Islands of Heaters: A Novel Thermal Management Framework for Photonic Network-on-Chip

Dharanidhar Dang, **Rabi Mahapatra**
Texas A&M University, College Station, TX, USA
{d.dharanidhar, rabi}tamu.edu
Sai Vineel Reddy Chittamuru, **Sudeep Pasricha**
Colorado State University, Fort Collins, CO, U.S.A.
{sai.chittamuru, sudeep}@colostate.edu

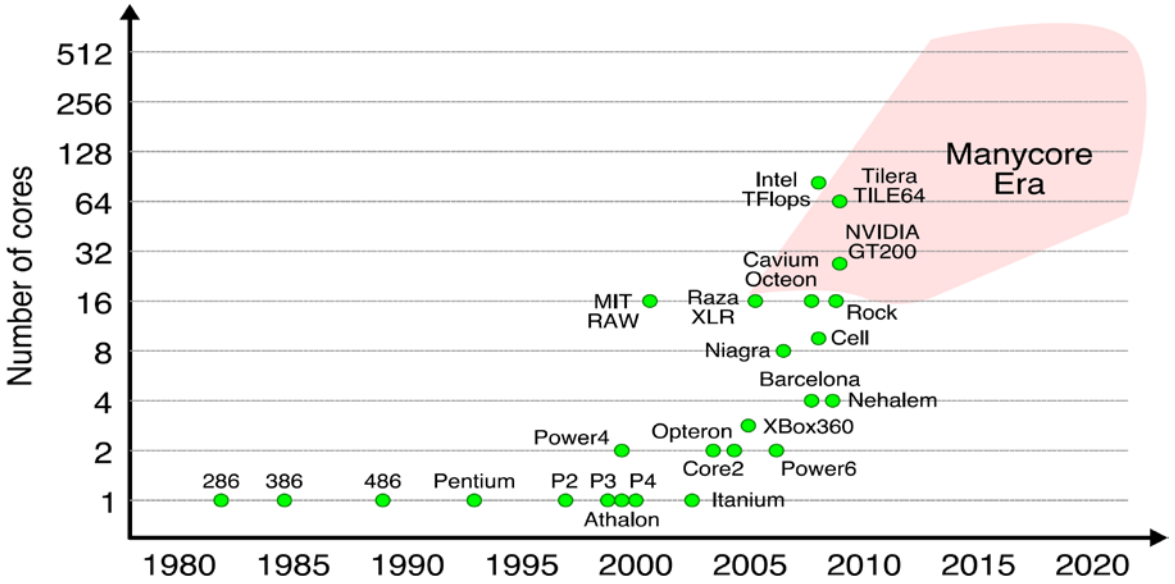
Dang and Sai are primary authors

Outline

- Introduction
- Motivation
- Related Work
- IHDTM Framework Overview
  - ❑Device-level Islands of Heaters Framework
  - ❑System-level Thread Aware Thermal Management (TATM)
- Experimental Results
- Conclusion

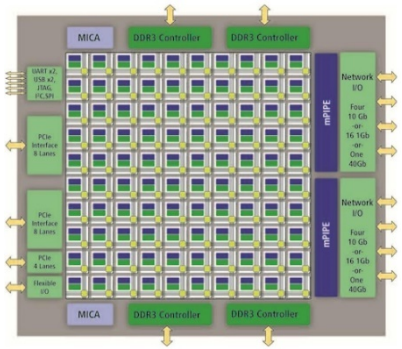
Introduction

The Manycore era…



- 512-core chip by 2020

Intel Xenon 7500 processor

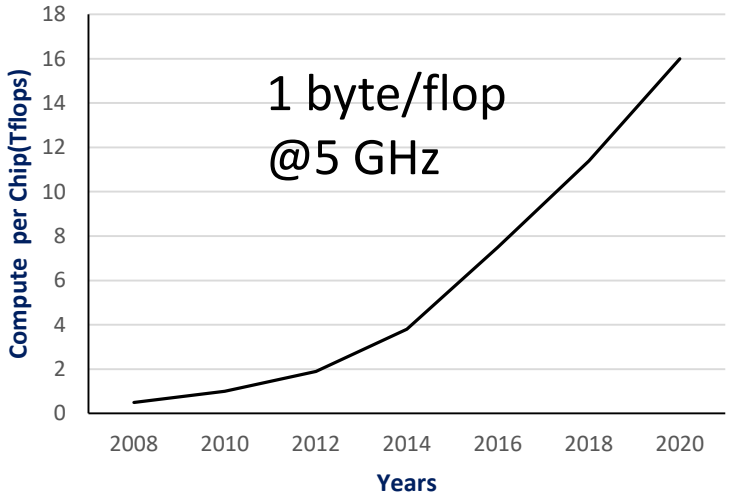


Intel Tilera 100-core chip

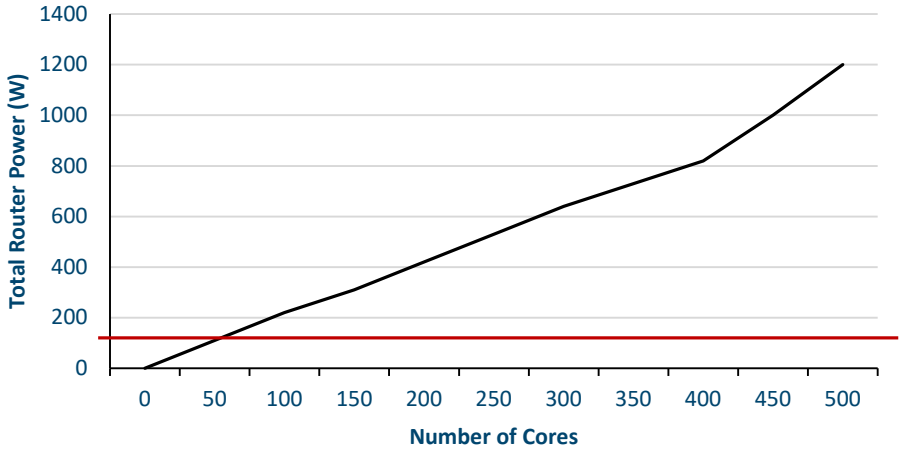
Source: Joshi et al. HOTI, 2008

Introduction

- How good is Electrical NoC?



- 10-15 Tflops/chip at 5 GHz
- Data intensive apps may take 1 byte/flop
- Requirement: 10-15 Tbps of NoC throughput



Limit for an entire HPC Chip

Source: Joshi et al. HOTI, 2008 and Results obtained from Booksim simulation

Introduction

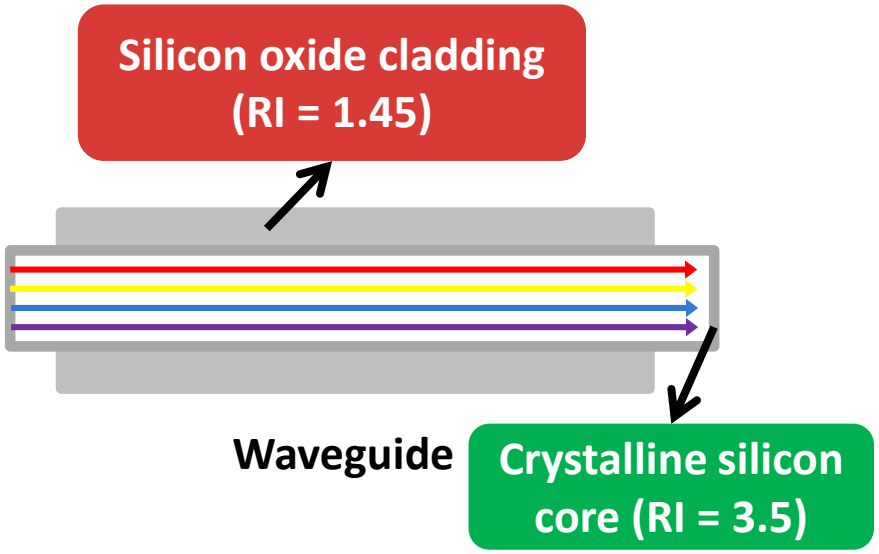
Electrical Interconnects vs Photonic Interconnects

Interconnects	Energy (pJ/bit)	Bandwidth Density (Gbps/µ)
On-chip Optimally Repeated Electrical Link	1.0	5.0
On-chip Photonic Link	0.25	160.0-320.0
Off-chip Electrical Link (100 µ pitch)	5.0	0.1
Off-chip Photonic Link (50 µ Coupler pitch)	0.25	13-26

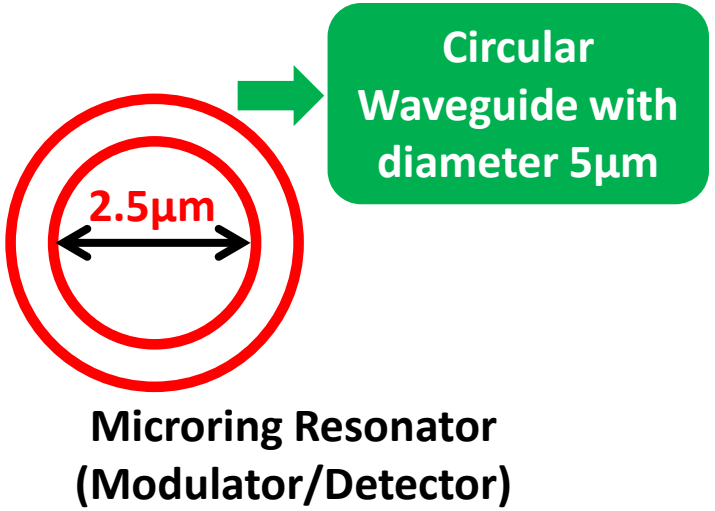
Photonic interconnects based Network-on-Chips (PNoCs) provide higher bandwidth with lower power consumption

Source: Joshi et al. HOTI, 2008 and

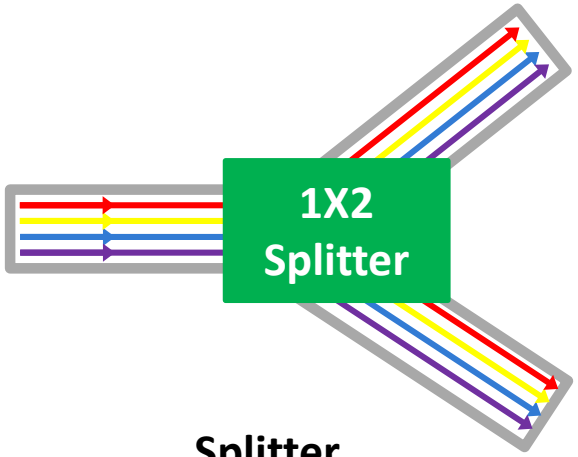
Introduction to Photonic Elements



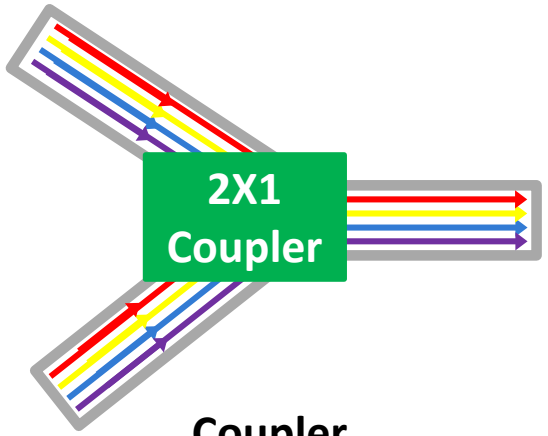
Waveguide



Microring Resonator
(Modulator/Detector)



