Thermal Modeling, Analysis and Management of 2D Multi-Processor System-on-Chip

Prof. David Atienza Alonso

Embedded Systems Laboratory (ESL)
Institute of EE, Faculty of Engineering

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

ARTIST Summer School 2010, Autrans (France)

Outline

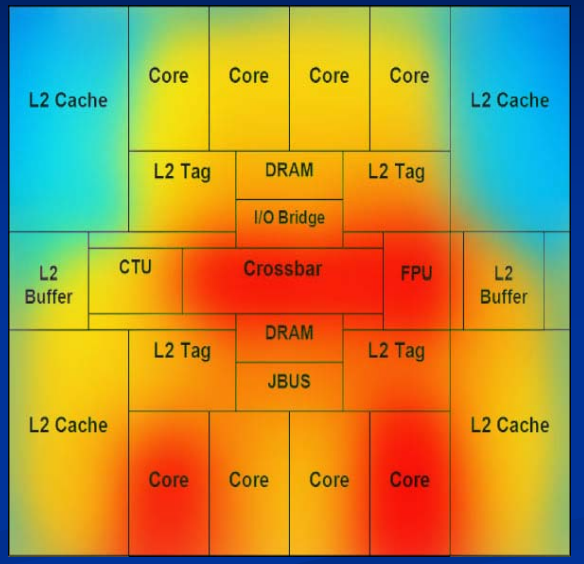
- MPSoC thermal modeling and analysis
- HW-based thermal management for MPSoCs
- SW-based thermal management for MPSoCs
- Conclusions

Outline

- MPSoC thermal modeling and analysis
- HW-based thermal management for MPSoCs
- SW-based thermal management for MPSoCs
- Conclusions

MPSoC Thermal Modeling Problem

- Continuous heat flow analysis
  - Capture geometrical characteristics of MPSoCs
  - Explore different packaging features and heat sink characteristics
- Time-variant heat sources
  - Transistor switching depends on MPSoC run-time activity (software)
  - Dynamic interaction with heat flow analysis



Very complex computational problem!

MPSoC Thermal Modeling State-of-the-Art

- MPSoC Modeling and Exploration

1. SW simulation: Transactions, cycle-accurate (~100 KHz)
   [Synopsys Realview, Mentor Primecell, Madsen et al., Angiolini et al.]

At the desired cycle-accurate level, they are too slow for thermal analysis of real-life applications!

2. HW prototyping: Core dependent (~50-100 MHz)
   [Cadence Palladium II, ARM Integrator IP, Heron Engineering]

Very expensive and late in design flow, no thermal modeling, only used for functional validation of MPSoC architectures!

- Heat Flow Modeling:

1. Software thermal/power models [Skadron et al., Kang et al.]

Too computationally intensive and not able to interact at run-time with inputs from MPSoC components!

MPSoC Thermal Modeling State-of-the-Art

- MPSoC Modeling and Exploration

1. SW simulation: Transactions, cycle-accurate (~100 KHz)
   [Synopsys Realview, Mentor Primecell, Madsen et al., Angiolini et al.]

At the desired cycle-accurate level, they are too slow for thermal analysis of real-life applications!

2. Combination of cycle-accurate MPSoC behavior and IC heat flow modeling at run-time is unheard of

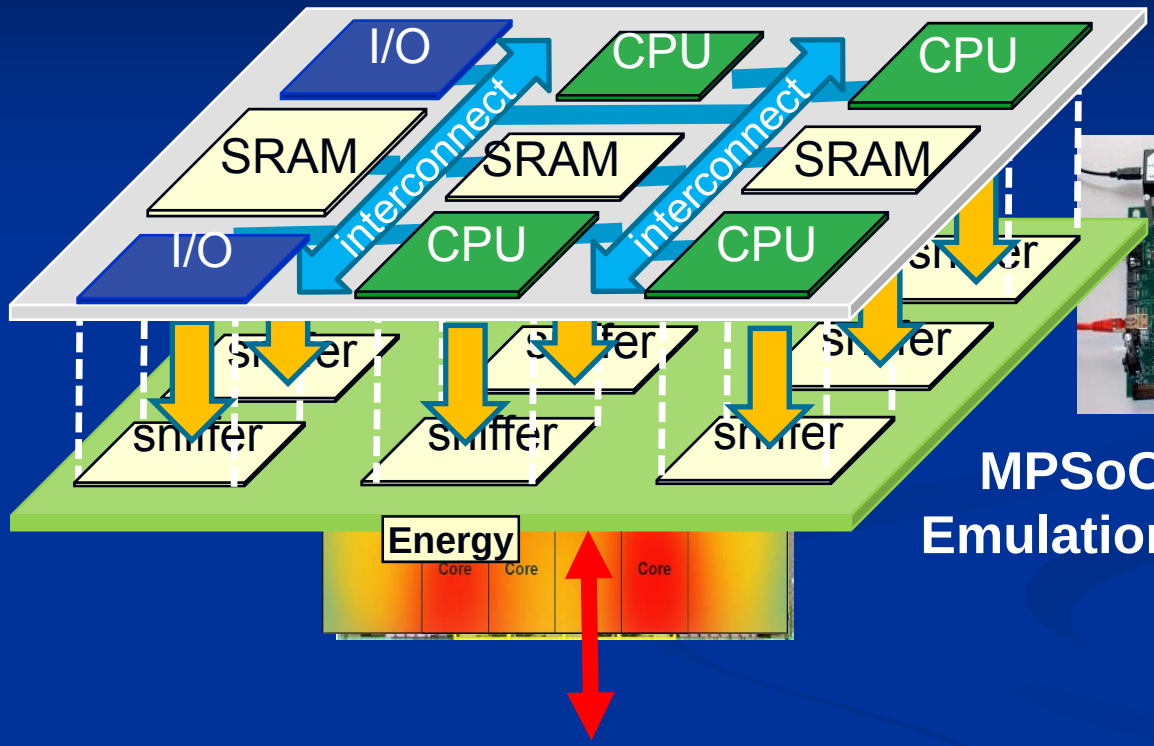
Very expensive and late in design flow, no thermal modeling, only used for functional validation of MPSoC architectures!

- Heat Flow Modeling:

1. Software thermal/power models [Skadron et al., Kang et al.]

Too computationally intensive and not able to interact at run-time with inputs from MPSoC components!

Orthogonalizing MPSoC Thermal Modeling and Analysis

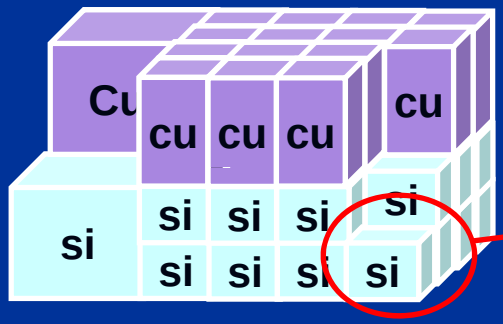


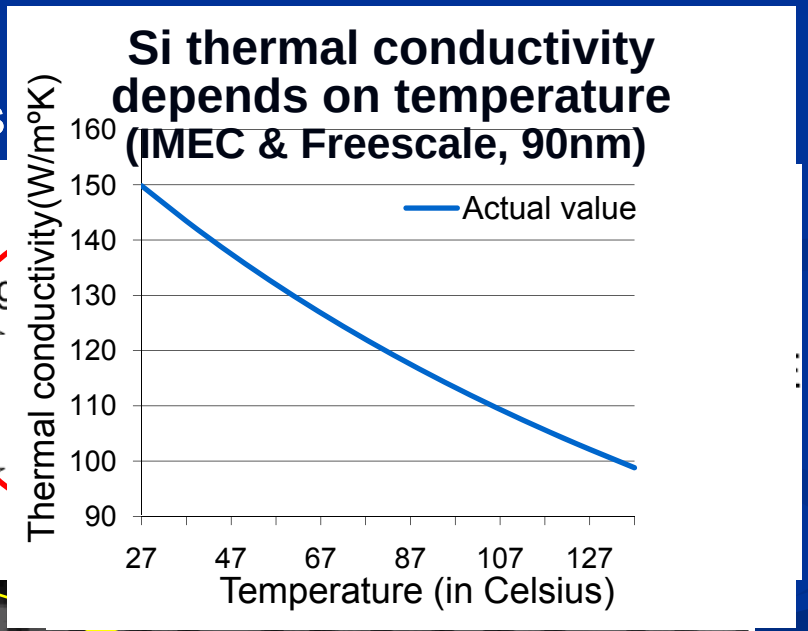
MPSoC Behavior Emulation on FPGA

Framework: MPSoC behavioral model on reconfigurable HW interacting with efficient thermal estimation

