip-scale networks

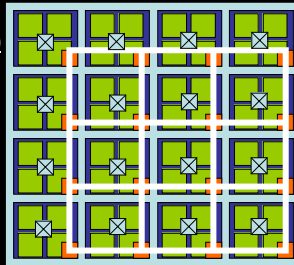


Austin TRIPS (0.1um, 1GHz)
 Four 5x5 networks
 Average: 14.1W
 Peak: 23.1W
 ~25% of avg chip power



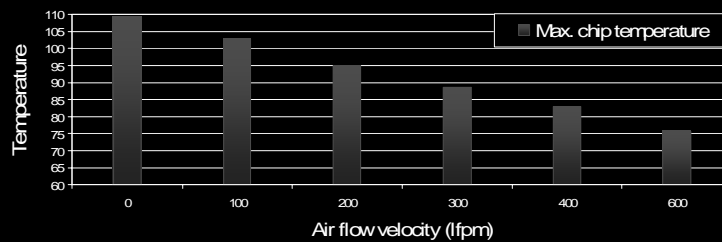
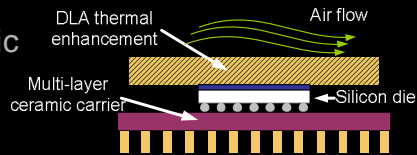
Stanford Smart Memories (0.1um, 1GHz)
 A 4x4 network + xbars
 Average: 12.0W
 Peak: 22.5W

MIT Raw (0.18um, 300MHz)
 Four 4x4 networks
 Average: 7.2W
 Peak: 14.8W
 36% of avg chip power



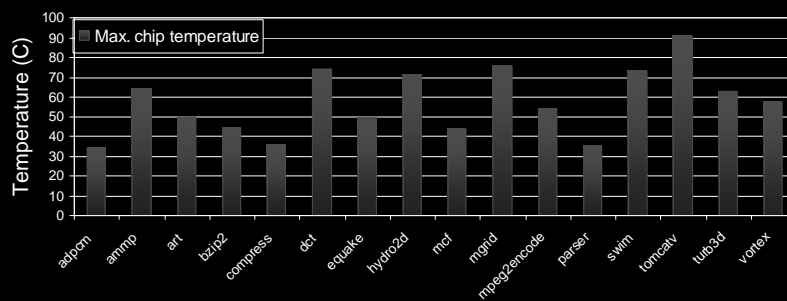
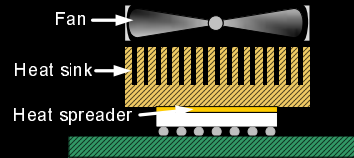
Thermal Impact: MIT RAW

MIT RAW on-chip network with IBM 42.5mm x 42.5 mm Ceramic Column Grid Array (CCGA) package with Direct Lid Attach (DLA) thermal enhancement



