NoRD: Node-Router Decoupling for Effective Power-gating of On-Chip Routers

Ming Hsieh Department of Electrical Engineering

Lizhong Chen, Timothy M. Pinkston

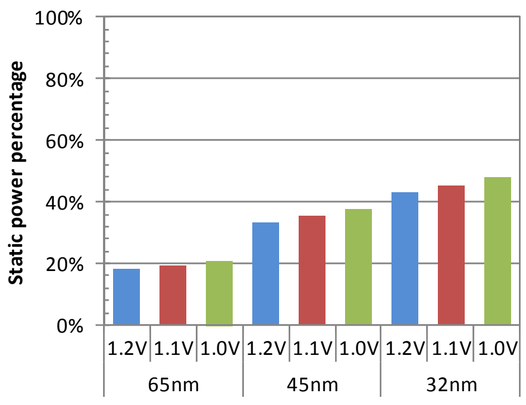
***SMART* Interconnects Group**

USC Viterbi School of Engineering

Abstract

A fundamental requirement of power-gating is for the idle periods to be sufficiently long to compensate for the power-gating and performance overhead. However, packet arrivals in multicore chips are quite intermittent, breaking the potentially long idle period of on-chip routers into fragments. In this work, we propose NoRD (Node-Router Decoupling), a novel power-aware on-chip network approach that provides for power-gating bypass to decouple the node's ability for transferring packets from the powered-on/off status of the associated router, thereby maximizing the length of router idle periods. Full system evaluation shows that, compared to an optimized conventional power-gating technique applied to on-chip routers, NoRD can further reduce the router static energy by 29.9% and improve the average packet latency by 26.3%, with only 3% additional area overhead.

Motivation

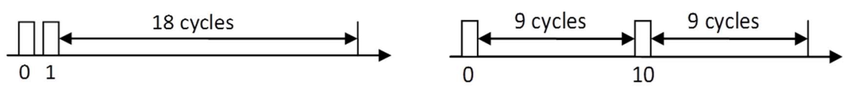


NoC Power

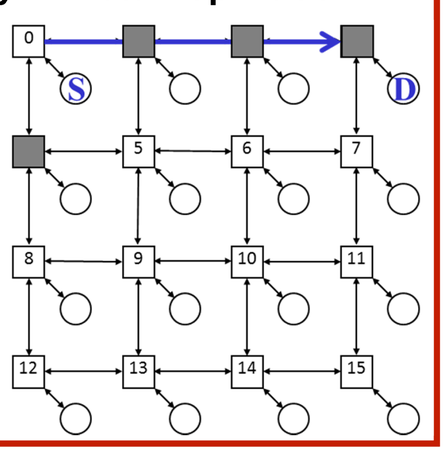
- Static power of on-chip routers increasing

Problems in Applying Power-gating

- Intensified Breakeven-time (BET) limitation
  - Intermittent packet arrivals break long idle periods into fragments (61% in PARSEC)



- Cumulative wakeup latency in multi-hop NoCs
  - Worse in larger networks
- Disconnection problem
  - Idle period is upper-bounded by local node's traffic
  - Disconnected network