Chip and Package Heat Flow Modeling

- Model interface
  - Input: power model of MPSoC components, geometrical properties
  - Output: temperature of MPSoC components at run-time
- Thermal circuit: 1st order RC circuit
  - Heat flow ~ Electrical current ;
  - Heat spreader and IC compos





Thermal conductance matrix

Thermal capacitance matrix

$$\begin{bmatrix} C_{si,1} & & & \\ & C_{si,2} & & \\ & & \ddots & \\ & & & C_{cu,n} \end{bmatrix} \begin{bmatrix} G_{1,2}\text{-}G_{1,2} & & \\ -G_{2,1} & G_{2,1} & \\ & \ddots & \end{bmatrix}$$

Temperature change

$$C \dot{t}_k = - G (t_k) t_k + p_k \; ; \; k = 1..m$$

Temperature vector at instant k

Power consumption vector

SW Thermal Estimation Tool for MPSoCs

$$C \dot{t}_k = - G (t_k) t_k + p_k \; ; k = 1..m$$

- Creating linear approximation while retaining variable Si thermal conductivity:
  - Si thermal conductivity linearly approx. : $G_{i,j} (t_k) = l + q\, t_k$
  - Numerically integrating in discrete time domain the

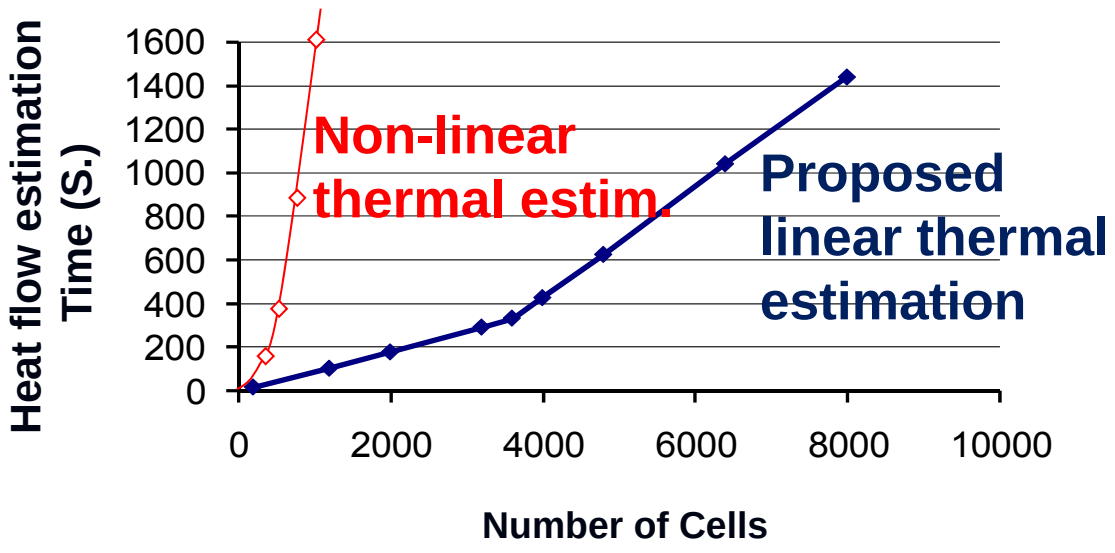
$$t_{k+1} = A(t_k) t_k + B$$

Complexity scales linearly with the number of modeled cells (simulated on P4@ 3GHz)

thermal library validated against 3D finite element model (IMEC & Freescale)

Si thermal conductivity

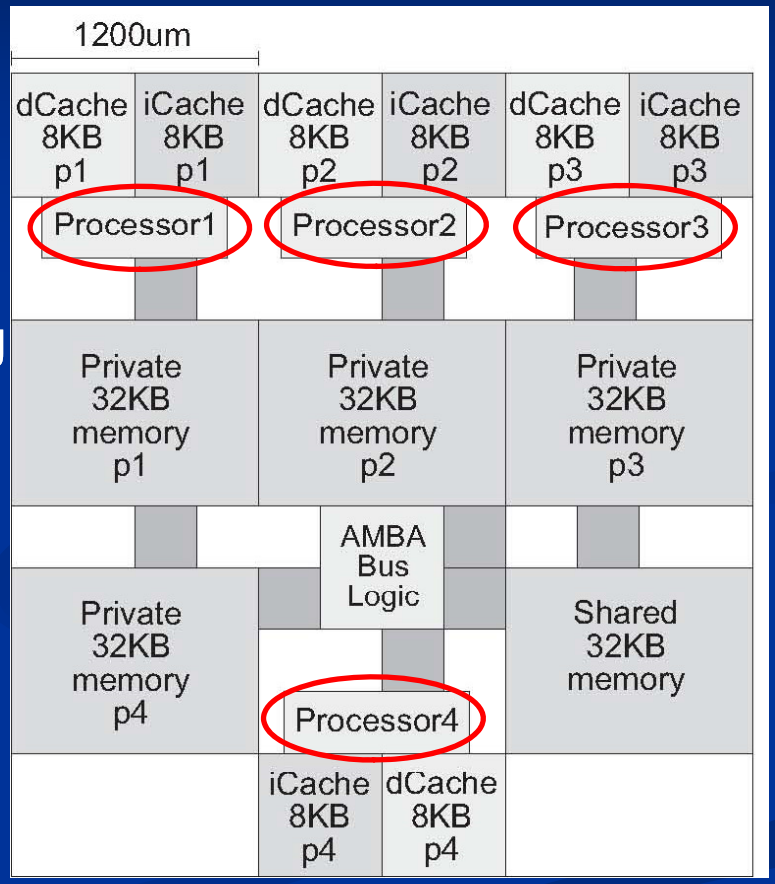
60 sec of MPSoC heat flow analysis



Case Study: HW 4-Core MPSoC

- MPSoC Philips board design:
  - 4 processors, DVFS: 100/500 MHz
  - Plastic packaging
- Software:
  - Image watermarking, video rendering
- Power values for 90nm:

Element	Max Power (mW) 100 MHz	Max Power (mW) 500 MHz
Processor	2,92 x 10^2	1,02 x 10^3
D-Cache	1,42 x 10^2	7,10 x 10^2
I-Cache	1,42 x 10^2	7,10 x 10^2
Priv Mem	0,61 x 10^2	2,75 x 10^2
AMBA	0,31 x 10^2	0,68 x 10^2



Results: Thermal Validation 4-core Philips MPSoC

- MPARM: Cycle-accurate SW architectural simulator
  - Complete power/thermal models tuned to Philips/IMEC figures
  - Simulations too slow: 2 days for 0.18 real sec (12 cells)
- HW thermal emulation able to validate policies at run-time
  - Dynamic Voltage and Frequency Scaling (DVFS) based on thresholds

Many weeks of Simulation!!

