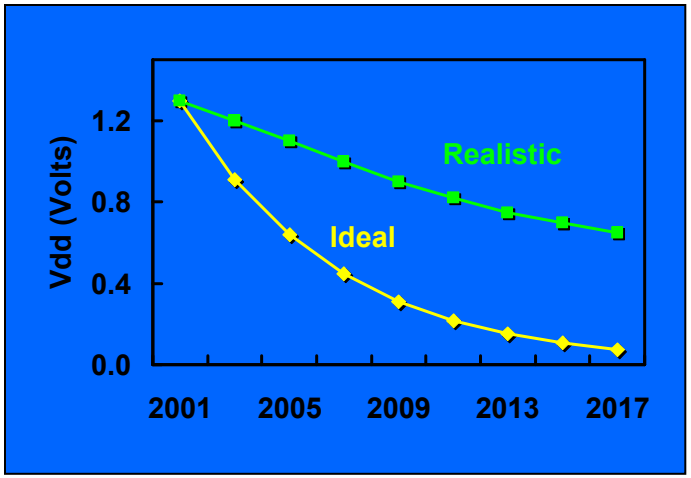
Dynamic Thermal Management in Modern Processors

Shervin Sharifi

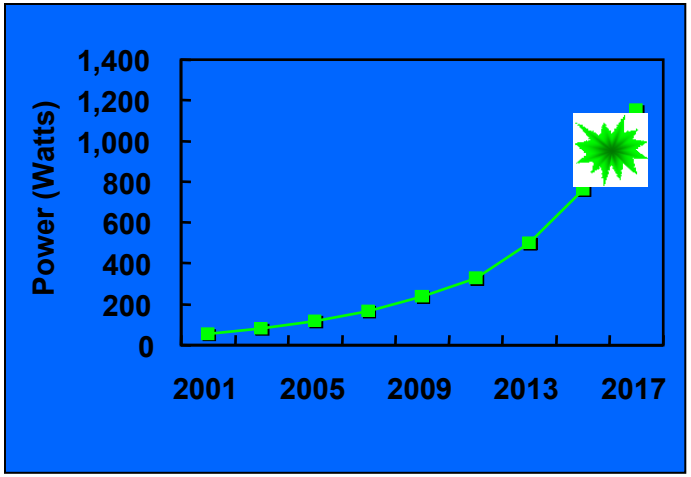
PhD Candidate
CSE Department, UC San Diego

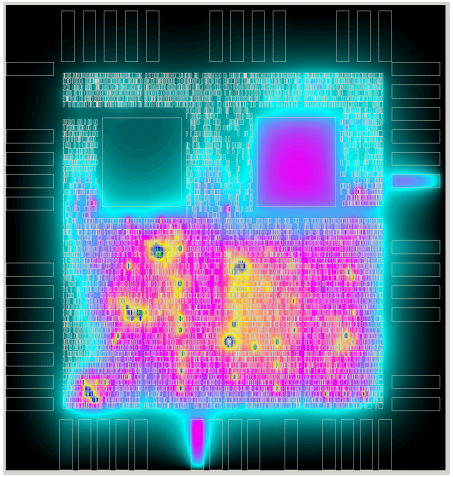
UCSD | Computer Science and Engineering

Power Outlook



- V_{dd} scaling will slow down
- Power will increase constantly
- Feature sizes decrease
- **Significant increase in Power densities**



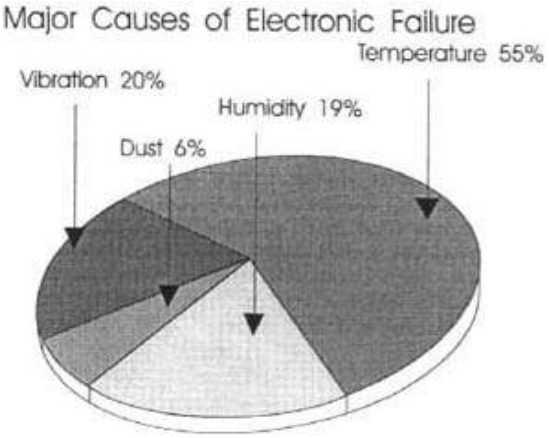


S. Borkar, "Thousand Core Chips: A Technology Perspective," DAC07

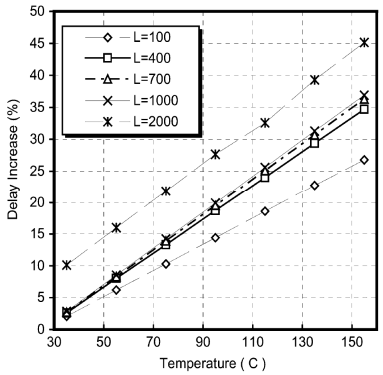
[B. Charlot & K. Torki, TIMA]

Temperature Induced Problems

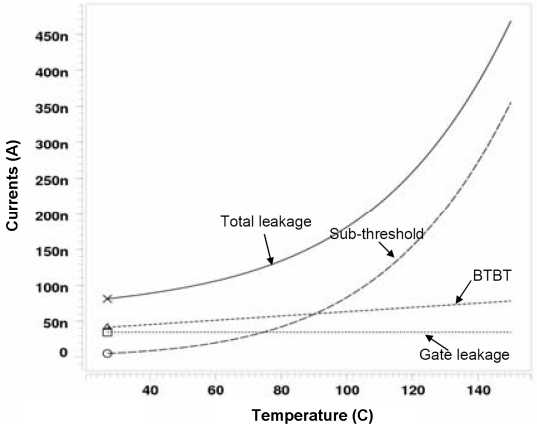
- Thermal hot spots
  - Accelerates failure mechanisms
    - Exponentially dependent on temperature (Electromigration, etc.)
  - Performance loss
  - Higher leakage power
- Spatial variations
  - Performance mismatch, Clock skew
  - Mechanical stress
- Temporal variations
  - Thermal cycling



www.aitechnology.com



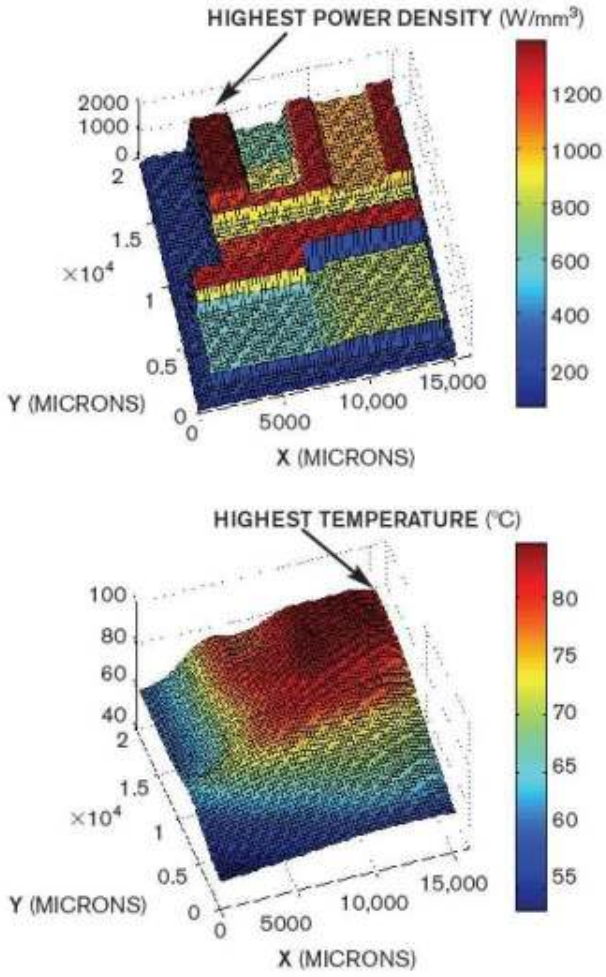
Ajami, et al, ICCAD 01



Meterelliyoz, et al, ITC 2005

Thermal Management

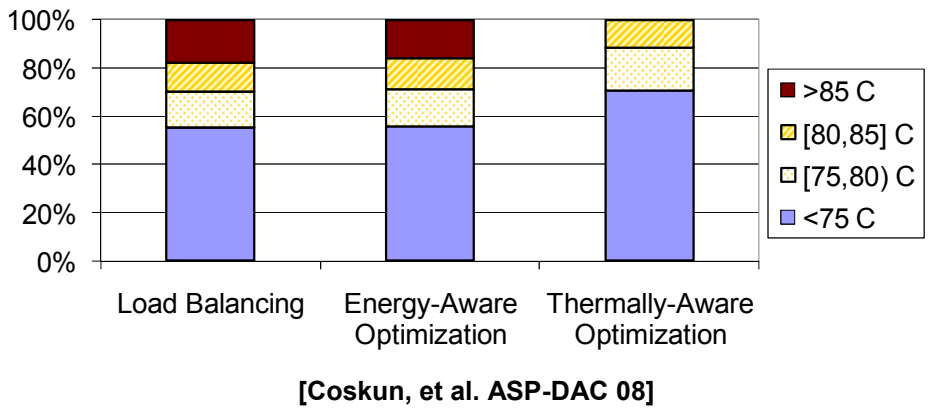
- Techniques to control the chip temperature
- Power management is not enough for thermal management
  - TM techniques are concerned with thermal hotspots and temperature variations
  - PM techniques usually concerned about the overall power consumption
- Off-chip : e.g. Cooling techniques
- On-chip : e.g. Temperature aware task scheduling
- Static : e.g. Temperature-aware floorplanning
- Dynamic : e.g. Thread migration



