Run-time Strategies for Energy-efficient Operation of Silicon-Photonic NoCs

Ajay Joshi

joshi@bu.edu

Integrated Circuits and Systems Group (ICSG)

Department of ECE, Boston University, Boston MA

BOSTON
UNIVERSITY

Acknowledgement

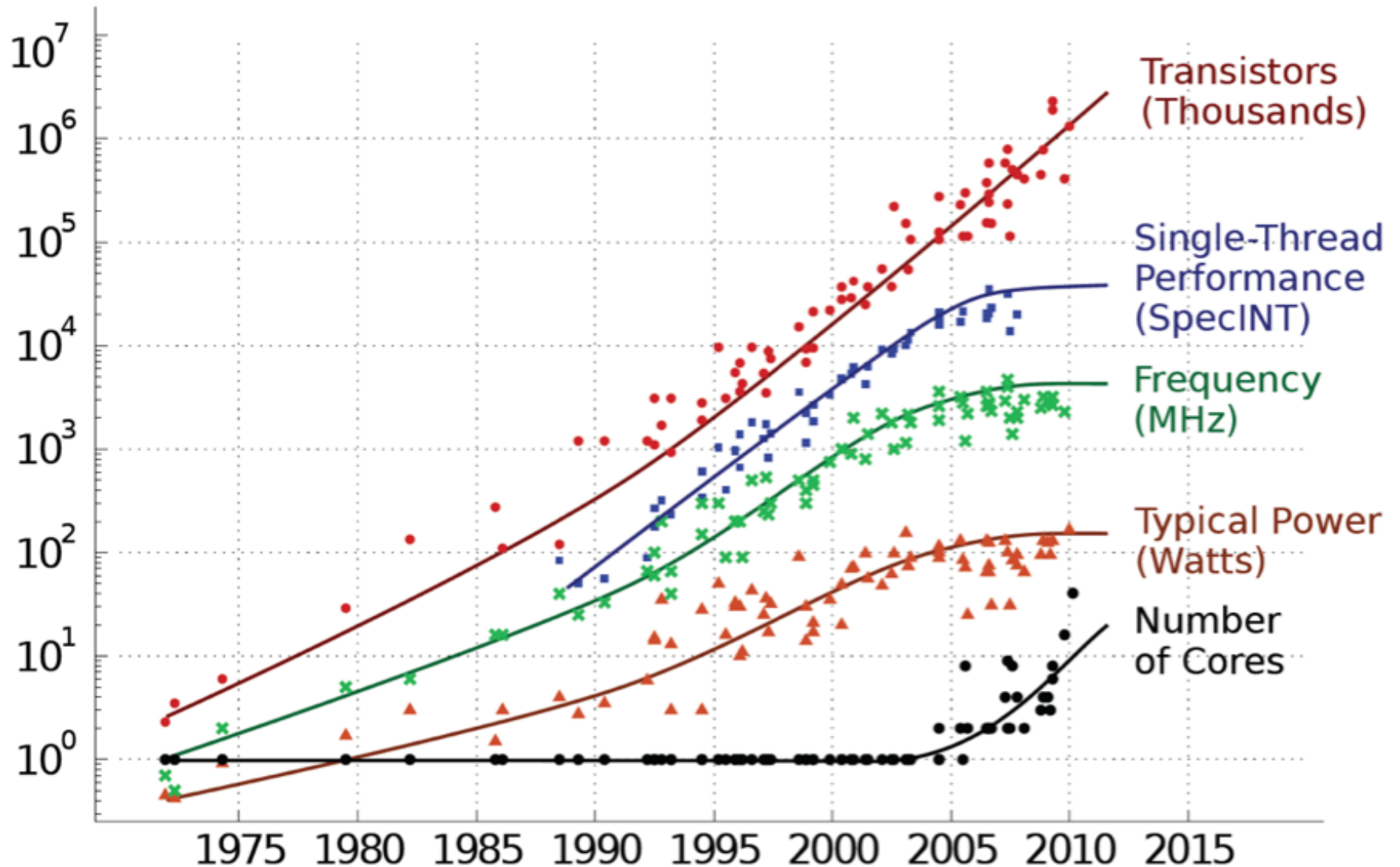- ## Graduate Students/Postdocs
  - Chao Chen, Jose Abellan, Tiansheng Zhang

- ## Faculty
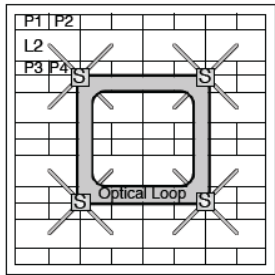 - Ayse Coskun, Jonathan Klamkin, Milos Popovic

- ## This research has been funded by the NSF grant CCF-1149549

2
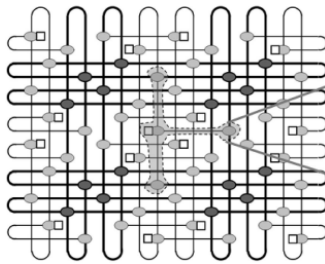
Joshi

Historic Trends in VLSI Systems

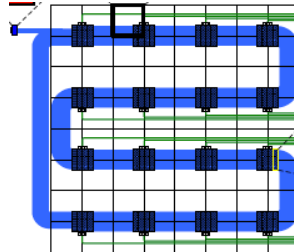Data collected by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, C. Batten

3
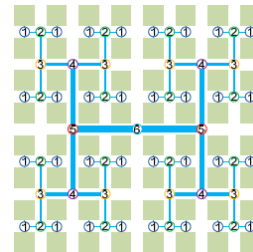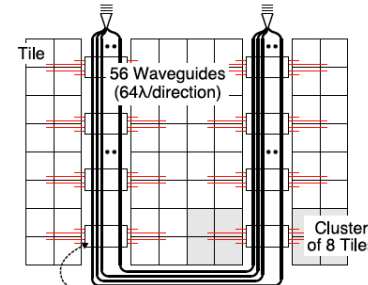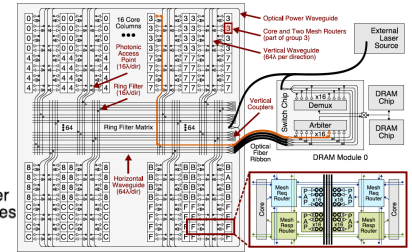
Joshi

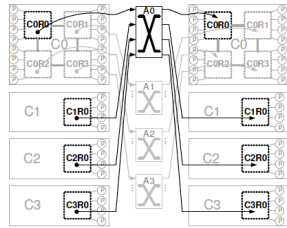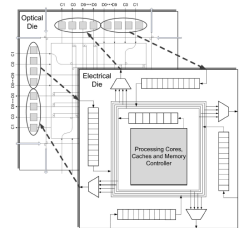# Silicon-Photonic NoC Research

[Kirman'06] [Shacham'08] [Vantrease'08] [Gu'09] [Joshi'09] [Batten'09]

[Pan'09] [Cianchetti'09] [Koka'10] [Pan'10] [Beamer'10] [Xue'10]
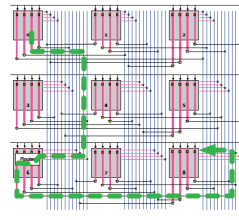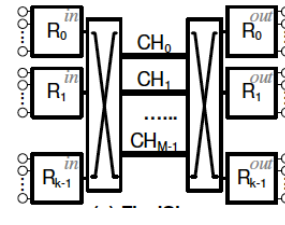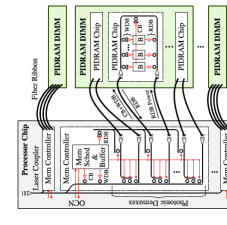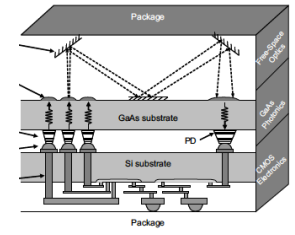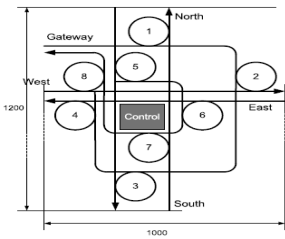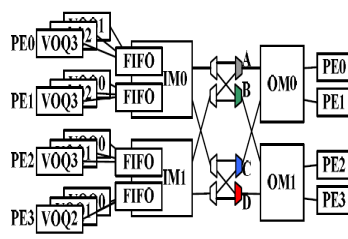
[Hendary'11] [Kao'11] [Li'13] [Zulfiqar'13] [Chen'14]

4

Joshi

Silicon-Photonic NoC Challenges

❑ **Bandwidth utilization**

- Current applications/architectures do not need Tbps on-chip bandwidth

❑ **Packaging/Integration**

- Interface electrical and photonic devices
- Coupling tens of off-chip laser sources to photonic NoC is challenging