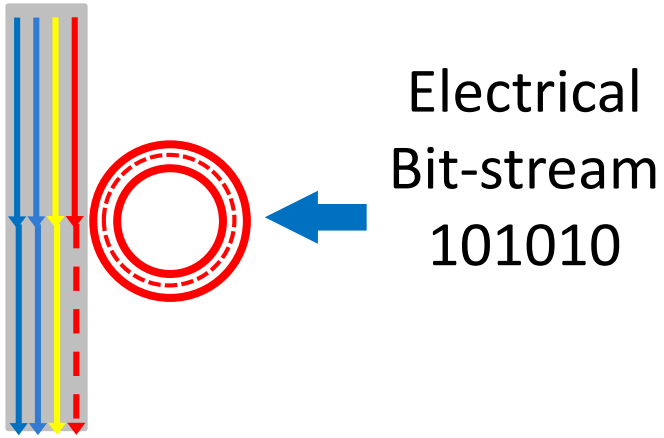
Splitter



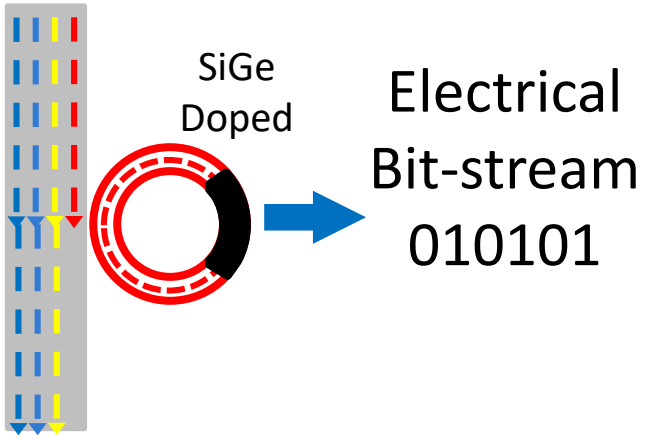
Coupler

Introduction to Photonic Elements

Modulator



Detector

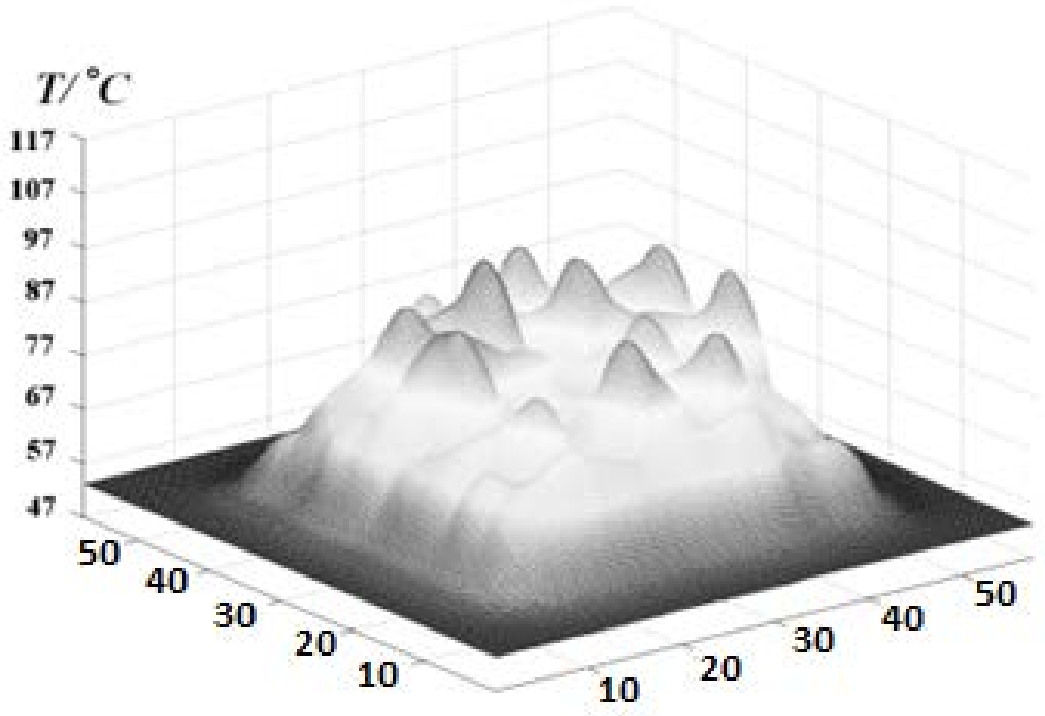


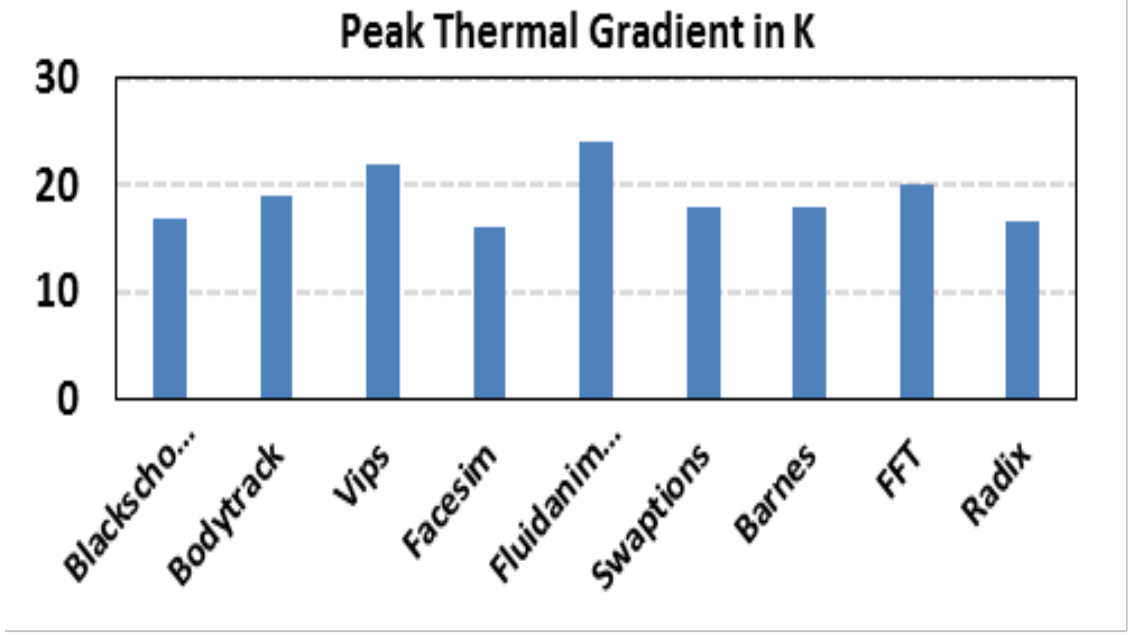
- **Microring resonator operation**
  - Modulator to write data
  - Detector to read data

Modulators and detectors are used to traverse data in photonic links of PNOCs

Motivation

64-core PNoC
22 nm PTM
HotSpot 5.0
Total Time Slice: 1000 Sec
PARSEC and Splash Suites





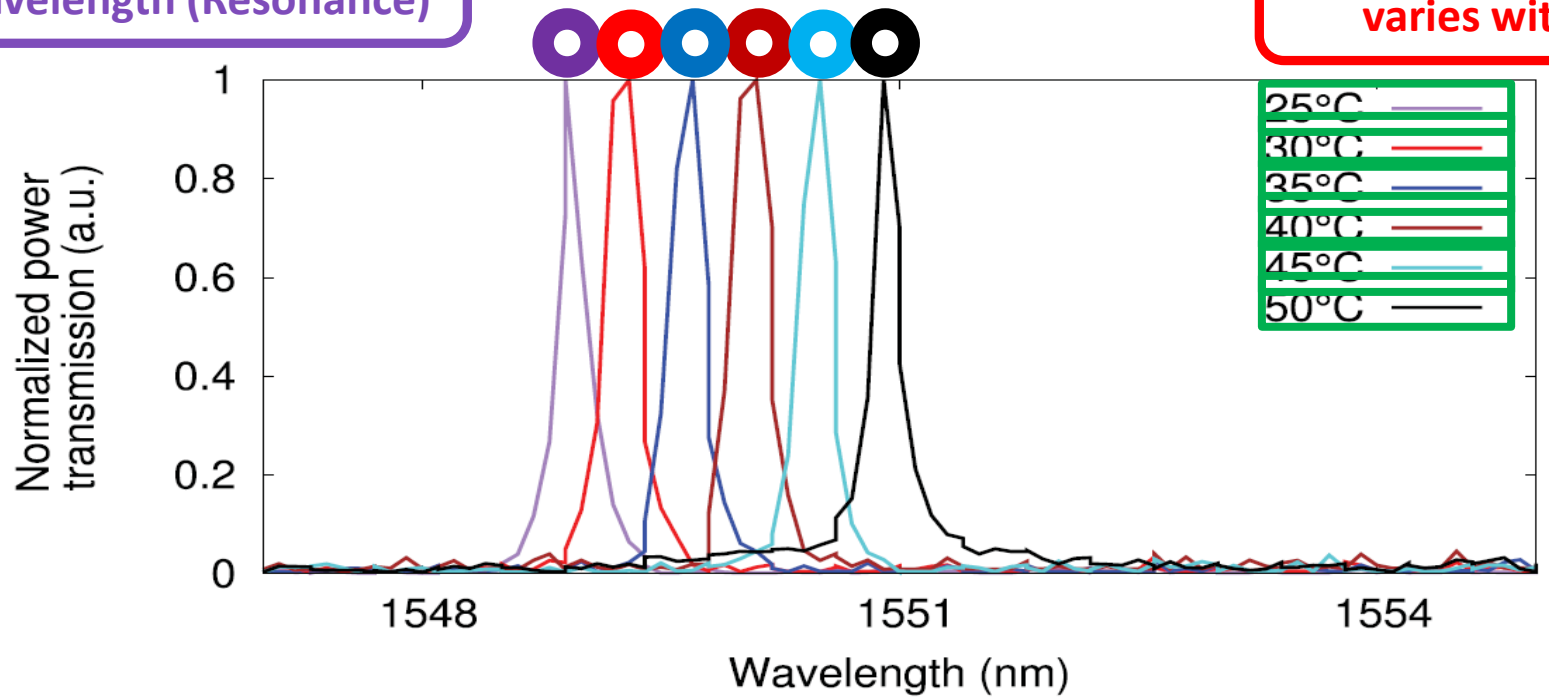
On an average 15-20^{0}C of temperature gradient

Motivation

Temperature variation impact on MRR

MRR will receive this wavelength (Resonance)