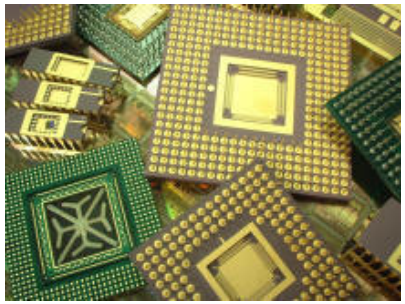
M. Santarini, EDN, Sep 2005

Dynamic Thermal Management

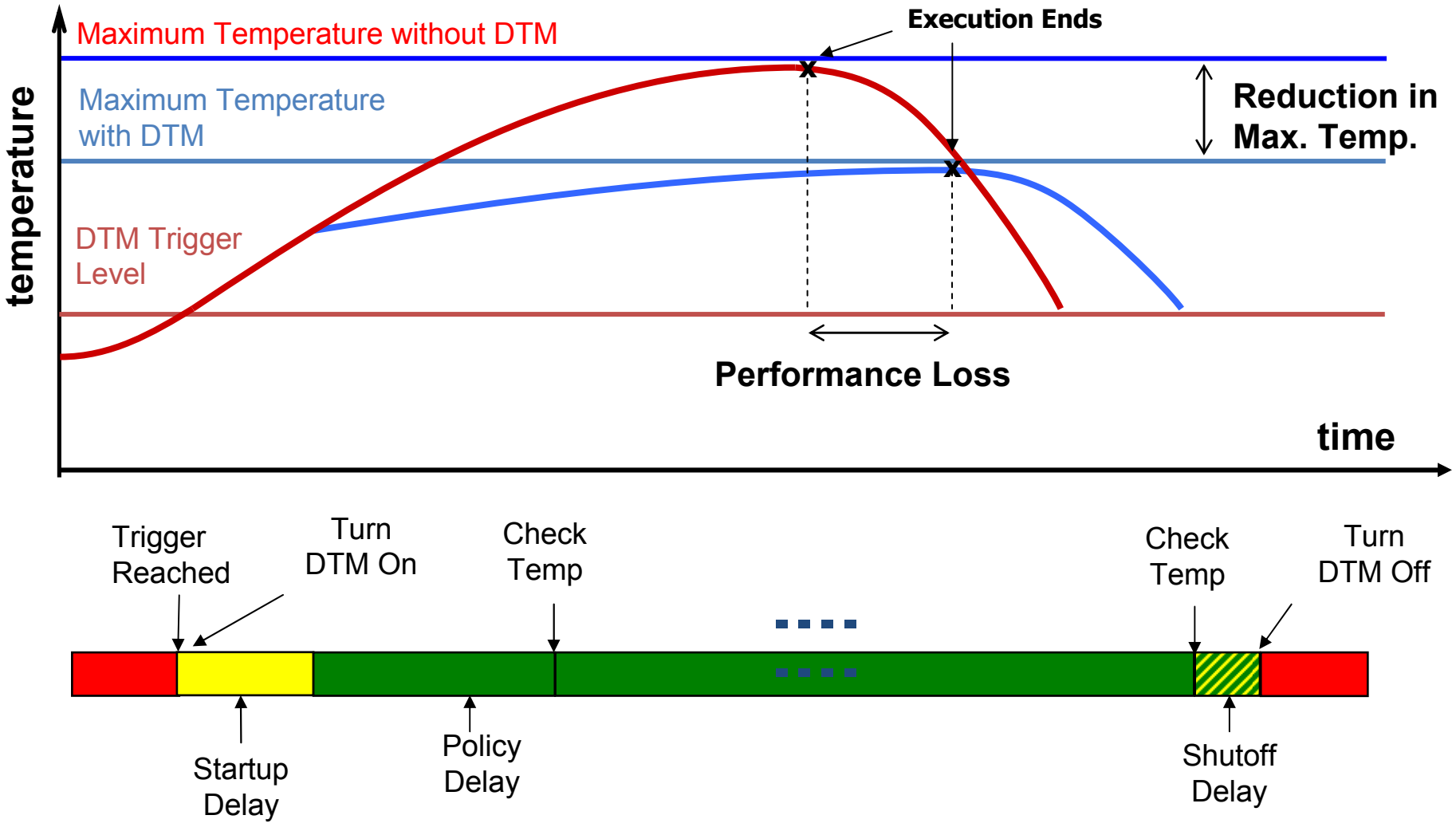
- Low-Power design techniques are not sufficient
- Goals of DTM
  - Address thermal hotspots
  - Some recent techniques also address spatial and temporal temperature variations
- Design for typical temperature instead of worst case temperature
- Achieve the highest performance under thermal constraints
- DTM techniques respond to thermal emergencies by
  - Reducing the heat generation
  - Distributing the heat generation
- DTM incurs overhead
  - Performance
  - Hardware overhead



[Coskun, et al. ASP-DAC 08]



Sequence of Events in DTM



D.Brooks, M. Martonosi, "Dynamic Thermal Management for High-Performance Microprocessors." HPCA01

Classification of DTM Techniques

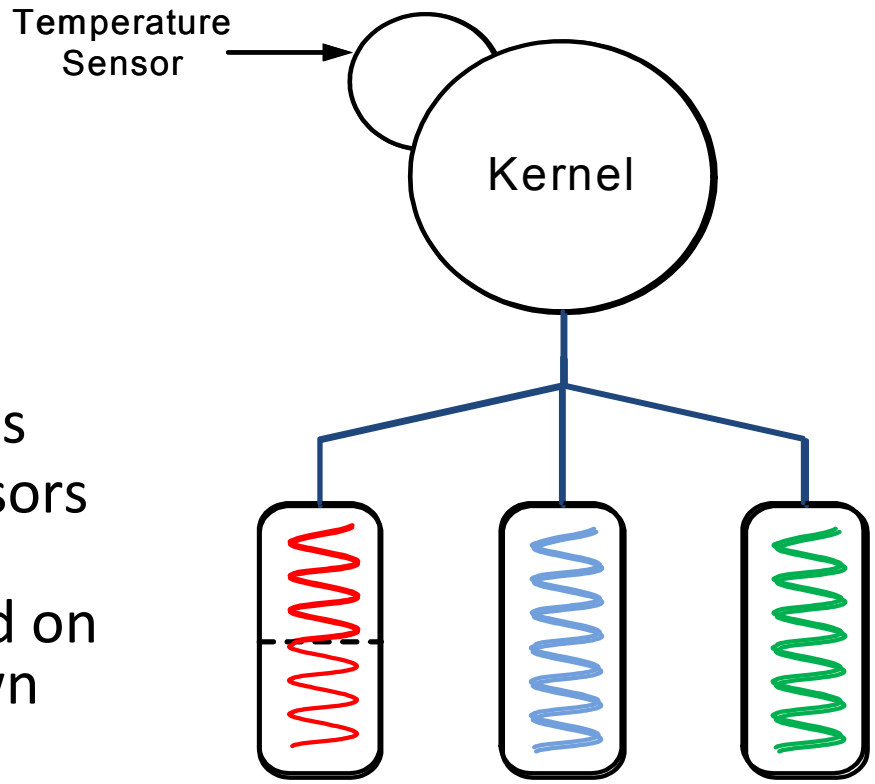
- Software
  - Implemented only in software
  - Example: Temperature aware scheduling
- Hardware throttling
  - Global
    - Hardware support only for throttling the whole chip
  - Local
    - Hardware support for throttling parts of the chip
  - Examples: Clock gating
- Hybrid
  - Combinations of previous classes

Classification of DTM Techniques

- Software
- Hardware throttling
  - Global
  - Local
- Hybrid

Temperature Aware Scheduling for a Single-threaded Processor

- OS-level scheduler for a single processor
  - Process-level control
  - Access to hardware statistics
- Reacts to the thermal emergencies detected by the temperature sensors
- Hot processes are identified based on their CPU activity and slowed down



E. Rohou,et al., "Dynamically Managing Processor Temperature and Power", Workshop on Feedback-Directed Optimization, 1999

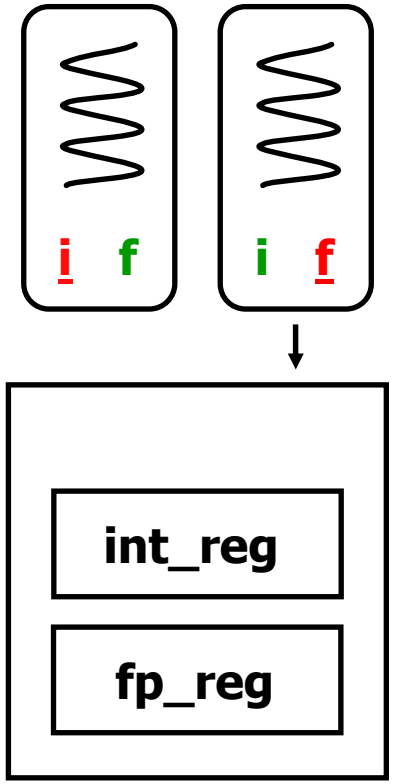
Temperature Aware Scheduling for a Single-threaded Processor

- Advantages
  - No hardware overhead
  - Only penalizes the hot processes, the rest run at full speed
  - Fine granularity in adjusting the temperature
- Disadvantages
  - Limited cooling capability
  - Reduces the performance for the most demanding jobs

E. Rohou, et al., "Dynamically Managing Processor Temperature and Power", Workshop on Feedback-Directed Optimization, 1999

Temperature Aware Scheduling for Simultaneously Multithreading Processors

- Assumption
  - Program's hotspot behavior is characterized by intensity of its accesses to int and fp register files
- Approach
  - Picks instruction from the thread that is likely to cool or less quickly heat the register files
- Hardware performance counters
  - To identify int (fp) intensive threads
  - To detect thermal danger
- To keep processor temperature below 85°C
  - 30% performance increase compared to fetch gating



J. Donald, et al., "Leveraging Simultaneous Multithreading for Adaptive Thermal Control", Workshop on Temperature-Aware Computing Systems, 2005.

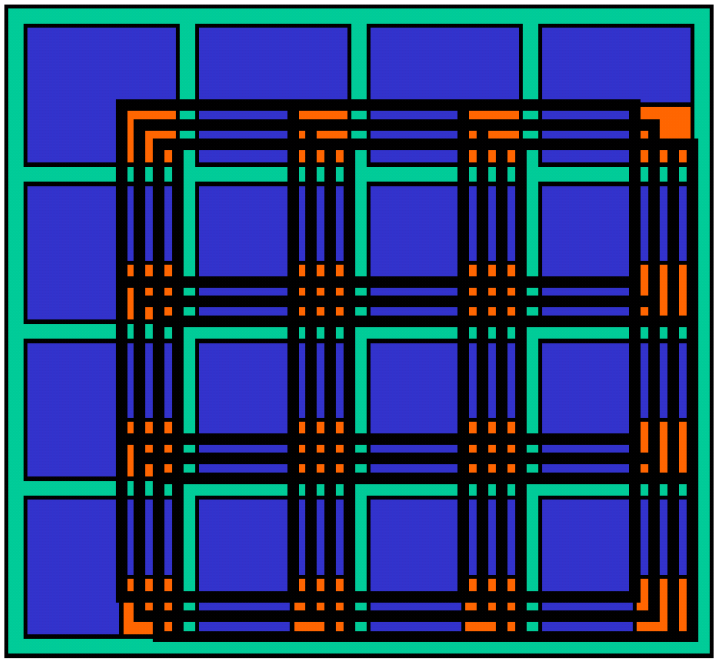
Temperature Aware Scheduling for Simultaneously Multithreading Processors

- Advantages
  - No hardware overhead
  - Doesn’t slow down the whole processor
- Disadvantages
  - Limited cooling capability
  - Works only for SMT processors
  - Reduced thread fairness

J. Donald, et al., "Leveraging Simultaneous Multithreading for Adaptive Thermal Control", Workshop on Temperature-Aware Computing Systems, 2005.

