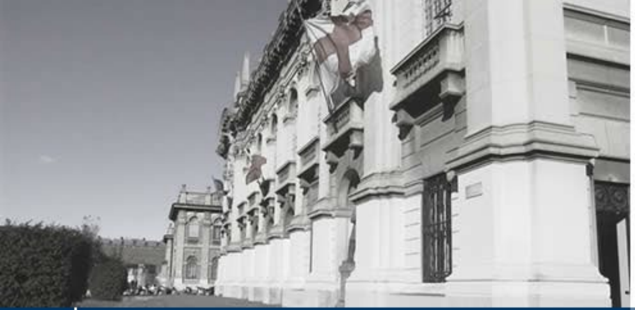


AA 2023 - 2024

POLITECNICO DI MILANO

Thermal and reliability issues in multi-core processors

William Fornaciari

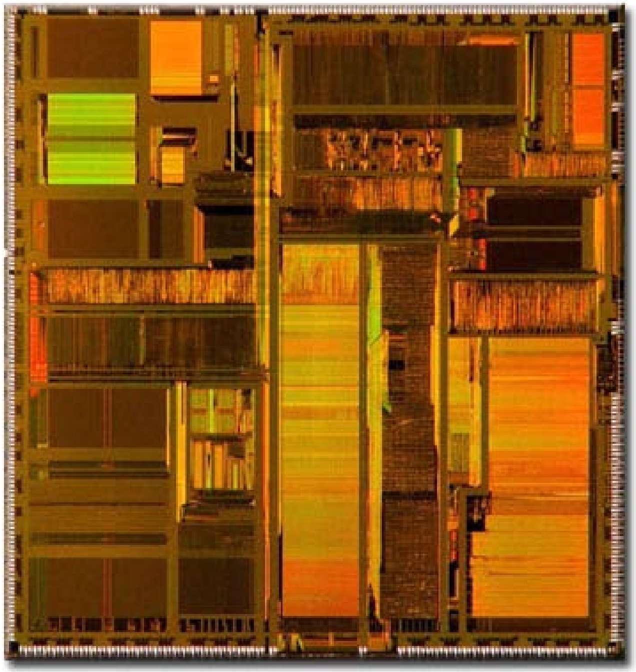
Politecnico di Milano
Dipartimento di Elettronica e Informazione
Via Ponzio 34/5, 20133, Milano, ITALY

Outline

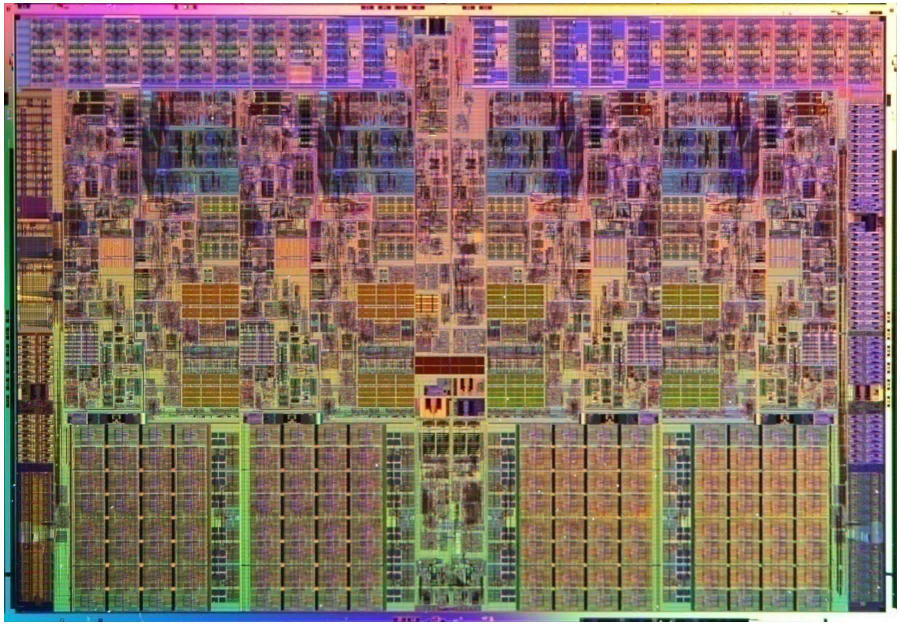
- **Introduction**
  - Technology scaling and reliability challenges
- **Goals and issues of a simulation flow**
  - Design time vs run-time optimization
  - Design of management policies
- **Thermal profile optimization**
  - State-of-the-art
  - Proposed approach
- **NBTI mitigation**
  - State-of-the-art
  - Proposed approach
- **Conclusions and lessons learnt**

Technology scaling at Intel

- **Intel processor evolution**



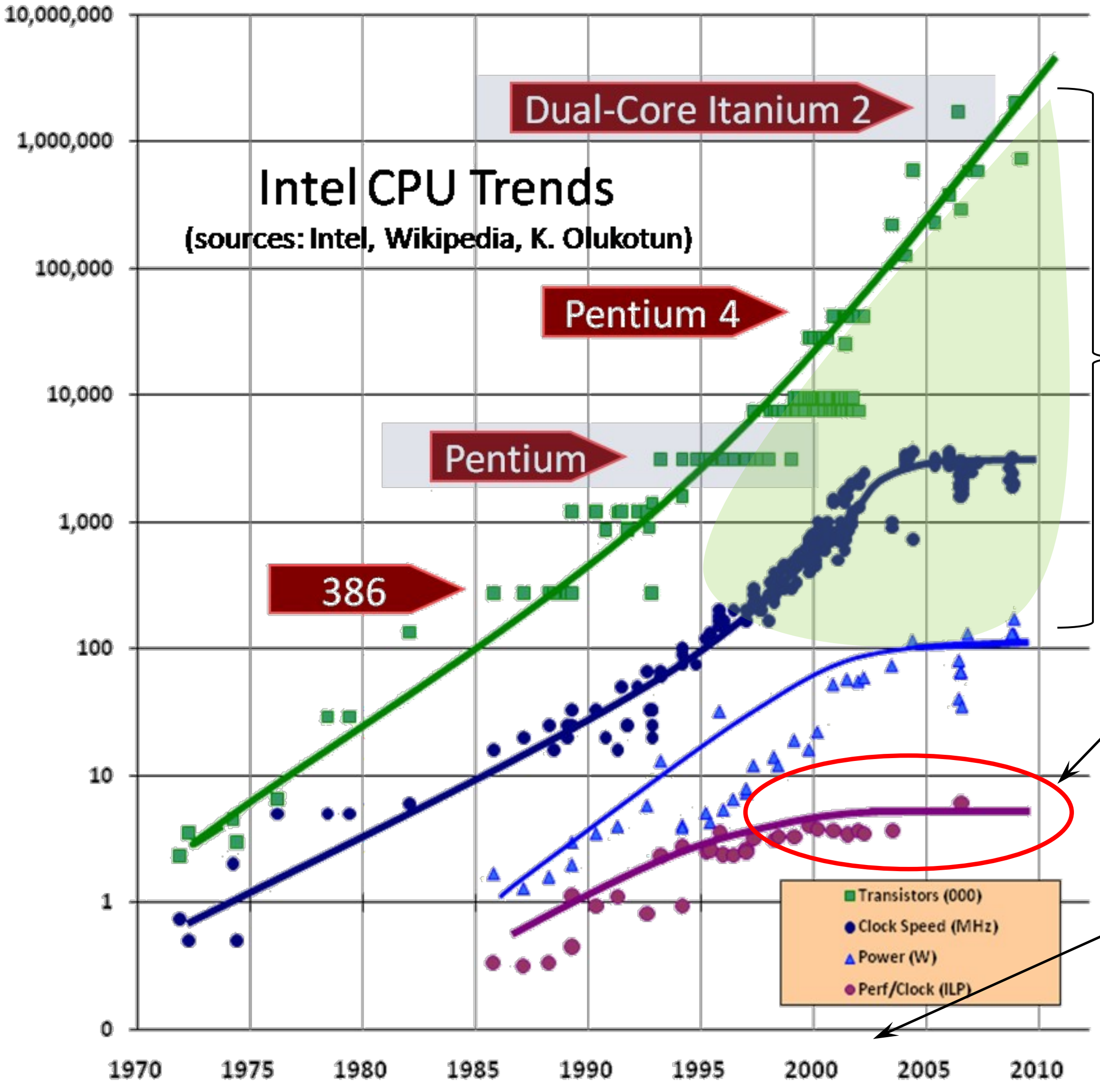
Pentium P5, 1993
60-66MHz, 5V
3.1M transistors, 294mm^2
0.8µm



i7 quad-core, 2008
1.73GHz @0.8-1.4V
730M transistors, 296mm^2
45nm

Source: www.intel.com

Technology scaling at Intel (cont'd)



Increasing density

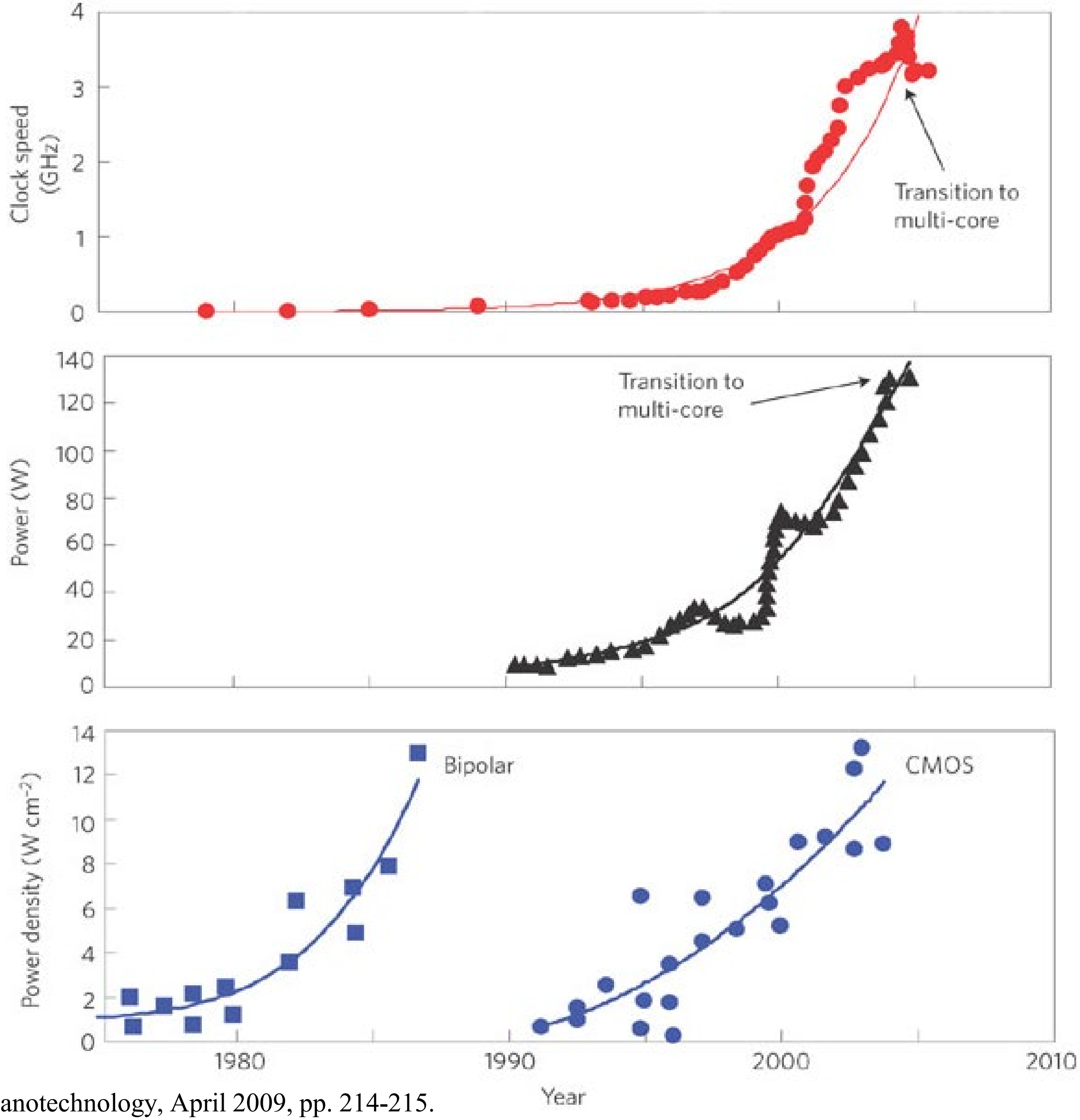
No more benefits from ILP

Turning point

Power trend in microelectronics industry

Performance and power

- Performance gain and multi-core era
- Power consumption increases
- Power density causes higher temperatures