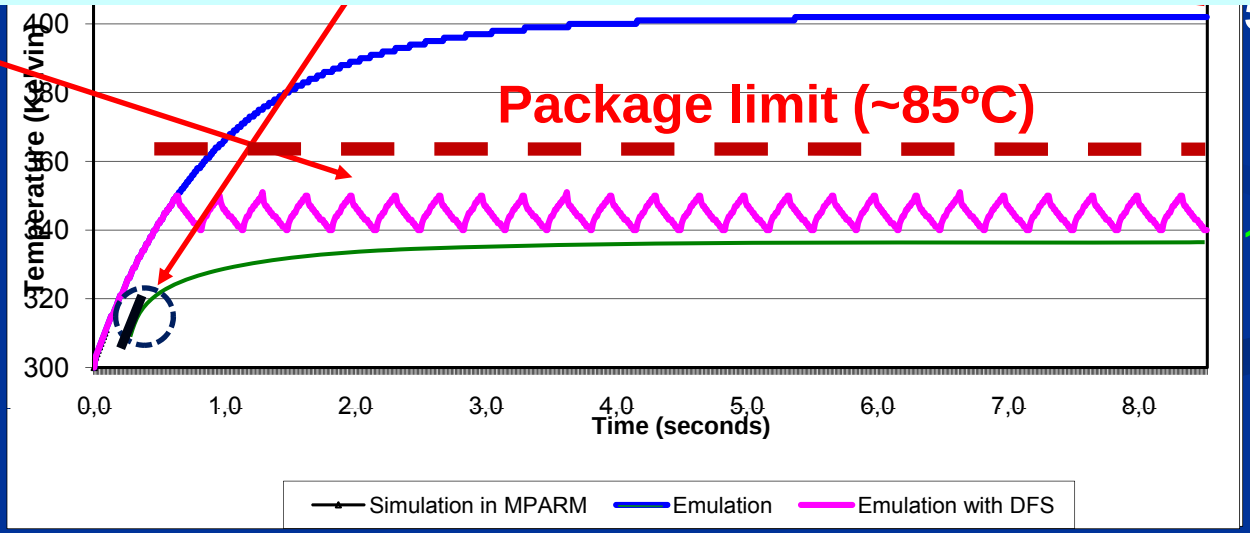
Emulation time 45 sec. (128 cells)!

Very fast validation of MPSoC run-time thermal behavior and management

500/100 MHz.

500 MHz.

100 MHz.

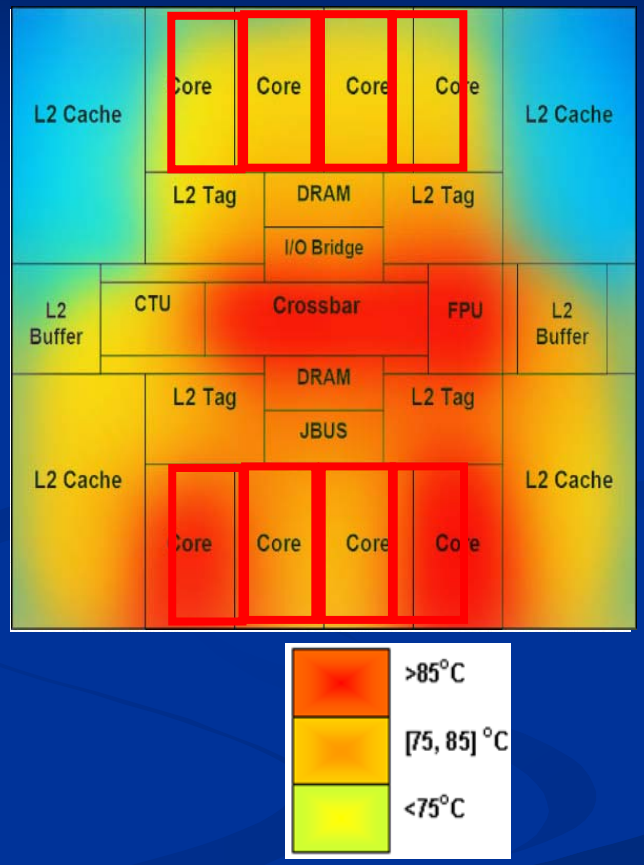


Outline

- MPSoC thermal modeling and analysis
- HW-based thermal management for MPSoCs
- SW-based thermal management for MPSoCs
- Conclusions

Temperature Management is Power Control under Thermal Constraints

- Power consumption of cores determines thermal behavior
  - Power consumption depends on frequency and voltage
  - Setting frequencies/voltages can control power and temperature
- Optimization problem: frequency/voltage assignment in MPSoCs under thermal constraints
  - Meet processing requirements
  - Respect thermal constraint at all times
  - Minimize power consumption



HW-Based Thermal Management State-of-the-Art

- Static approach: thermal-aware placement to try to even out worst-case thermal profile [Sapatnekar, Wong et al.]
  - Computationally difficult problem (NP-complete)

Not able to predict all working conditions, and leakage changing dynamically, it is not useful in real systems

No formalization of the thermal optimization problem

- Dynamic approach: HW-based dynamic thermal management
  - Clock gating based on time-out [Xie et al., Brooks et al.]
  - DVFS based on thresholds [Chaparro et al, Mukherjee et al,]
  - Heuristics for component shut down, limited history [Donald et al]

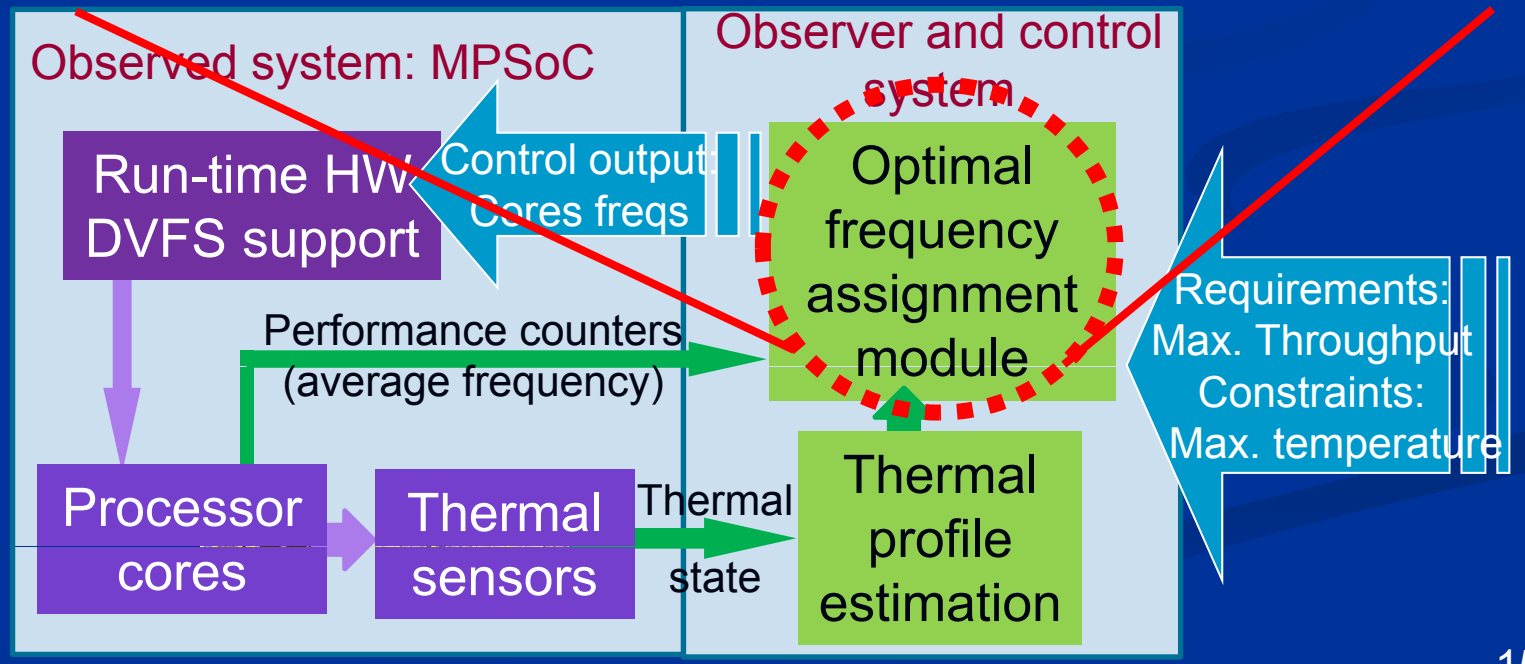
Techniques to minimize power, they only achieve thermal management as a by-product