ThermalHerd

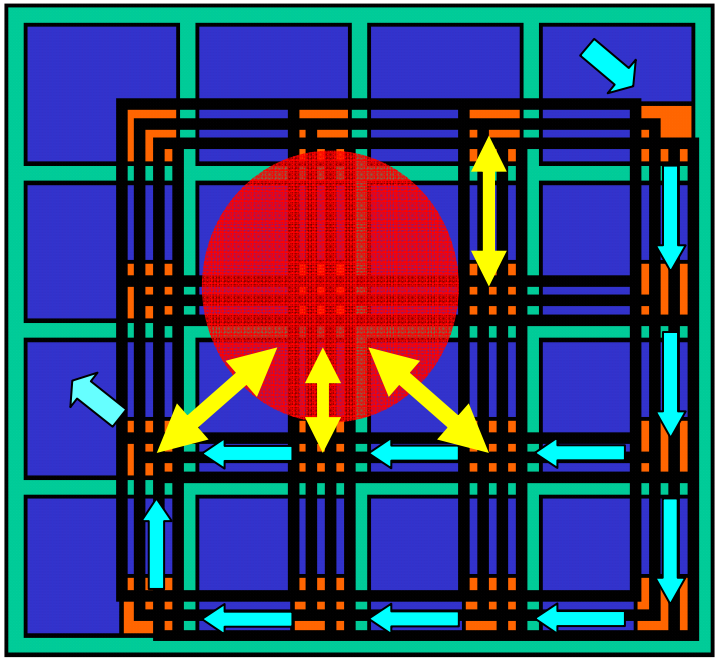
- DTM for on-chip networks
- Dynamically steers network traffic to avoid hotspots
  - Distributed traffic throttling
  - Thermal correlation based traffic routing
- Distributed traffic throttling
  - Quota reduces exponentially as temperature rises



Shang et al, “Temperature Aware On-Chip Networks,” IEEE Micro 2006

ThermalHerd

- Thermal correlation based routing
- Thermal correlation
  - The mutual thermal effect of two units
- Choose minimal path with thermal correlations to hotspot less than a threshold
- Reduced network peak temperature by 10°C
  - Throughput degradation of less than 1%
  - Latency overhead of less than 1.2%



Shang et al, "Temperature Aware On-Chip Networks," IEEE Micro 2006

ThermalHerd

- Advantages
  - No hardware overhead
  - Low performance degradation
- Disadvantages
  - Specific to the chips with on-chip networks
  - Computation and communication are treated independently, which is not optimal

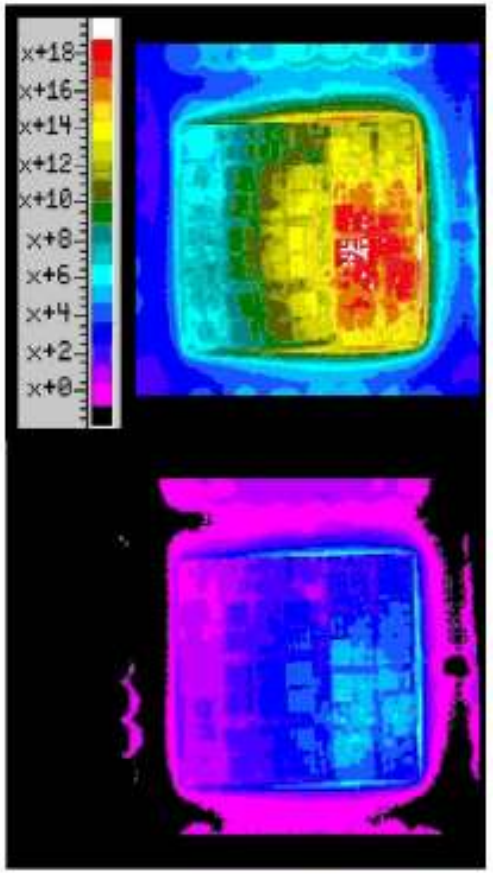
Shang et al, “Temperature Aware On-Chip Networks,” IEEE Micro 2006

Classification of DTM Techniques

- Response Mechanism
  - Software
  - Hardware throttling
    - Global
    - Local
  - Hybrid

Global Clock Gating

- The clock signal to the bulk of the processor logic is stopped for a short time period
- Dynamic thermal management for Pentium4
  - Limited to a few microseconds



Jacobson et al. HPCA 2005

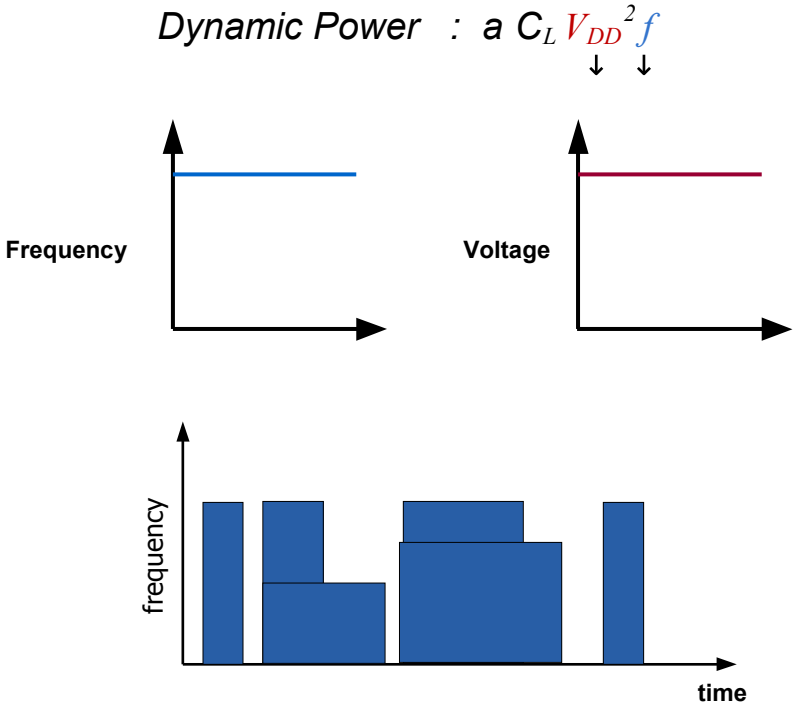
Gunther, et al. "Managing the impact of increasing microprocessor power consumption." Intel Technology Journal. 2001

Global Clock Gating

- Advantages
  - High cooling capability
    - Disables the whole processor
    - Eliminates clock tree power
  - Low invocation time
- Disadvantages
  - Performance impact is high
    - Slows down all processes
  - Hardware overhead

Gunther, et al. "Managing the impact of increasing microprocessor power consumption." Intel Technology Journal. 2001

Global Dynamic Voltage and Frequency Scaling



- DVFS for DTM
- For DTM, only two voltage steps is enough
  - Low voltage → Low frequency but less time to reduce temperature
- To keep temperature below 85°C
  - 20-30% slowdown

Skadron, et al., "Temperature-Aware Computer Systems: Opportunities and Challenges", IEEE Micro, 2003.

Global DVFS

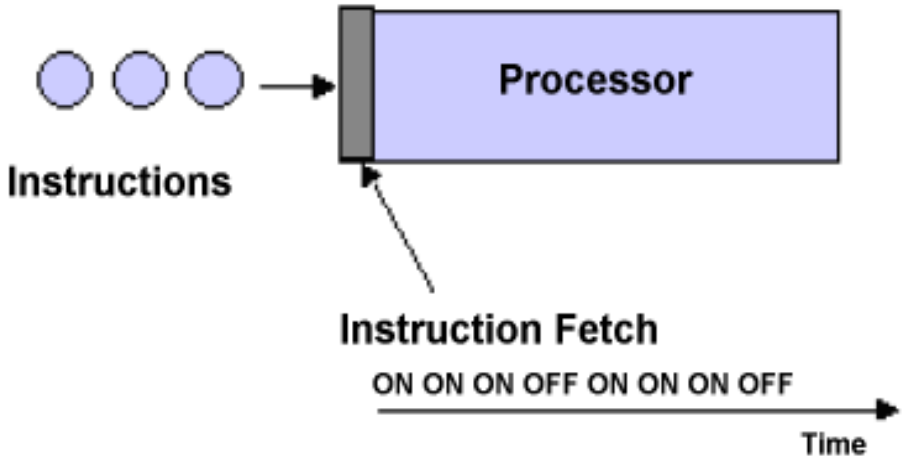
- Advantages
  - Fast reduction of temperature
  - The processor can continue running
- Disadvantages
  - All processes are slowed down
  - Hardware overhead

Classification of DTM Techniques

- Response Mechanism
  - Software
  - Hardware throttling
    - Global
    - Local
  - Hybrid

Fetch Gating

- Utilizes ILP to hide the performance impact
- The important decision
  - Duty cycle
- ILP helps only with mild duty cycles
- At more aggressive duty cycles, slowdown becomes proportional to duty cycle
- To keep temperature below 85°C
  - 10-20% slowdown



Skadron, et al., "Temperature-Aware Computer Systems: Opportunities and Challenges", IEEE Micro, 2003.

Fetch Gating

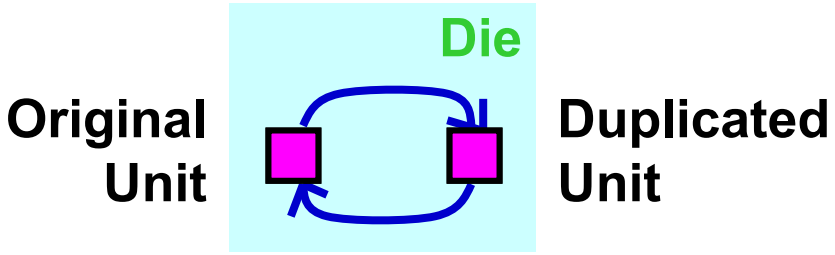
- Advantages
  - Processor is not disabled completely, available ILP compensates for the wasted cycles
  - Low invocation time
  - Hardware overhead is relatively low
- Disadvantages
  - Moderate cooling capability
  - Choice of optimal duty cycle is not easy

Skadron, et al., "Temperature-Aware Computer Systems: Opportunities and Challenges", IEEE Micro, 2003.