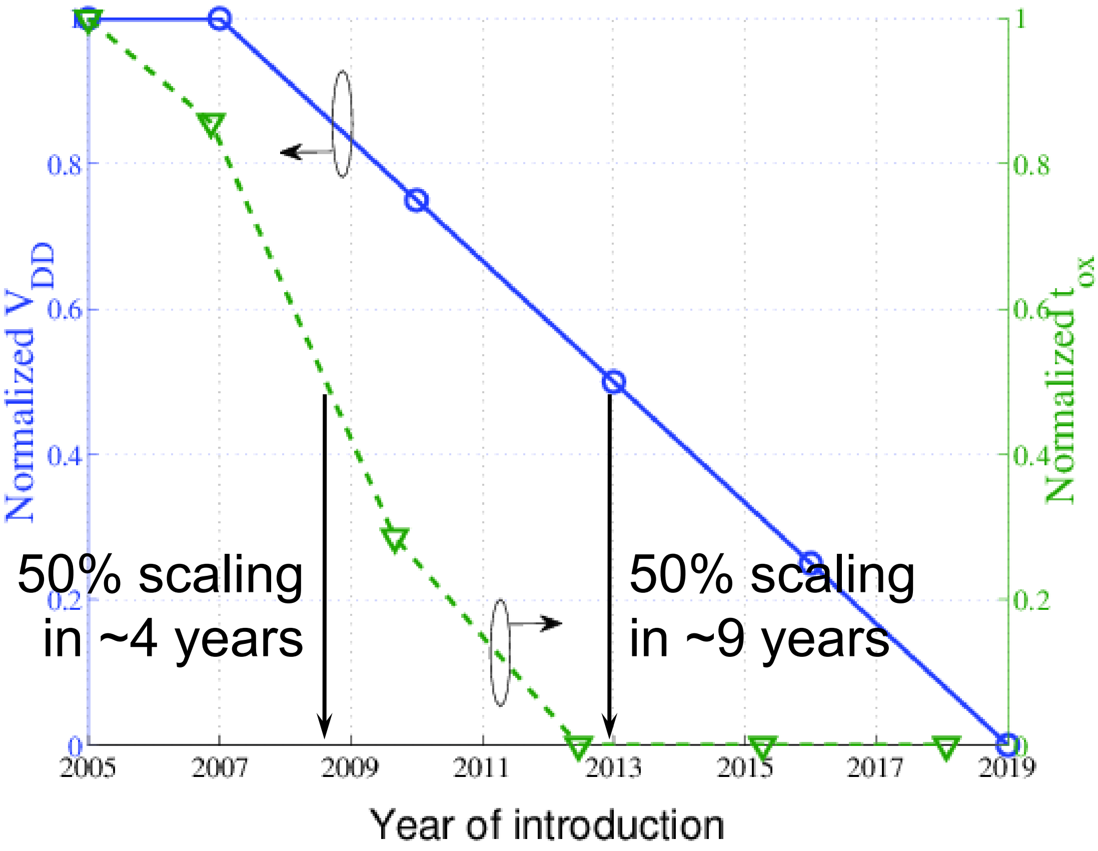


Source: A. Majumdar. “Helping chips to keep their cool”. Nature Nanotechnology, April 2009, pp. 214-215.

Technology scaling: reliability challenges

$$MTTF \propto k \exp\left\{ {}^{E_a}\!\big/_{kT} \right\}$$

- **Increasing temperature**
  - Electromigration
  - Stress-migration
  - TDDB, NBTI, HCI
- **Increasing electric field at the oxide**
  - NBTI, TDDB, HCI



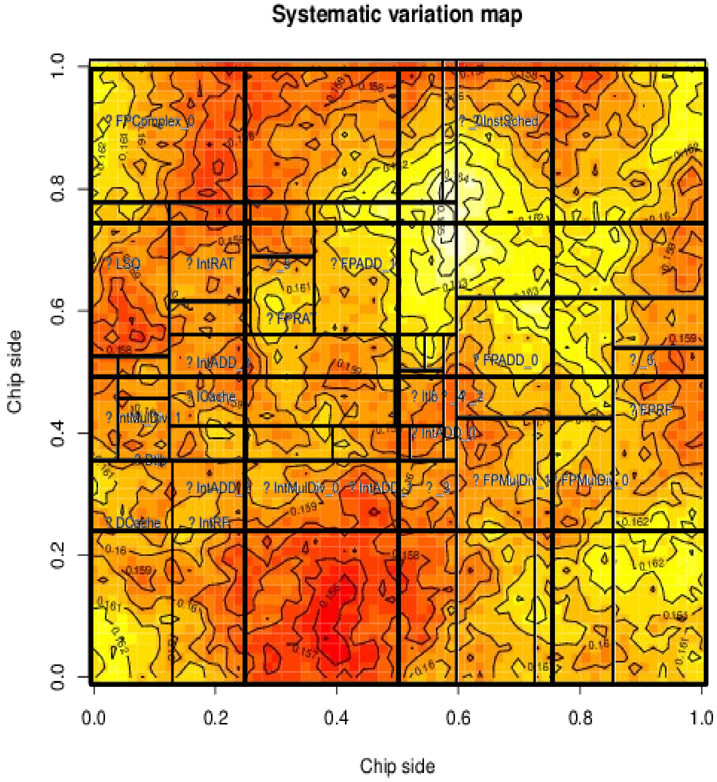
50% scaling in ~4 years

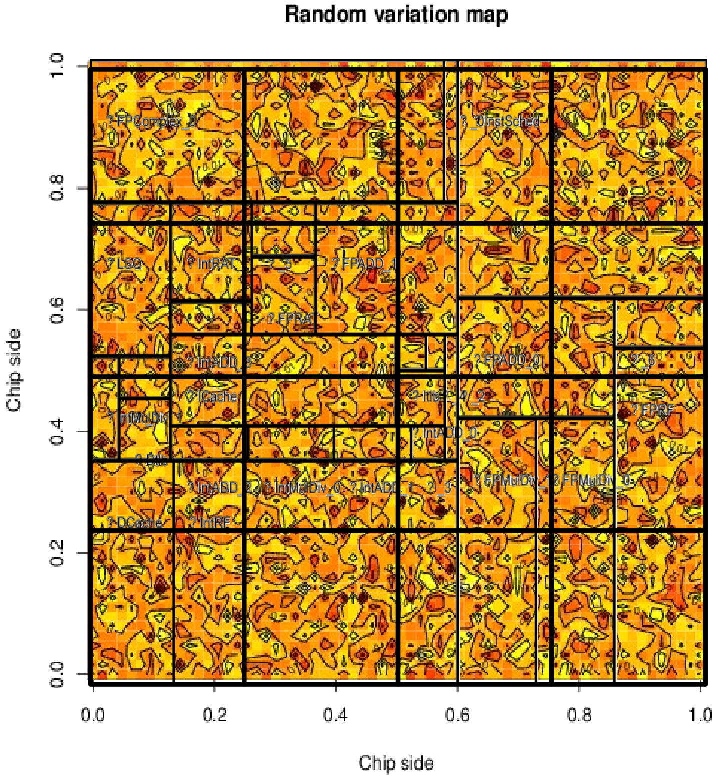
50% scaling in ~9 years

Source: MASTAR model from STMicroelectronics

Technology scaling: process variation

- **Process variability is the deviation of sensible parameters from their nominal value**
  - Systematic and random variation


Systematic variation map


Random variation map

Source: Smruti R. et al. "VARIUS: A Model of Parameter Variation and Resulting Timing Errors for Microarchitects". IEEE Transactions on Semiconductor Manufacturing (IEEE TSM), February 2008.

Design-time and run-time domains

- **Key observations from technology scaling trends**
  - Increasing complexity of VLSI circuits
    - Multi-core VS single-core
  - Reliability/performance trade-off challenges
    - Power consumption VS temperature
  - Variability "*at time 0*" → process variation
  - "*Run-time*" variability → operating conditions + aging

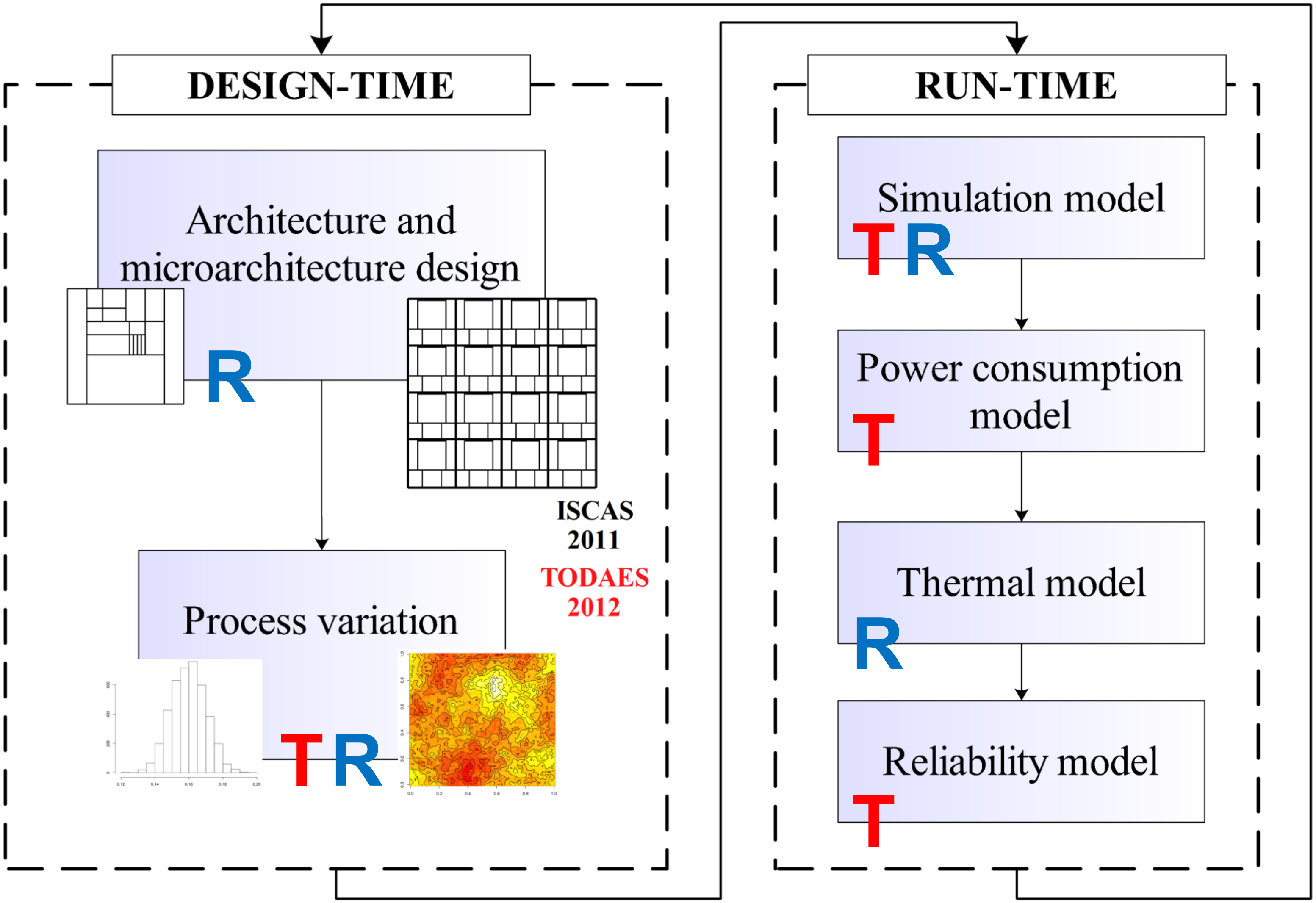
- **Need for design-time approaches**
  - To *estimate* subset of operating conditions beforehand
- **Need for run-time approaches**
  - To *adapt* to changing execution environment

Need of a proper toolset/simulation flow

Two-steps iterative approach

GLSVLSI 2012

MICRO 2012



DESIGN-TIME

Architecture and microarchitecture design

R

Process variation

T R

ISCAS 2011

TODAES 2012

RUN-TIME

Simulation model

T R

Power consumption model

T

Thermal model

R

Reliability model

T

GLSVLSI 2012

ISLPED 2012

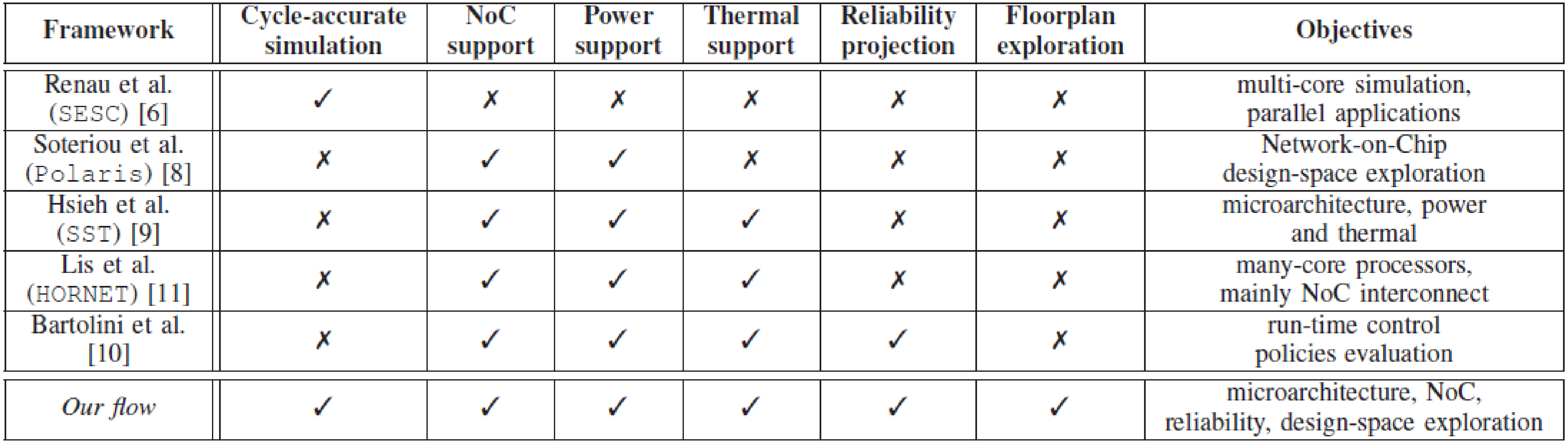
ISVLSI 2012

MICRO 2012

Simulation flows

- **Simulation flows are essential for system-level analysis of design choices**
  - Power/performance metrics
  - Thermal/reliability projections
- **Advantages of simulation flows**
  - Scalability and flexibility
  - Design space exploration
- **Drawbacks of simulation flows**
  - Computational requirements
  - Models are required for different metrics, e.g. power