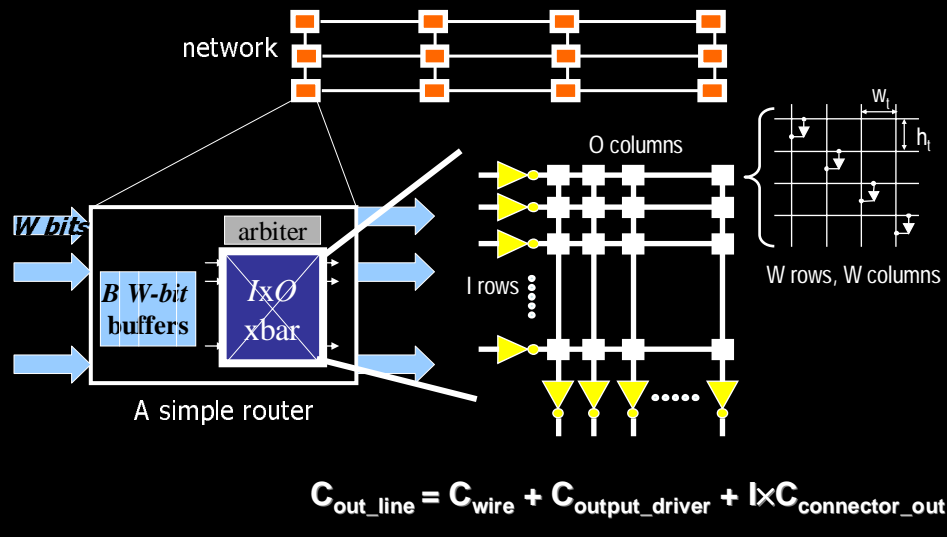
Thermal Impact: Austin TRIPS

A TRIPS-like 5x5 mesh on-chip network with complete cooling package



System-level power and thermal modeling of chip-scale networks

Power modeling

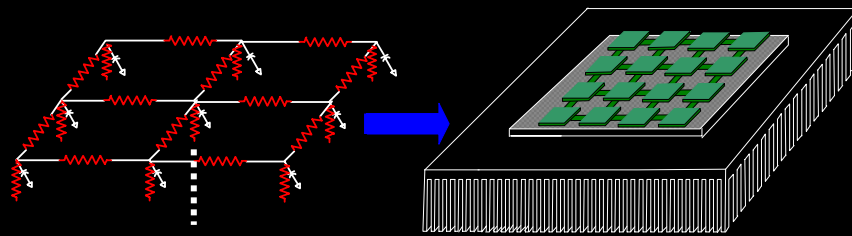


Validation of power models

- Dynamic power models
 - Academic on-chip networks
 - MIT RAW
 - 3-10% error against ChipBench tool
 - Commercial routers for off-chip networks
 - IBM InfiniBand 8-port 12X switch
 - Alpha 21364 router
 - Close to designers' estimates
- Leakage power models
 - HSPICE simulation
 - 95-97% accuracy

Thermal modeling

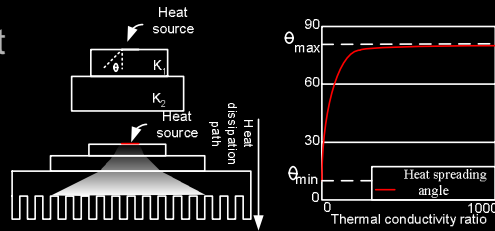
- Architecture-level thermal model
 - Multi-layer thermal RC network
 - Inter-router thermal correlation modeling
 - Link thermal model



Thermal modeling

- Heat spreading effect
 - Heat spreading angle

$$q = \tan^{-1}(k_1/k_2)$$



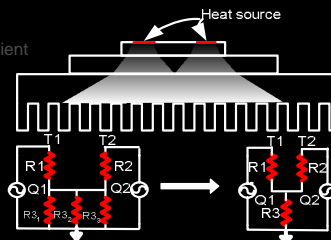
- Router thermal resistance

$$R_i = R_{i_silicon} + R_{i_spreader} + R_{i_sink} + R_{i_ambient}$$

- Thermal correlation

$$T1 = Q1R1 + (Q1+Q2)R3$$

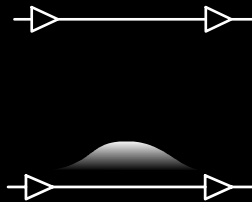
$$T2 = Q2R2 + (Q1+Q2)R3$$



Thermal Modeling of On-Chip Links

- Thermal impact of on-chip interconnects increasing
 - Reduced metal pitch and increased metal layers
 - High thermal resistance of silicon dioxide layers
 - High thermal resistance of low-k insulator materials
- Modeling methodology:

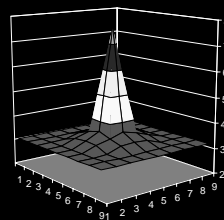
- Break links into multiple segments assuming optimal repeatering
- Estimate temperature along each segment, factoring in underlying silicon temperature



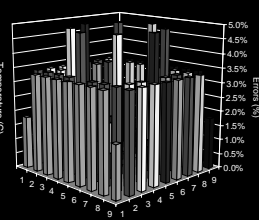
Thermal Model Validation

Compared with FEMLAB
Single heat source

- Evaluate the accuracy of modeling single heat source
- [3.2°C, 7.3°C], error < 5%



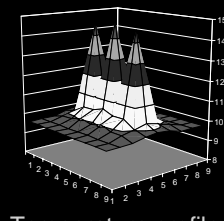
Temperature profile



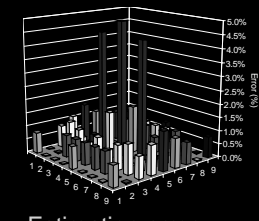
Estimation error

Three heat sources

- Evaluate the accuracy of modeling thermal correlation
- [9.7°C, 14.8°C], error < 5%