❑ **Power consumption**

- Large laser power and thermal tuning power could negate bandwidth benefits

Joshi

Silicon-Photonic NoC Challenges

❑ Bandwidth utilization

  ▪ Current applications/architectures do not need Tbps on-chip bandwidth

❑ Packaging/Integration

  ▪ Interface electrical and photonic devices

  ▪ Coupling tens of off-chip laser sources to photonic NoC is challenging

❑ Power consumption

  ▪ Large laser power and thermal tuning power could negate bandwidth benefits

Joshi

Outline

❑ Background

❑ Laser Power Management using NoC and Cache Reconfiguration

❑ Thermal management using job allocation

❑ Summary

Joshi

Silicon photonic link

Joshi

Silicon photonic link

Joshi

Silicon photonic link

Laser source could consume 10-15 watts of power

10

Joshi

Related Work

- Low loss photonic devices [plenty of efforts in place]

- Channel sharing [Pan 2010] [Li 2013]

- NoC bandwidth scaling [Zhou 2013] [Chen 2013] [Demir 2014]

- Sharing/Placement of laser sources [Chen 2014]

We use a combination of NoC and Cache reconfiguration to save laser power

BOSTON UNIVERSITY

Joshi

Cache Reconfiguration Process

L2 cache bank deactivation

L1$-a   L1$-b   L1$-c   L1$-d

Dirty (L1$-c)

3.2 Req

2.1 Req

1.2 Flush

1.1

Owner (c)
Shared (c,d)
L2$-b
Dirty

L2$-a

2.2 Nack

1.X WB   3.1 Ack

Memory
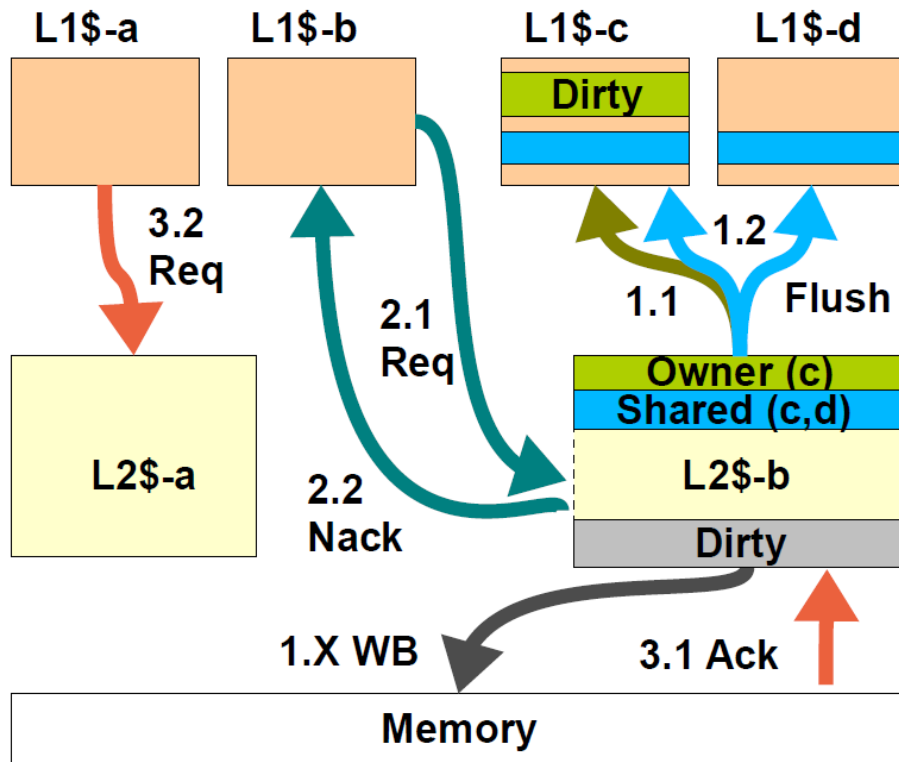
L2 cache bank activation

L1$-a   L1$-b   L1$-c   L1$-d

Dirty (L1$-a)   Dirty (L1$-b)   Dirty (L1$-c)

2. Req/Nack   1.1 Flush   3.2 Req

a
b
a
b
a
b

L2$-a

L2$-b

3.1 Ack   1.X WB

Memory
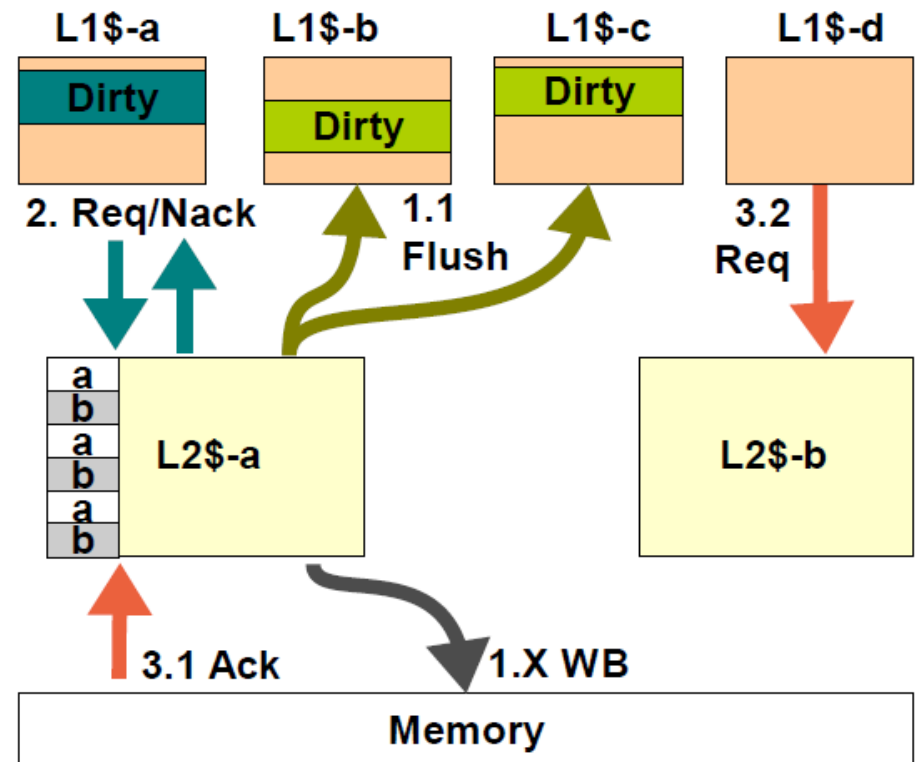
[Musalappa'06] [Sim'12] [Qureshi'07] [Wu'11]

12

Joshi

Decision on Cache Reconfiguration

❑ L2 cache replacement rate is used for making decision on increasing or decreasing L2 cache bank count

❑ Dual threshold approach is adopted

  ▪ We use T_{high} (log) = -3 to keep the average performance degradation across all benchmarks less than 10%

  ▪ We use T_{low} (log) = -4.5 to minimize fluctuations in L2 cache bank count

Joshi
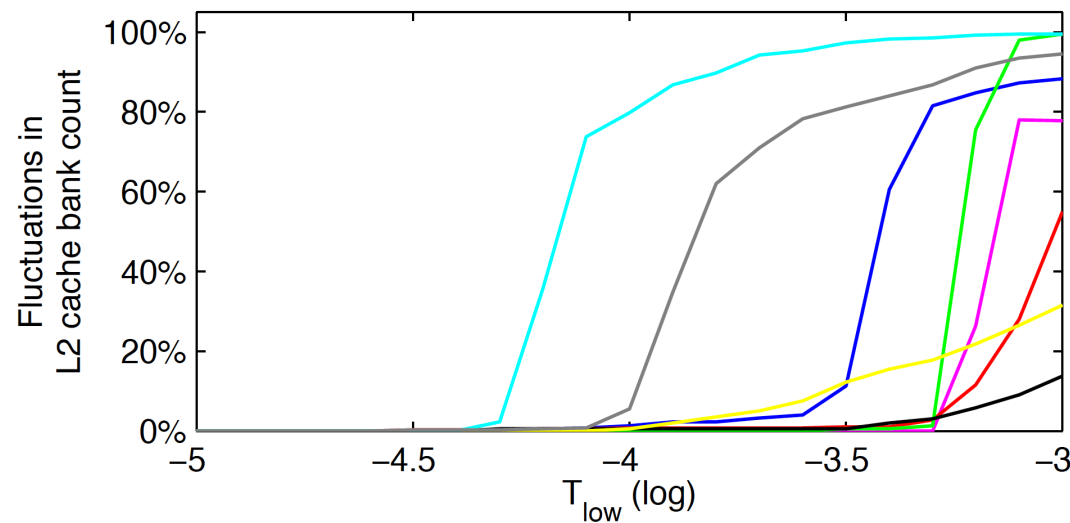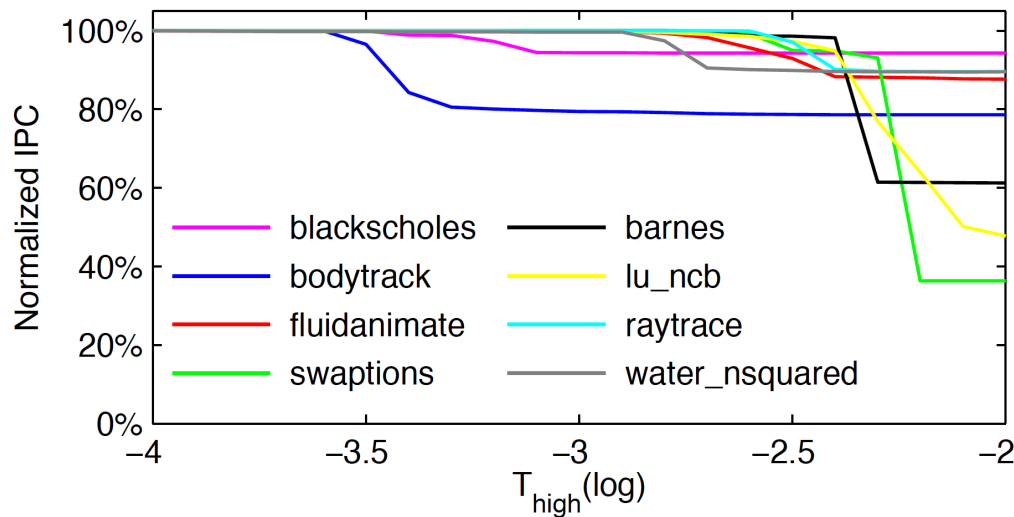
Reconfiguration Policy – Flow Chart