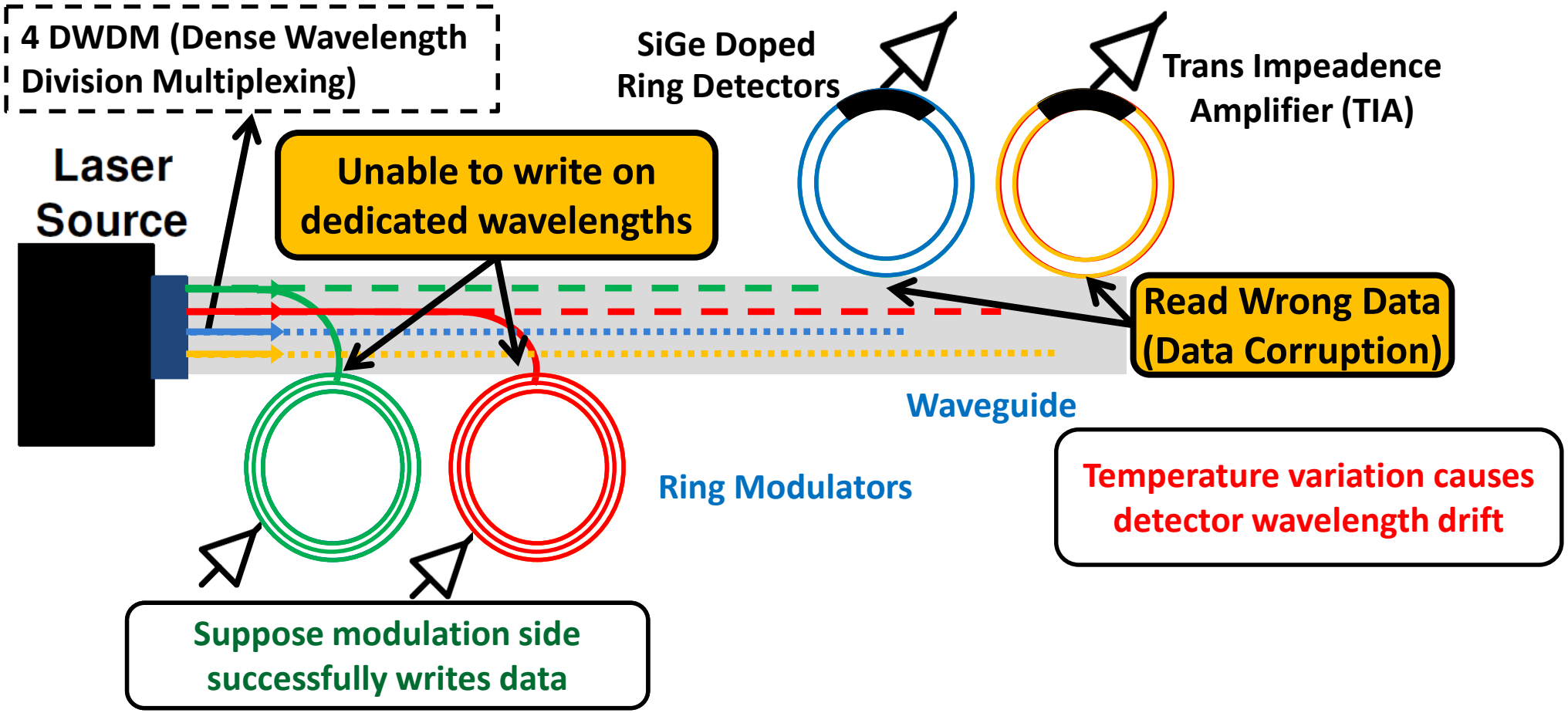
Reason: Refractive index of ring varies with temperature



Pass band shift with thermal variation

- **Temperature versus pass band shift (MRR resonance wavelength shift)**
  - Linear dependency ($\Delta\boldsymbol{\lambda}_{shift} = K\ \Delta T_{max}$)

Motivation



These drawbacks of MRRs motivate us to propose a dynamic thermal management for PNoCs

Source: Joshi et al. HOTI, 2008 and

Related Work

Device-level thermal management in PNoCs :

- **[Z. Li et al. IEEE TVLSI 2012]** Presents athermal photonic devices to reduce the localized tuning/trimming power in MRs
- **[Joanna et al. Materials 2010]** Use of liquid crystal cladding to reduce the effect of temperature variations

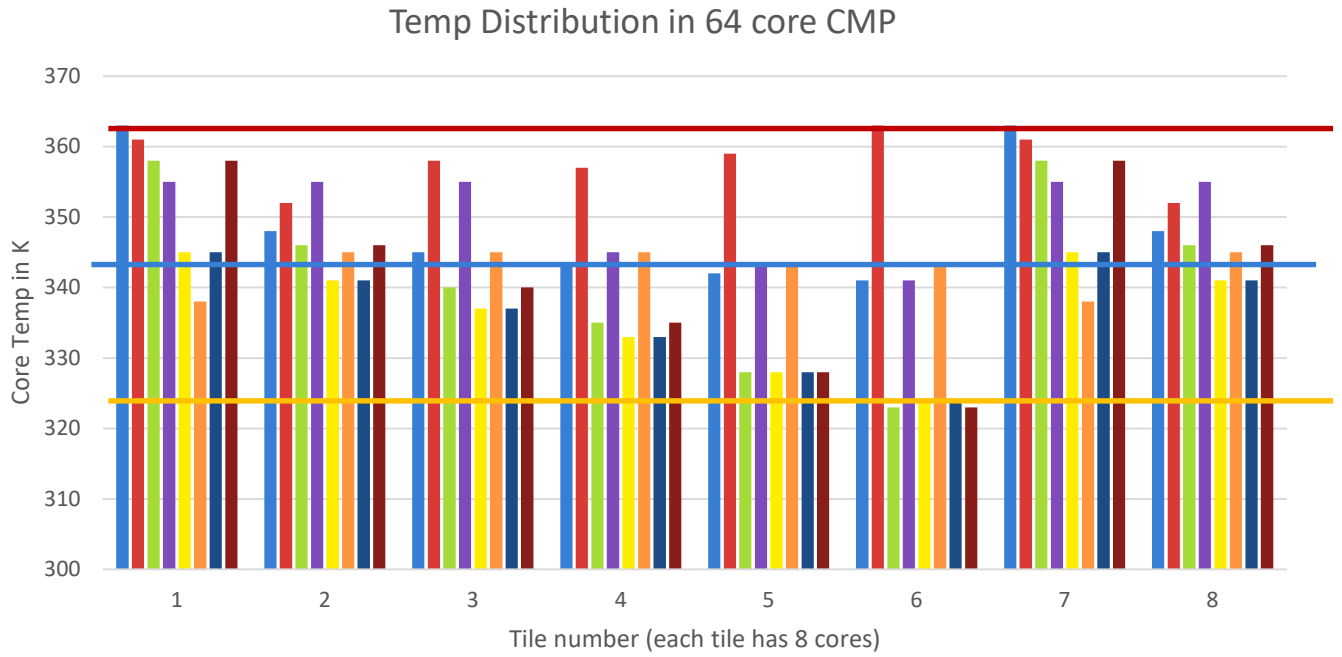
These works:
1)High power and area overhead
2)Require costly changes in the manufacturing process

System-level thermal management in PNoCs:

- **[C. Nitta et al. HPCA 2011]** overhead associated with localized tuning of MRRs is reduced in using the group drift property of co-located MRs
- **[T. Zhang et al. DATE 2014]** a ring aware thread scheduling policy (RATM) is proposed to reduce on-chip thermal gradients in a PNoC

These works do not consider:
1) impact of run-time workload variations
2) relationship between thermal hotspots and transmission reliability

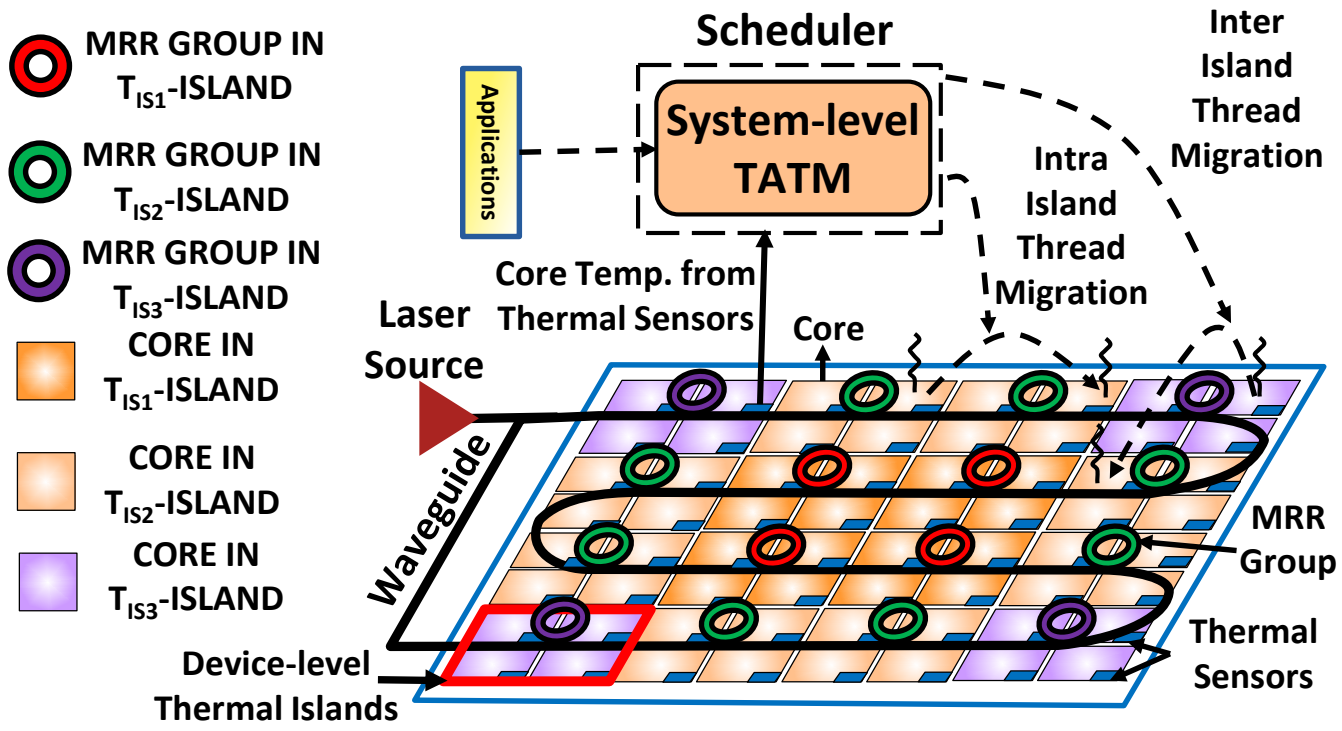
Going Deeper into Thermal Distribution



64 core PNoC
Tool: 3D-ICE
Benchmark: PARSEC and SPLASH

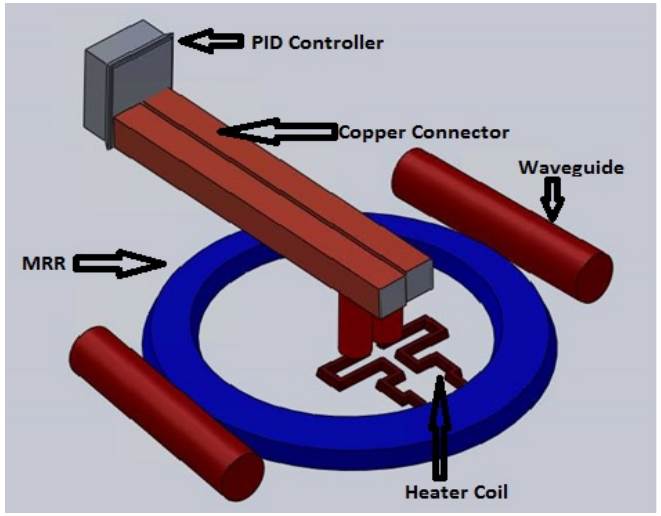
Simulating PARSEC and SPLASH suite on 3D-ICE shows:
Three major temperature zones i.e. 363K, 343K, 323K

Cross-layer: IHDTM Framework



PID Controlled Heater

- Each MRR is integrated with a PID controlled heater
- Thermal sensor of corresponding core feeds temperature data
- Heater is set to work at the corresponding island temperature.