Activity Migration

- Computations are migrated to spare units in cold areas of the chip
- Power density is reduced by distributing the heat generation

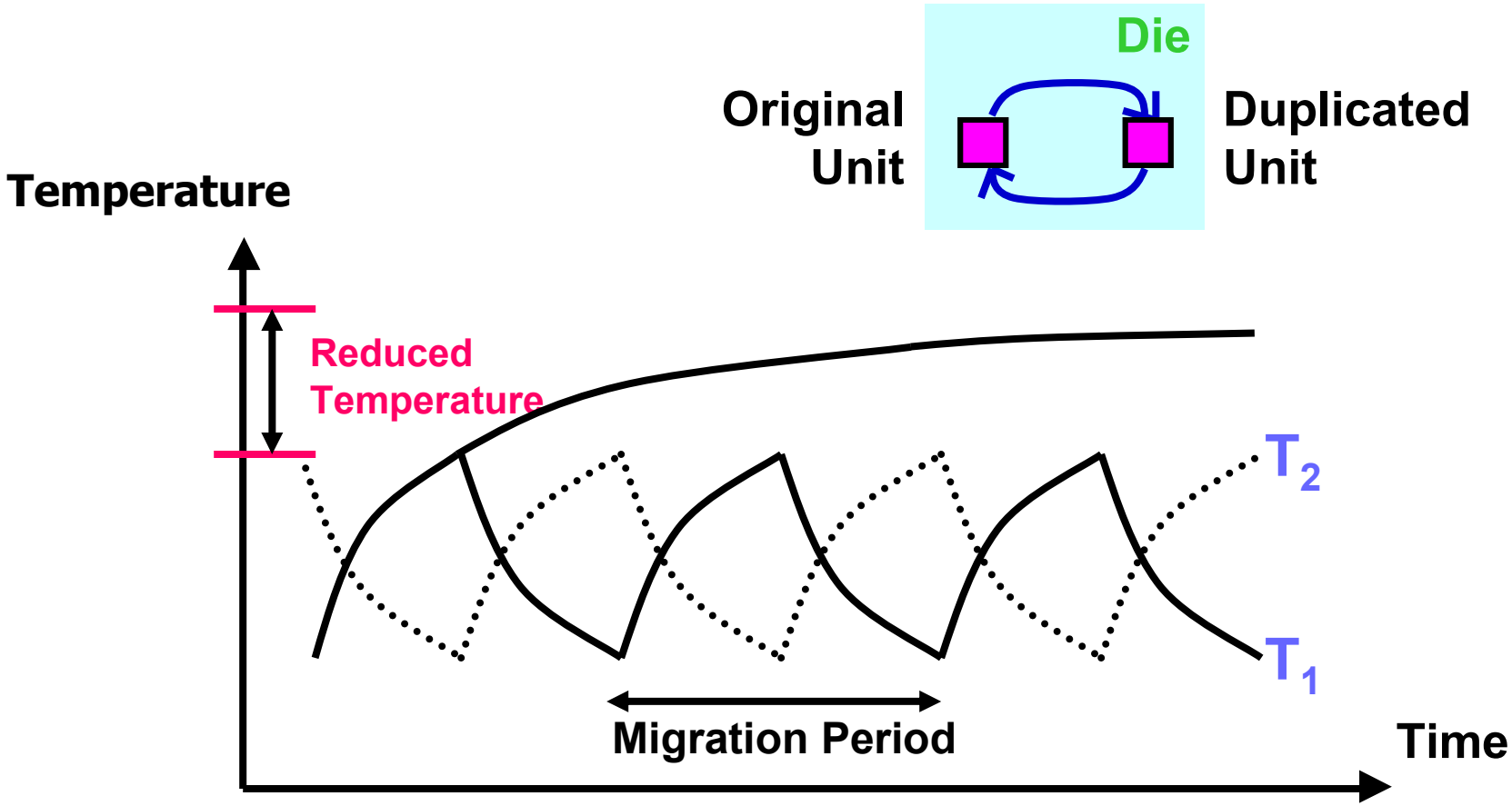
- Activity ping-ponging
  - One unit is disabled when the other one is active
- Register file
  - About 12°C Max. temperature reduction with about 2% IPC loss



Activity Ping-Ponging

Heo, et al. "Reducing Power Density through Activity Migration ", ISLPED 2003

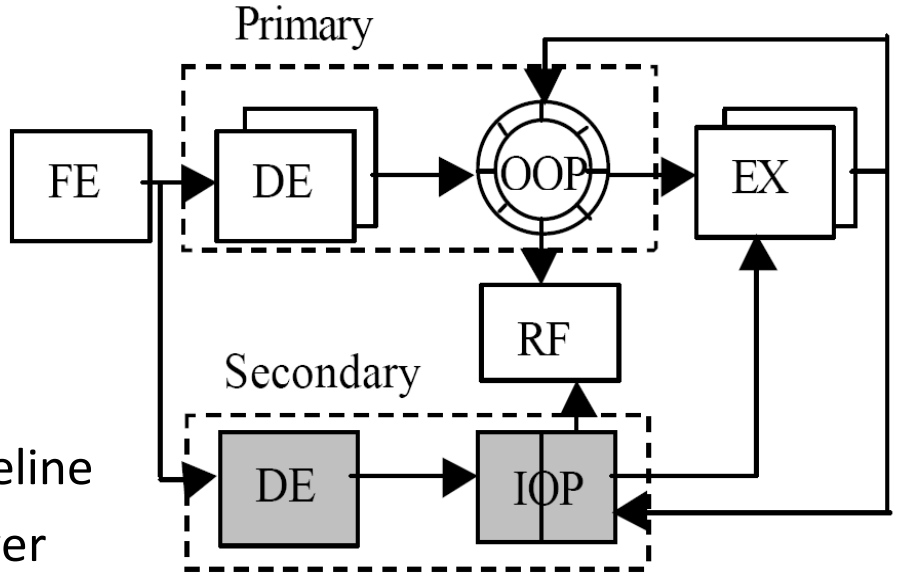
Activity Ping-ponging



Heo, et al. "Reducing Power Density through Activity Migration " ISLPED 2003

A Thermal-Aware Superscalar Microprocessor

- A secondary pipeline
  - Architecturally simple
  - Ultra low power
- Response
  - Clock gates the primary pipeline
  - Secondary pipeline takes over
- To keep temperature below 85°C at least 11% improvement in energy-cpu time product compared to global DVFS



Lim, et al. "A thermal-aware superscalar microprocessor." ISQED 2002.

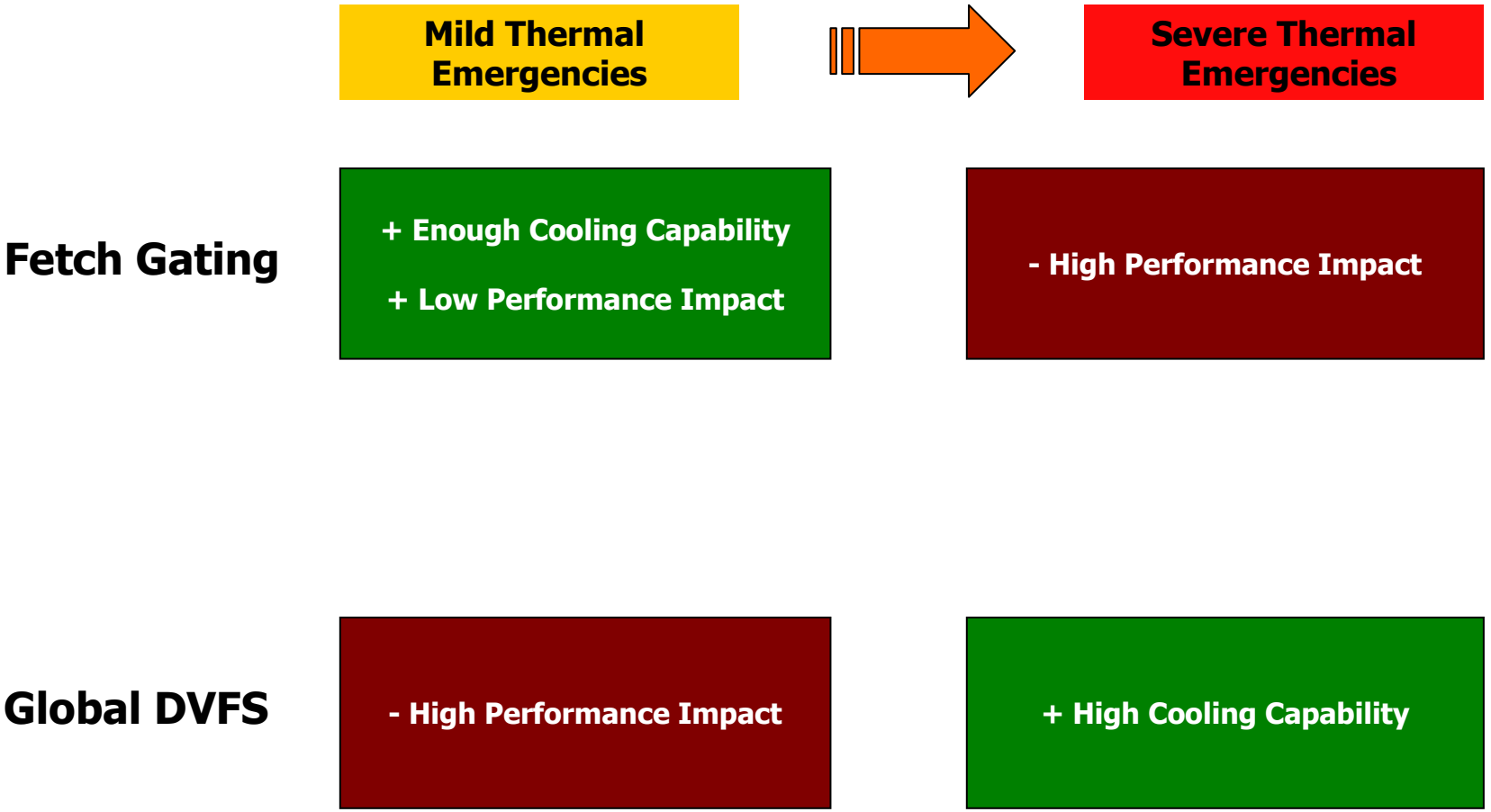
Activity Migration

- Advantages
  - Effective in reducing local hotspots
  - Processor is not disabled completely
- Disadvantages
  - Extra hardware for the redundant units
  - Longer interconnects lead to higher delays and higher power
  - Overhead of copying data to the new unit

Classification of DTM Techniques

- Response Mechanism
  - Software
  - Hardware throttling
    - Global
    - Local
  - Hybrid

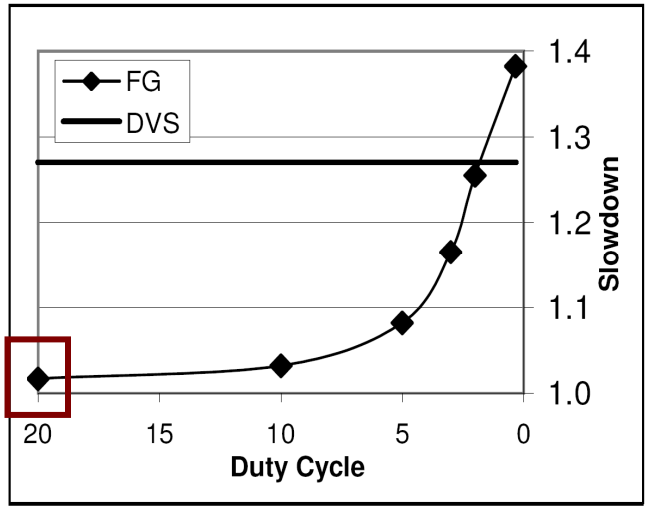
Hybrid Architectural DTM



K. Skadron, "Hybrid architectural dynamic thermal management", DATE 2004.

Hybrid Architectural DTM

- Combines
  - Local Hardware Throttling
    - Fetch gating
  - Global Hardware Throttling
    - Global DVFS
- To keep temperature below 85°C
  - 25% Reduction in DTM overhead compared to Global DVFS