Formalization of Thermal Management Problem in MPSoCs

- Control theory problem
  - Obse
    - H
    - P
  - Cont
    - Tu

Optimal frequency assignment module, 2-phase approach:

1) **Design-time phase:** Find optimal sets of frequencies for the cores for different working conditions
2) **Run-time phase:** Apply one of the predefined sets found in phase 1 for the required system performance



Pro-Active HW-Based Thermal Control: Phase 1 – Design-Time

- Predictive model of thermal behavior given a set of frequency assignments

Allowed core power values and frequencies

Chip floorplan

Packaging, heat spreader information

Phase inputs

Optimization problem: $\text{minimize} \quad \sum_{k=1}^{m} 1^T p_k$

Minimize sum of power ption of cores

Non-linear offline problem

Optimal frequency assignment module

Performance constraint: on average, freq. is f_{avg}

$T_{k+1} = A(t_k)t_k + B p_k, \quad k = 1, \ldots, m$

Thermal equation: Thermal equation depends on temp

$t_k \le t_{\max}, \quad k = 1, \ldots, m$

Meet temp. constraints at all time points

$p_{\max} f_{i,k}^2 / f_{\max}^2 = p_{i,k}, \quad i = 1, \ldots, n, \quad \forall k$

Power equation: quadratic dependence on freq. Equation: quadratic dependency

$f_{\min} \le f_k \le f_{\max}, \quad k = 1, \ldots, m.$

Frequency in predefined range

Method

outputs

Table of cores frequencies assignments

Making Power and Thermal Constraints Convex

- Power constraint adaptation

Solve convex problem and get table of optimal frequencies for different working conditions in polynomial time (number of processors)

Required average frequencies	Starting Temperatures <= 30 °C	35 °C	▪ ▪ ▪	100 °C
<= 100 MHz	<120,80,80,120>			
150 MHz				
▪ ▪ ▪				
1000 MHz				

Pro-Active HW-Based Thermal Control: Phase 2 - Run-Time, Putting It All Together

- Use table of frequencies assignments and index by actual conditions at regular run-time intervals

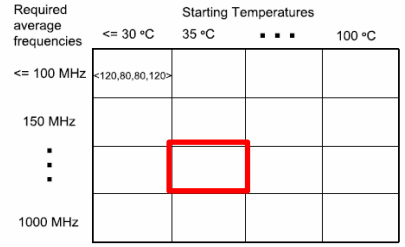
Targeted operating frequency of cores

Current temperature of cores

Method inputs

Run-time optimal DVFS assignment HW module

1) Index table output of phase 1 with current working conditions



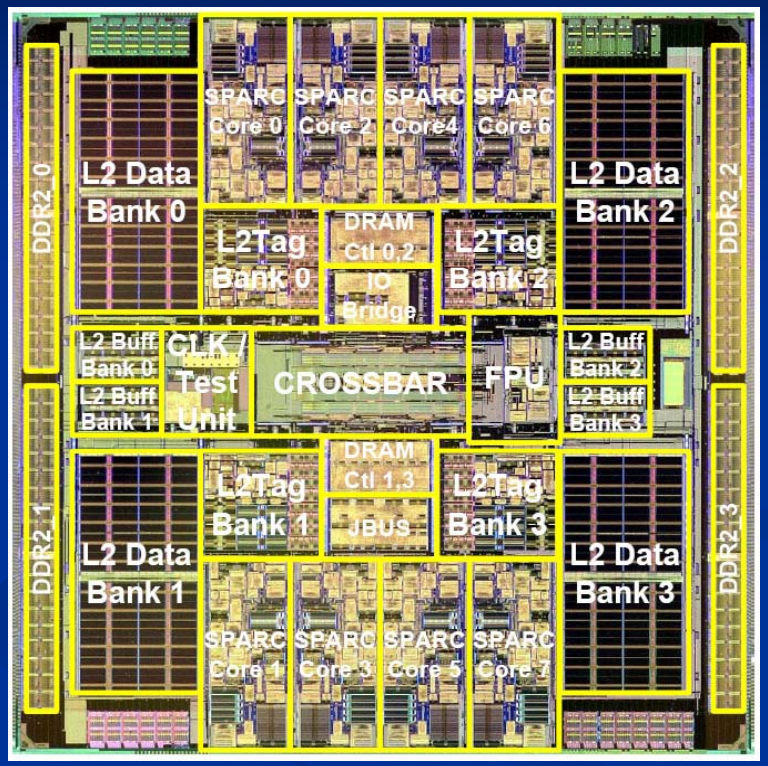
2) Compare to current assignment to cores and generate required signaling to modify DVFS values

Phase output

Run-time DVFS changes for processors

Case Study: 8-Core Sun MPSoC

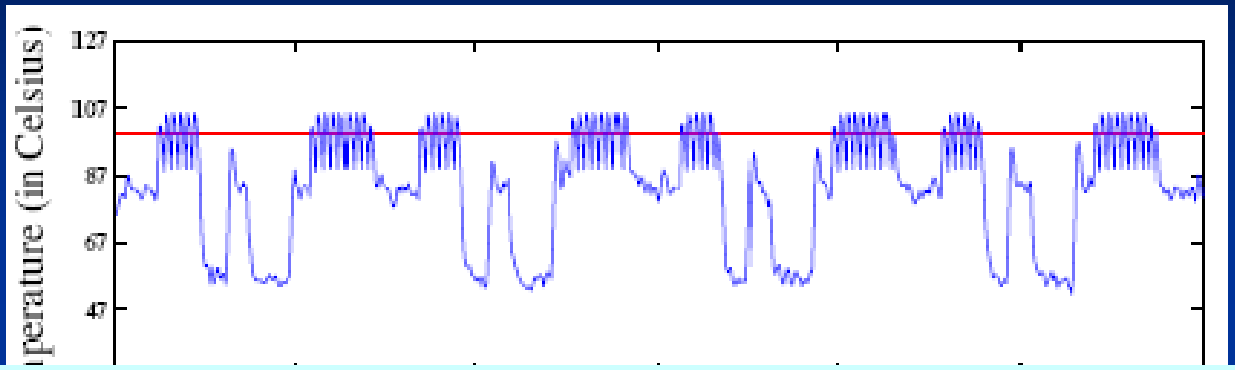
- MPSoC Sun Niagara architecture
  - 8 processing cores SPARC T1
- Max. frequency each core: 1 GHz
  - 10 DVFS values, applied every 100ms
- Max. power per core: 4 W
- Execution characteristics of workloads [Sun Microsystems]:
  - Mixes of 10 different benchmarks, from web-accessing to multimedia
  - 60,000 iterations of basic benchmarks, tens of seconds of actual system execution



Sun's Niagara MPSoC

Results: Thermal Constraints Respected

DVFS:

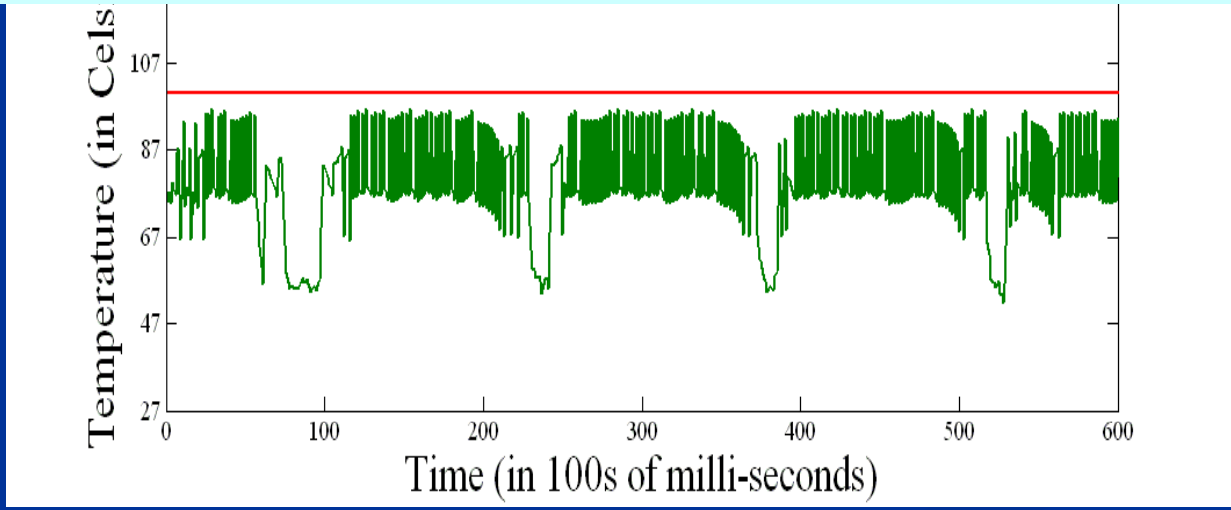


Total run-time of benchmarks

180 sec

Proposed method achieves better throughput than standard DVFS while satisfying thermal constraints

2-phase Convex method:

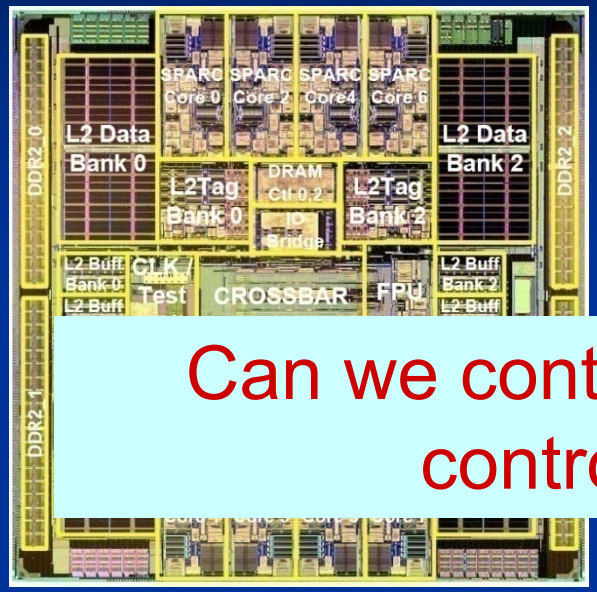


106 sec (45% less exec. time)

Outline

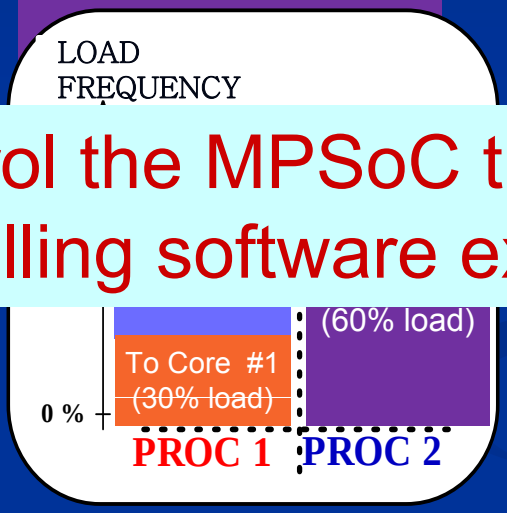
- MPSoC thermal modeling and analysis
- HW-based thermal management for MPSoCs
- SW-based thermal management for MPSoCs
- Conclusions

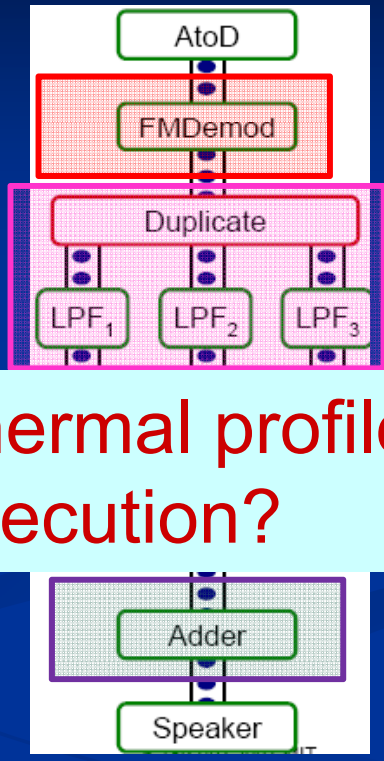
MPSoC System-Level Architecture: HW and SW Layers



MPOS

Task Manager





To Core #1 (30% load)

To Core #2 (60% load)

To Core #2 (35% load)

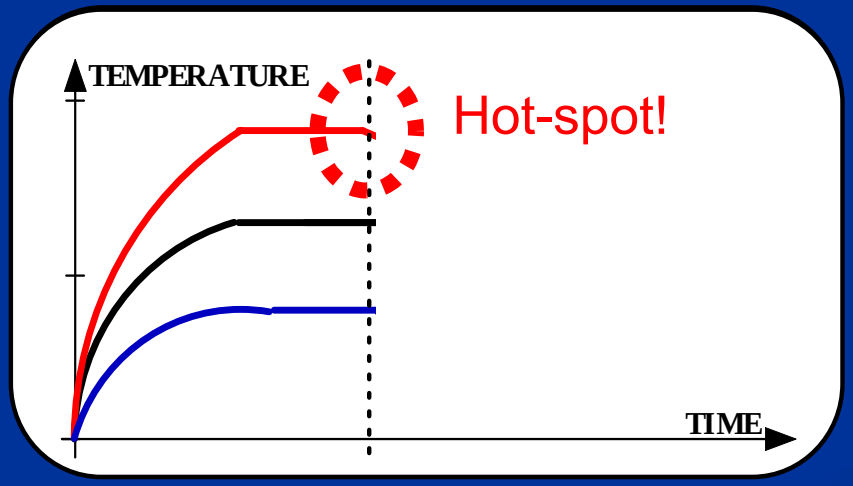
Can we control the MPSoC thermal profile by controlling software execution?