Switch to new configuration and continue execution

Start

Wait till the current time interval is complete

Activate L2 cache bank(s)

Switch OFF laser source

Continue execution with current L2 bank count till laser source stabilizes

Get statistics from performance counters

Deactivate L2 cache bank(s)

Switch ON laser source

Decide on L2 bank count

Decrease L2 cache bank count

Increase L2 cache bank count

No change in L2 cache bank count

14

Joshi

Target System

- 64-core in 22 nm @ 1.25 GHz

- 16 KB Private L1, 4 MB Shared L2 (8 banks)

- Crossbar NoC topology with 512-bit channels

  - L1-to-L2 communication uses Multiple Write Single Read (MWSR) arbitration

  - L2-to-L1 communication uses Single Write Multiple Read (SWMR) arbitration

- Evaluation

  - GEM5

  - McPat + Cacti + Inhouse setup for NoC

8 shared L2 banks

64 cores with private L1 caches

Couplers driven by on-chip laser sources

Joshi
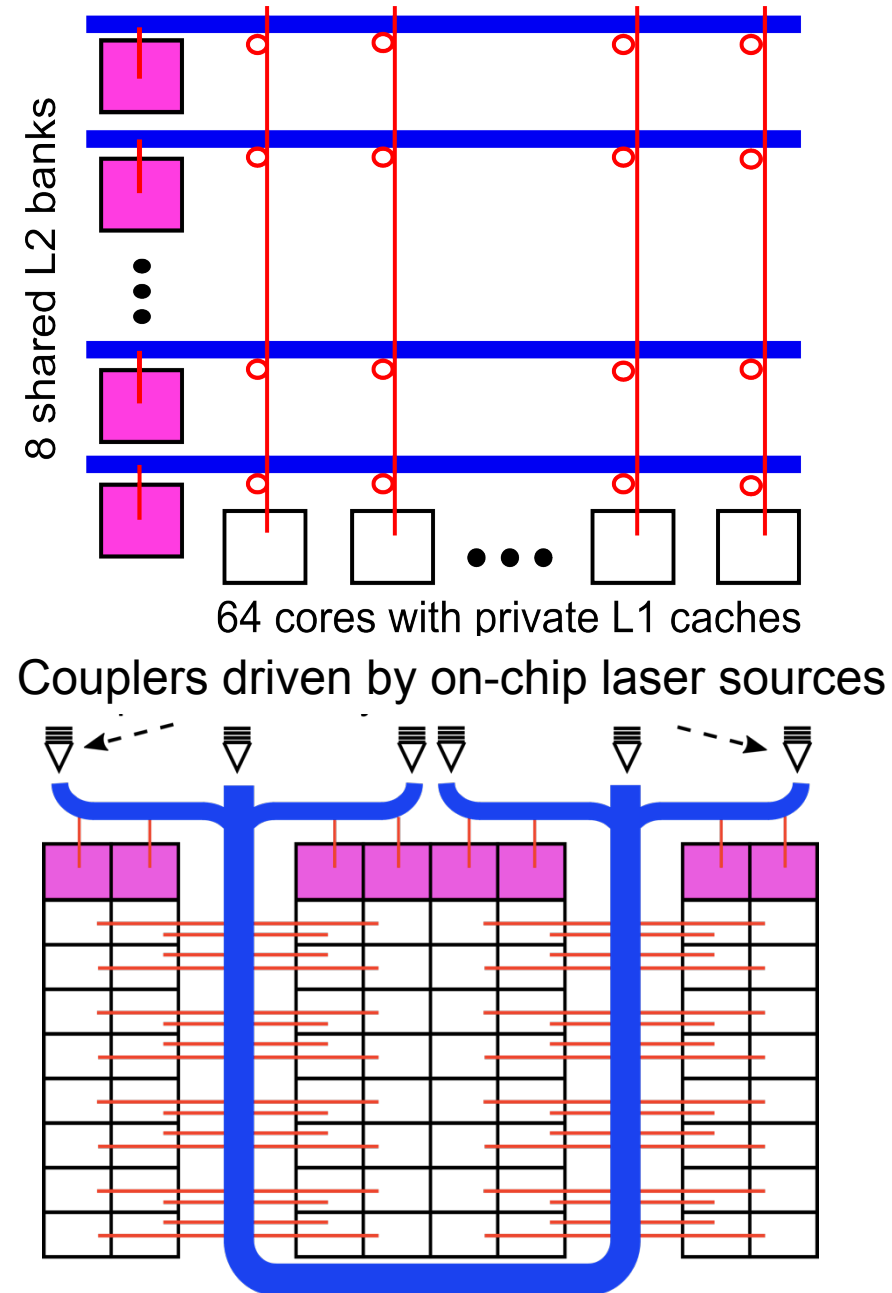
Evaluation – IPC, Replacement rate, Bank count

Joshi
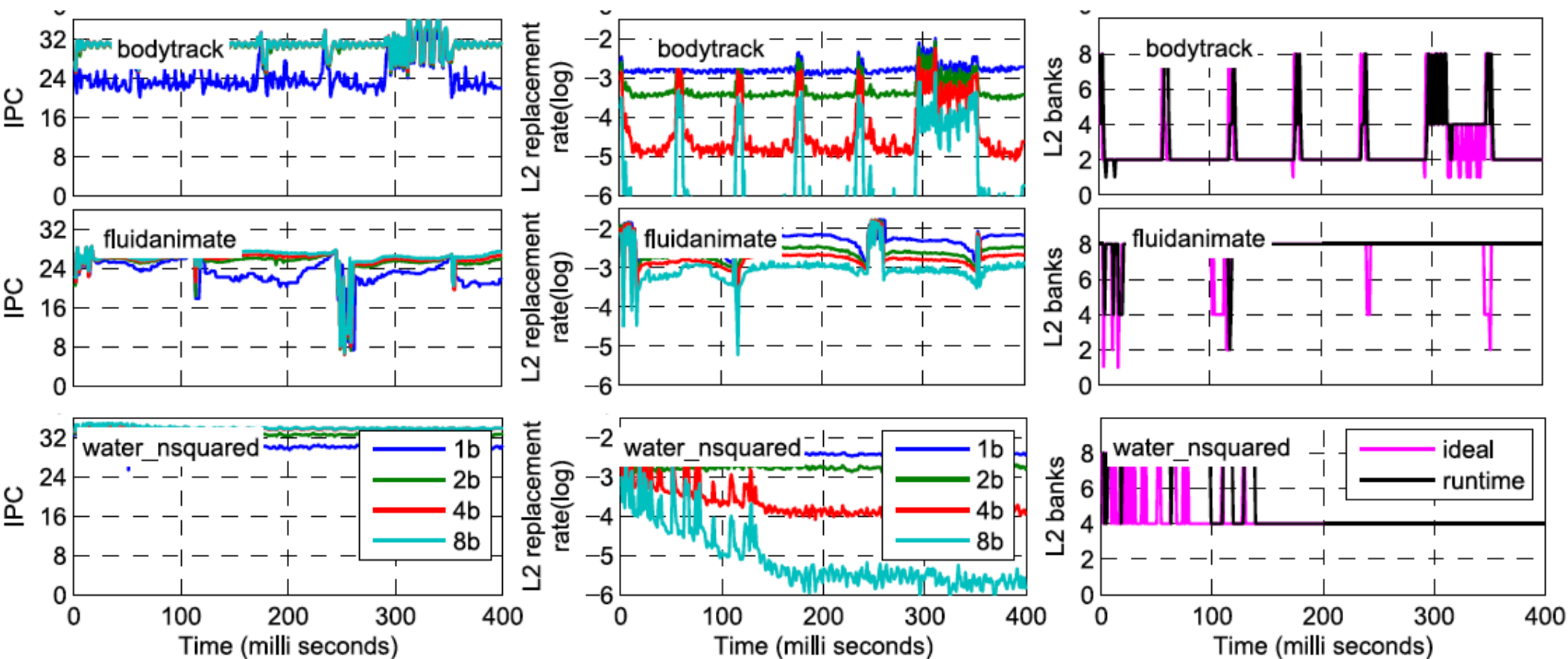
Evaluation – IPC, Power

❑ Target system: 64 cores and 8 L2 banks

- 23.8% saving in laser power on average (74.3% peak) with 0.65% IPC degradation on average (2.6% peak)
- 9.9% reduction in system power on average (30.6% peak) with 9.2% improvement in EDP on average (26.9% peak)