State-of-the-art

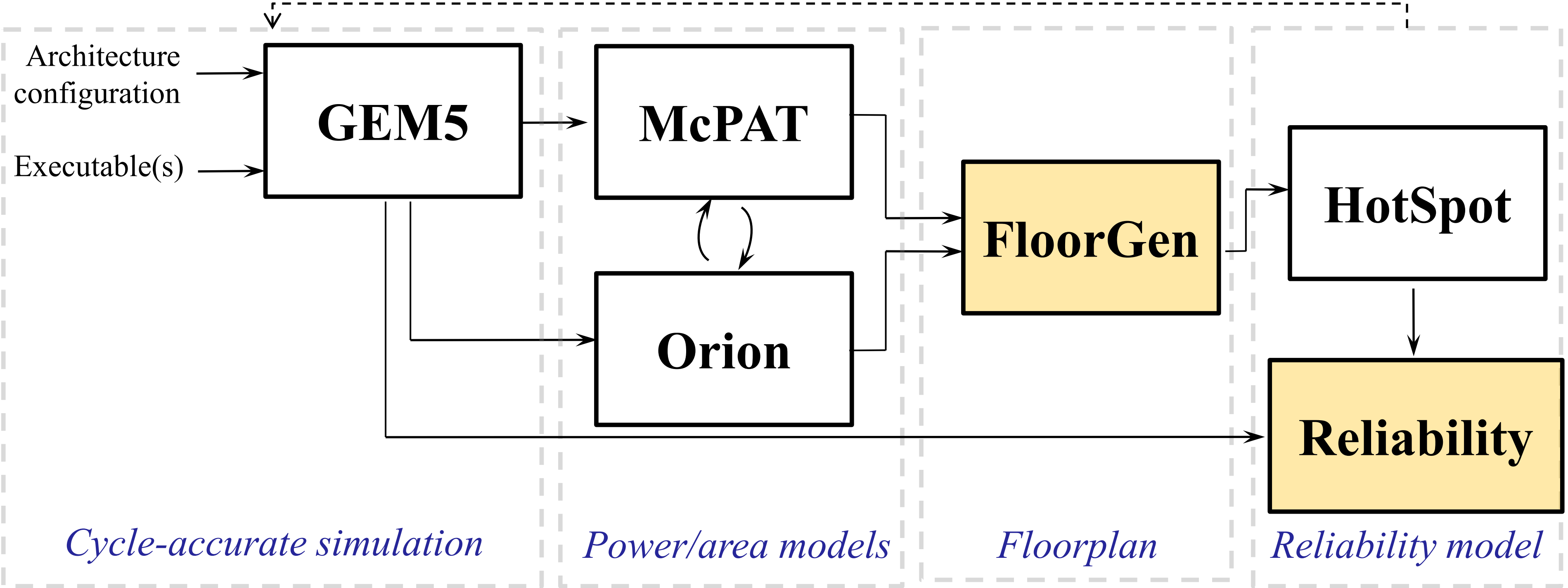


Framework	Cycle-accurate simulation	NoC support	Power support	Thermal support	Reliability projection	Floorplan exploration	Objectives
Renau et al. (SESC) [6]	✓	✗	✗	✗	✗	✗	multi-core simulation, parallel applications
Soteriou et al. (Polaris) [8]	✗	✓	✓	✗	✗	✗	Network-on-Chip design-space exploration
Hsieh et al. (SST) [9]	✗	✓	✓	✓	✗	✗	microarchitecture, power and thermal
Lis et al. (HORNET) [11]	✗	✓	✓	✓	✗	✗	many-core processors, mainly NoC interconnect
Bartolini et al. [10]	✗	✓	✓	✓	✓	✗	run-time control policies evaluation
Our flow	✓	✓	✓	✓	✓	✓	microarchitecture, NoC, reliability, design-space exploration

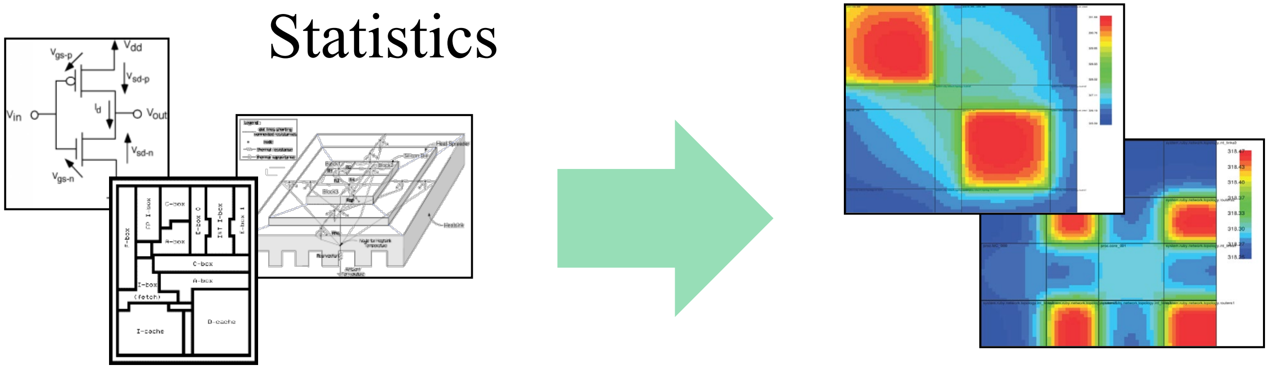
Limitations of state-of-the-art approaches

- No joint thermal/performance estimation is generally achieved
- No joint computational + communication power is achieved
- Little flexibility for DSE-like analysis (e.g., floorplan exploration)

Design-time proposal

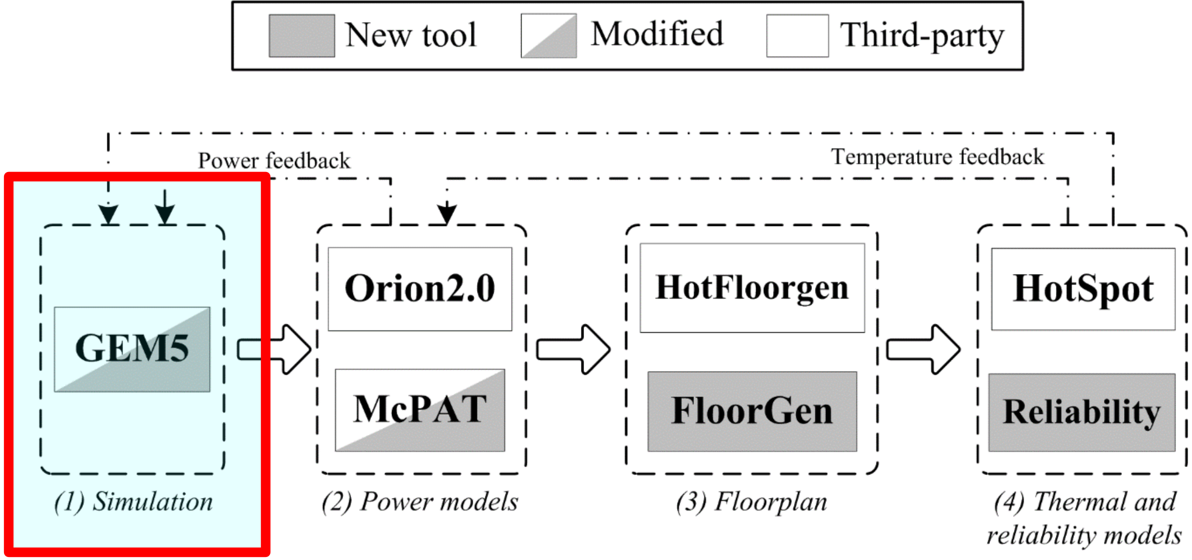


Statistics



[C1] Corbetta, S.; Zoni, D. and Fornaciari, W. "A Temperature and Reliability Oriented Simulation Framework for Multi-Core Architectures". In ISVLSI 2012.

HANDS: Performance metric



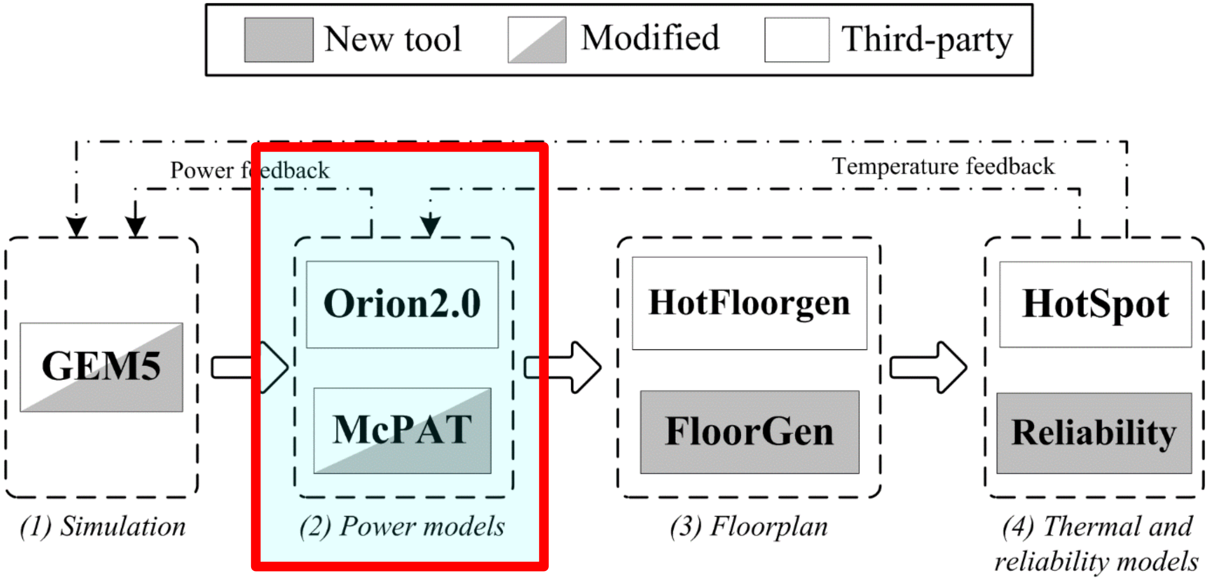
Functionalities

- Cycle-accurate simulation is required to get access statistics to microarchitecture blocks, based on GEM5 simulator, NoC model from Garnet
- Core, Router and Links information at micro-architectural level
- Functional and detailed CPU models are available, In-order and out-of-order models for different ISAs
- 3-stages pipelined detailed router model, 2D-mesh topology only

Improvements

- Out-of-order and in-order CPU models binded to the NoC simulator (2d-mesh)
- Clock toggling emulation
- OS-less shared memory hardware primitives support

HANDS: Power metric



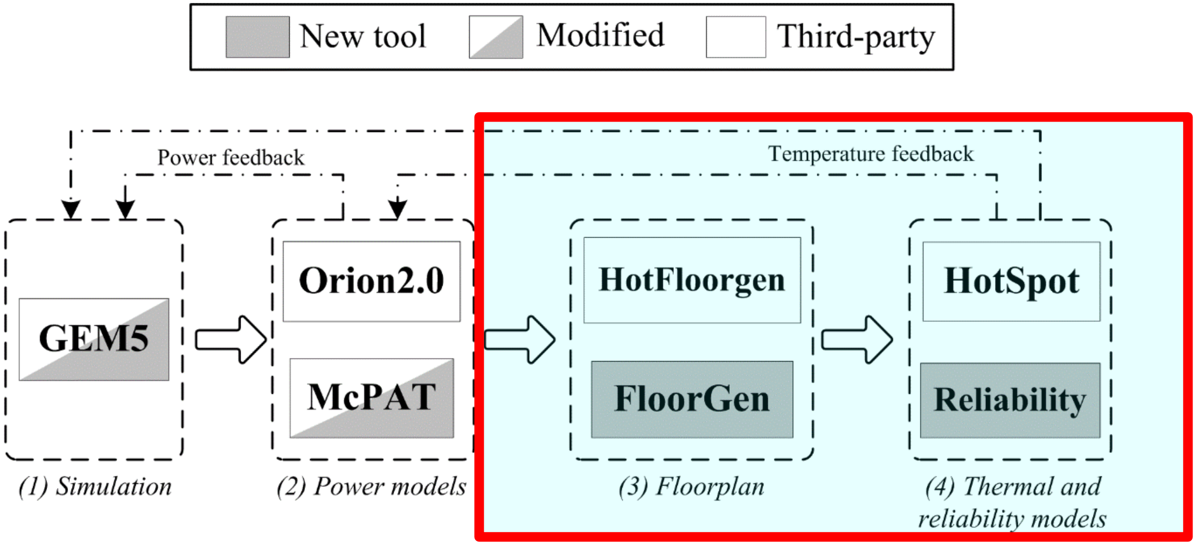
Functionalities

- Core and memory area and power estimates (McPAT)
- Router and links area and power estimates (Orion2.0)