Heater Type	Power Req (mW/nm)	Heating τ (µs)	Cooling τ (µs)
Doped-Si	3.138	21.3	66.0
Silicide	3.462	19.1	75.8
Tungsten	3.6	38.2	45.11
Doped WG	3.369	43.4	39.8

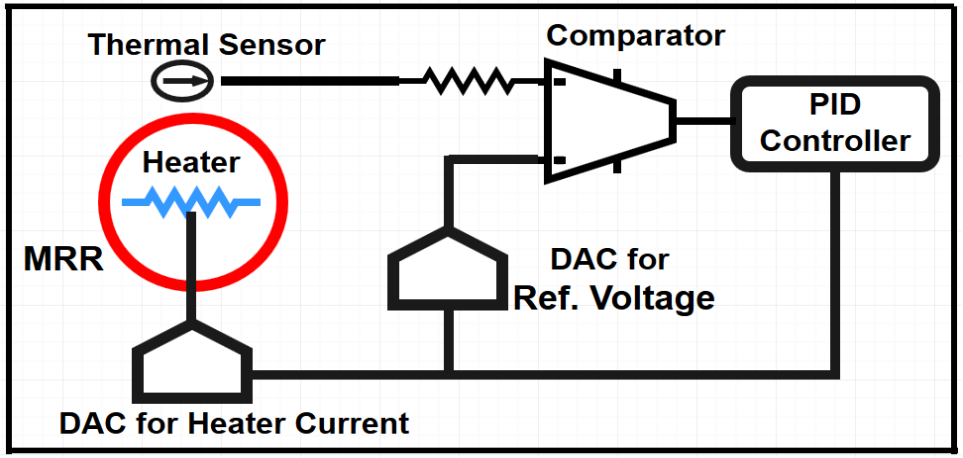
Source: D Dang et al, ICCD 2015

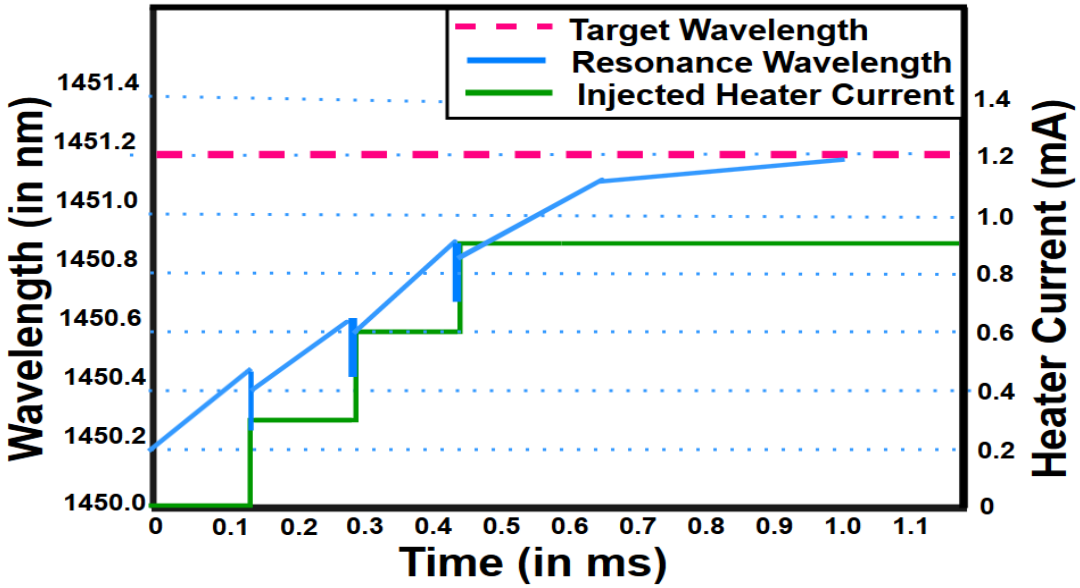
PID Controlled Heater

PID Algorithm

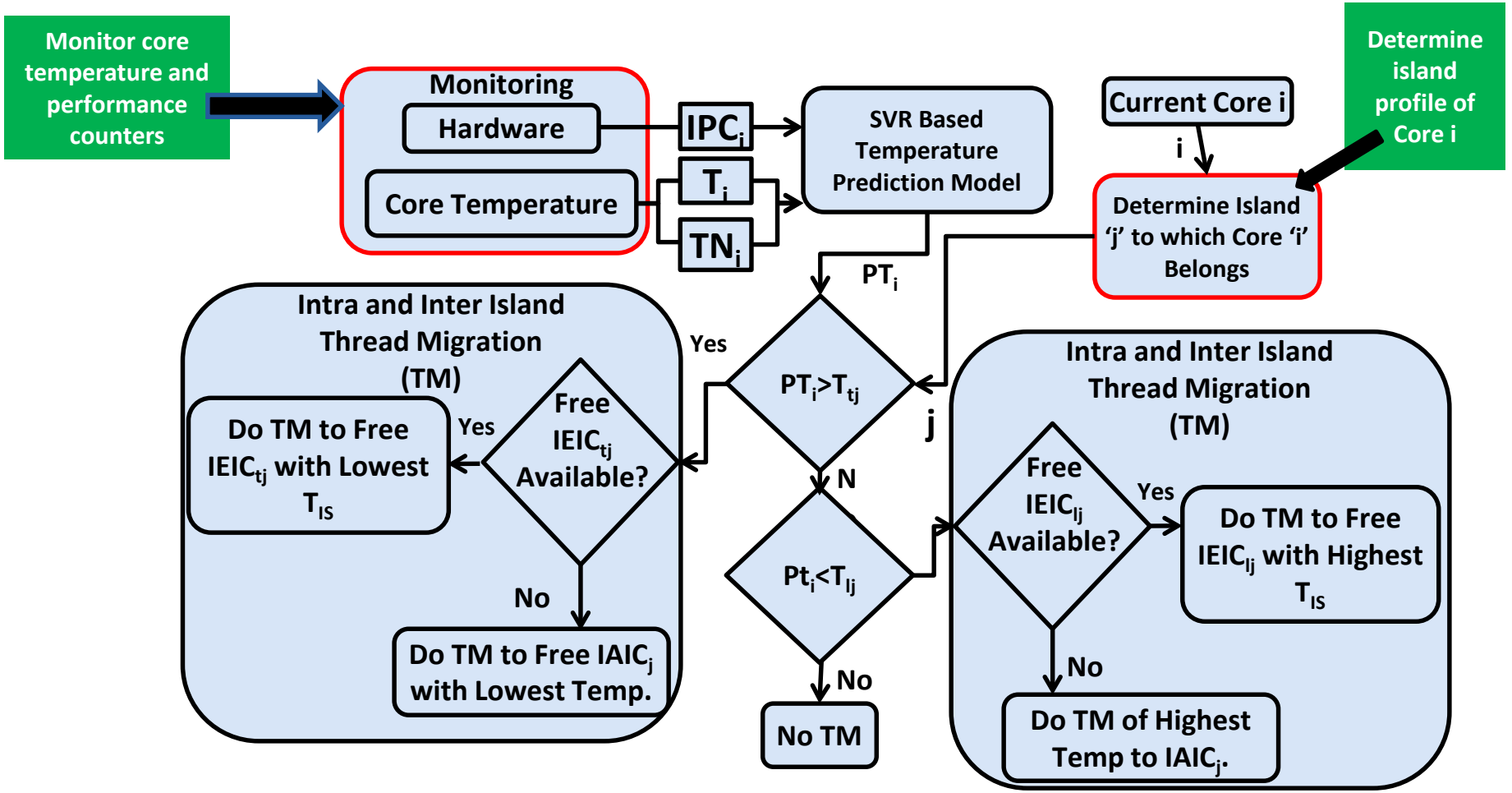
$$T_{island} = Temperature\ of\ the\ Island$$

1. ***Sensor-start***
2. $if\ (T\ !=\ T_{island})$
   a. $dT = |T_{island} - T|$
   b. $P_{Heat} = \frac{dT}{\rho} * H_{eff}$
3. $if(T \geq T_{island})$
   c. $i_{Heat} = i_{Current} - \sqrt{P_{Heat}/R_{Heat}}$
4. $if\ (T < T_{island})$
   c. $i_{Heat} = i_{Current} + \sqrt{P_{Heat}/R_{Heat}}$
5. $Delay\ (1 * 10^6)$ **// Delay of 1 mili-second**
6. **Loop continue**