Joshi
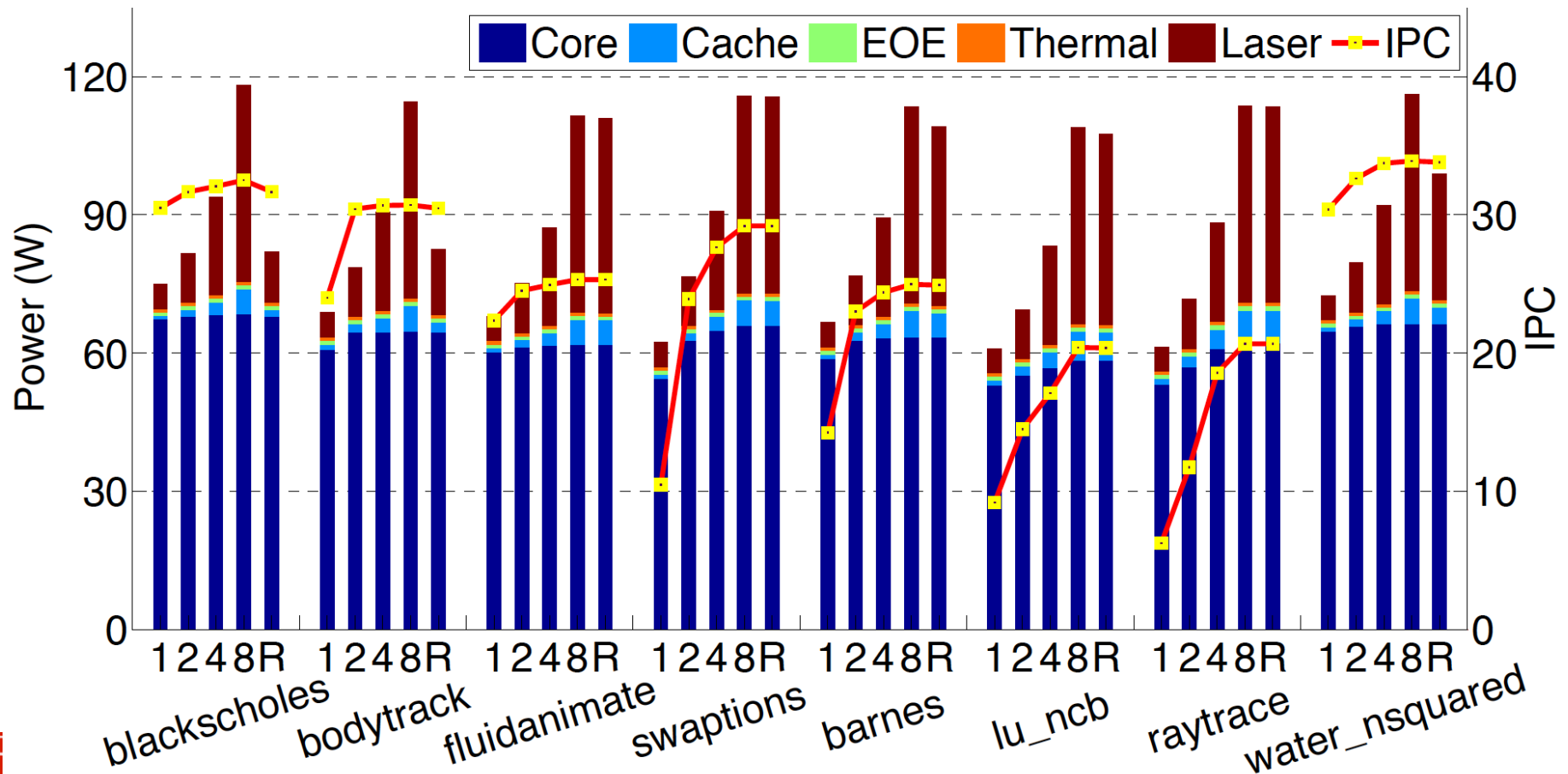
Evaluation – Reconfiguration Overhead

❏ Reconfiguration involves flushing L1 and L2 back to memory or fetching memory blocks to activated L2 banks

  ▪ Upto 18,000 cycles required for reconfiguration
  ▪ 150 uJ energy overhead for DRAM accesses

Joshi
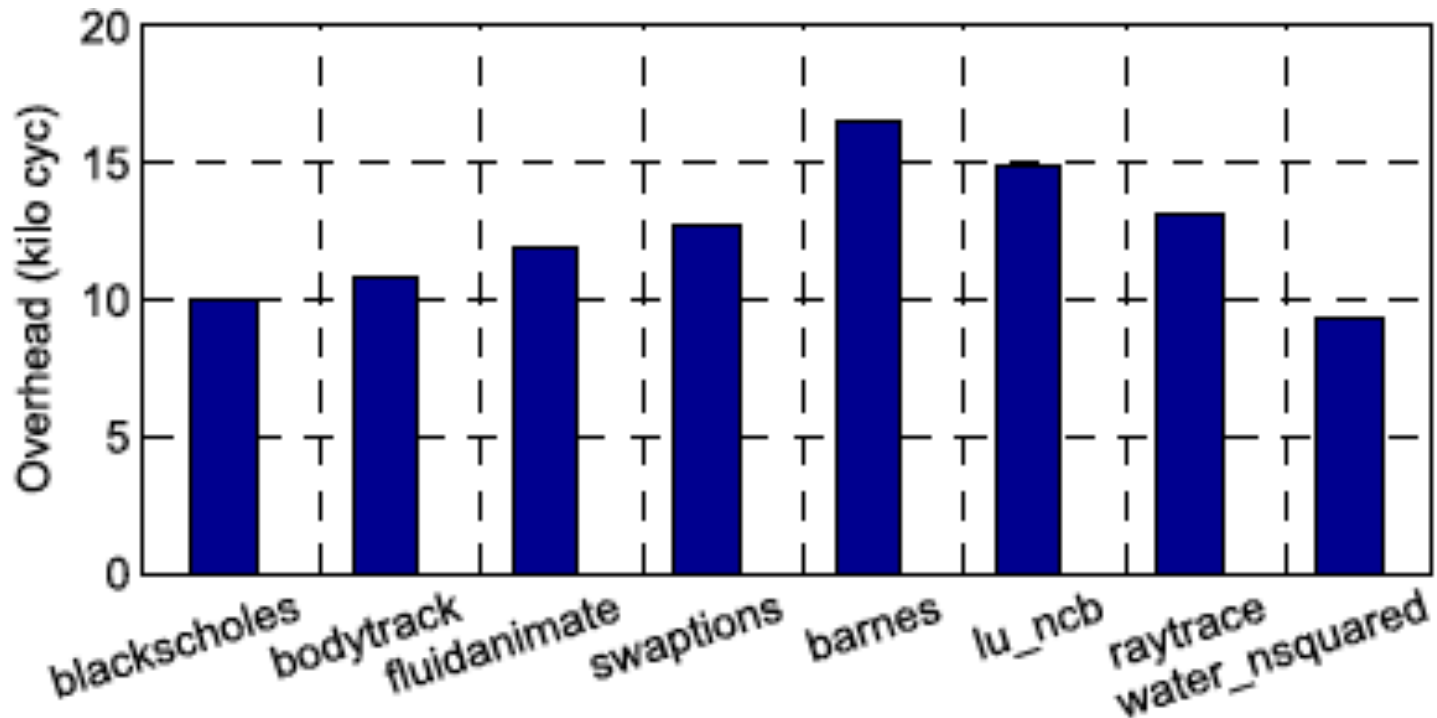
Summary I

❑ We proposed to manage the laser power by reconfiguring the NoC bandwidth based on the temporal and spatial variations in the cache size required by applications