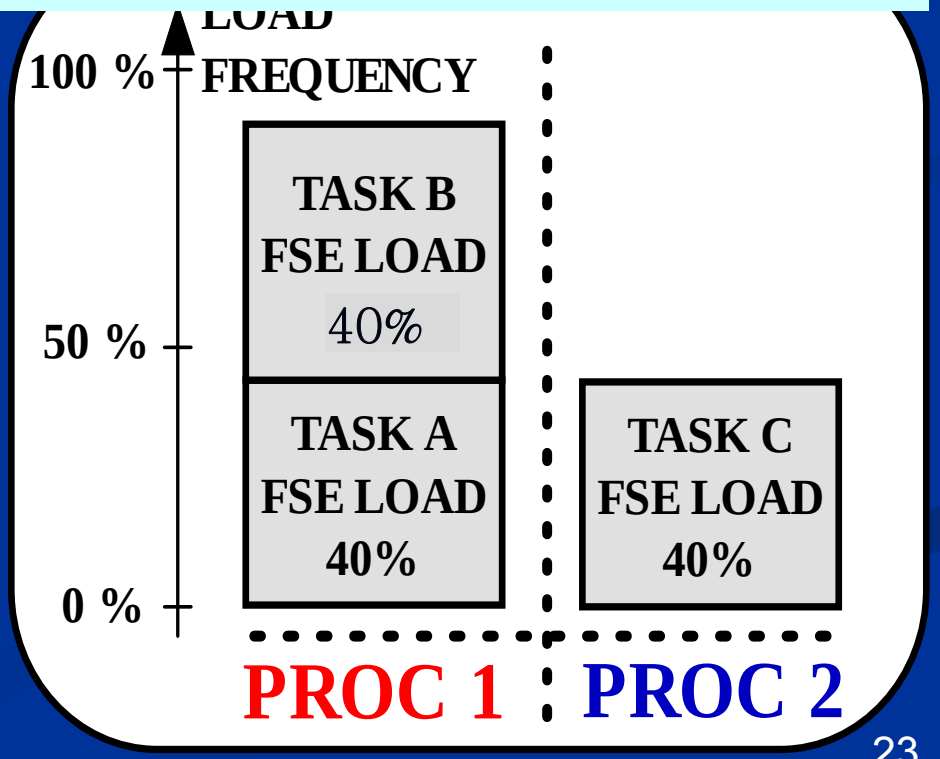
- SW layers introduced to better exploit the HW of MPSoCs
  - Applications divided in tasks: blocks of operations to be executed
  - Multi-processor Operating System (MPOS) distributes the tasks
    - Load balancing: equal distribution of work between processors

Task Migration for Load vs. Thermal Balancing

- Plain load balancing

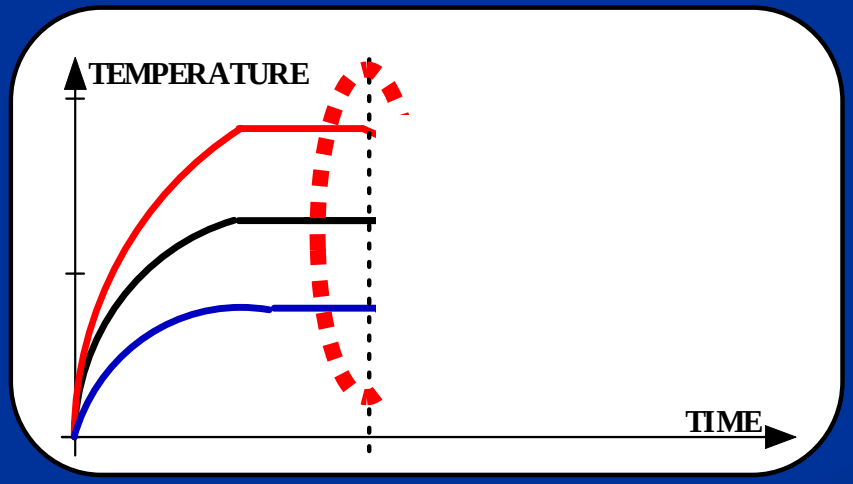
No improvement in workload distribution possible: no migration

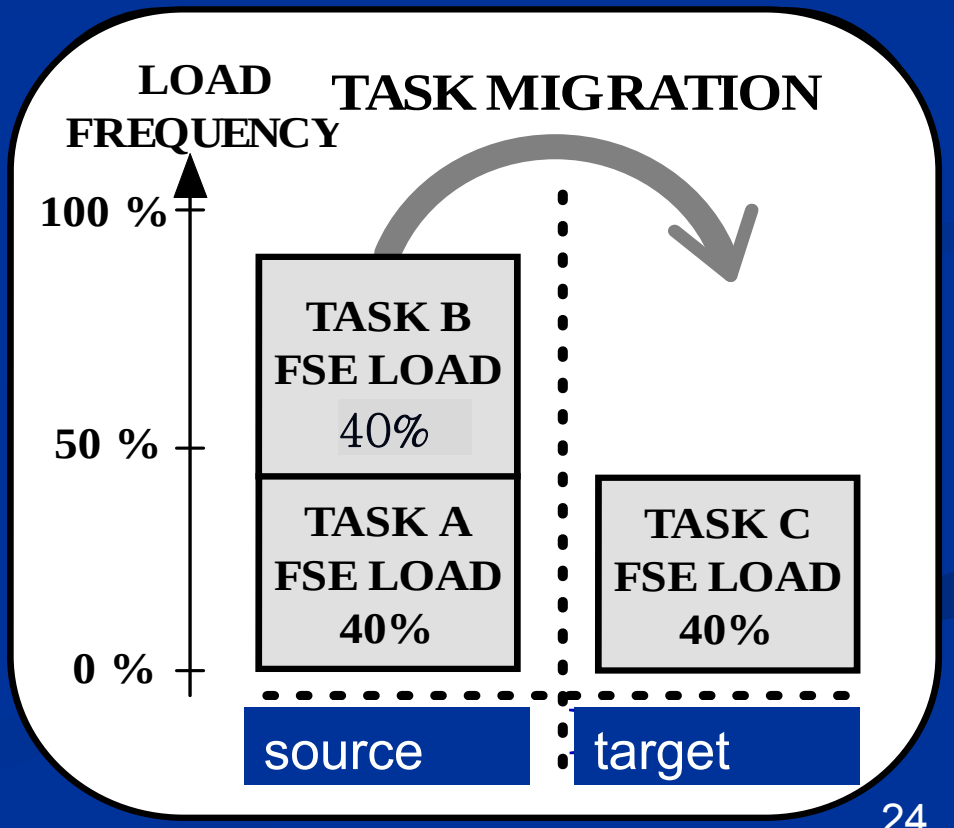




Task Migration for Load vs. Thermal Balancing

- Heat&Run: Load balancing with local knowledge of temperature in MPSoC components