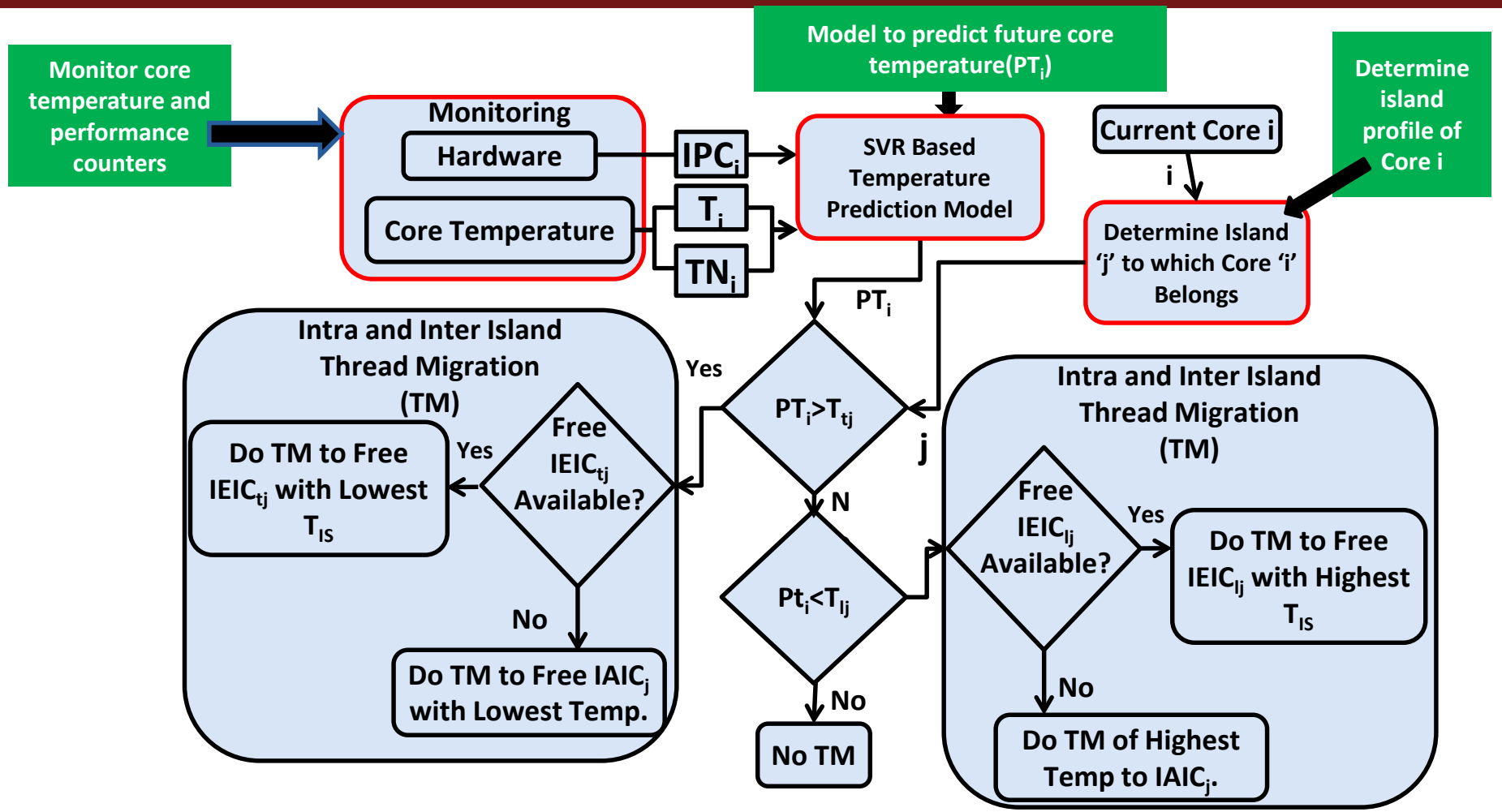




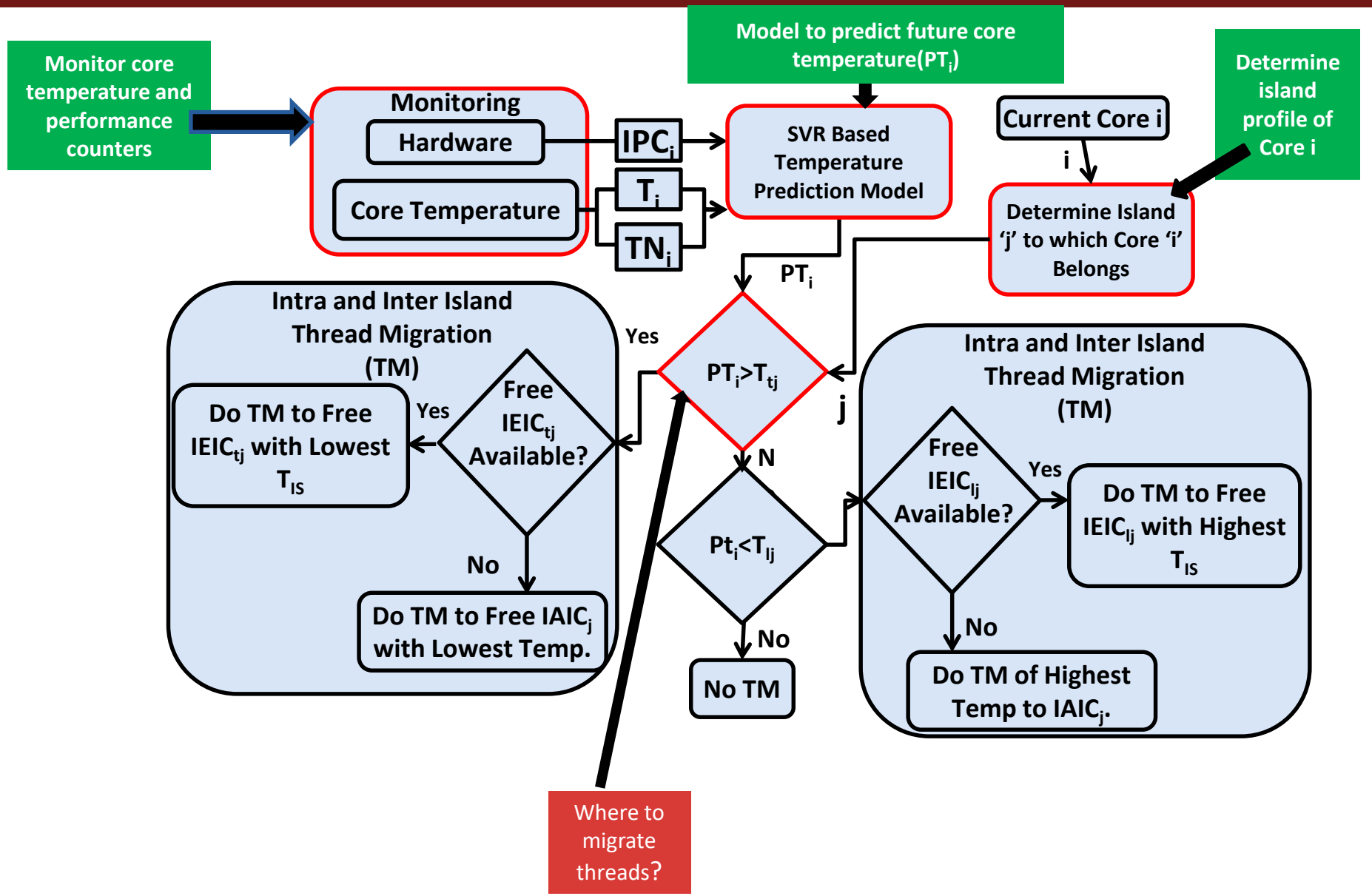
System-level TATM: Working Principle



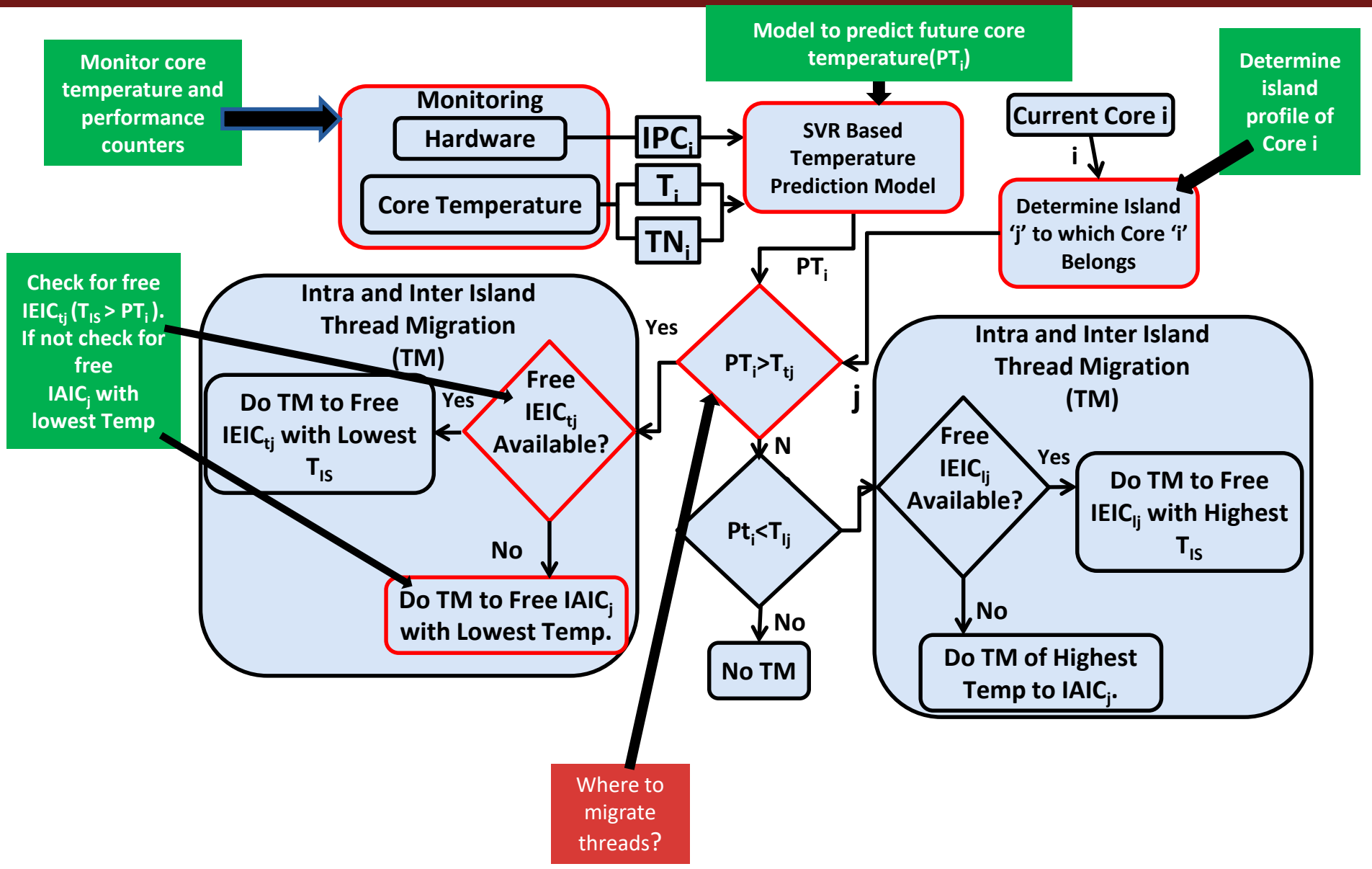
System-level TATM



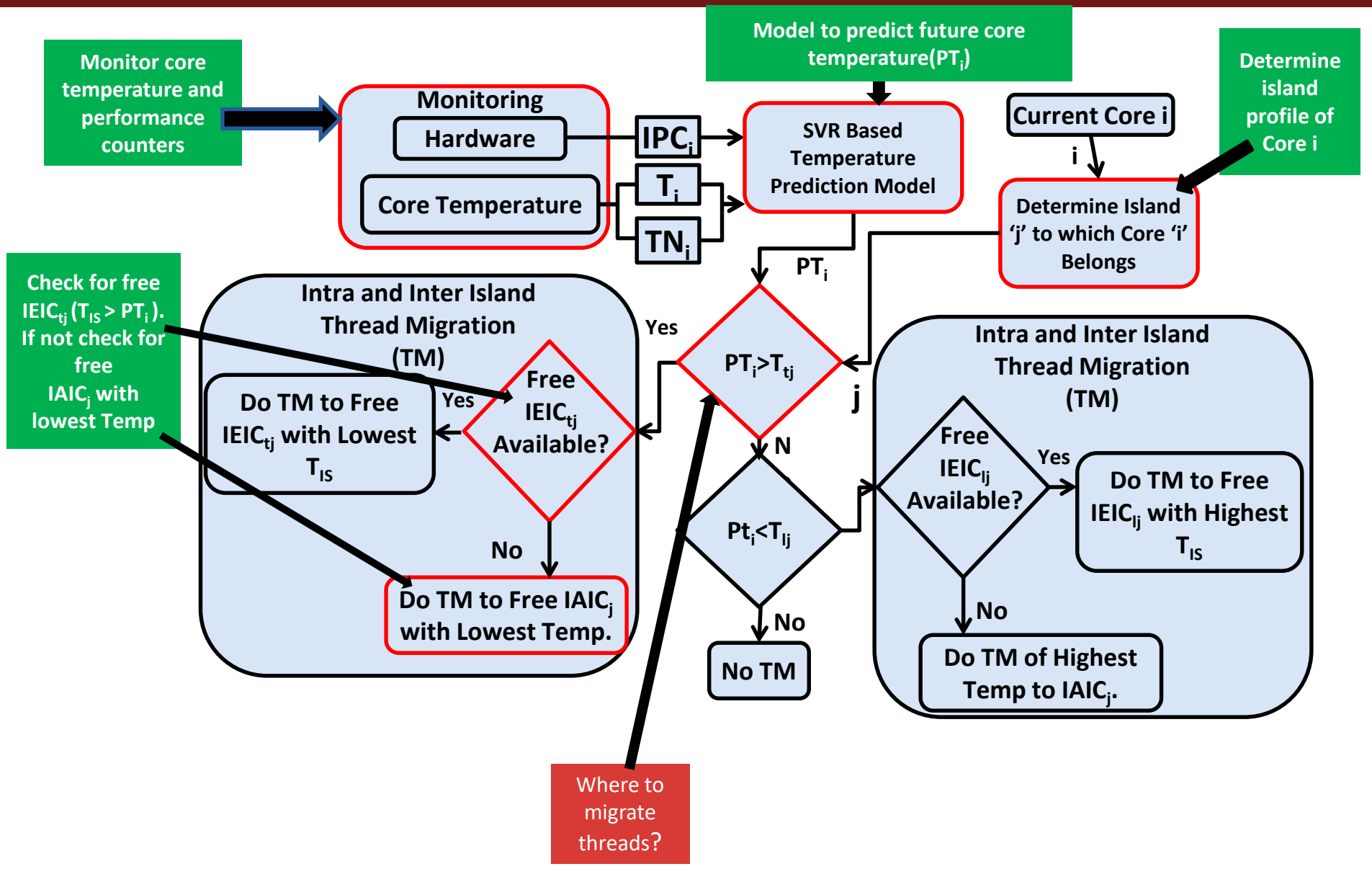
System-level TATM



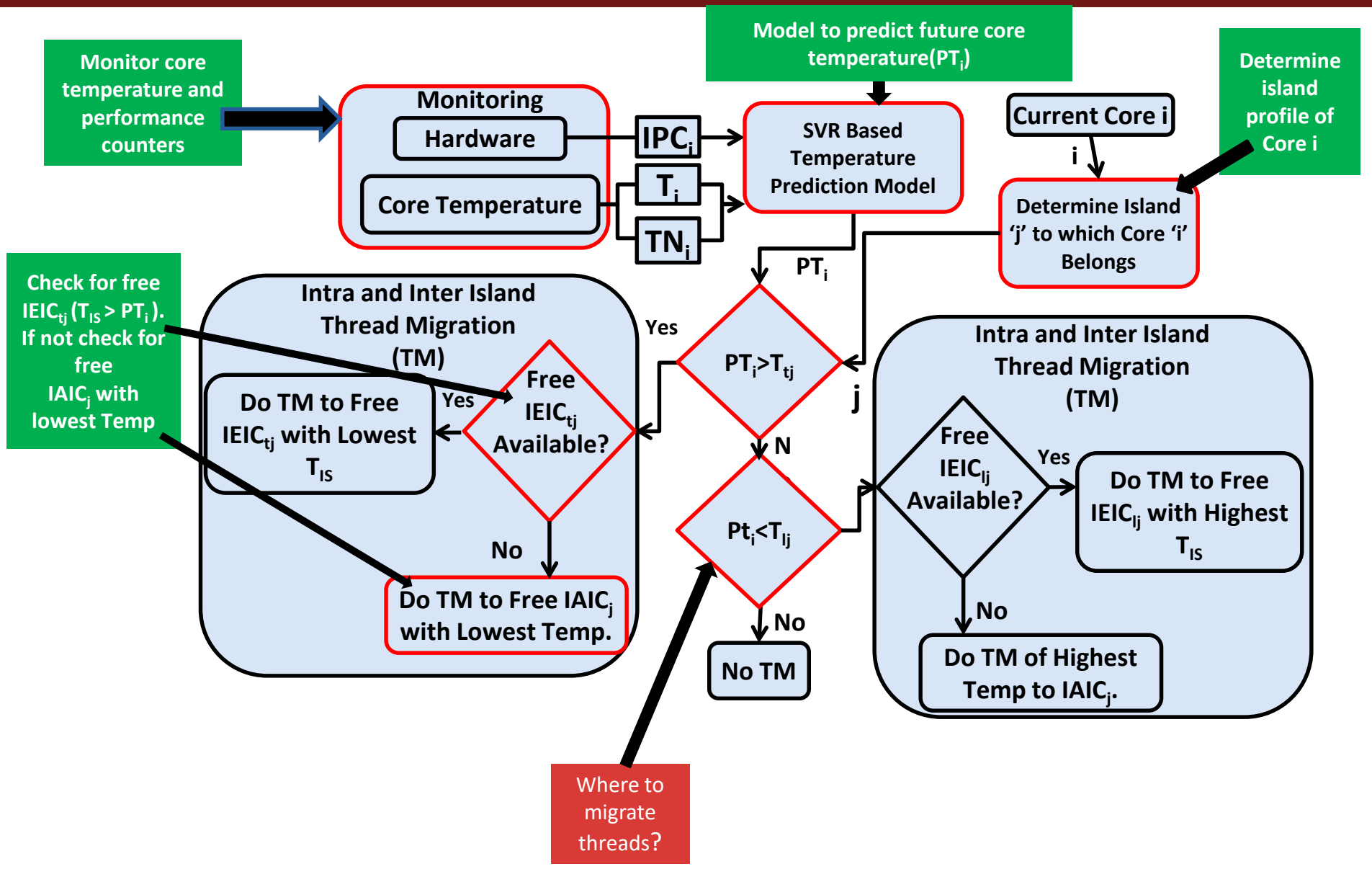
System-level TATM



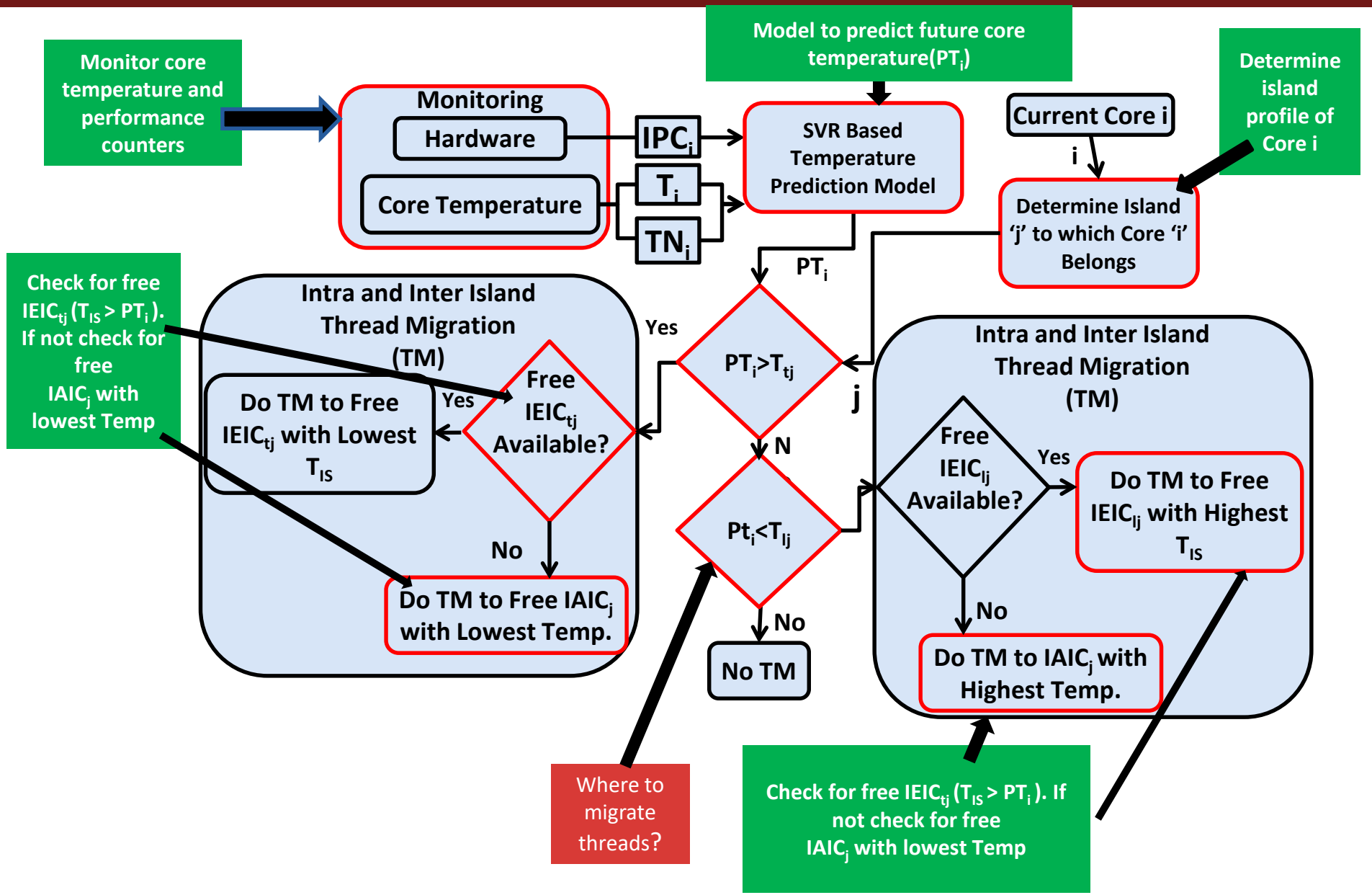
System-level TATM



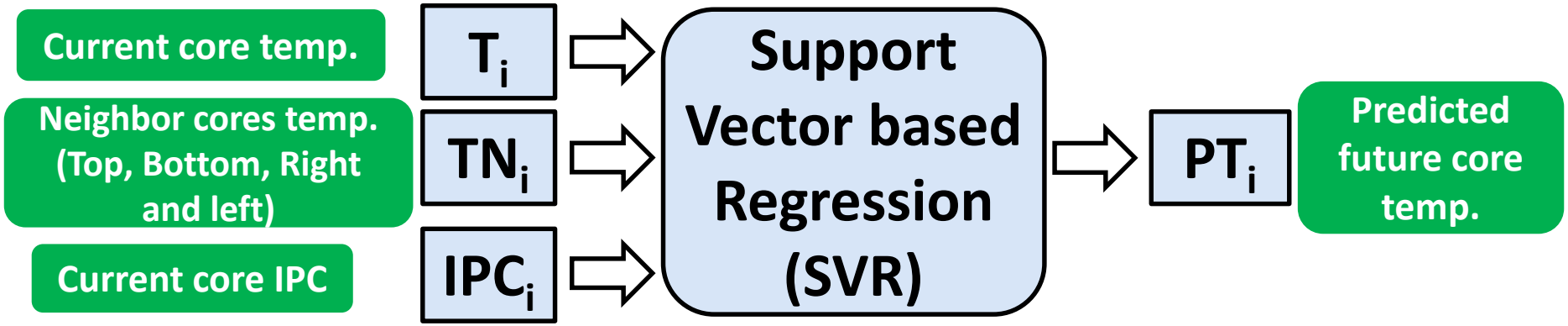
System-level TATM



System-level TATM



SVR based Temperature Prediction



- SVR model employs a kernel based regression
  - Kernel is Radial Basis Function (RBF)
- Training of SVR
  - 9-core (3×3) platform is used to generate training sets
  - with different thread mappings of PARSEC and SPLASH-2 benchmark apps
    - Running 2, 4 and 8 threads
- Determination of accuracy of SVR
  - 6000 floor plans are generated
    - 70% are used for training
    - 30% are used for testing
  - Accuracy of our SVR model is over 95%