❑ We adopted dual threshold approach to determine the L2 bank count at runtime

❑ On a 64-core target system, our proposed technique reduces laser power and system power by 23.8% and 9.9%, respectively, and improves EDP by 9.2% on average

BOSTON UNIVERSITY

19

Joshi

Outline

❑ Background

❑ Laser Power Management using NoC and Cache Reconfiguration

❑ Thermal management using job allocation

❑ Summary

20

Joshi

Silicon photonic link - DWDM

Ring Filter
with λ2 Resonance

Laser
Source

Ring Modulator
with λ_1 Resonance

Ring Modulator
with λ_2 Resonance

Ring Filter
with λ_1 Resonance
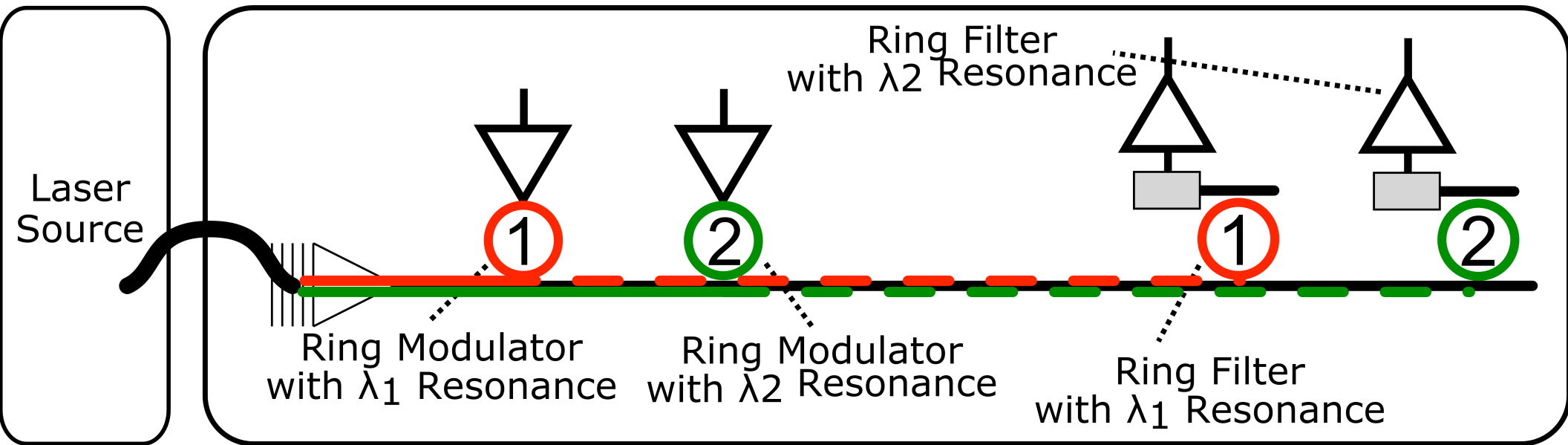
- ❑ Dense WDM (as much as 64 λ/wg, 10 Gbps/λ) improves bandwidth density

- ❑ Thermal management becomes very challenging

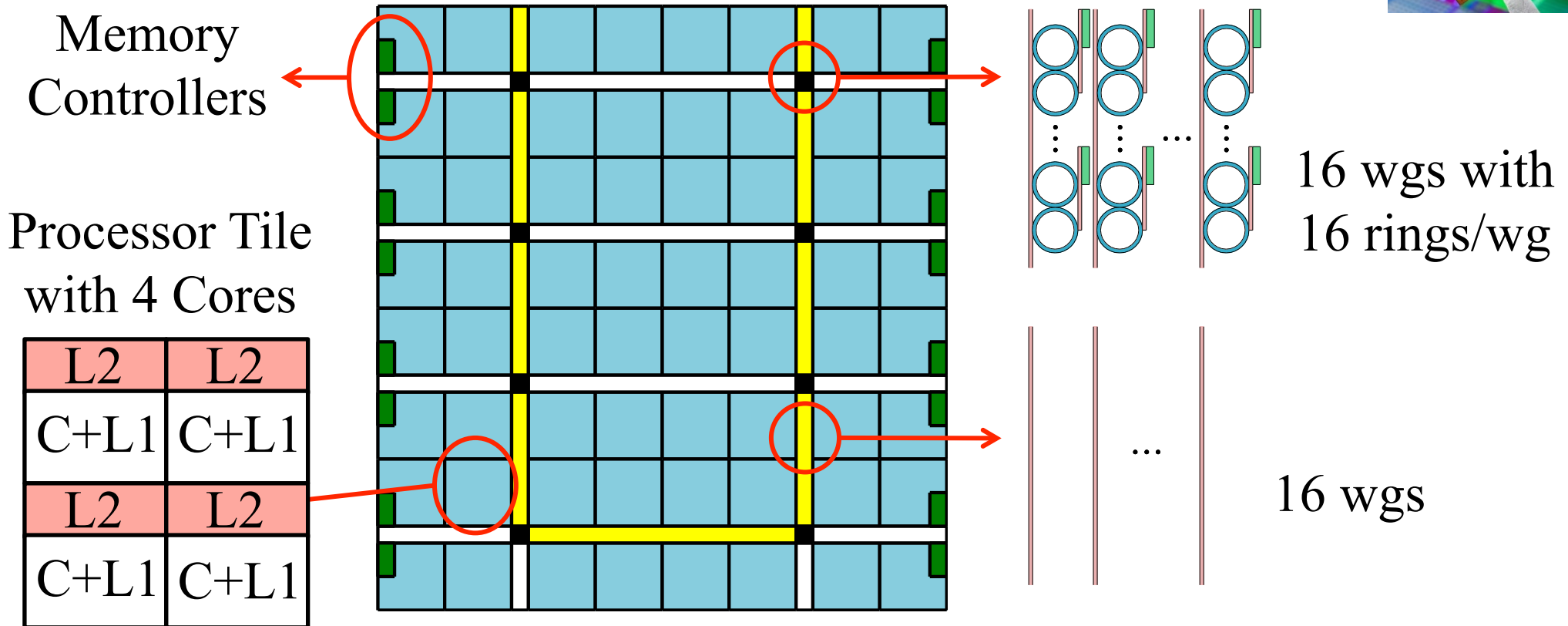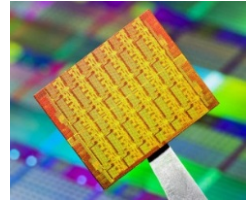   - ▪ Thermal tuning could cost more than 10 W of power

Joshi

Related Work

❑ Thermal control through novel ring filters/modulators designs

- Cladding [Djordjevic, Optical Exp.' 2013]
- Heaters [Zhou, TACO'2010; Li, TVLSI'2012]
- Mach-Zehnder interferometers [Biswajeet, Optical Exp.'2010]

❑ Techniques for thermal management in manycore systems

- DVFS [Quan, DAC'2001]
- Workload migration [Zhou, TACO'2010]
- Liquid cooling [Coskun, VLSI-SoC'2009]