Improvements

- McPAT leakage power discretized model refinement for full-scale temperature range
- Temperature-dependent leakage estimation
  - Temperature feedback from thermal model
- Power feedback to drive power-management decisions

HANDS: Thermal metric



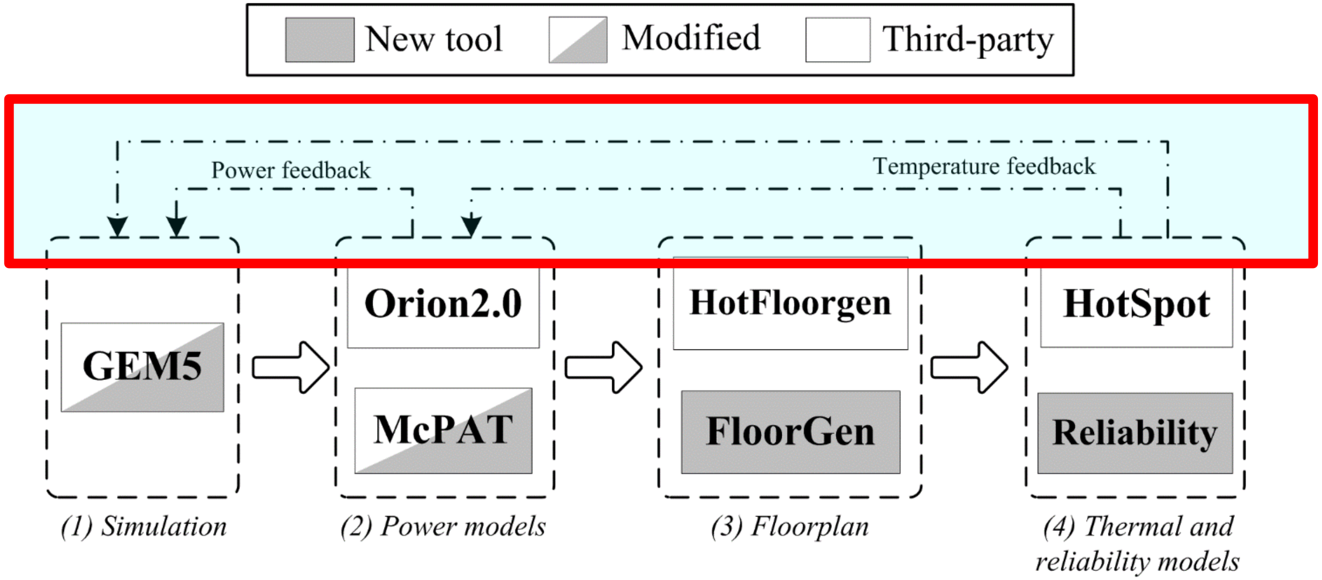
Functionalities

- Core, memory and router thermal estimates for single core-architectures (HotSpot)
- Floorplan optimization on area or power objective functions (HotFloorplan)

Improvements

- FloorGen for flexible and scalable multi-core floorplan design
  - Enables design space exploration of thermal-aware floorplans
- Core, memory and router thermal estimates for multi-core architectures
  - Multiple granularity levels, i.e. functional units (fine), routers and cores (coarse)
- Reliability metrics introduced and computed along the flow
  - Thermal-dependent MTTF metrics
  - Thermal- and access-dependent NBTI estimation

HANDS: feedback paths



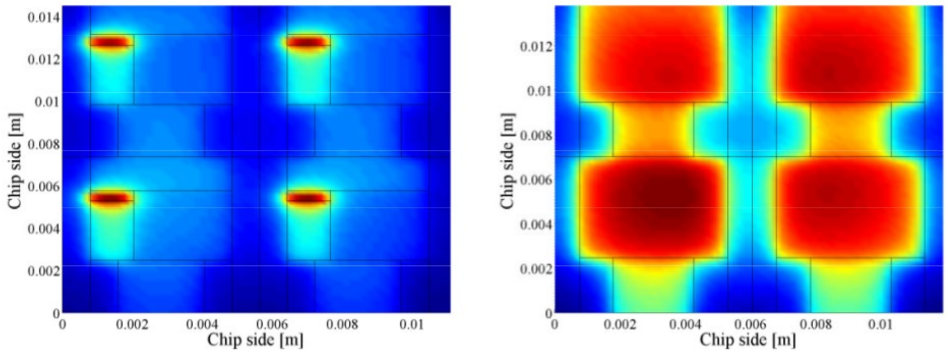
Functionalities/Improvements

- Evaluation and optimization of thermal and power management methodologies
- Block leakage power is now dependant on the block temperature itself

Example: Flexible floorplan design VS thermal

Suitable for aggressive thermal management evaluation

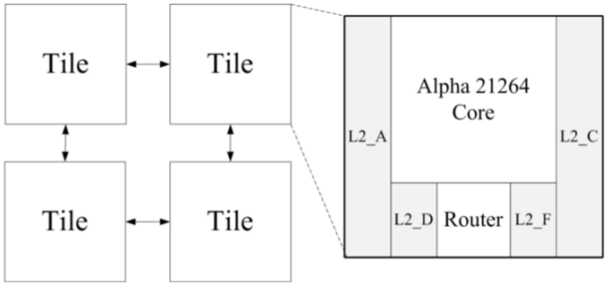
(The same is valid for the power profile also)



(a) Highest detail level

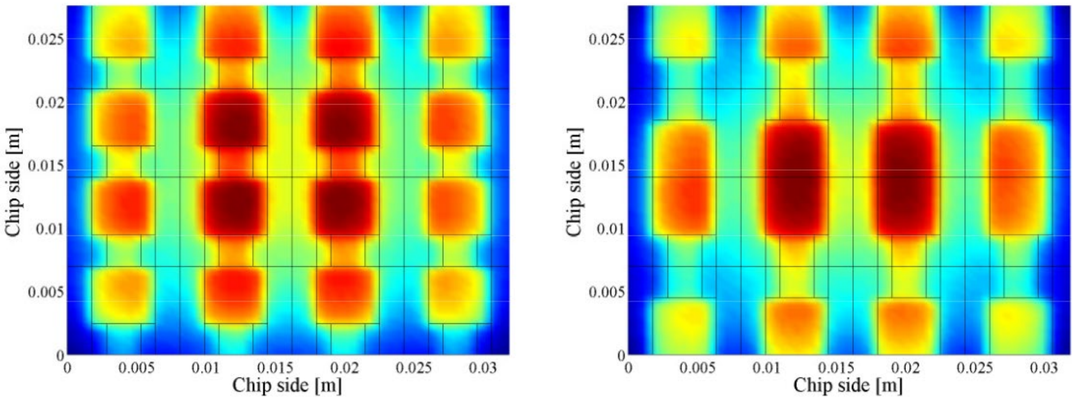
(b) Lowest detail level

Our tile logic structure



Flexible floorplan design at chip level

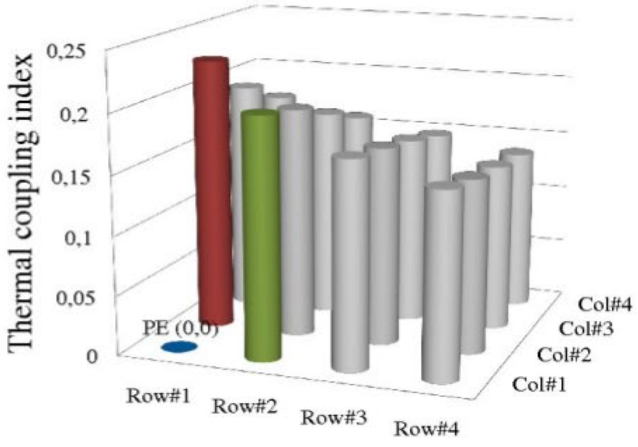
(The same is valid for power profile also)



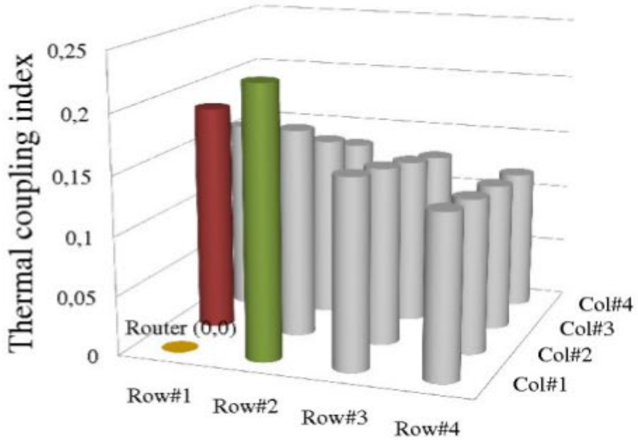
(a) Classical floorplan

(b) Floorplan with rotation

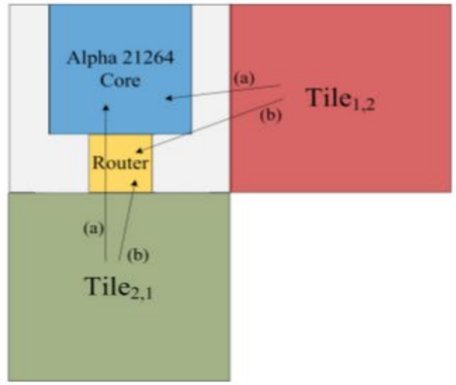
Example : Thermal coupling effects



(a) PE coupling



(b) Router coupling



(c) Configuration

Compute the effect of an isolated component on other tiles

$$\psi_{ij}(p) = \frac{T^p_{ji} - T^p_i}{T^p_{ji} - T_{amb}}$$

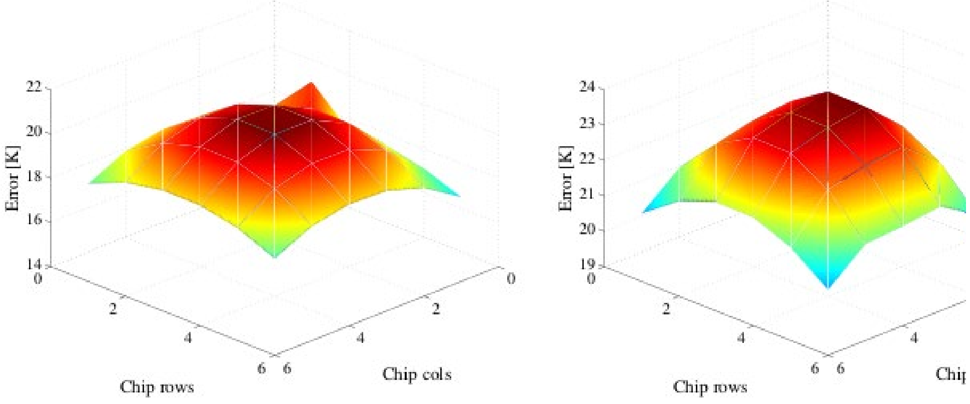
Coupling formula

T_{ji} induced temperature from j to i

T_i quite state temeprature for block I

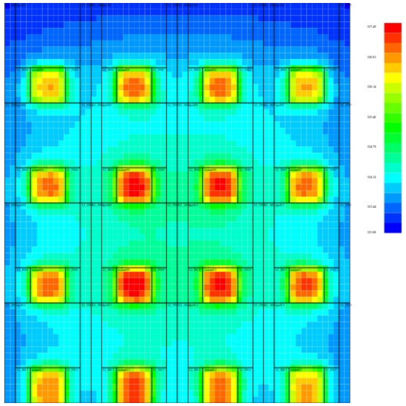
T_{amb} ambient temperature

Example : Thermal coupling effects (cont’d)