K. Skadron, "Hybrid architectural dynamic thermal management", DATE 2004.

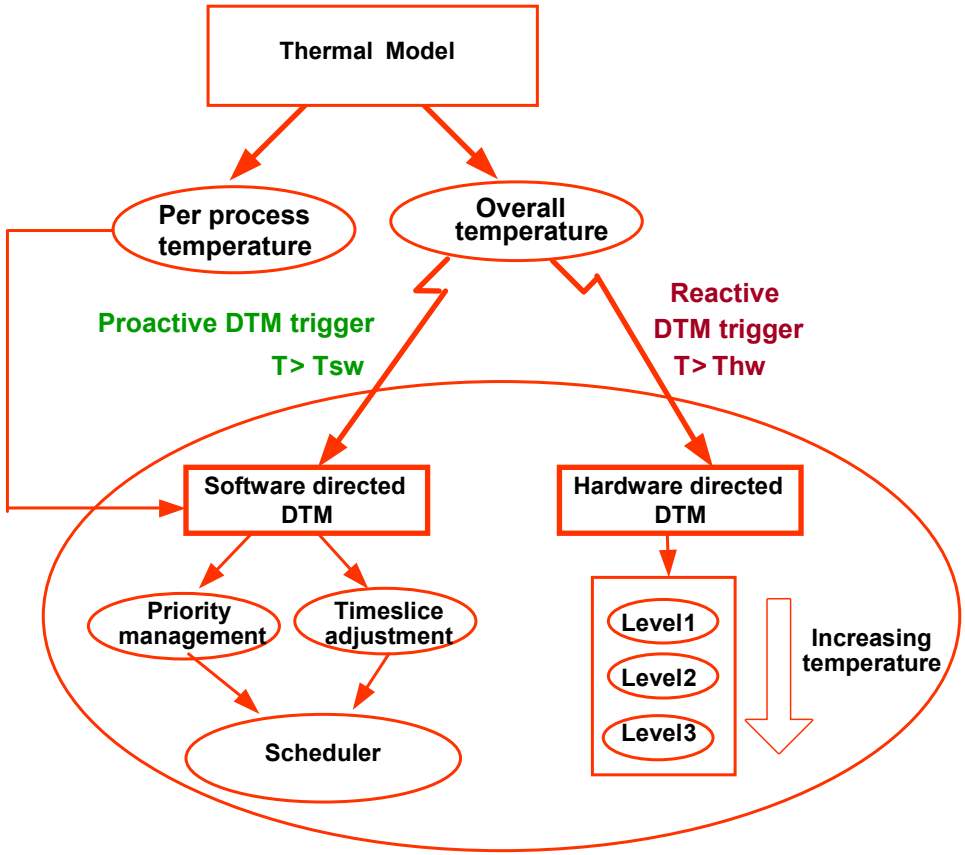
Hybrid Architectural DTM

- Advantages
  - Low performance impact
  - Fast temperature reduction
  - Ability to adjust the response to the severity of thermal emergency
- Disadvantages
  - Design complexity
  - Hardware support for both fetch gating and DVFS

K. Skadron, "Hybrid architectural dynamic thermal management", DATE 2004.

HybDTM

- Combines
  - Temperature-aware Scheduling
    - Mild thermal adjustments and preventing hotspots
  - Global clock gating
    - Severe thermal emergencies
- Execution time overhead to keep temperature below 65°C
  - 10% compared to 20% in global clock gating



Kumar et al., "HybDTM: A Coordinated Hardware-Software Approach for Dynamic Thermal Management" , DAC 2006.

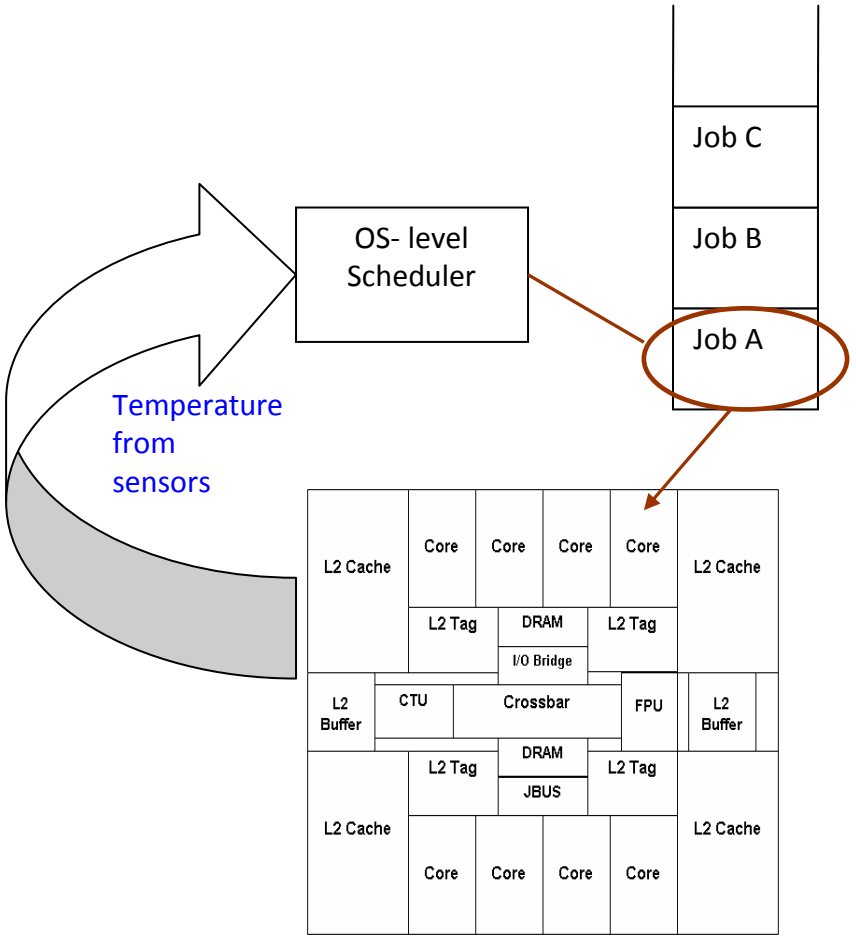
HybDTM

- Advantages
  - Low performance impact
  - Fast temperature reduction
  - Finely adjusts the response to the severity of thermal emergency
- Disadvantages
  - Design complexity

Kumar et al., "HybDTM: A Coordinated Hardware-Software Approach for Dynamic Thermal Management," , DAC 2006.

Dynamic Temperature Aware Scheduling

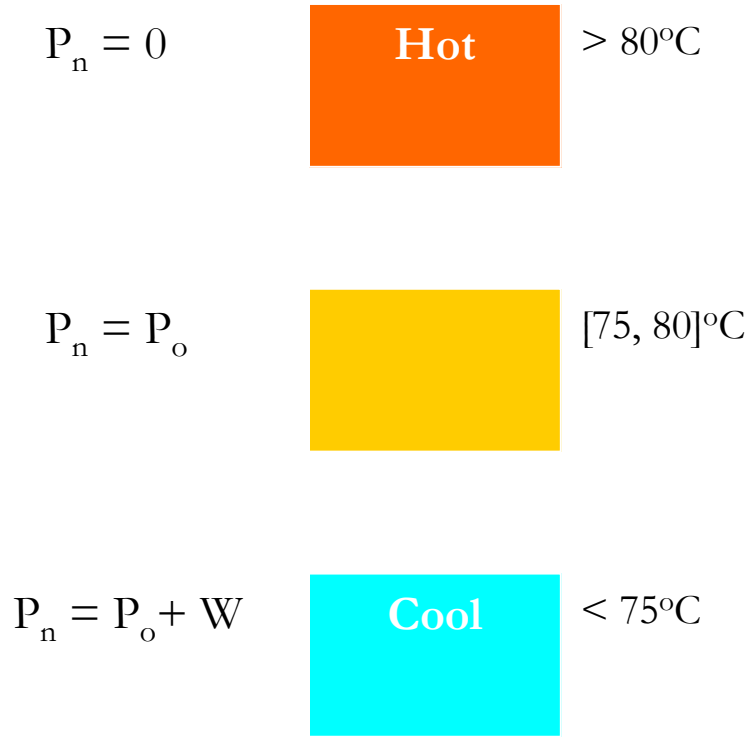
- For multiple processors, there is no optimal dynamic scheduling algorithm for dynamically changing tasks [Liu'73]
  - Heuristics are required
- Temperature information is passed to scheduler at each interval
- Scheduler makes decisions based on temperature
  - Fast and simple implementation



Adaptive-Random Policy

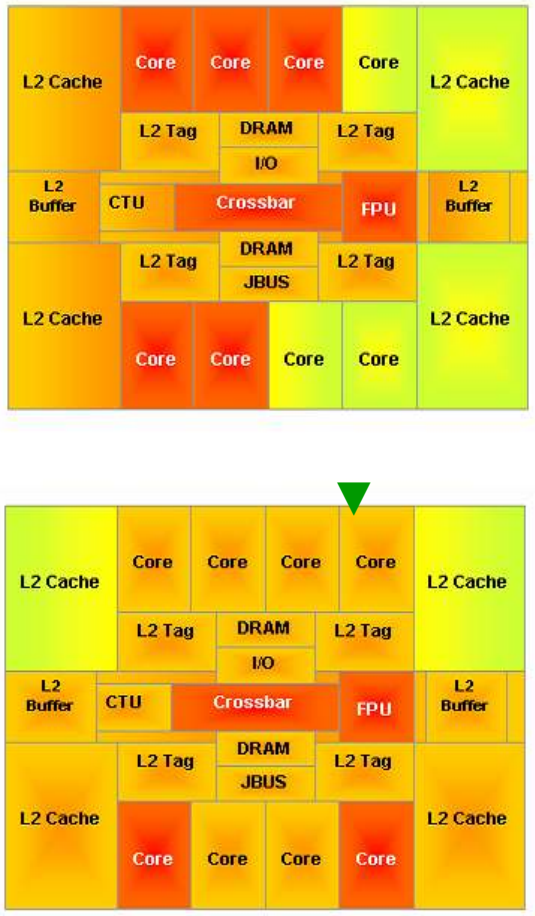
- Goal: To balance workload, minimize & balance temperature with low scheduling complexity
- Updates the probability of sending workload to a core based on its temperature history
- P_n : Probability a core receiving workload
  - Evaluated at each job arrival
- W: Evaluated periodically
  - Interval length: 1 sec
  - $W = \beta / Av_{thr}$

Av_{thr} : (Avg. T below T_{thr}) / T_{thr}
T_{thr} : Threshold temperature
β : Empirically set constant (system dependent)



Adaptive Random

- OS level scheduler for MPSoCs
  - Takes floorplan into account
  - Addresses temperature variations
- Adaptive Random
- When scheduling is not enough, it is combined with thread migration or local DVFS
- Reduces the hotspots to ~1%
- >10X reduction in temporal variations
- Performance impact
  - With thread migration : ~3%
  - With DVFS : ~7%



A. Coskun, T. S. Rosing and K. Whisnant. "Temperature Aware Task Scheduling in MPSoCs," DATE07

Adaptive Random

- Advantages
  - Low overhead
  - Local control of temperature
  - Ability to reduce temperature variations
- Disadvantages
  - Design complexity
  - Large hardware overhead if local per-core DVFS is used

A. Coskun, T. S. Rosing and K. Whisnant. “Temperature Aware Task Scheduling in MPSoCs,” DATE07

Hybrid Control-theoretic DTM for MPSoCs

- Combines
  - Thread migration
    - Balancing the heat and optimizing the performance
  - Local per-core DVFS
    - Fine-grained local adjustments
- To keep temperature below 85°C
  - 2.5X instruction throughput compared to distributed clock gating

Donald et al., "Techniques for Multicore Thermal Management: Classification and New Exploration," ISCA 2006.