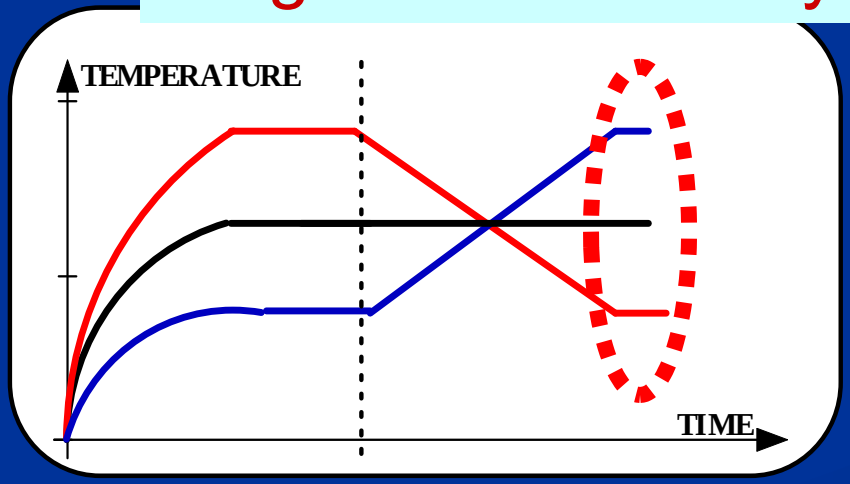


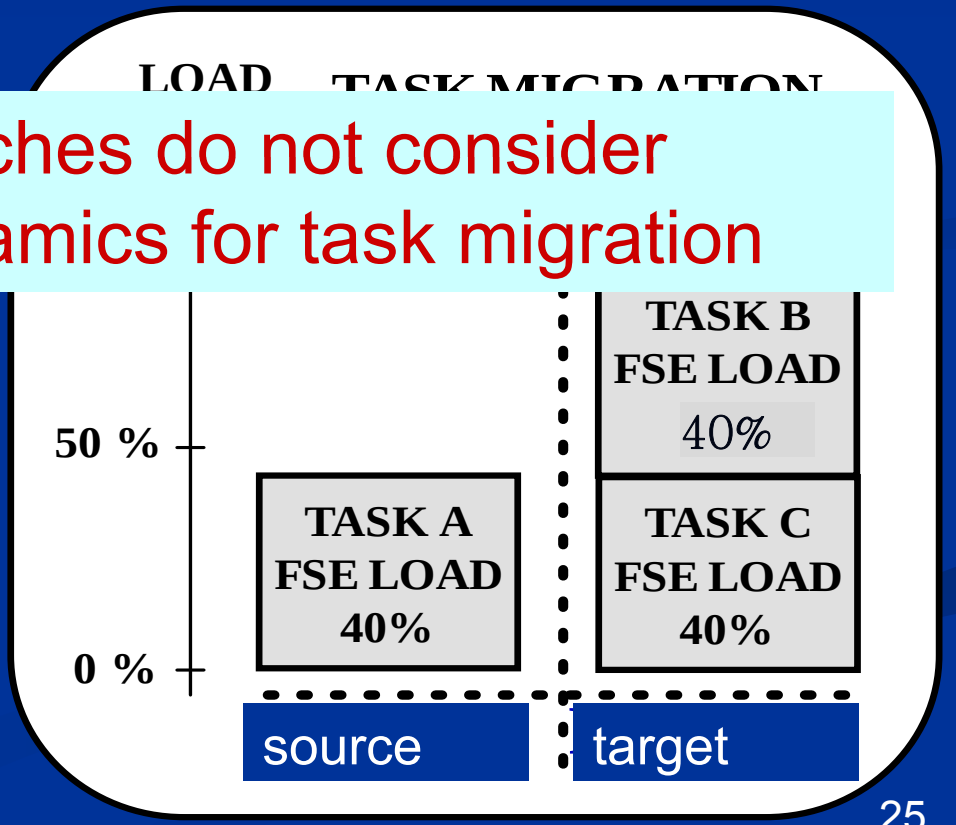


Task Migration for Load vs. Thermal Balancing

- Heat&Run: Load balancing with local knowledge of temperature in MPSoC components
  - Helping with hot-spots, but no thermal balancing

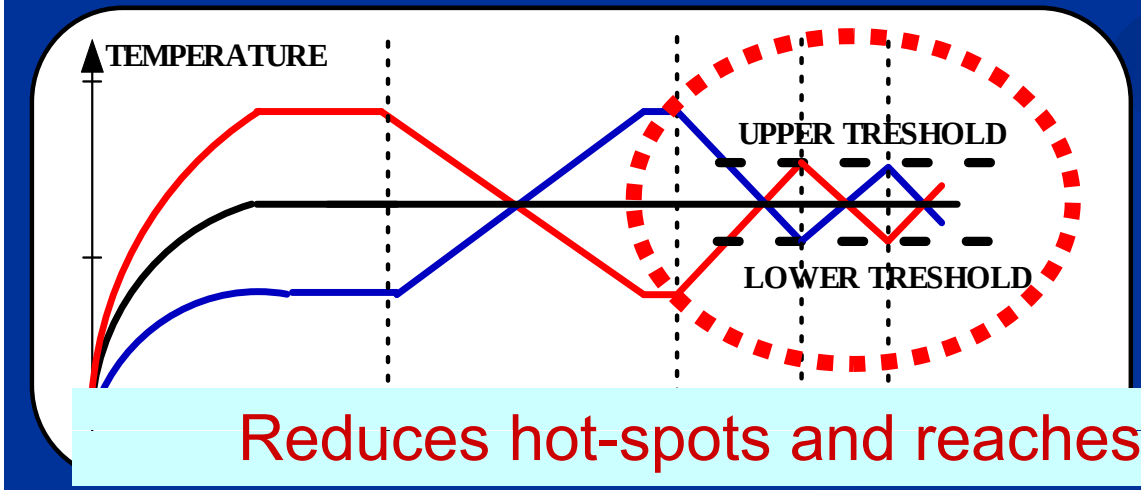
Existing approaches do not consider global thermal dynamics for task migration





Task Migration for Load vs Thermal Balancing

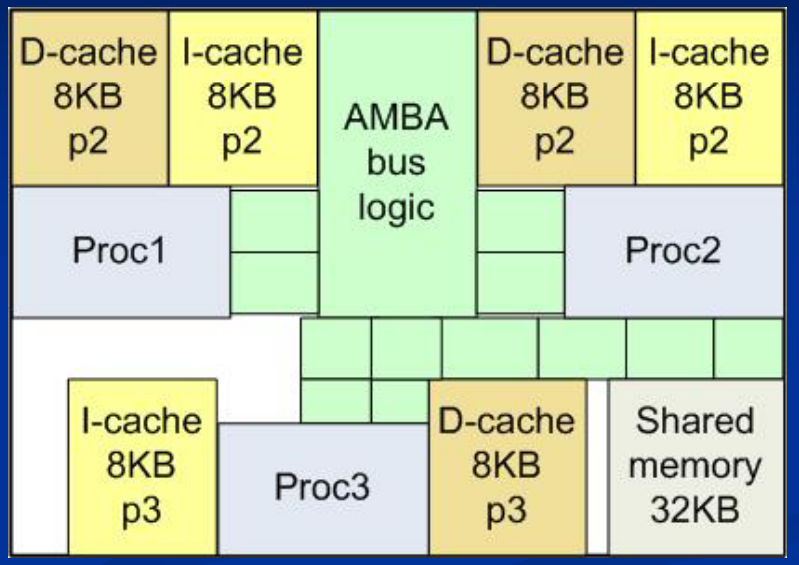
- Contribution: Migration strategy for thermal balancing
  - Global knowledge of temperature at MPOS level
  - Adjusted to particular thermal dynamics of each platform
- Formalization
  - Dynamic number of tasks, no control theory formalization possible
  - Knapsack problem, move N largest tasks between cores: estimated increase in temperature and minimizing performance penalty