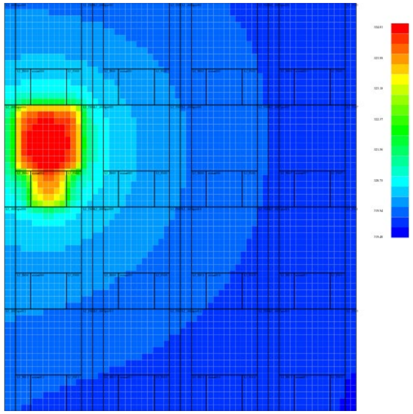


(a) 65nm

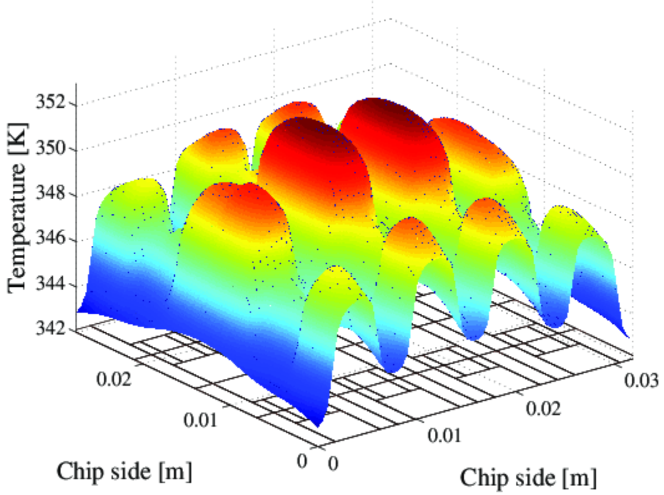
(b) 45nm



NoC only thermal impact

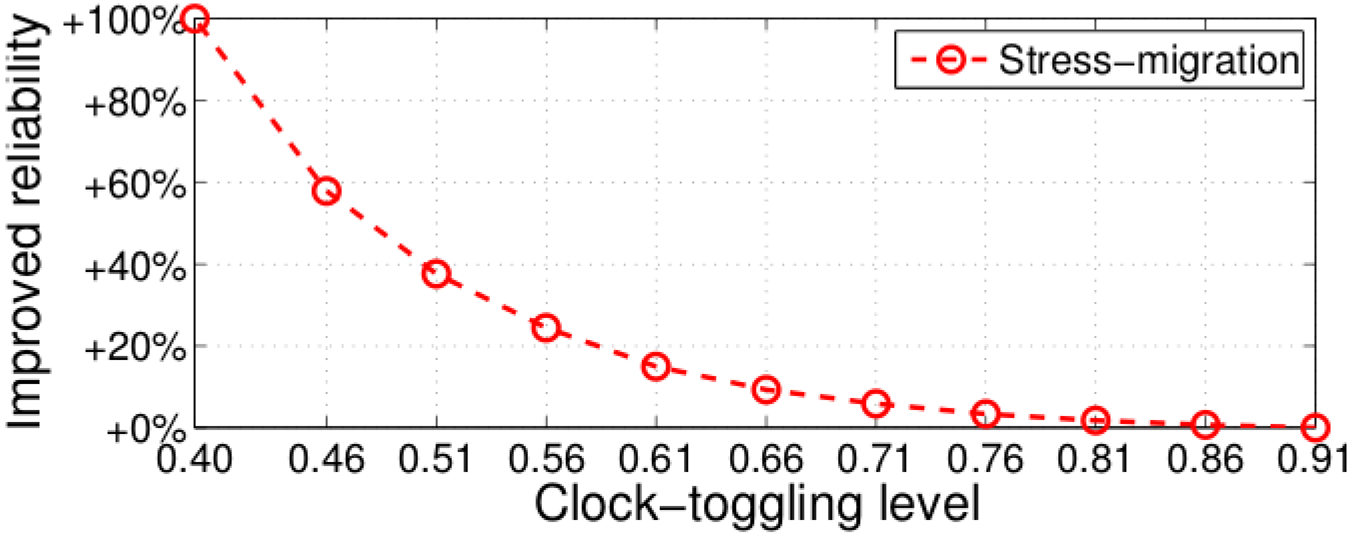


Custom tile pattern for thermal impact evaluation



Multi-core themal map

Example: Thermal/performance trade-off in NoC



$$MTTF_{SM} \propto |T_0 - T|^{-n} \cdot exp\left\{\frac{E_{SM}}{k\ T}\right\}$$

MTTF formula

k – *Boltzmann's constant*

T – *temperature*

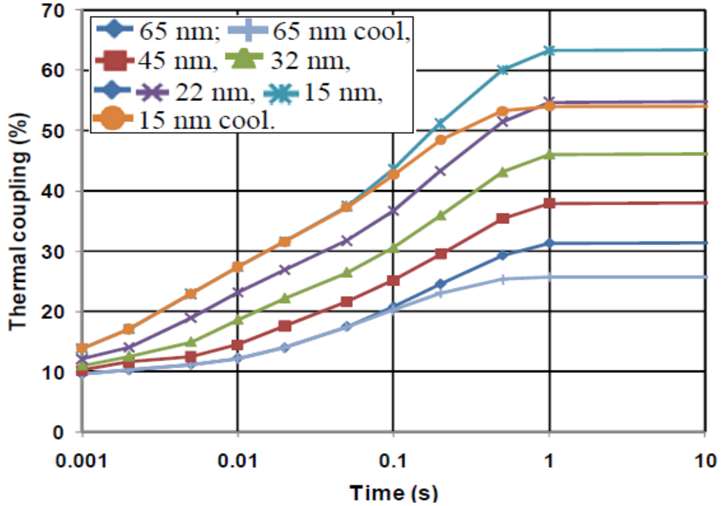
E_{SM} – *activation energy for stress migration*

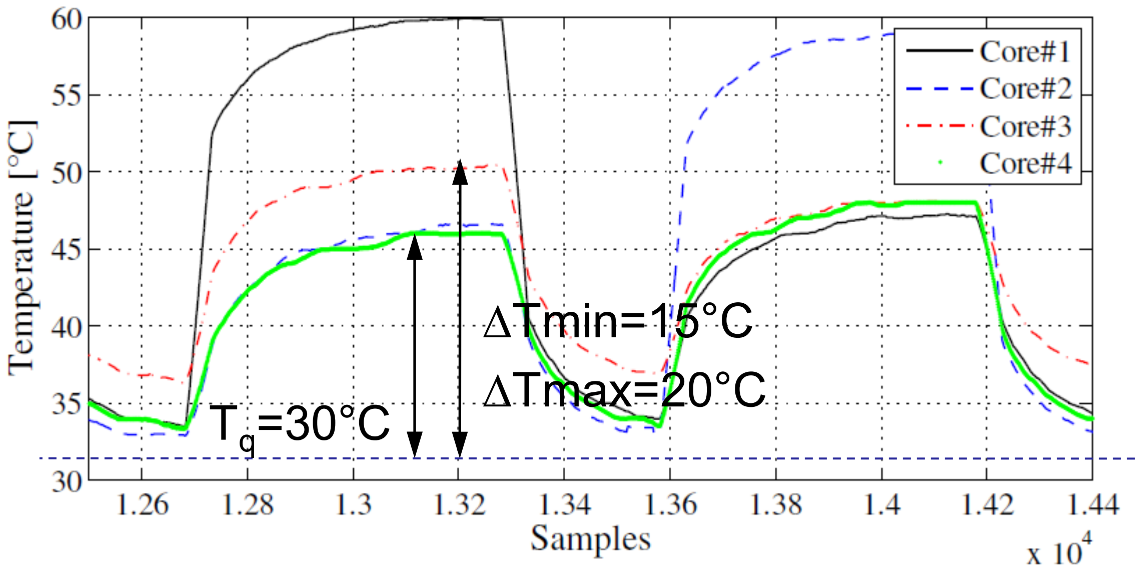
n – *technology parameter*

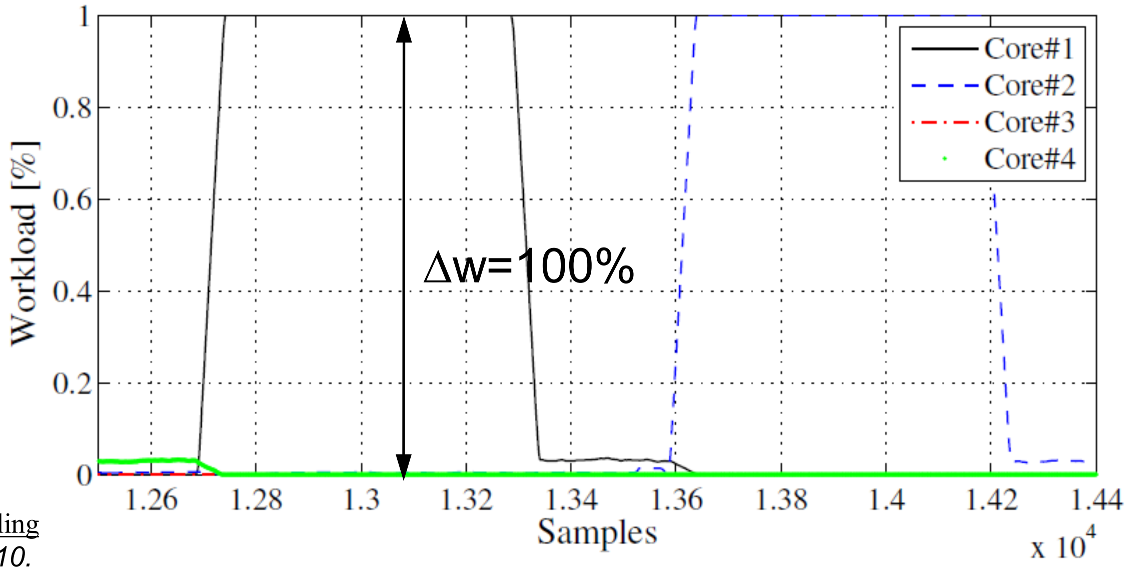
T_0 – *reference temperature*

Thermal coupling in multi-core processors

- Thermal coupling effects in quad-core Intel processor
- Effects increase with technology scaling



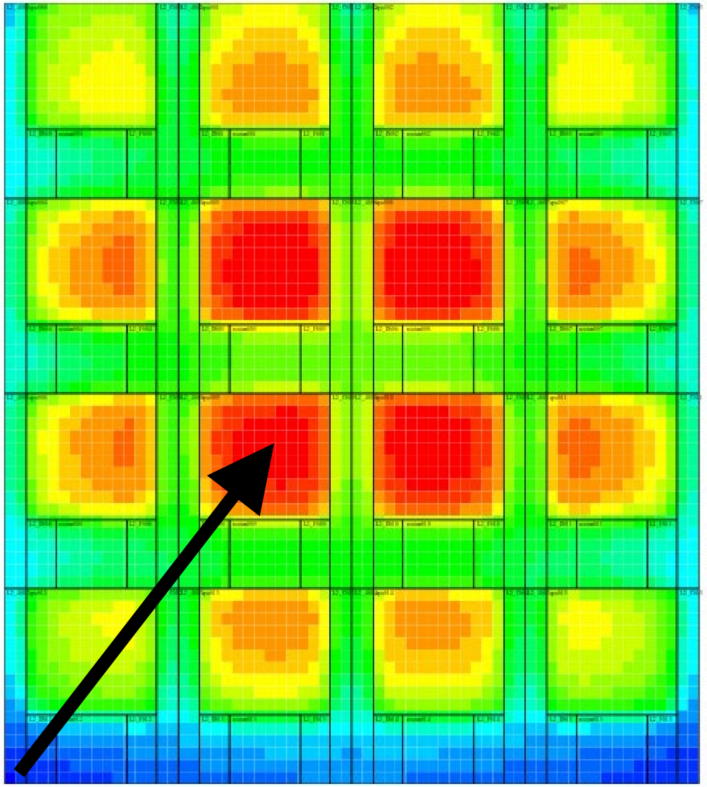




Source: M. Janicki et al. "Hot spots and core-to-core thermal coupling in future multi-core architectures". *SEMITHERM Symposium 2010.*

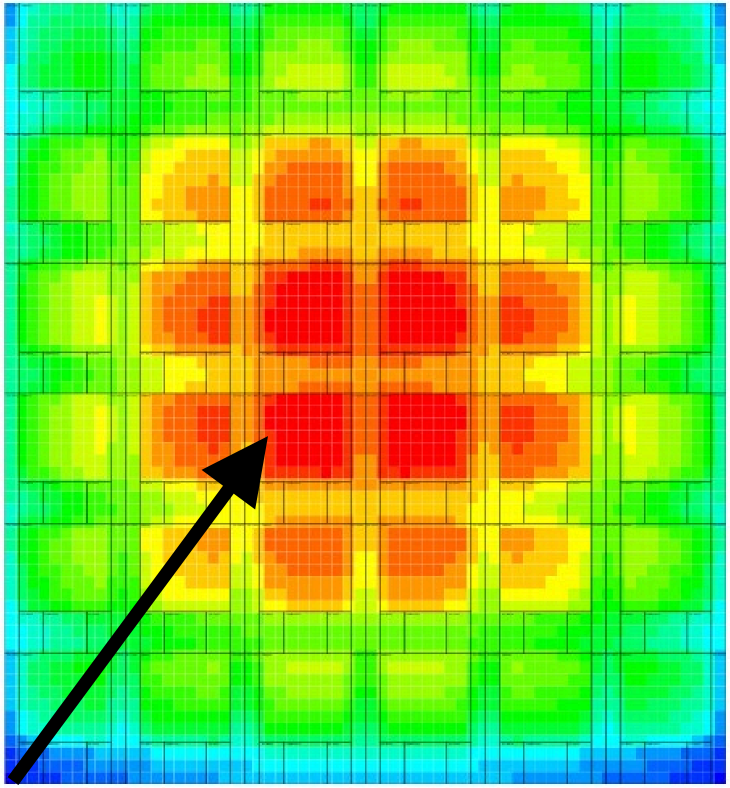
Thermal coupling in NoC

- **Computational power and interconnect power**
  - Thermal coupling at the centre of the chip!