Hybrid Control-theoretic DTM for MPSoCs

- Advantages
  - Low performance impact
  - Distributed local control of temperature
  - Formal control theory allows accurate design and provable guaranty for temperature control
- Disadvantages
  - Design complexity
  - High hardware overhead

Donald et al., "Techniques for Multicore Thermal Management: Classification and New Exploration," ISCA 2006.

Summary

		Performance Impact	Cooling Capability	Hardware Cost	Design Complexity
Software		Low	Low	None	Low
Hardware	Global Throttling	High	High	Low	Low
	Local Throttling	Average	Average-High	Low-High	Average-High
Hybrid		Low-Average	Average-High	Average-High	High

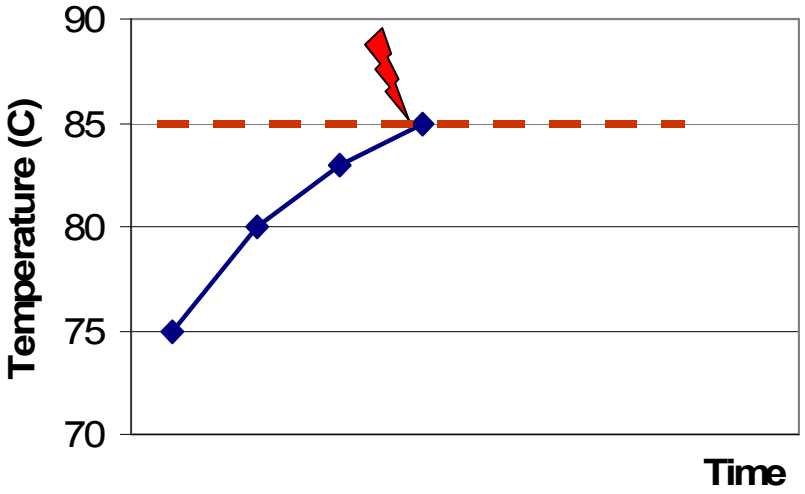
Proactive Dynamic Thermal Management

Reactive vs. Proactive

- Reactive
  - Not activated until the thermal emergency
  - Typically have high performance impact
- Proactive
  - Prevent thermal emergencies
  - Part of the normal system operation

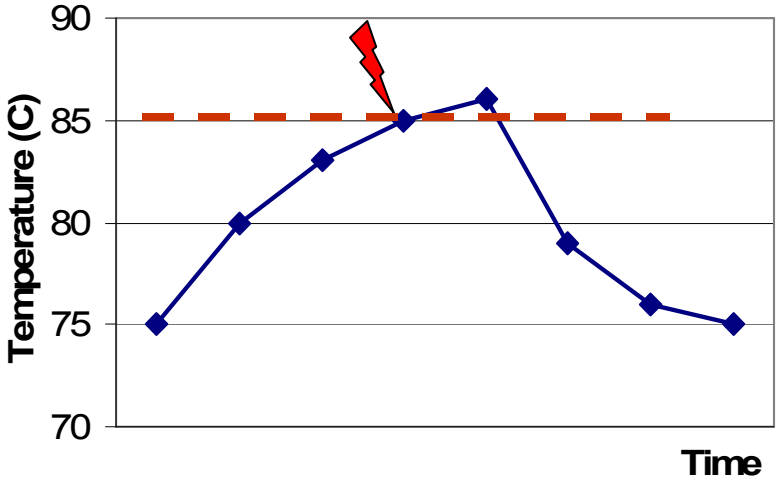
Reactive vs. Proactive Management

- **Reactive**

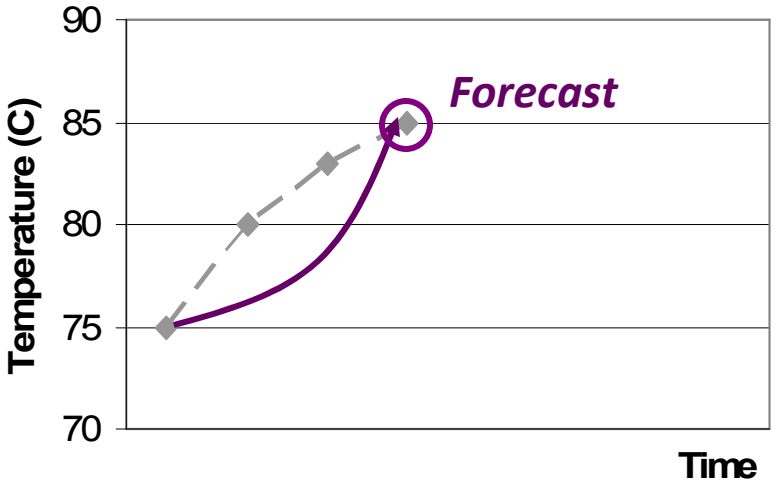


Reactive vs. Proactive Management

- **Reactive**



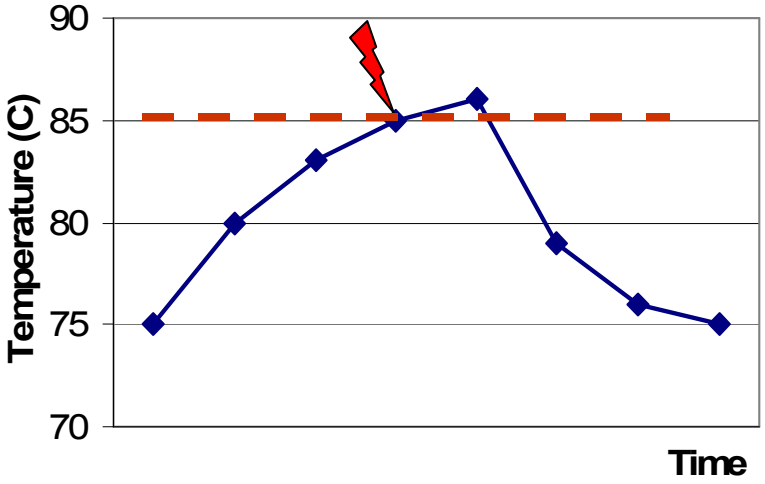
- **Proactive**



- e.g., DVFS, fetch-gating, workload migration, …

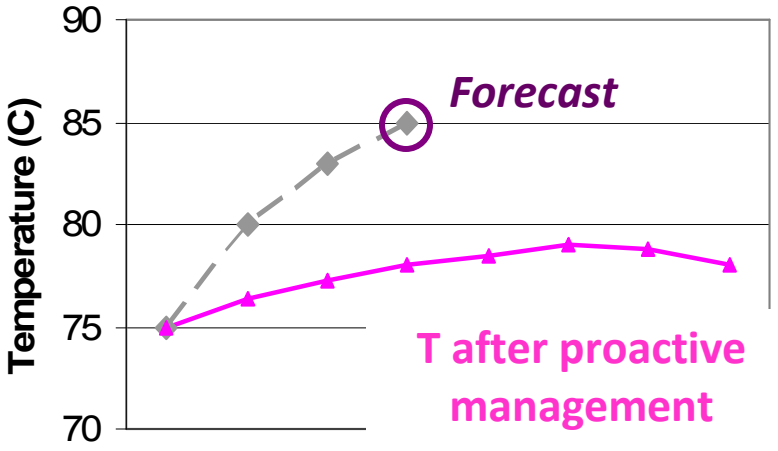
Reactive vs. Proactive Management

- **Reactive**



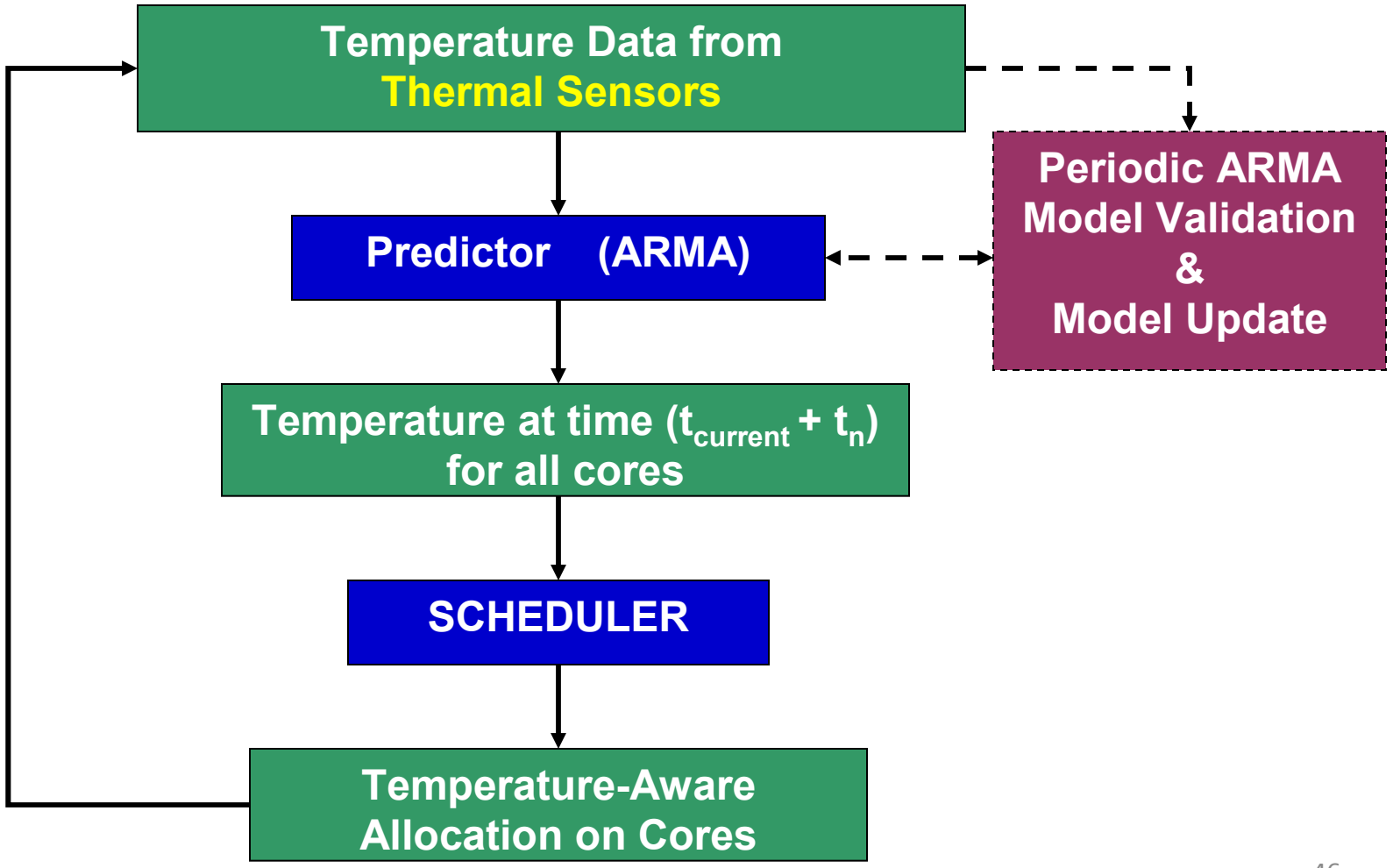
- e.g., DVFS, fetch-gating, workload migration, …

- **Proactive**



- Reduce and balance temperature
  - Adjust workload, V/f setting, etc.