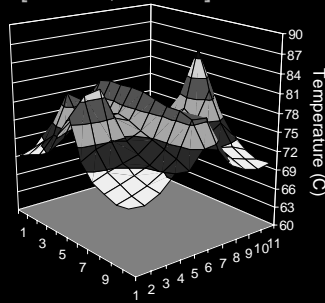
Temperature profile



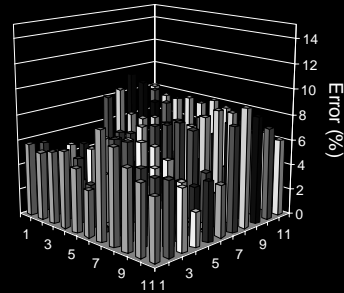
Estimation error

Thermal Model Validation

- Compared with an actual design from IBM
 - Power measurements
 - In-house finite-element based thermal simulator
 - [70.2°C, 85.4°C]



Chip temperature profile using our thermal model



Estimation error

Orion network simulator infrastructure

Network components:
Functional, timing, area, power, and thermal models.



Applications

Networked computer system simulator

Orion



Performance, power, and thermal statistics

Current Status:

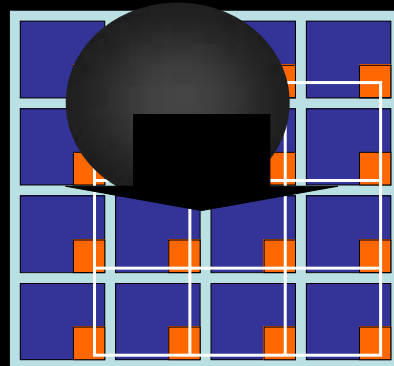
- User base includes Cambridge, CMU, Illinois, Intel, Stanford, Wisconsin
- Incorporates dynamic, leakage power, thermal models

Addressing the power and thermal impact of networks

- ThermalHerd: *Distributed, collaborative* thermal management of networks

Chip-level solution

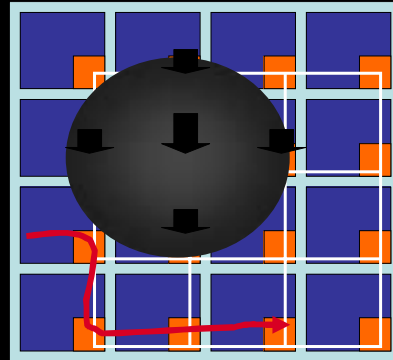
- Design-time
 - Design for worst-case
- Run-time
 - *Centralized* dynamic thermal management [Global throttling/clock gating, Pentium IV]



System-level solution: ThermalHerd

- *Distributed, Collaborative* thermal management

- Thermal monitoring
- Traffic estimation & prediction
- Collaborative traffic throttling
- Thermal-aware routing

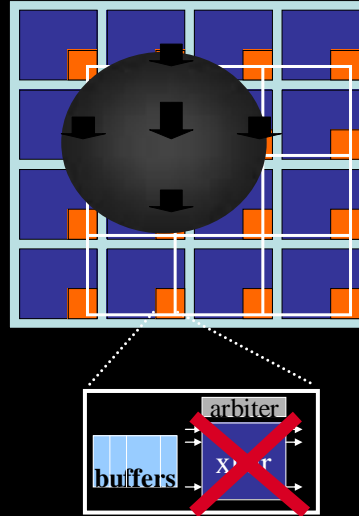


Run-time Thermal Monitoring, Estimation and Prediction

- Run-time Thermal Monitoring
 - Thermal sensors
 - Thermal models in dedicated hardware or PEs
- Traffic Estimation & Prediction
 - Hardware counters in routers track traffic
 - Guides throttling and routing to minimize performance impact