Experimental Setup

- **We Analyzed our IHDTM Framework by porting it to PNoCs**
  - [D. Vantrease et al. MICRO 2009] **Corona** PNoC architecture with token slot arbitration and 64X64 multiple write single read (MWSR) crossbar
  - [Y. Pan et al. HPCA 2010] **Flexishare** PNoC architecture with token stream arbitration and multiple write multiple read (MWMR) crossbar
- CMP configuration for implementation for **Corona** and **Flexishare** PNoCs

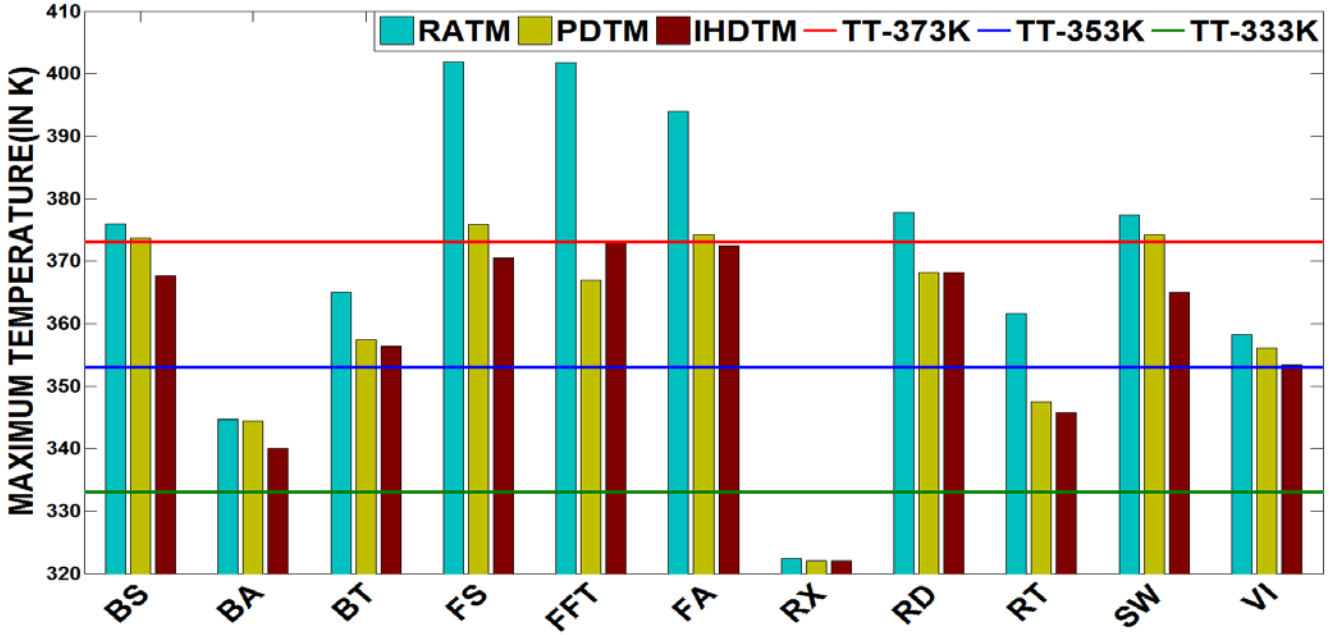
CMP Configuration	
Number of cores	64
Technology node	32nm
Memory controllers	8
Main memory	8GB; DDR4@30ns
Per Core:	
L1 I-Cache size/Associativity	32KB/Direct Mapped Cache
L1 D-Cache size/Associativity	32KB/Direct Mapped Cache
L2 Cache size/ Associativity	256KB/ Direct Mapped Cache
L2 Coherence	MOESI
Frequency	5 GHz
Issue Policy	In-order

Comparison with Prior Work

- **We compare IHDTM when ported to Corona and Flexishare with**
  - [T. Zhang et al. DATE 2014] a ring aware policy (RATM)
    - Distributes threads uniformly across cores that are closer to PNoC MRR clusters
    - then distributes the remaining threads in a regular pattern from outer cores to inner cores
  - [I. Yeo et al. DAC 2008] predictive dynamic thermal management (PDTM) framework
    - Uses a recursive least square based temperature predictor
    - When a core temperature is more than thermal threshold
      - Thread migration is performed to the coolest core that is not executing any threads

Max Temp for Corona with 48 Threads

Corona: D. Vantrease et al. MICRO 2009

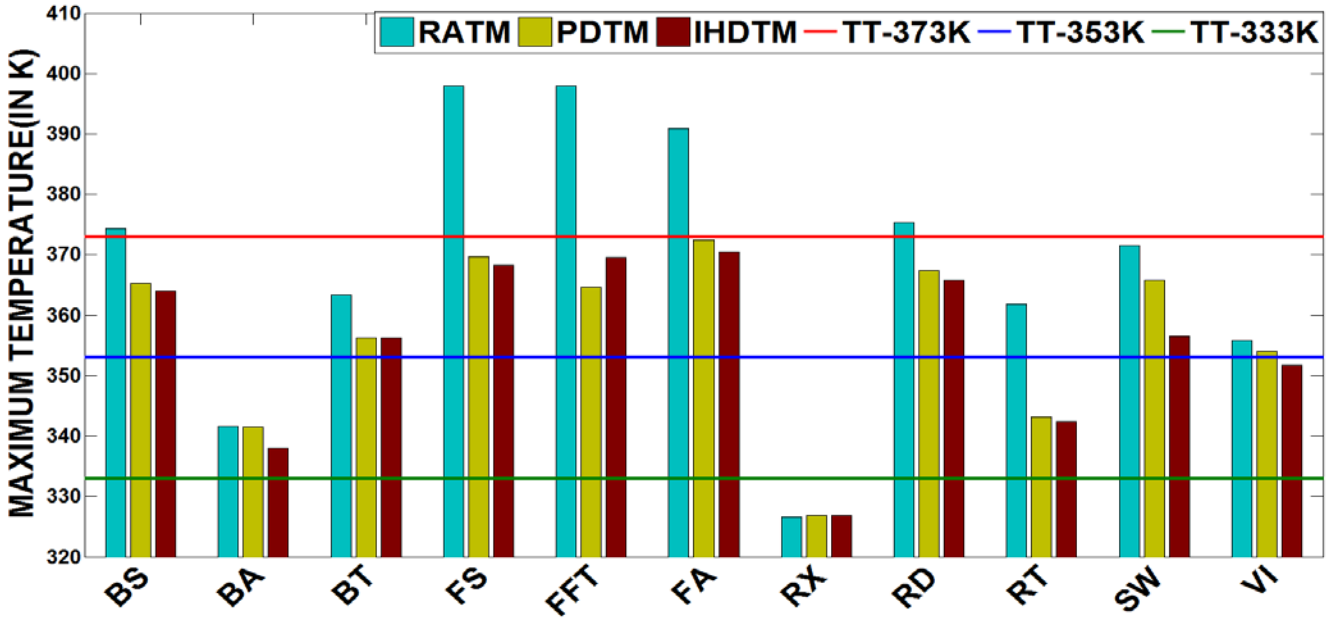


- With integrated localized tuning by PID heater, IHDTM lowers maximum temperature compared to RATM by

IHDTM has 13.2K and 2.37 K lower maximum temperatures compared to the RATM and PDTM policies respectively, for 48 threads

Max Temp for Corona with 32 Threads

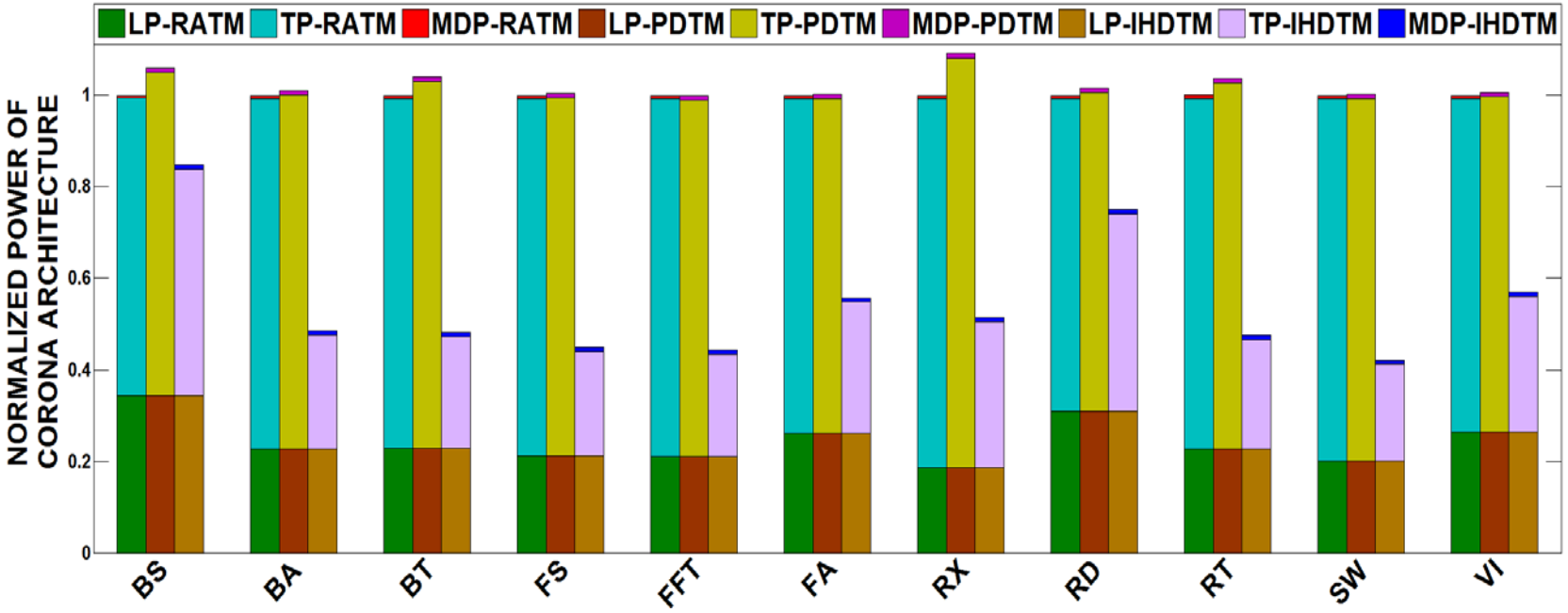
Corona: D. Vantrease et al. MICRO 2009



IHDTM has better maximum temperature control for 32 threads compared to 48 threads as more free cores are available for thread migration