Flow

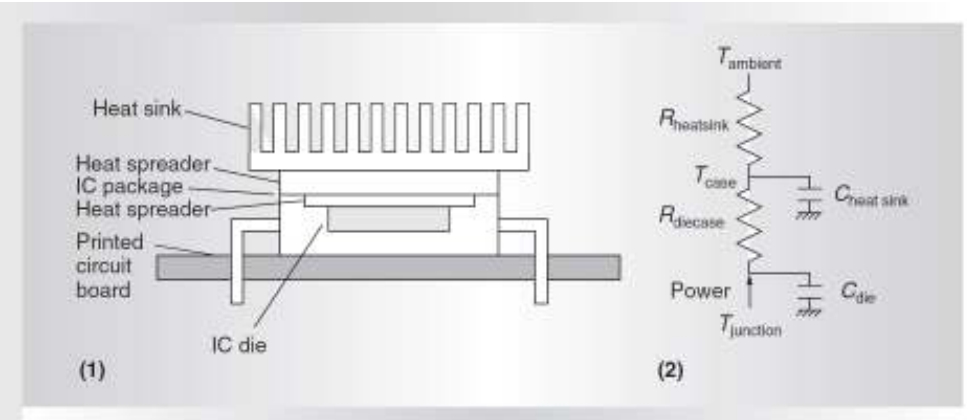


[Coskun, et al. ISLPED 08]

Temperature Modeling and Sensing

Thermal model

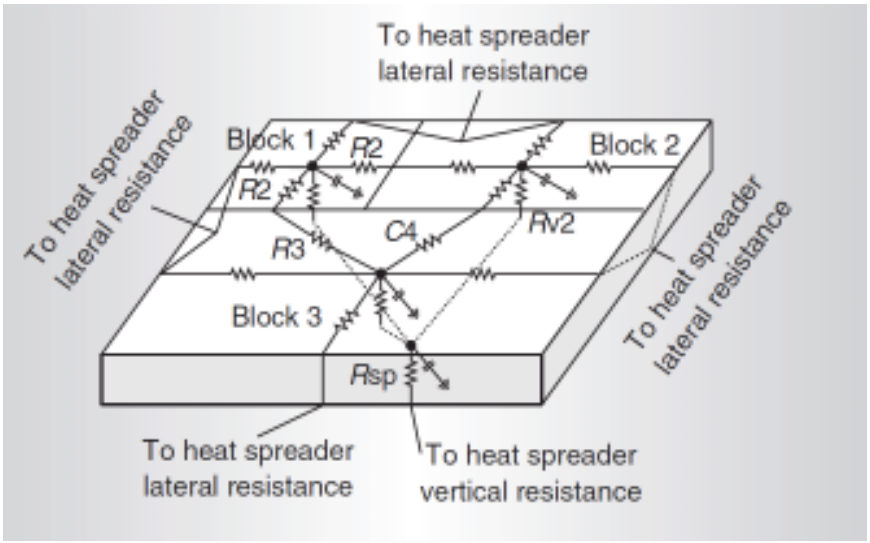
- Based on the *duality* between heat flow and electrical phenomena
- Heat flow could be considered as a current passing through a thermal resistance, leading to a temperature difference analogous to voltage
- Thermal *R* and *C*s are calculated based on the system characteristics

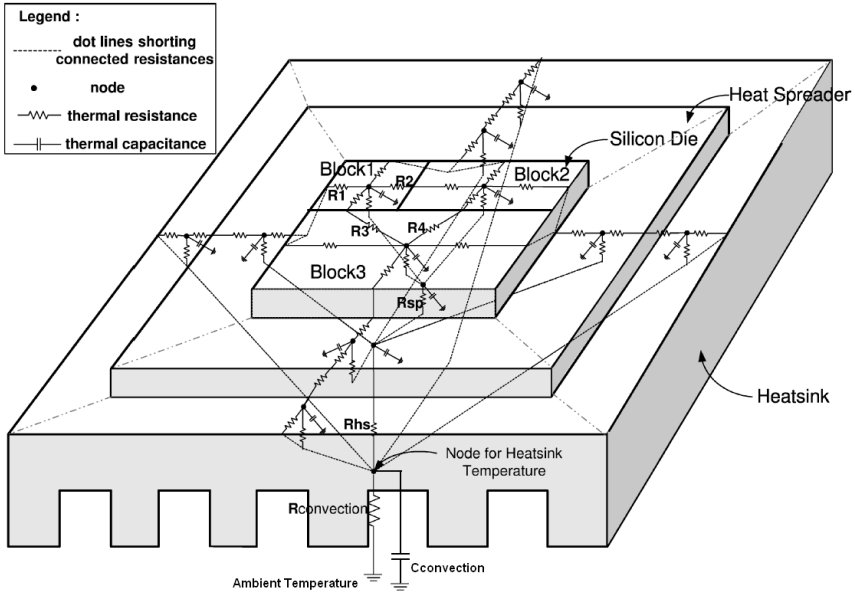


Duality between Thermal and Electrical Quantities	
Power	Current
Temperature Difference	Voltage Difference
(R_{th}) Thermal Resistance	(R) Electrical Resistance
(C_{th}) Thermal Capacitance	(C) Electrical Capacitance
$(R_{th} \times C_{th})$ Thermal RC Constant	$(R \times C)$ Electrical RC Constant

K. Skadron, et. al. Micro 03

Thermal Model

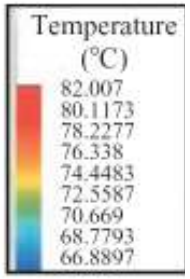


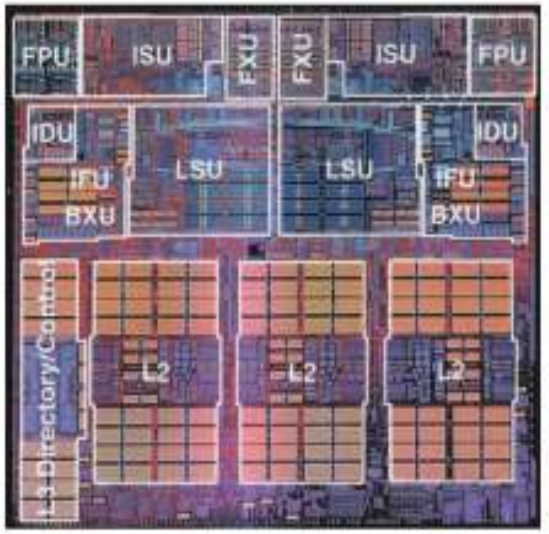


K. Skadron, et. al. ACM TACO 04, Vol. 1, No. 1

Challenges in using thermal sensors

- Limitations in sensor placement
  - Sensors may not be placed at locations of interest
  - Routing and I/O considerations
- Sensor overhead
  - Sensing circuitry
  - A/D converter
- Few sensors available
  - Silicon real estate
  - Low mean time to failure
- Sensor noise
  - Inaccuracies due to power variations, sensor degradation, etc.
- Dynamic change of hot spot locations
  - Static placement can not cover all locations

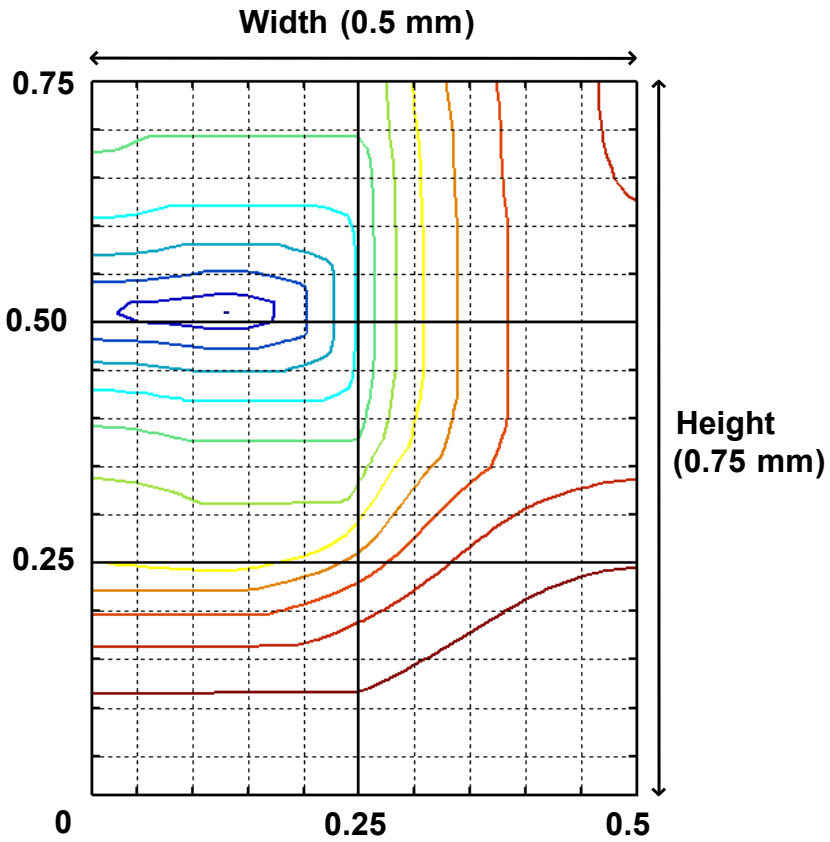




Maximum temperature differences

- Temperature differences were traditionally found by extensive simulations

- Long simulation times
- Workload dependent
- No guarantee on the maximum found temperature differences
- Under what situations does this maximum difference happen?