We use workload allocation technique to minimize the thermal gradients among photonic devices
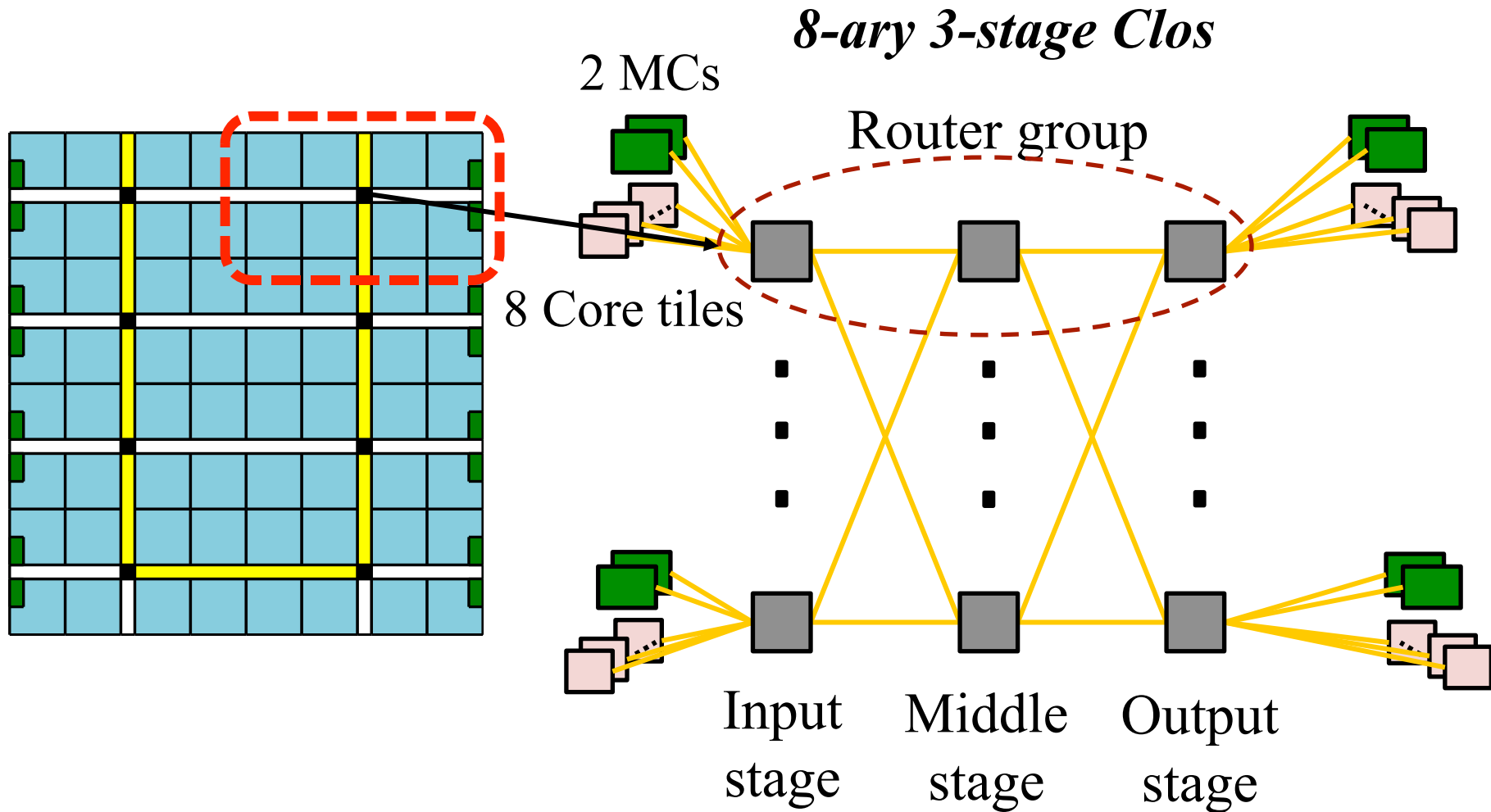
Target Manycore System Architecture

- 256-core system with Clos network

Core Architecture: IA-32 core in Intel SCC [*Howard,ISSCC2011*], 16KB I/D L1 cache & 256KB L2 cache; Average power consumption: 1.166W

Memory Controllers

Processor Tile with 4 Cores
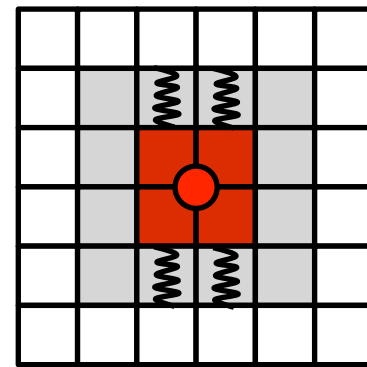
L2	L2
C+L1	C+L1
L2	L2
C+L1	C+L1

16 wgs with 16 rings/wg

...

16 wgs

Joshi

Target Manycore System Architecture

- 256-core system with Clos network

8-ary 3-stage Clos

2 MCs

Router group

8 Core tiles

Input stage Middle stage Output stage
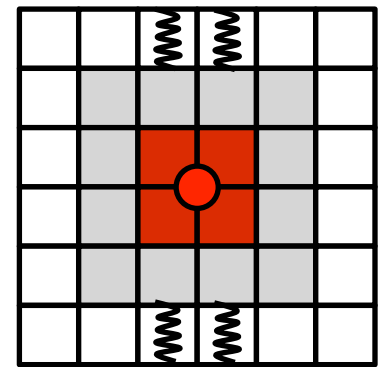
26

Joshi

Ring-Aware Thermal Management

❑ **Core power impacts on ring temperature**

❑ **Goals:**
  ▪ Minimize the difference among ring temperatures
  ▪ Reduce the overall chip temperature

❑ **Approach:**
  ▪ Classify the cores based on their distances to a ring group
  ▪ Determine the impact of each region on ring temperature
  ▪ Allocate threads so as to minimize thermal gradients

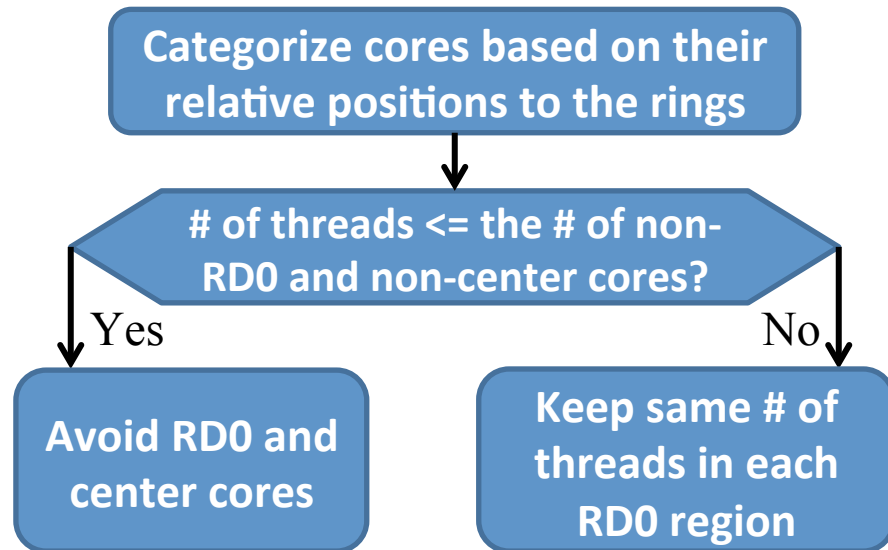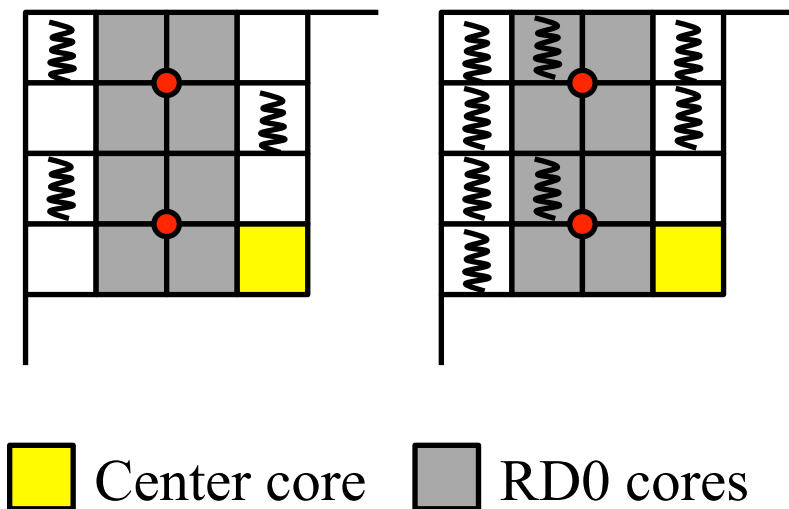● Rings   ■ RD0 cores   ▢ RD1 cores   ▢ RD2 cores   ⧤ Threads

Ring Temp. Gradient: *7.5°C*

<1°C   <1°C

Joshi
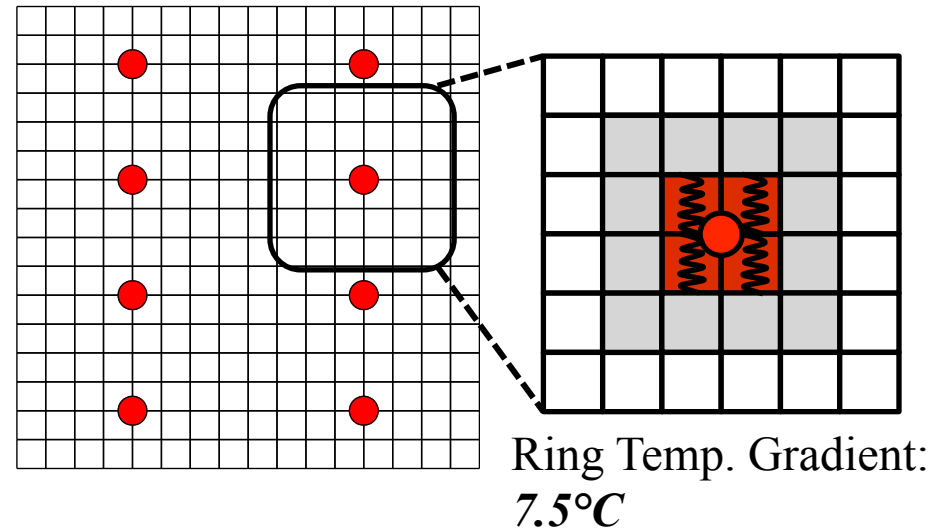
Ring-Aware Thermal Management

- ***Ring-aware workload allocation***

Center core   RD0 cores

Categorize cores based on their relative positions to the rings

# of threads <= the # of non-RD0 and non-center cores?

Yes

No

Avoid RD0 and center cores

Keep same # of threads in each RD0 region

- ***Multi-program support***
  - Sort the applications and threads based on their power dissipation & allocate high-power application first