Corona Power for 64-core CMP

Corona: D. Vantrease et al. MICRO 2009

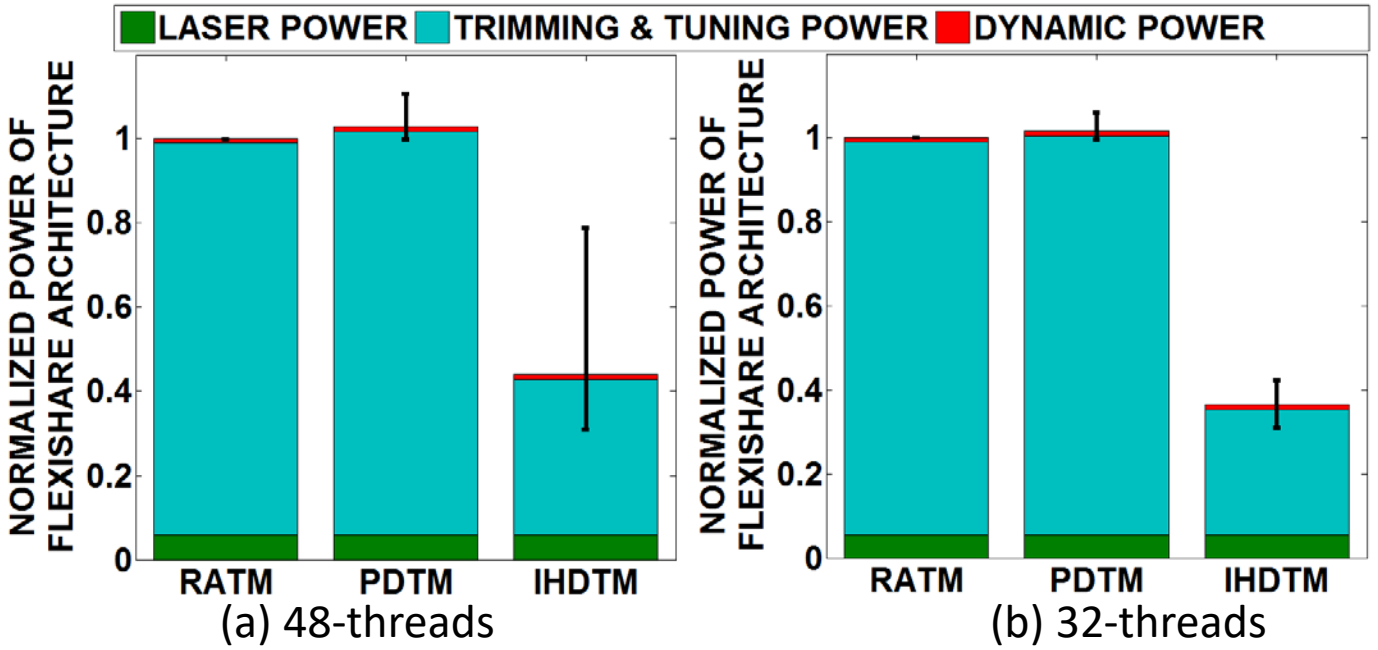


- IHDTM with Corona has lower power dissipation than RATM and PDTM

IHDTM with Corona has 45.5% and 46.8% lower total power consumption compared to Corona with RATM and PDTM respectively

Flexishare Power for 64-core CMP

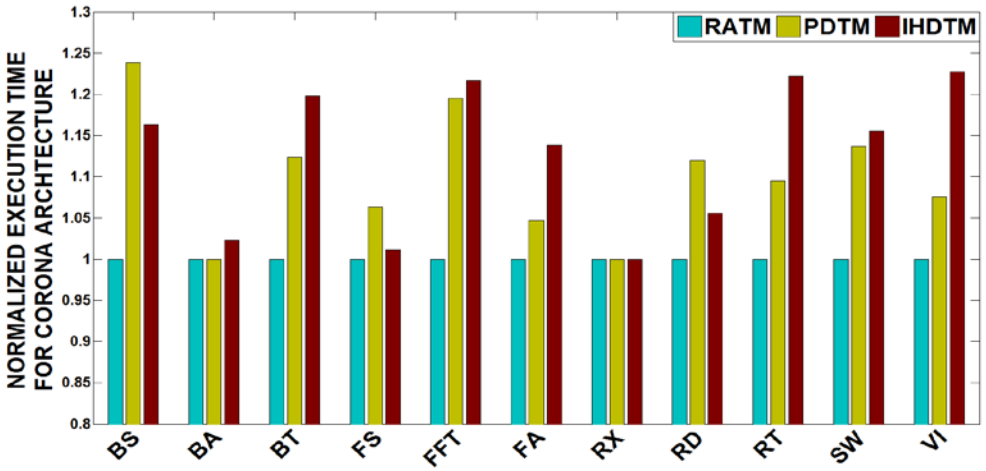
Flexishare: Y. Pan et al. HPCA 2010



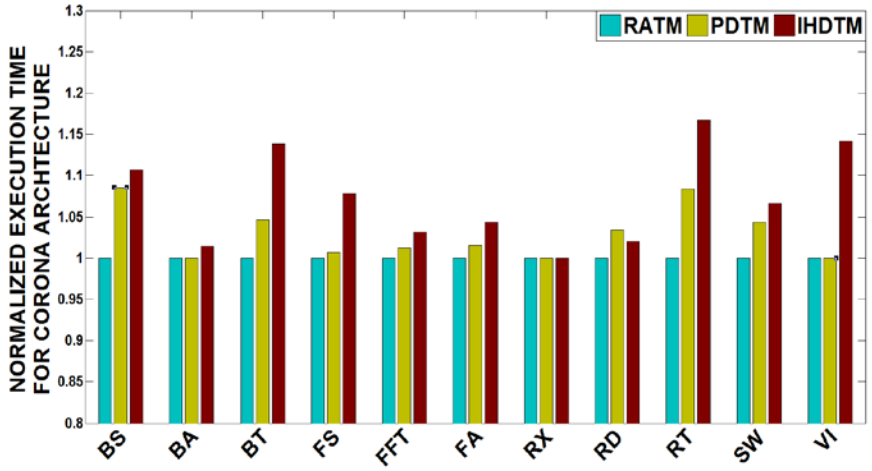
- Flexishare with IHDTM has 63.5% and 64.1% lower power dissipation compared to Flexishare with RATM and PDTM respectively

Flexishare with IHDTM has more power savings compared to Corona with IHDTM

Corona Execution Time for 64-core CMP



(a) 48-threads



(b) 32-threads

- Corona with IHDTM has 12.8% and 7.4% higher execution time compared to Corona with RATM for 48 and 32-threads respectively
  - ➢ IHDTM takes slightly extra time for migration whereas no migration in RATM
- Corona with IHDTM has 2.6% and 4.3% higher execution time compared to Corona with PDTM for 48 and 32-threads respectively
  - ➢ IHDTM performs more migration (inter and intra) when thermal emergency is predicted

Conclusion

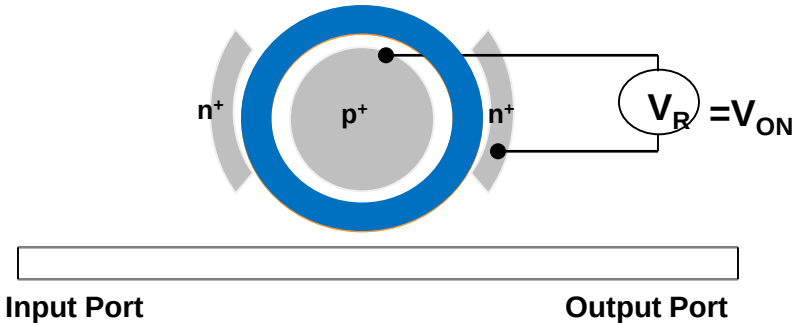
- Proposed IHDTM framework
  - ➢ combines a novel device-level framework with a new dynamic thermal management mechanisms to
    - ❑reduce maximum on-chip temperature
    - ❑conserve trimming and tuning power
- IHDTM improvements over state-of-the-art solutions
  - ➢ up to 64.1% (Total Power), 71% (Trimming/tuning power)
- IHDTM is more effective in reducing power for optimized PNoCs like Flexishare compared to Corona

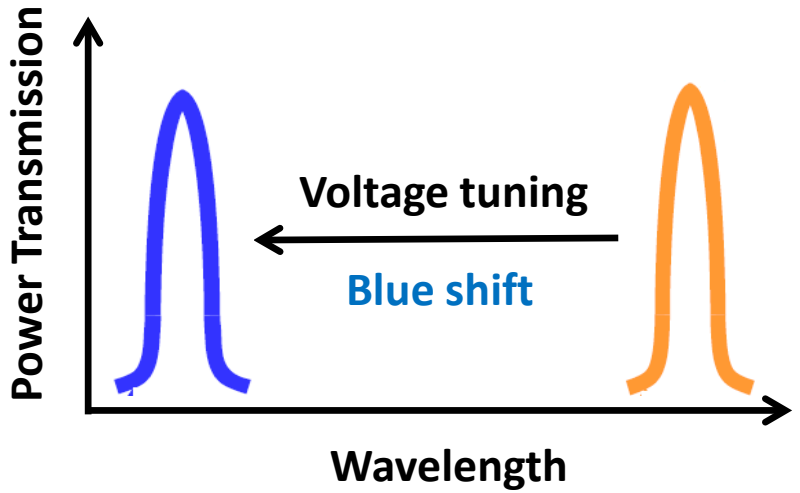
Thank You

- **Questions / Comments ?**

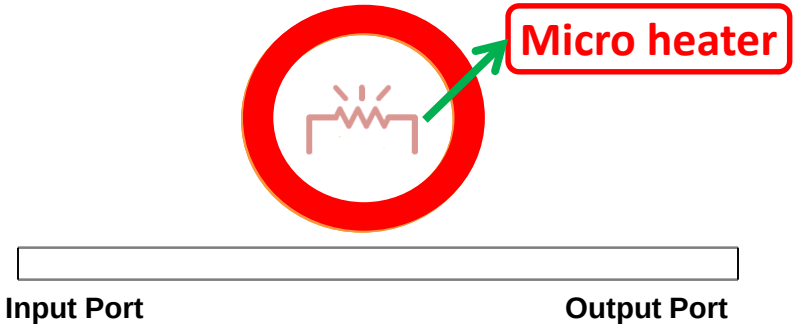
Extra:Tolerate Thermal Variations

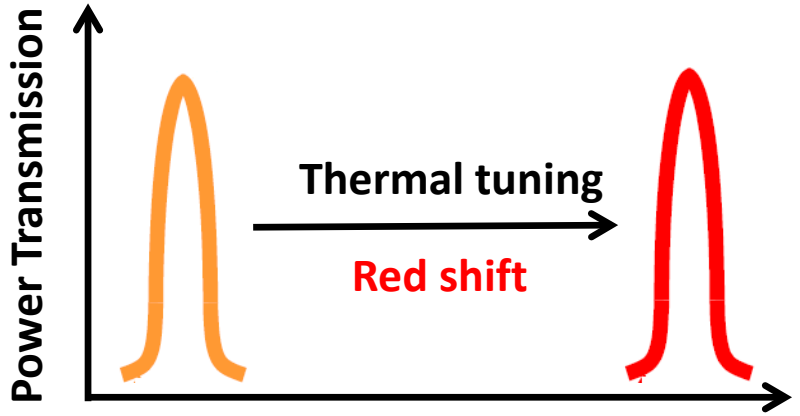
Voltage tuning:





Thermal tuning:





These solutions have higher power consumption and motivates to system level approach