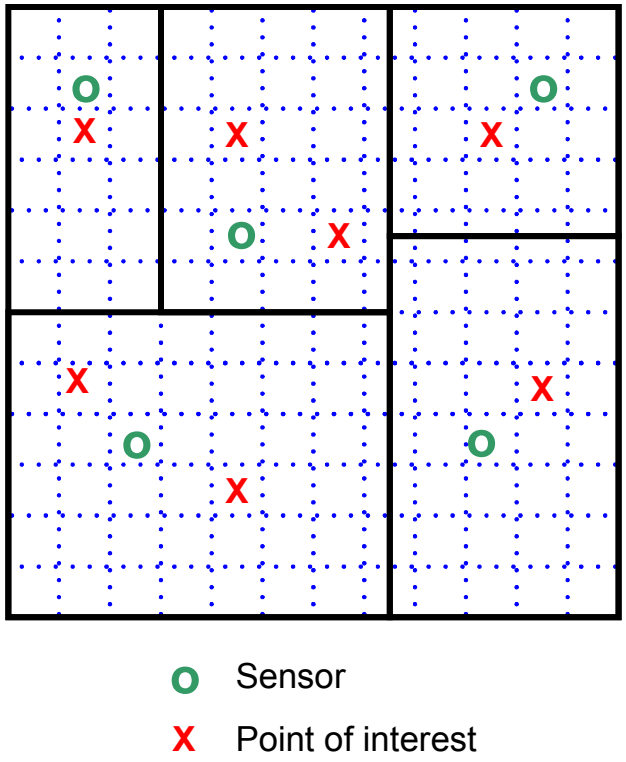
Contour map of temperature difference to a location of interest on die

[Sharifi, et al., TCAD'10]

Sensor placement

- Sensor overhead
- Sensor placement limitations
- Sensor placement error
  - Temperature difference between the hotspot and its corresponding sensor

- **Sensor placement technique**
  - Covering all locations of interest guaranteeing a maximum tolerable placement error
  - Minimum number of sensors

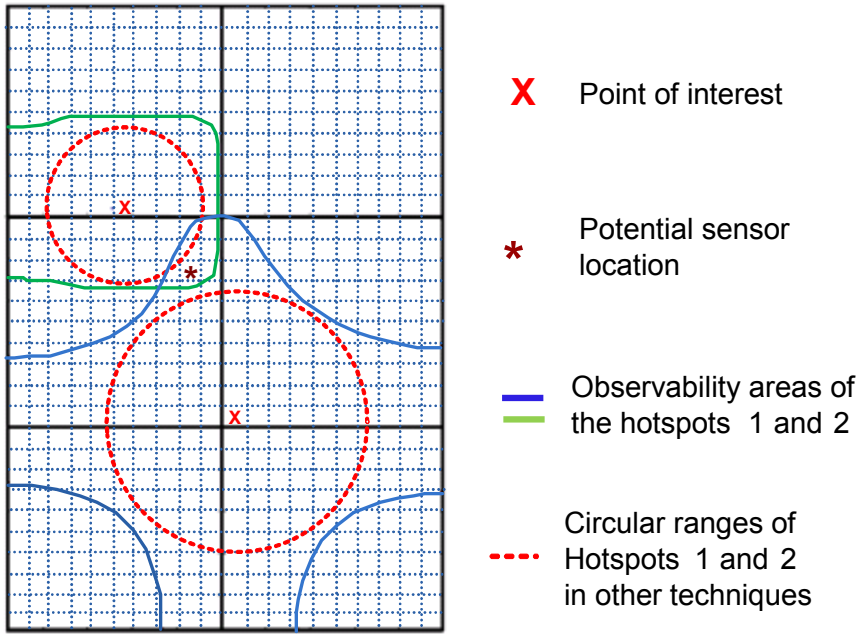


Sensor Placement Comparison

- Circular range

  [Lee. et. al. ICCD 05]

  Based on exponential dependence of temperature difference to the distance from a location of interest



[Sharifi, et al., TCAD'10]

	SOC1							SOC2						
Accuracy (°C)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Our technique	7	7	6	5	5	5	4	8	8	8	7	6	6	5
Circular range*	7	7	7	7	6	6	6	8	8	8	8	8	8	8

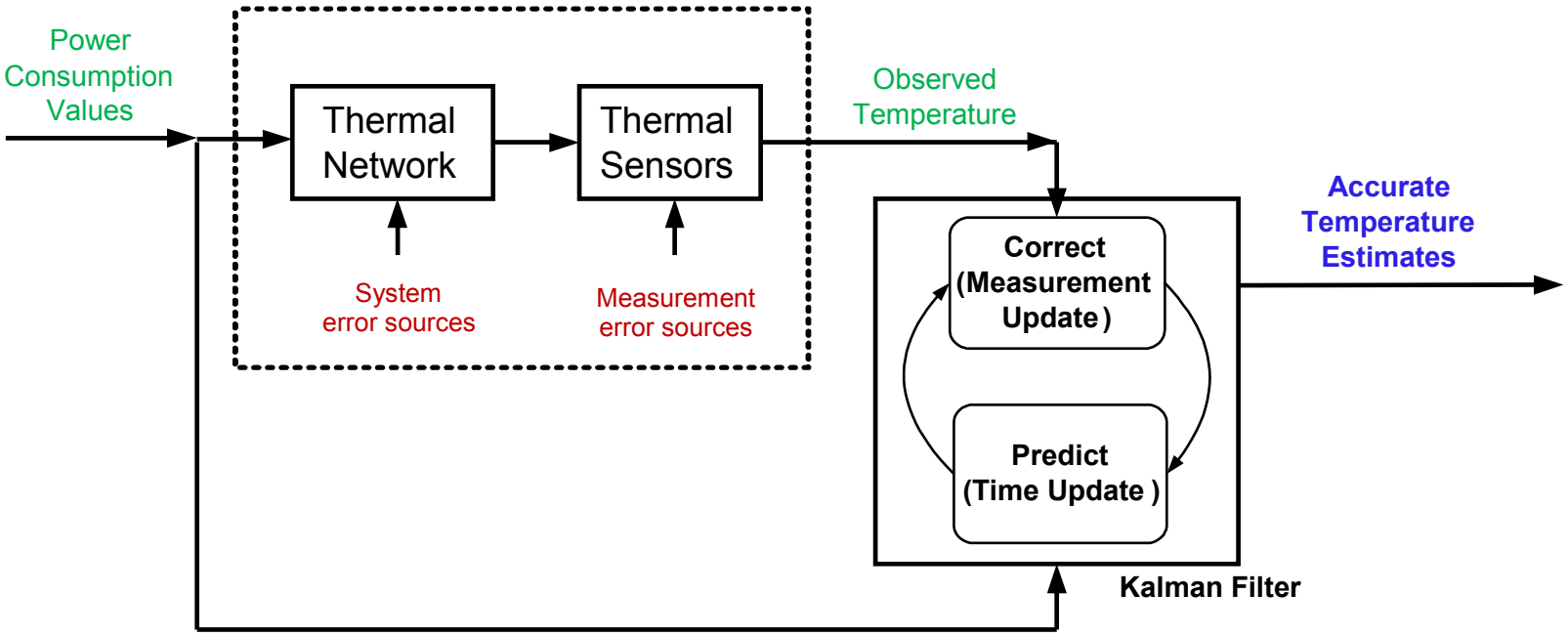
Indirect temperature sensing

- Runtime temperature information is needed
  - Few sensors available
  - Sensor noise
  - Dynamic change of hot spot locations

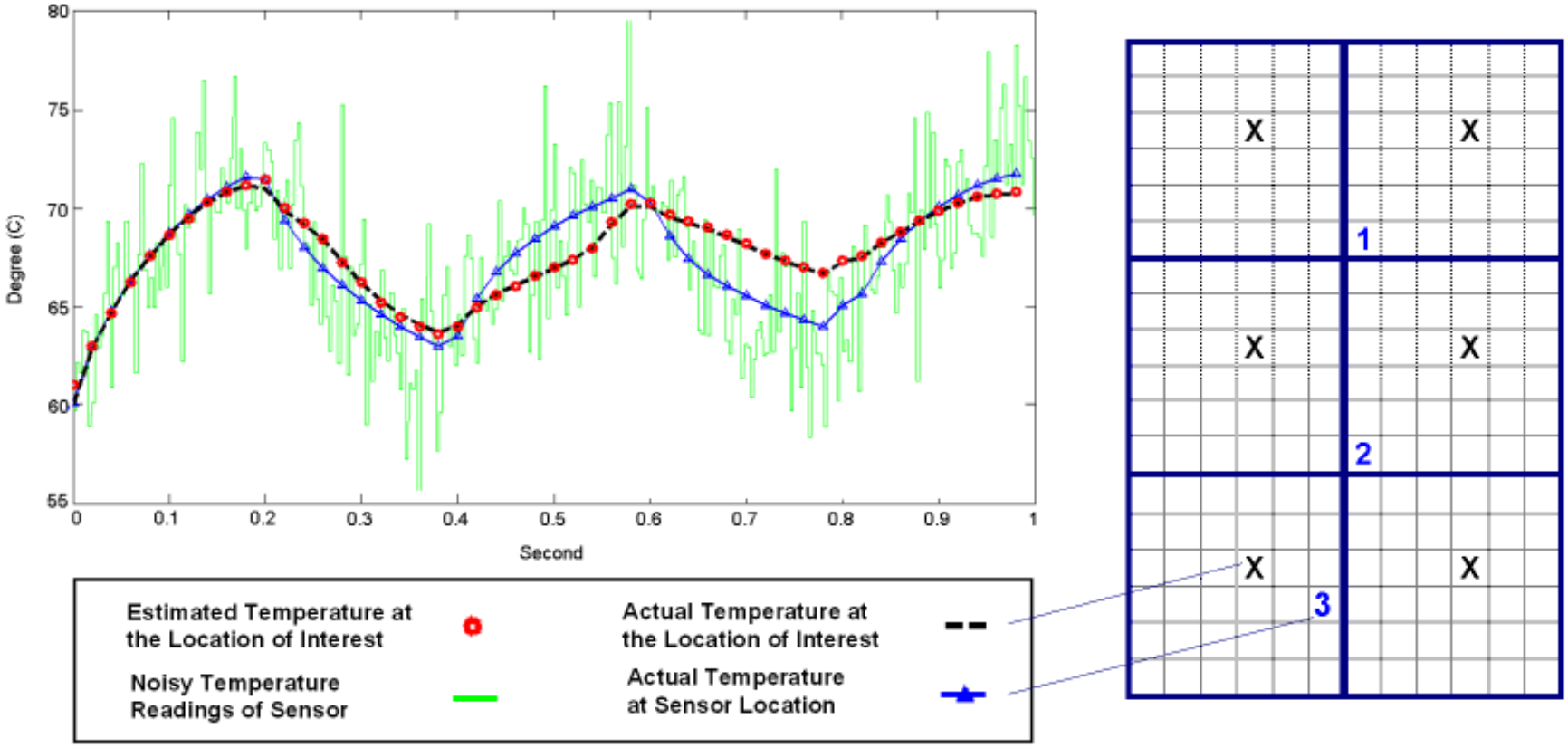
- Accurate temperature estimates are required at the points other than sensor locations

- Applications
  - Small number of deployed sensors due to overhead/ placement limitations
  - Operational systems with degraded/failed sensors
  - Changes in the locations of interest

Accurate Temperature Estimation



Accurate Temperature Estimation (Cont'd)



[Sharifi, et al., TCAD'10]

Questions?