Joshi

# Experimental Methodology

□ **Simulation Platform**

- Performance: Sniper simulator + SPLASH-2 & PARSEC

- Power: McPAT 0.8 + Temperature dependent leakage power model

- Thermal: HotSpot 5.02

□ **Simulated Systems:**

- 256-core system with silicon-photonic Clos NoC

- Tech. Node: 22nm; Area: 340 mm^2

- Single-application and multi-program workloads

- Utilization scenarios: 32, 64, 96, 128, 156, 180, 206, 230 and 256 threads

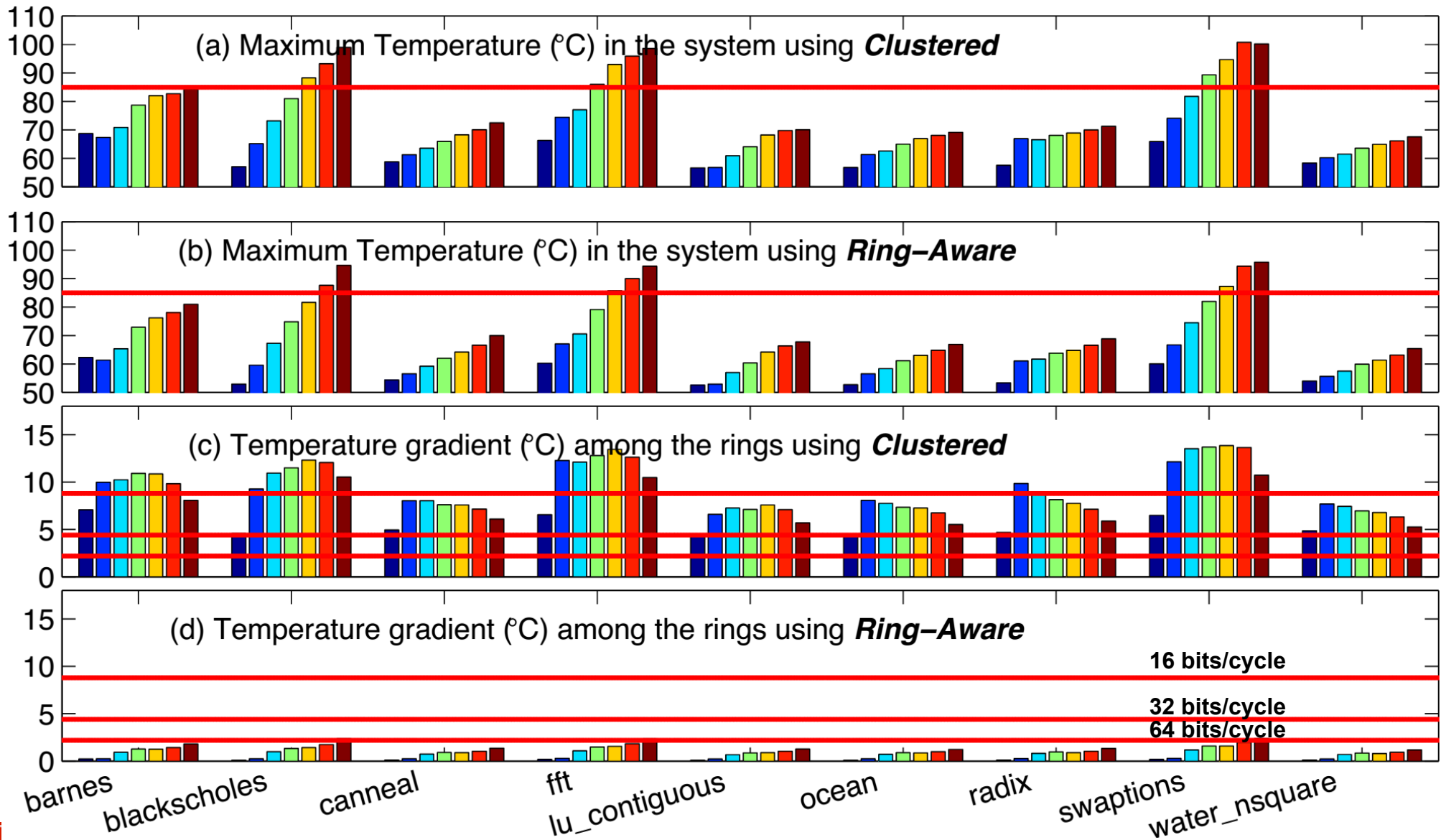Single application workload   Multi-program workload
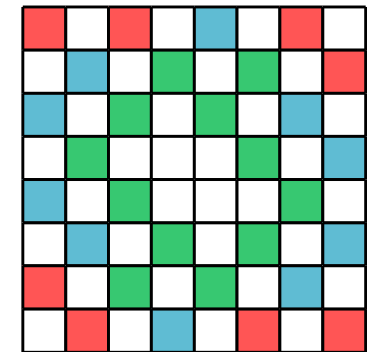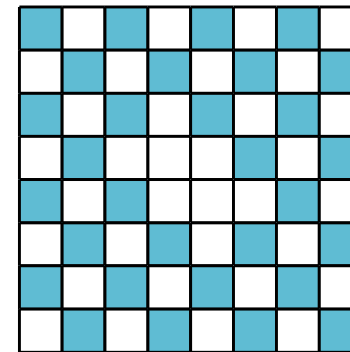
– *Allocation Policies:*