Avg: 334.6K

Max delta: 5.8K



Avg: 382.8K

Max delta: 15.5K

Run-time proposal

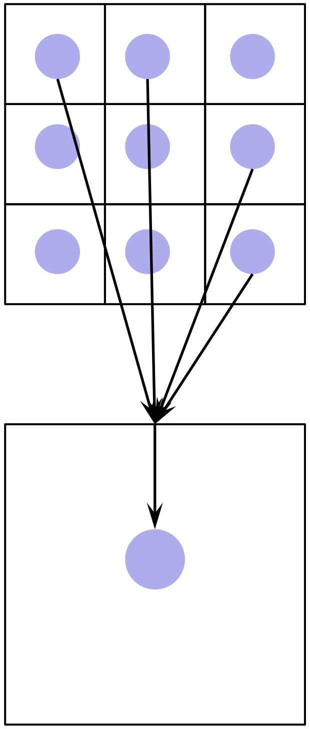
- **Thermal status as a function of different contributions**
  - Self-heating contribution (local)
  - Mutual effects contribution (system-wide)

Core-level view (local)

$$\begin{cases} g_j(t) = \zeta_{dyn} \cdot \left(\left(\dfrac{V_j(t)}{V_{MAX}} \right)^2 \dfrac{f_j(t)}{f_{MAX}} w_j(t) \right) + \zeta_{st} \cdot \exp\{-1/T_j(t)\} \\ s_j(t) = \alpha_j(t) g_j(t) + \sum\limits_{k \in N_j} \theta_{jk} \alpha_k(t) g_k(t) \end{cases}$$

System-level view (global)

$$\begin{cases} \Lambda = \dfrac{s_1 T_1 + s_2 T_2 + \ldots + s_N T_N}{T_1 + T_2 + \ldots + T_N} = \dfrac{1}{NT_p} \sum\limits_{j=1}^{N} s_j T_j \\ \dfrac{\mathrm{d}\Lambda}{\mathrm{d}s_j} \qquad \dfrac{\mathrm{d}\Lambda}{\mathrm{d}g_j} \end{cases}$$

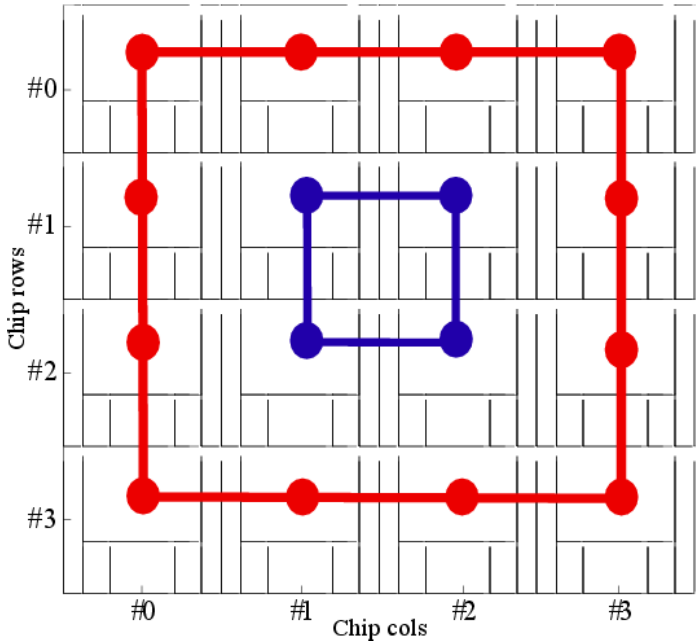


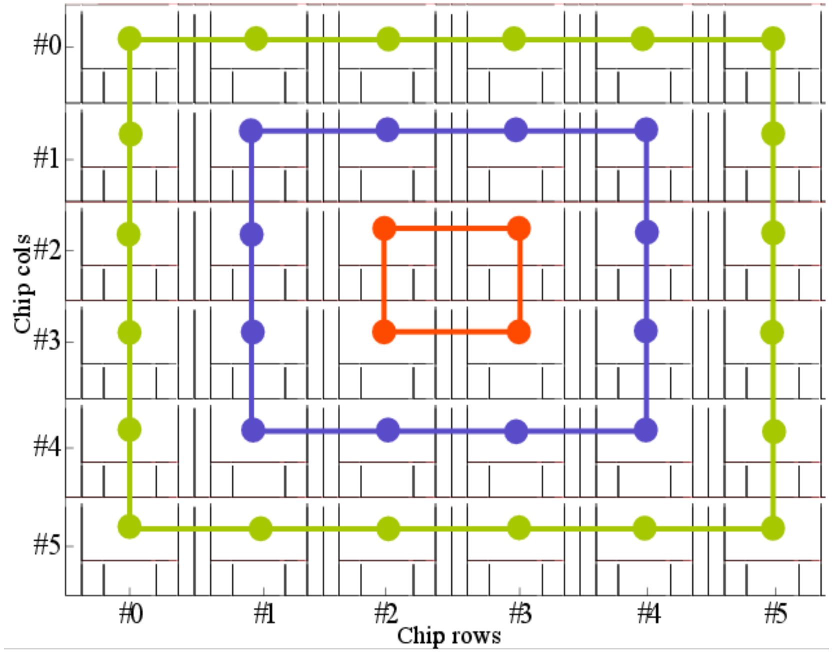
[C2] Corbetta, S. and Fornaciari, W. "Estimation of thermal status in multi-core systems". In ISCAS 2011

[J3] Corbetta, S. and Fornaciari, W. "Exploiting thermal coupling information in MPSoC dynamic thermal management"

Run-time proposal (cont'd)

- **Performance/reliability trade-off in NoC architectures**
  - Control thermal profile through clock-toggling mechanisms
  - Trade-off reliability versus performance





[C4] Zoni, D.; Corbetta, S. and Fornaciari, W. "Thermal/Performance Trade-off in Network-on-Chip Architectures". In SOC 2012

NBTI mitigation

NBTI mitigation at architecture level in literature

- Input vector control has been suggested in **[4]**
- Processor idle periods are used to recover NBTI stress **[5]**

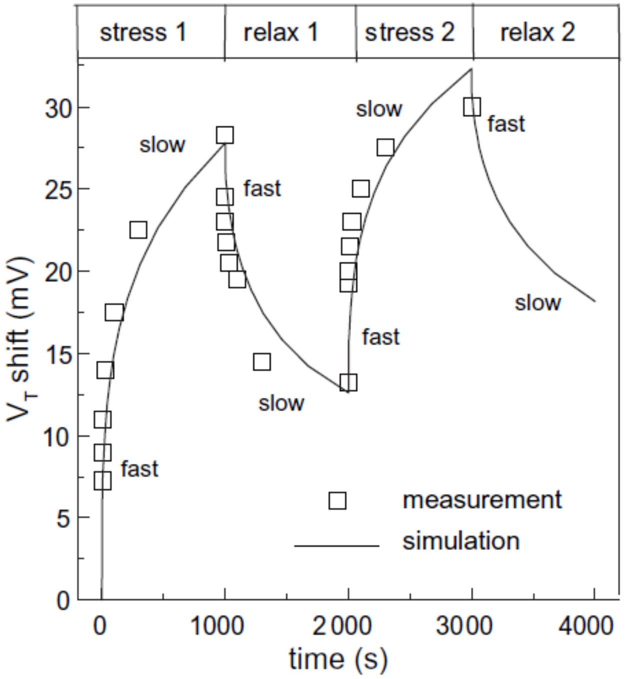
Our contributions

- NBTI-aware dynamic instruction scheduling strategy for superscalar processors (**GLSVLSI2012**)
- Design-time analysis of NBTI mitigation strategies, with process variation and technology scaling (**GLSVLSI2012**)
- Power gating of functional units for aggressive NBTI mitigation
- Reliability analysis tool for MPSoC/NoC (**ISVLSI2012**)

[4] Y. Wang et al. "On the efficacy of input vector control to mitigate NBTI effects and leakage power". *ISQED'09*, pp. 19-26.

[5] J. Sun et al. "NBTI aware workload balancing in multi-core systems". *ISQED'09*. pp.833-838.

NBTI mitigation at microarchitecture level



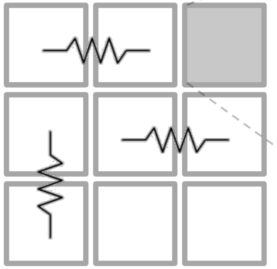
← NBTI stress/recovery phases **[6]**

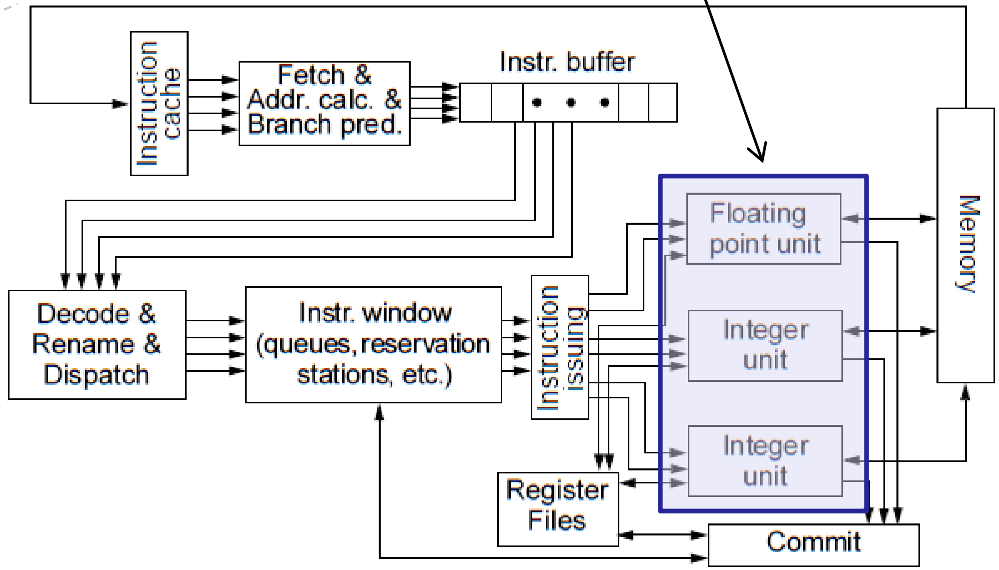
⇩

Long-term reliability model **[7]**

$$\boxed{\Delta V_{th}^{t_{ref}} \approx \left(\frac{n^2 K_v^2 \alpha C t_0 t_{ref}}{\xi_1^2 t_{ox}^2 (1-\alpha)}\right)^n}$$

Duty cycle control of FUs on superscalar processor



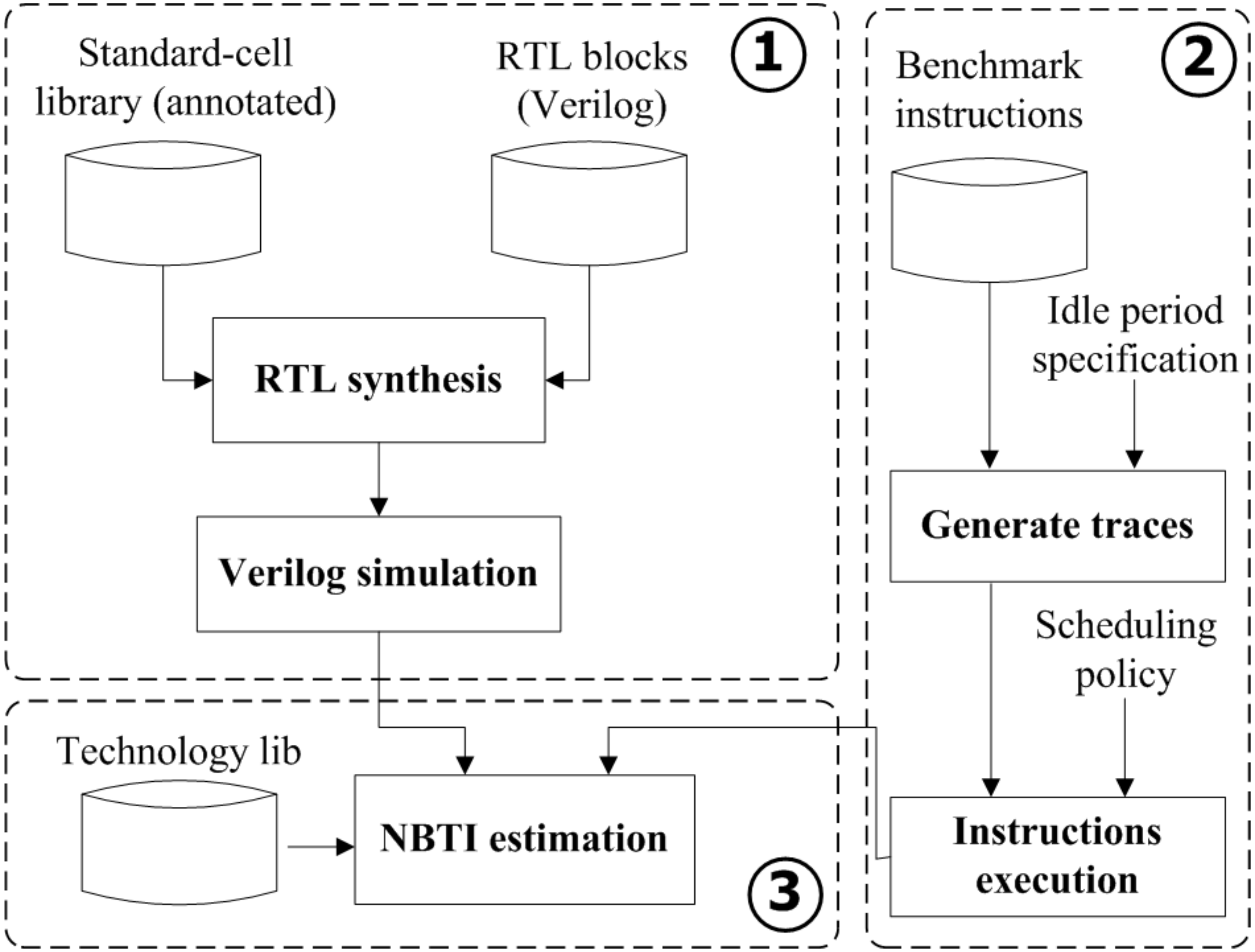


[6] M. Alam et al. "A comprehensive model of PMOS NBTI degradataion". *Microelectronics Reliability 2005*, 45:1, pp. 71-81.

[7] R. Vattikonda et al. "Modeling and minimization of PMOS NBTI effect for robust nanometer design". *DAC 2006.* pp.1047-1052.