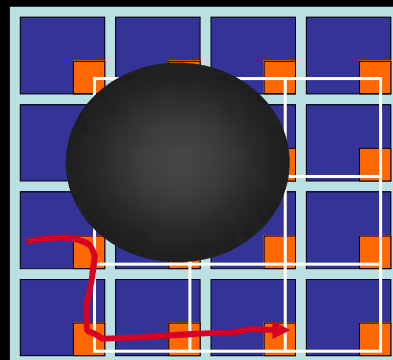
Collaborative Traffic Throttling

- Guided by traffic predictors
- Local traffic throttling
 - Local injected traffic has priority over traffic from neighborhood
- When local traffic throttling is not sufficient, neighboring routers are asked to help
 - Special messages prompt throttling at neighboring nodes
- Low-cost throttling through crossbar arbitration

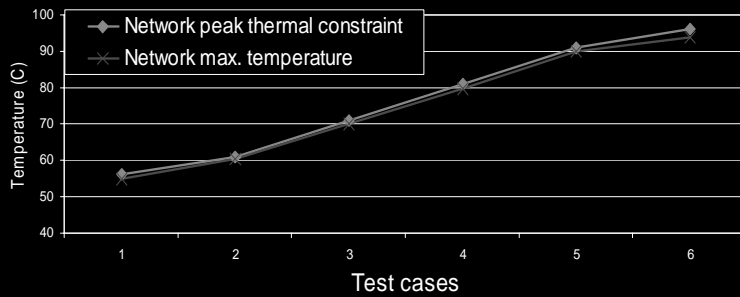


Thermal Correlation-Based Routing

- Reactive
 - Aggressively redirects traffic to avoid hotspot regions
- Proactive routing
 - Proactively smoothes temperature profile to avert thermal emergencies
- Guided by traffic predictors + thermal correlation matrix (resistances)



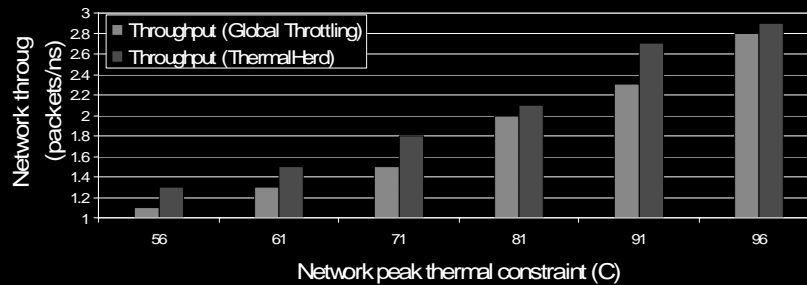
Effectiveness of ThermalHerd



ThermalHerd ensuring safe operation

Impact on network performance

- ThermalHerd (System-level) vs. Global Throttling (Chip-level)
 - Higher network performance (up to 20% with self-similar traffic that is fairly balanced across chip)
 - More performance gains expected with actual traffic (bursty TRIPS traces)

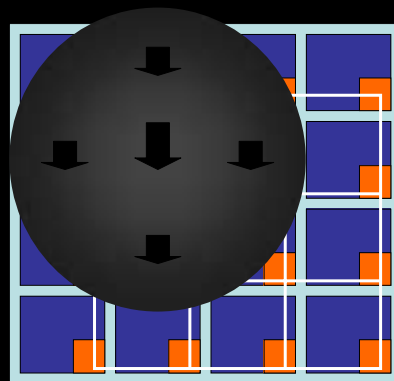


Power and thermal-aware networks as foundation for power and thermal-aware multicore chips?

ThermalHerd for multicore chips?

- *Distributed, Collaborative* thermal management

- Thermal monitoring at each core
- Workload estimation & prediction
- Collaborative core throttling
- Thermal-aware placement



Conclusions

- Chip-scale networks form backbone of future multi-core processors
- They have high power and thermal impact
- We can address this at the systems-level: ThermalHerd
- We see the *distributed, collaborative* approaches in power and thermal-aware networks forming the foundation for future power and thermal-aware multi-core chips

Acknowledgments

- **Academia**
 - MIT RAW team
 - Austin TRIPS team
 - K. Skadron of Virginia (HotSpot microprocessor thermal model)
- **Industry**
 - IBM InfiniBand switch: C. Stunkel
 - Intel Alpha 21364: A. Baum, S. Mukherjee
 - IBM chip thermal validation: H. Chen
- **Students**
 - Xuning Chen
 - Li Shang (w. Prof. Niraj Jha)
 - Hang-Sheng Wang (w. Prof. Sharad Malik)