  - *Clustered*

  - *Chessboard*

  - ***Ring-Aware***

BOSTON UNIVERSITY

Joshi

Evaluation of Single-Application Workloads

Joshi

Performance Results

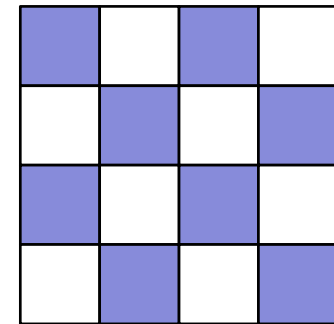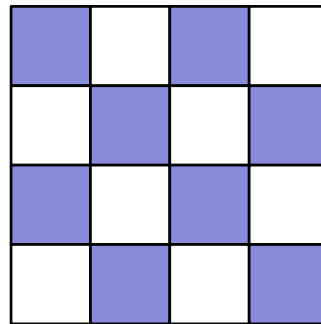Ring Temp. Gradient: >8.8°C   <=8.8°C   <=4.4°C   <=2.2°C
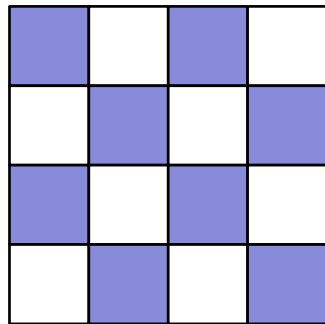
Channel bandwidth: ■ 8bits/cyc   ■ 16bits/cyc   ■ 32bits/cyc   □ 64bits/cyc
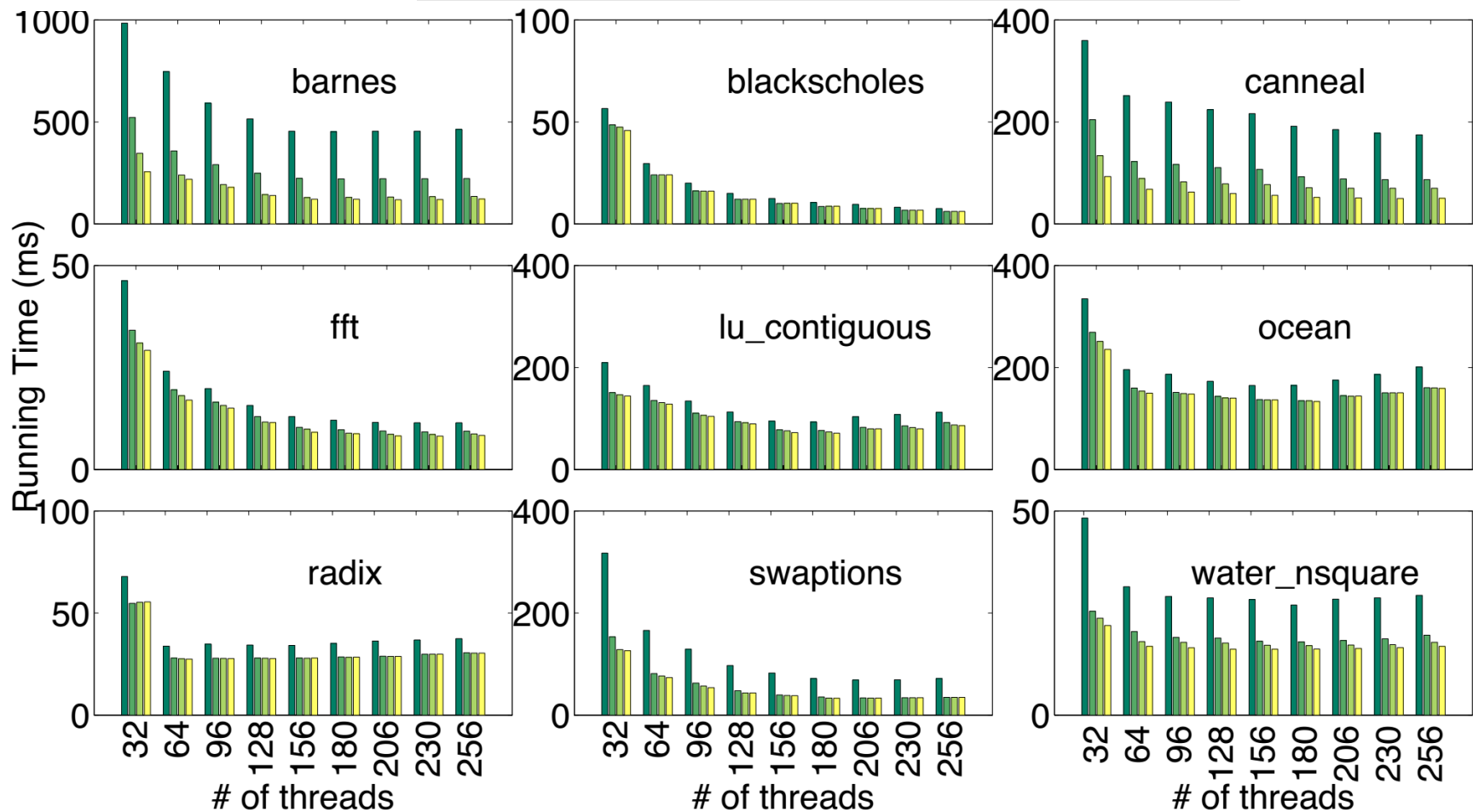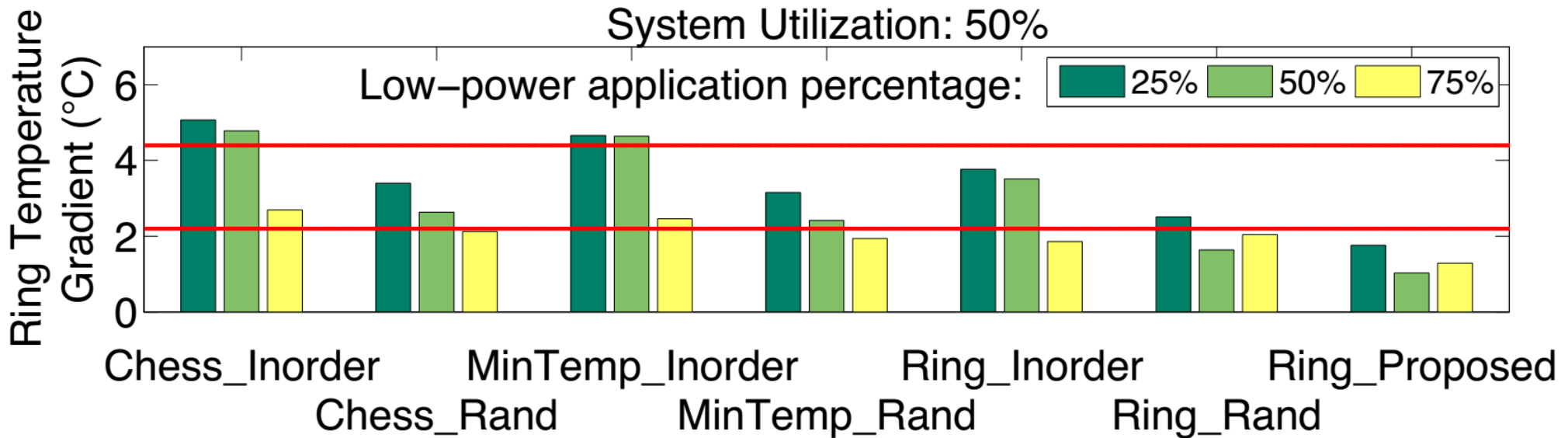

When using more than 50% of the cores, several applications have significantly better performance with the Ring-Aware approach

Joshi

Evaluation of Multi-Program Workloads

❑ Mapping Policies: In-order left (***Inorder***), random (***Rand***),
Proposed

Low-pow app   High-pow app
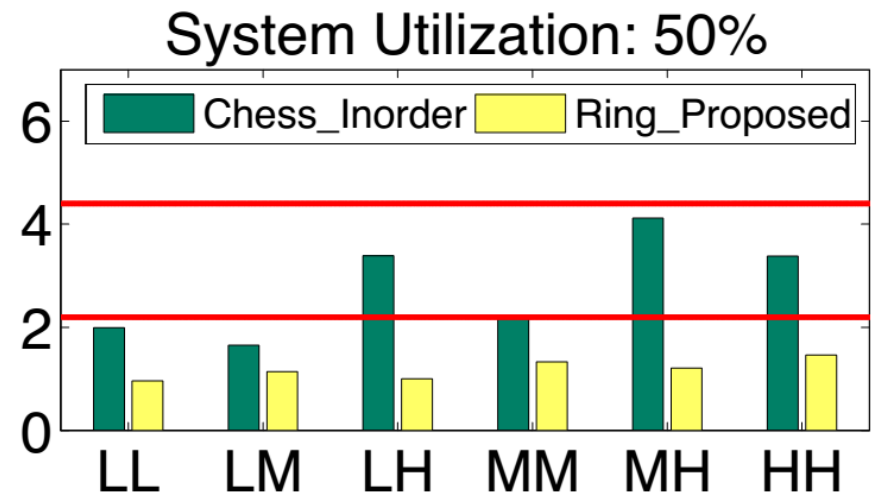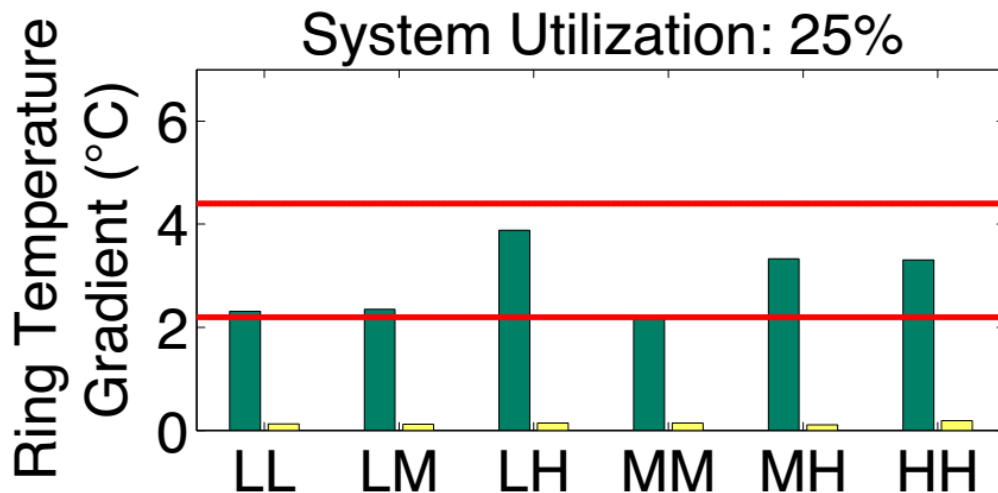
Joshi

Evaluation of Multi-Program Workloads

❑ **Diverse multi-program workloads**

L: low-power   *M*: medium-power *H*: high-power

LL	water_nsquare (L), lu_contiguous (L)	**HH**	barnes (H), fft (H)
LH	barnes (H), lu_contiguous (L)	**LM**	canneal (M), ocean (L)
MM	radix (M), blackscholes (M)	**MH**	radix (M), swaptions (H)

Chess_Inorder results in 6.1%, 13.8% and 8% lower performance compared to Ring_Proposed for LH, MH and HH, respectively.

Joshi

Summary II

- We used a cross-layer approach for thermal analysis & design of silicon photonic NoC

- We proposed a Ring-Aware job allocation policy to reduce thermal gradients among photonic devices

- Our policy enables us to operate the photonic links at their maximum bandwidth and in turn maximize application performance

Joshi

Summary

❑ Numerous challenges need to be overcome to make photonic NoC viable

- Bandwidth utilization

- Packaging/Integration

- Power consumption

  – NoC and Cache reconfiguration can be used to lower laser power

  – Software-based workload allocation policy can be used to reduce thermal tuning power

Joshi