Design-time proposal

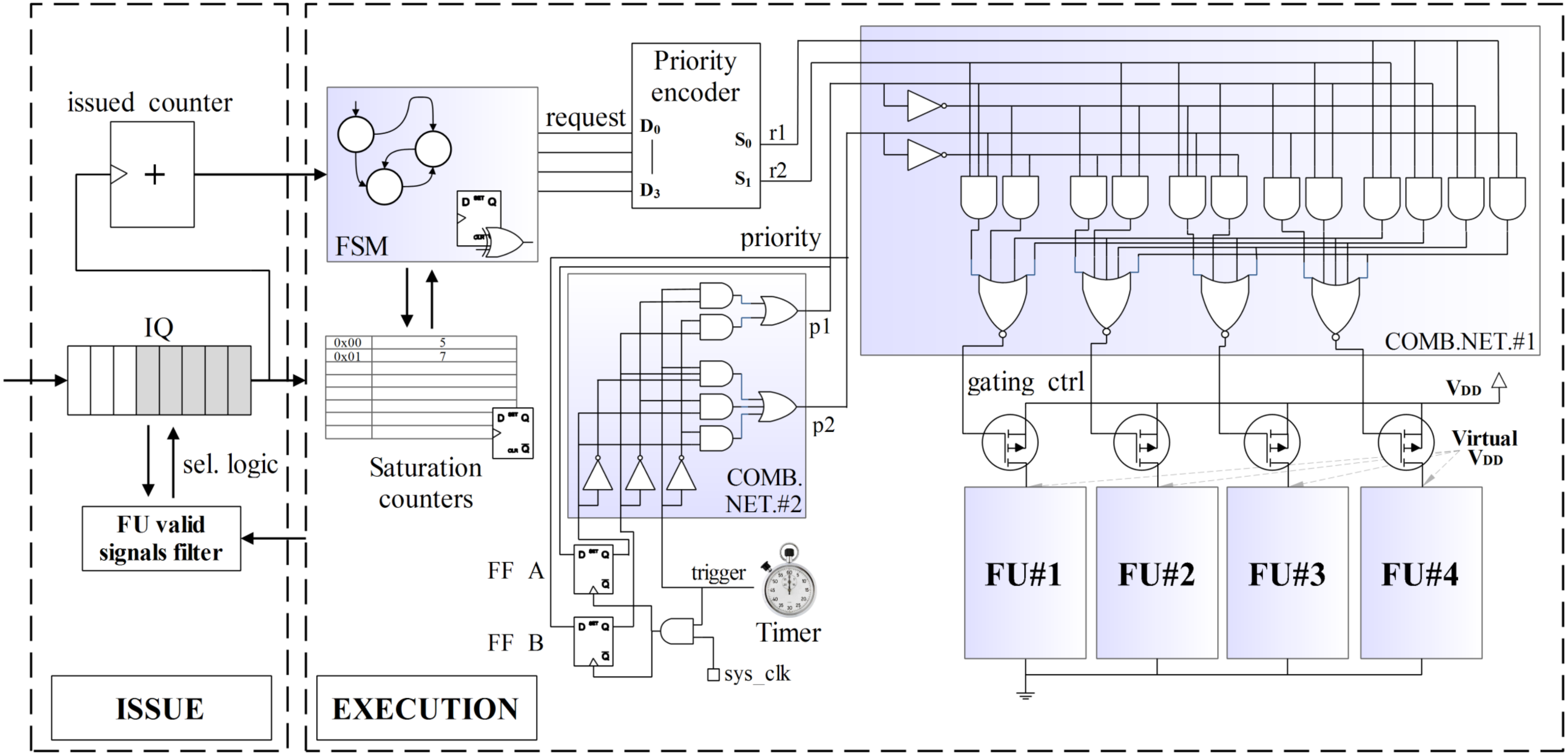
- **NBTI estimation of RTL designs**



[C5] Corbetta, S. and Fornaciari, W. "NBTI mitigation in microprocessor design". In GLSVLSI 2012

Run-time proposal

- **Proactive mitigation through power-gating**



[C6] Corbetta, S. and Fornaciari, W. "Power gating of functional units for proactive NBTI mitigation in superscalar processors". Submitted to MICRO 2012

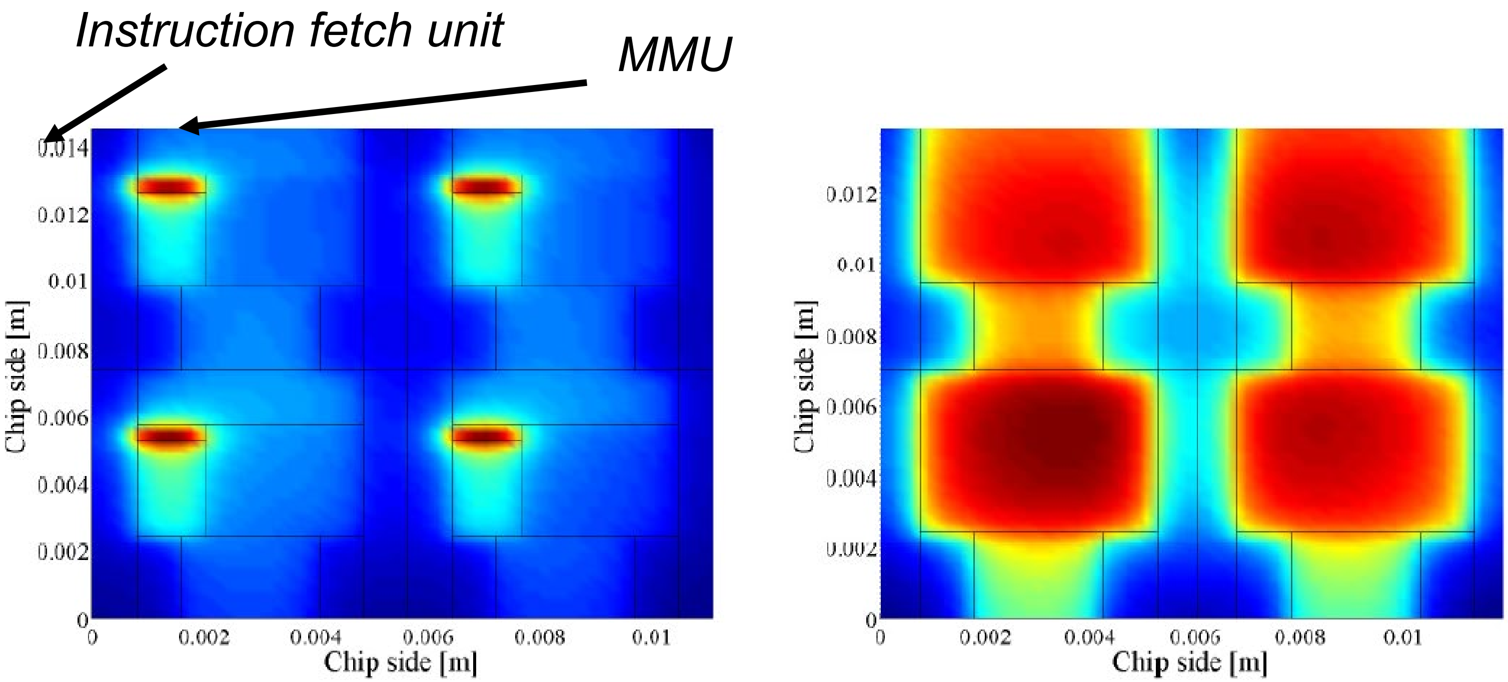
Design-space exploration

- **Objective**
  - To explore the flexibility and adaptation of the proposed flow for the estimation of sensible metrics beforehand
  - To show methodological aspects

- **Experimental setup**
  - **Processor core**: 3GHz, in-order model, Alpha-21264 ISA
  - **Functional units**: 4x Int-ALU, 4x Mul/Div, 4x FP-ALU
  - **Memory hierarchy**: 64kB, split I/D L1 cache and 1.75MB per-bank shared distributed L2 cache
  - **Router**: 3-stages wormhole
  - **Topology**: 2D-mesh
  - **Technology**: 45nm at 1.1V

Thermal-aware floorplan

- **Different levels of detail can be achieved, according to desired result**
  - Be aware of the error induced by averaging process
  - Absolute error up to 8 degrees for simple example!

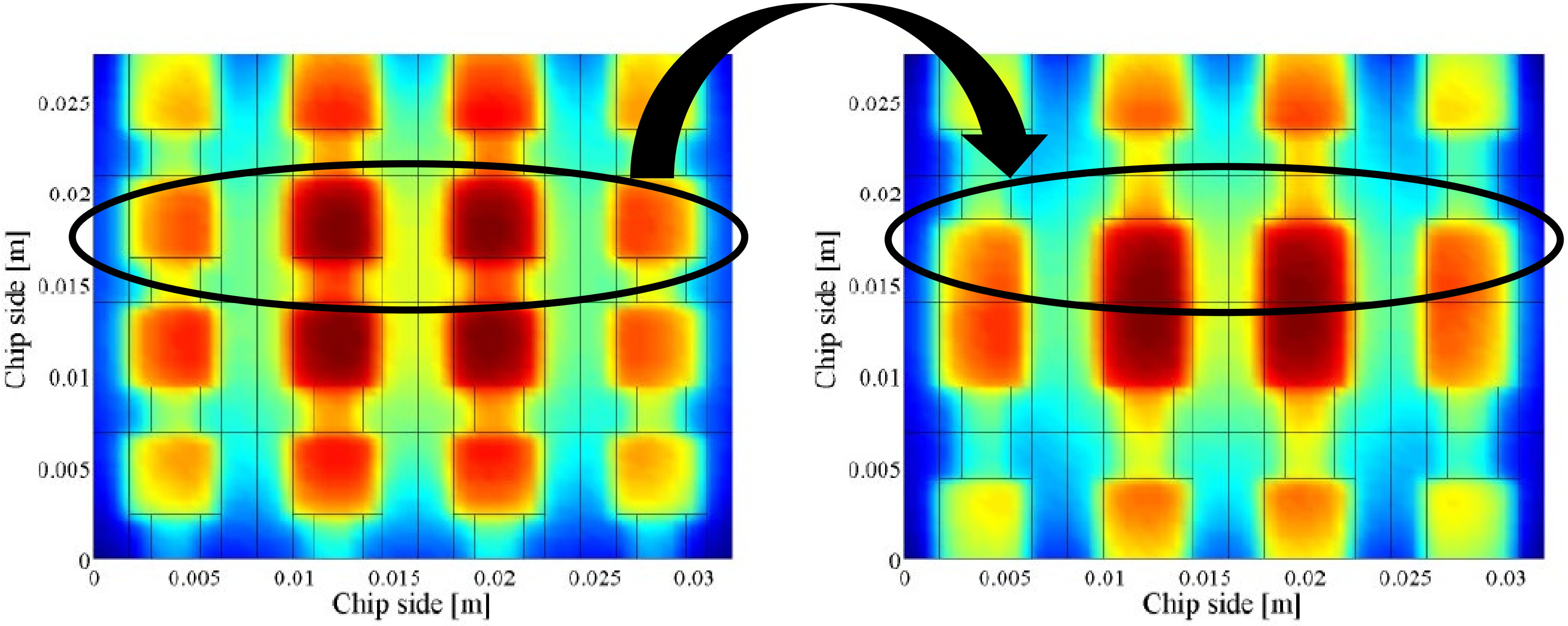


(a) Highest detail level

(b) Lowest detail level

Thermal-aware floorplan (cont'd)

- **Exploring the floorplan has never been so easy!**
  - Re-simulate the design is not required
  - Floorplan changes are applied to the input power consumption vector (under some specific conditions)