Reduces hot-spots and reaches thermal balancing

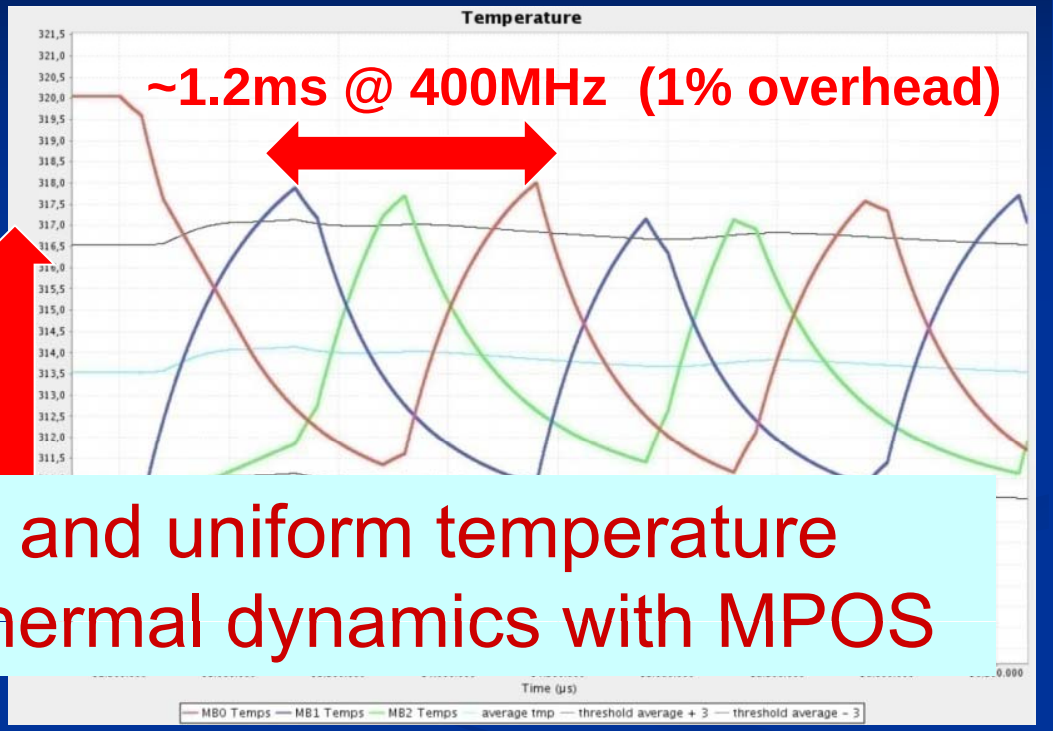
Case Study: Freescale MPSoC Board

- Hardware
  - 3 RISC processor cores
  - 16KB caches, 32KB shared mem.
  - AMBA bus, 2GB ext. mem
- Software
  - uCLinux-based MPOS
  - Multimedia applications: audio and video
- Two packaging options
  - Mobile embedded SoCs (slow temperature variations)
  - High performance SoCs (fast temperature variations)



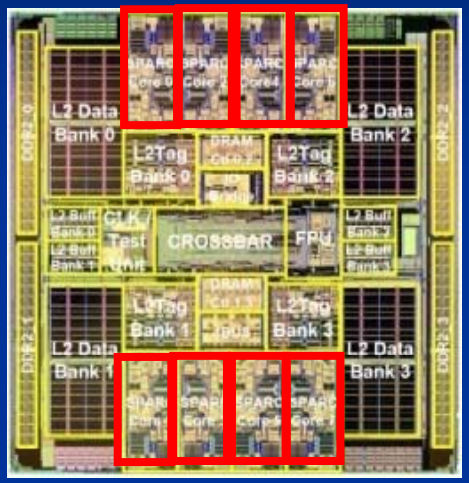
Results and Comparisons

- Good thermal balancing
  - Average: 40.5ºC, variations of < 3ºC
  - Small performance overhead (2 migrat/s)
- Comparison with
  - inefficient (>7ºC diffs)
  - Heat&Run inefficient or causes many deadline misses (40% below performance requirements)
  - Contribution: ***performance requirements met for both types of packaging***

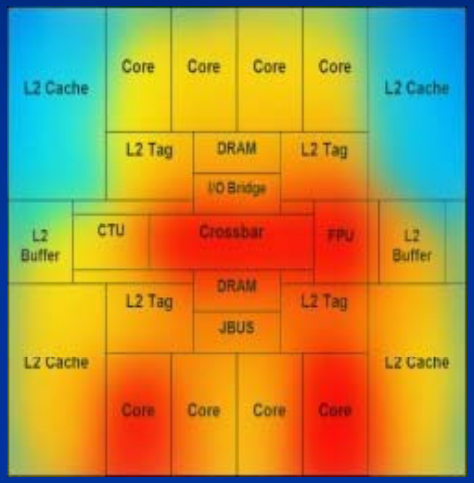


Good performance and uniform temperature adjusting globally to thermal dynamics with MPOS

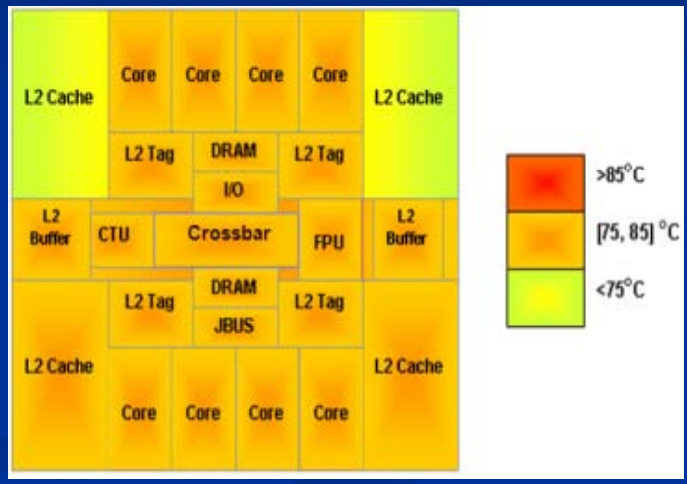
Adapt2D-MIGRA: Combination of HW and SW-Based Pro-Active Thermal Management



Initial: Large gradients



New: Thermal balancing



- HW-based management: Convex-based dynamic voltage and frequency scaling (DVFS) exploration
- SW-based management: Proactive task scheduling and migration
  - Support of multi-processor operating system: Solaris Multi-Core

Good thermal control in commercial MPSoCs in 90nm, what about 3D integration?

Outline

- MPSoC thermal modeling and analysis
- HW-based thermal management for MPSoCs
- SW-based thermal management for MPSoCs
- Conclusions

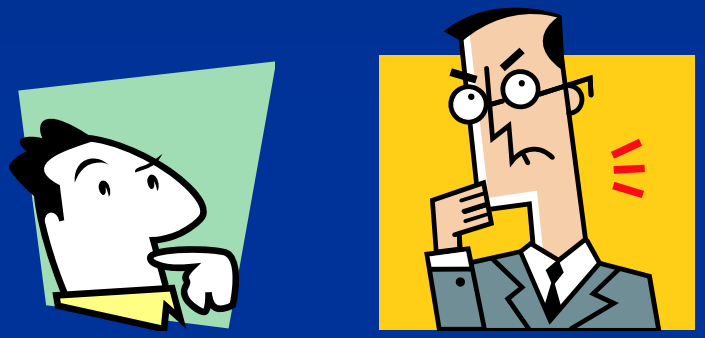
Conclusions

- Progress in semiconductor technologies enables new MPSoCs
  - Thermal/reliability issues must be addressed for safe human interaction
  - Thermal monitoring and control are key
- Clear benefits of thermal-aware design methods for MPSoCs
  - Novel, fast and low-cost thermal modeling approach at system-level
  - Formalization of HW-based thermal management problem as convex, and solved in polynomial time
  - New SW-based thermal balancing method with very limited overhead
- Validation on commercial 2D- MPSoCs (Sun, Freescale, Philips)
  - Fast exploration of thermal behavior of complex MPSoCs
  - Effective HW- and SW-based pro-active thermal management

Key References and Bibliography

- Thermal modeling and FPGA-based emulation
  - "**HW-SW Emulation Framework for Temperature-Aware Design in MPSoCs**", D. Atienza, et al. *ACM TODAES*, Vol. 12, Nr. 3, pp. 1–26, August 2007.
- Thermal management for 2D MPSoCs
  - "**Thermal Balancing Policy for Multiprocessor Stream Computing Platforms**", F. Mulas, et al., *IEEE T-CAD*, Vol. 28, Nr. 12, pp. 1870-1882, December 2009.
  - "**Processor Speed Control with Thermal Constraints**", A. Mutapcic, S. Boyd, et al. *IEEE TCAS-I*, Vol. 56, Nr. 9, pp. 1994-2008, Sept 2009.
  - "**Inducing Thermal-Awareness in Multi-Processor Systems-on-Chip Using Networks-on-Chip**", E. Martinez, et al., Proc. *ISVLSI* 2009.
  - "**Temperature Control of High-Performance Multi-core Platforms Using Convex Optimization**", S.Murali, et al., Proc. *DATE,* 2008.

Thank you!



QUESTIONS ?

Acknowledgements:

Swiss National Science Foundation

European Commission

UCSD / Sun Microsystems

IMEC / Philips

IBM Zürich

Bologna / Freescale semiconductors