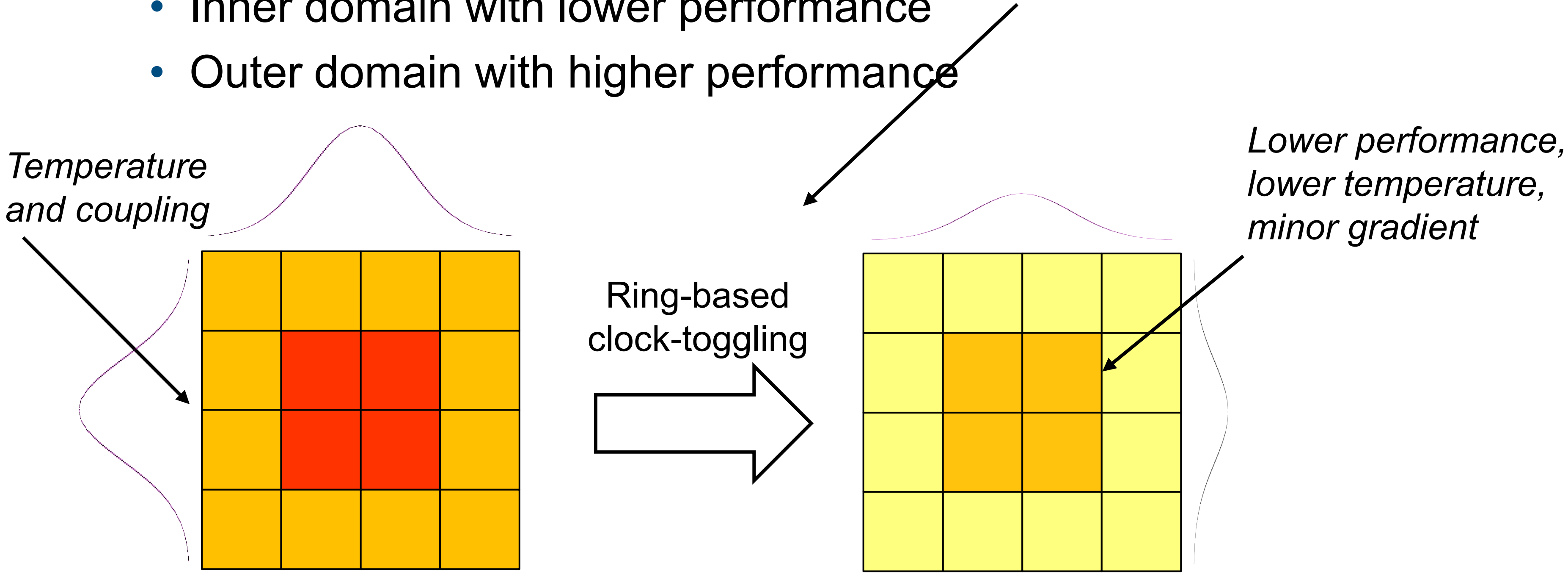


(a) Classical floorplan

(b) Floorplan with rotation

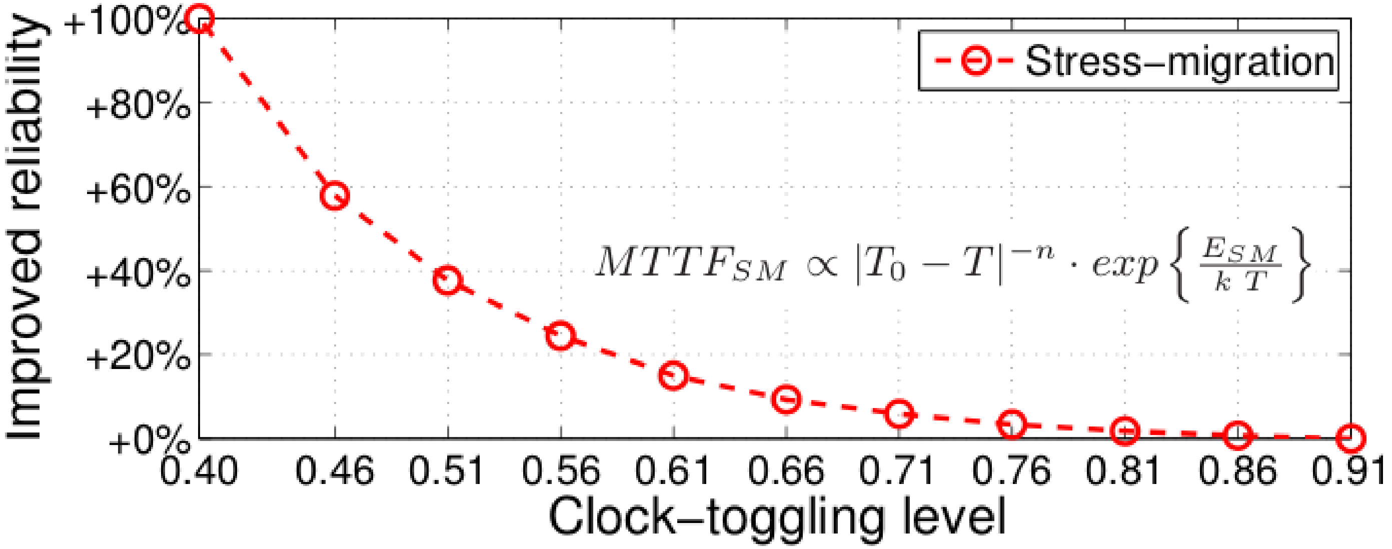
Thermal/performance trade-off

- **Clock-toggling can be employed to address thermal-induced reliability in Network-on-Chip architectures**
  - An insight of what has been developed and published **[1]**
    - Lower performance → lower temperature
  - Concentric performance domains
    - Inner domain with lower performance
    - Outer domain with higher performance


Temperature and coupling
Ring-based clock-toggling
Lower performance, lower temperature, minor gradient

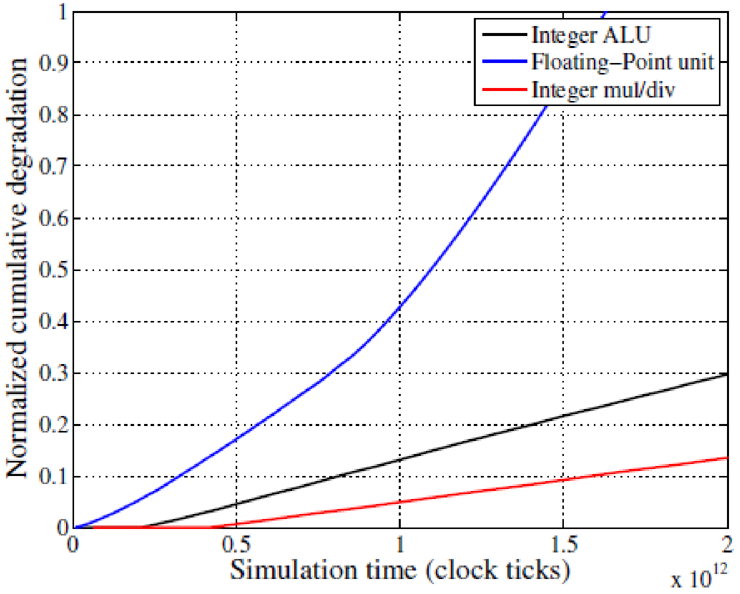
Thermal/performance trade-off (cont'd)

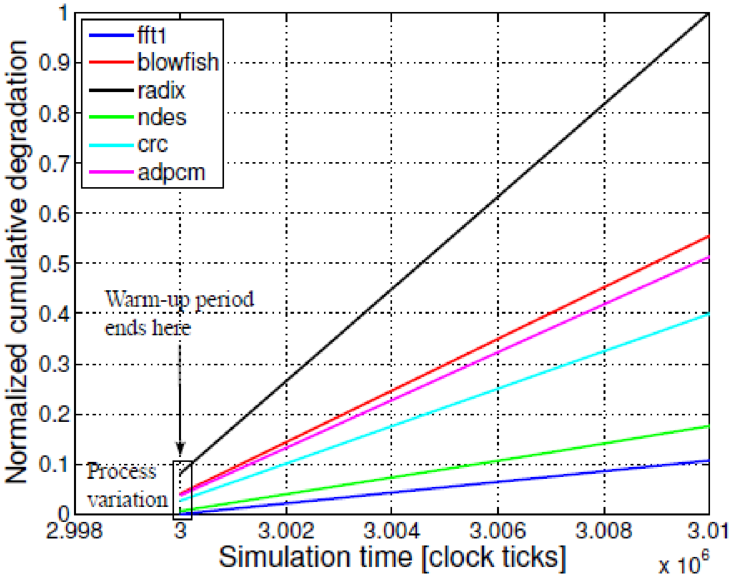
- **Lower temperature for higher reliability**
  - MTTF of stress migration is highly dependent on temperature (and thermal cycles)



NBTI estimation

- **Negative-Bias Temperature Instability (NBTI) is one of the major concerns in reliable scaled technology**
  - Is a function of temperature and usage (duty-cycle)
- **Estimation is done along the flow**
  - Considers also (random and systematic) process variation



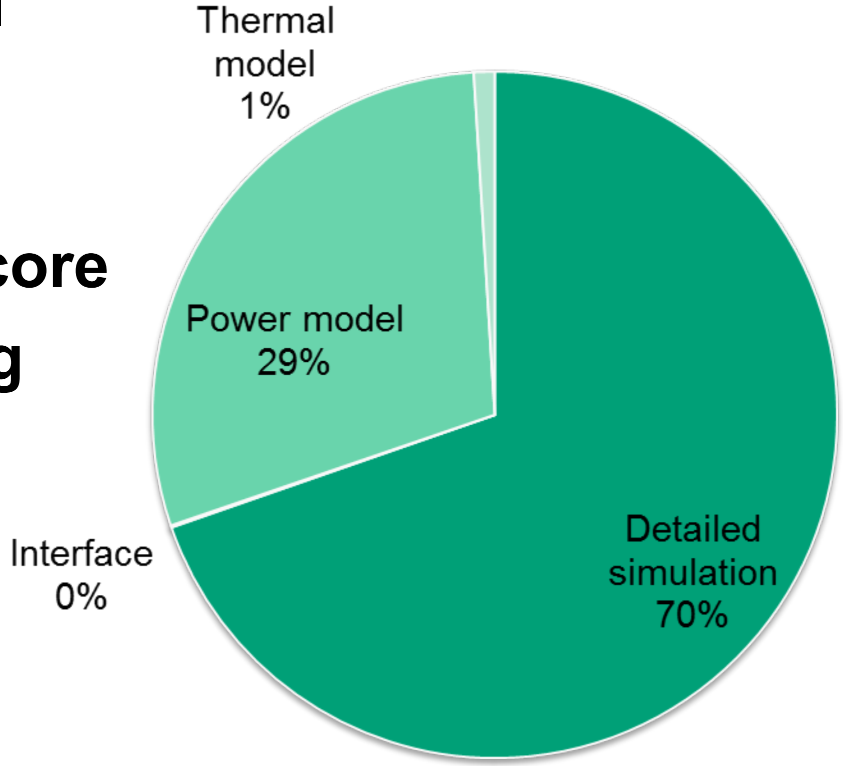


Performance – Intel core duo 2.15 GHz (1/3)

- **Time requirements (DETAILED simulation)**
  - In-order model of the CPU
  - Host instruction rate
    - 52kIPS simulating 16-cores processor (total of 8h of simulations while executing 1B instructions *EACH CORE*)
    - 26kIPS simulating 64-cores processor
  - Host memory usage
    - Depends on bench requests→ run different simulations in parallel
      - 4-cores up to 1GB almost all wcet, splash2 and SPEC-CPU
      - 16-cores up to 2GB
- **Complete flow example (64-cores, 0.5B instructions)**
  - Detailed simulation: 1560s
  - Power model: 655s   Thermal model: 22s
  - Interface tools: 1s

Performance – Intel core duo 2.15 GHz (2/3)

- **Simulation-Analysis Time - Breakdown**
  - **Example with 64-cores architecture, 45nm technology, 0.5B instructions grand total**



- **Thermal model 1% for the thermal map of the entire m-core**
- **Evaluate the thermal coupling requires more time, around 10% up to 16 cores**

Performance – Intel i5 2nd Gen 2.5 GHz (3/3)

- **Time requirements (DETAILED simulation, 2 vcs, 32bits ch width)**
  - Out-of-order model of the CPU
  - 16-cores 30min 1ms ~350/450 million instr/s
  - 4-cores 5min 1ms ~100 million instr/s
- **Time requirements (Synthetic traffic 32bits ch width)**
  - 2 vcs
    - 4-cores 2.5min 1ms 0.1 flit/port/cycle
    - 4-cores 6.83min 1ms 0.3 flit/port/cycle
  - 4 vcs
    - 4-cores 3.73min 1ms 0.1 flit/port/cycle
    - 4-cores 9.2min 1ms 0.3 flit/port/cycle

(few simulated milliseconds are useful for NoC and cache warm up)

Conclusions

- **A simulation flow is needed to**
  - Improve the architecture at design time
  - Analyse the impact of thermal and power management, joined with performance analysis
- **NBTI and process variation are hidden enemies**
  - NBTI is related to the temperature and recovers slowly
  - Process variation force to consider the worst case for the entire platform
    - PVT sensors can be useful (Process, Voltage, Temp.)
  - Hot spots makes worse the scenario
- **Speed of the control**
  - Fast thermal transients are hard to be catched

Lessons learnt

- **Power and thermal are related**
  - But they are not the same, thermal problems can exists even in the case of low power systems
  - DFS, DVFS are valuable actuators, the problem is to have the big picture of the thermal status during execution
  - Influence of rack and environment-> run-time management
- **NBTI**
  - When possible switch totally off system/subsystems
  - Distribute the workload to the core so to be thermal-aware
- **Process variation**
    - Exploitation of PVT sensors, if any
    - Distributed control
- **Reliability is becoming a real showstoppper**
  - For both embedded and High-end platforms

On-going and future works

- **Dynamic thermal management**
  - Exploitation of the thermal-coupling for effective DTM
  - Policy specification and integration in BBQ
- **Process variation impact on thermal profile of MPSoC architectures**
  - Design-time evaluation and run-time exploitation
- **NBTI mitigation**
  - NBTI mitigation in FPGA devices through dynamic reconfiguration

Readings on Dynamic thermal management

Dynamic Thermal Management in literature

- Application profiling and temperature history **[1,2]**
- Linear and non-linear models to predict temperature **[3]**

Aging: Aging Effects: From Physics to CAD, chapter 3, Springer

Targeting MPSoC devices

- Thermal coupling accounts for neighbors; transient information allows for proactive approach (**ISCAS2012**)
- A methodology to automatically generate thermal coupling coefficients from the target architecture
- Performance/reliability trade-off in NoC
- Reliability analysis tool for MPSoC/NoC (**ISVLSI2012**)

[1] J. Srinivasan et al. “Predictive dynamic thermal management for multimedia applications”. *ICS’03*, pp. 109-120.

[2] I. Yeo et al. “Predictive dynamic thermal management for multicore systems”. *DAC’08.* pp.734-739.

[3] R.Z. Ayoub et al. “Predict and Act: dynamic thermal managemtn for multi-core processors”. *ISLPED’09.* pp.99-104.