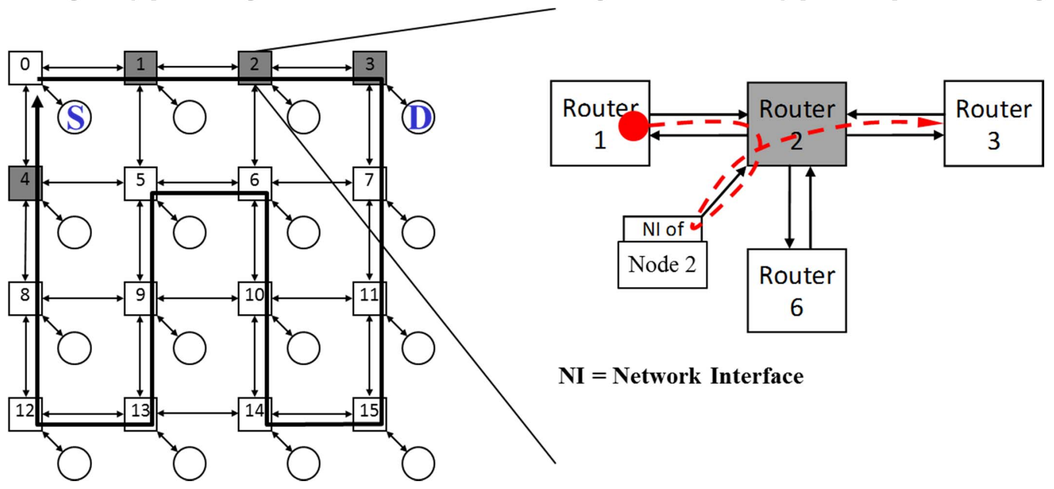
Node-Router Decoupling

Basic Idea

- Breaks the node-router dependence via decoupling bypass paths

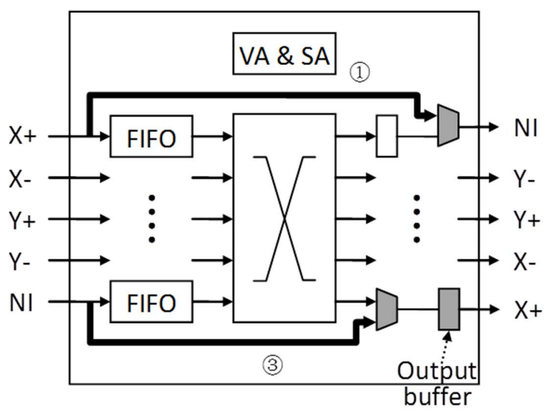
Chip Level

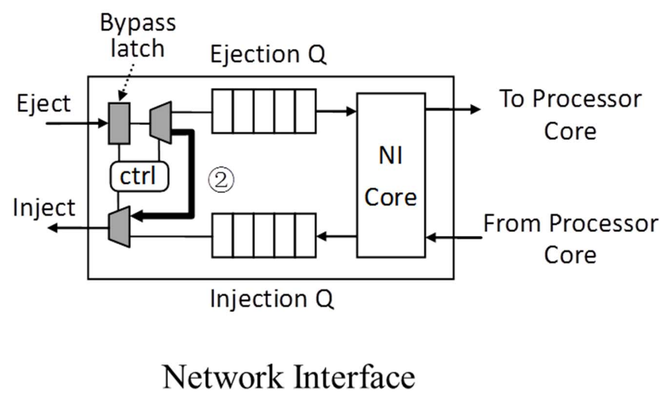
- A bypass ring connecting all nodes
- Receiving : add a bypass path from Bypass Inport to the NI ejection
- Sending: add a bypass path from the NI injection to Bypass Outport
- Forwarding: bypass gated-off router using the two bypass paths together



Router & Network Interface Level

- Two bypass paths and control logic are added
- Low implementation cost (3.1% of router area)

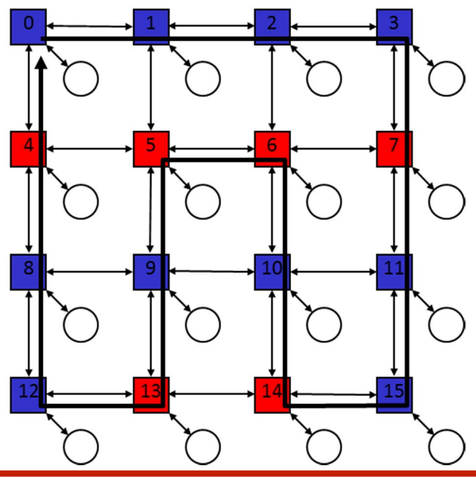




Network Interface

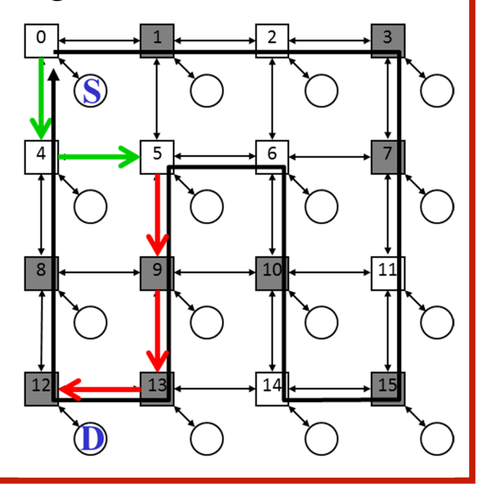
Shortcuts for Faster Packet Routing

- Classify routers in to performance-centric class and power-centric class
- Wake up early a few performance-critical routers to improve performance by adding "shortcuts" in routing
- Wake up late the rest (majority) of the routers to save more static power by allowing those routers to stay in gated-off state for a longer time



Routing

- Duato's Protocol
- Adaptive: minimal
- Escape: bypass ring



Evaluation Methodology & Results

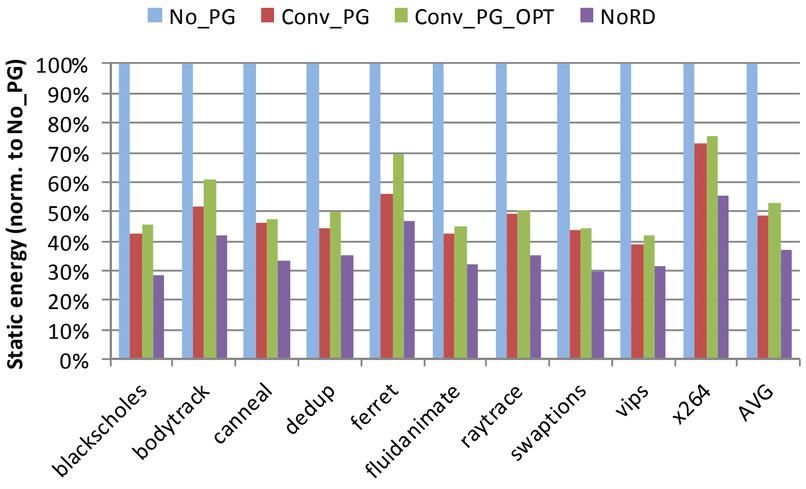
Simulation Platform

- Full system simulation: Simics + Gems
- Benchmark: PARSEC 2.0

Key parameters for simulations	
Core model	Sun UltraSPARC III+, 3GHz
Private I/D L1$	32KB, 2-way, LRU, 1-cycle latency
Shared L2 per bank	256KB, 16-way, LRU, 6-cycle latency
Cache block size	64Bytes
Coherence protocol	MOESI
Network topology	4x4 and 8x8 mesh
Router	4-stage, 3GHz
Virtual channel	4 per protocol class
Input buffer	5-flit depth
Link bandwidth	128 bits/cycle
Memory controllers	4, located one at each corner
Memory latency	128 cycles

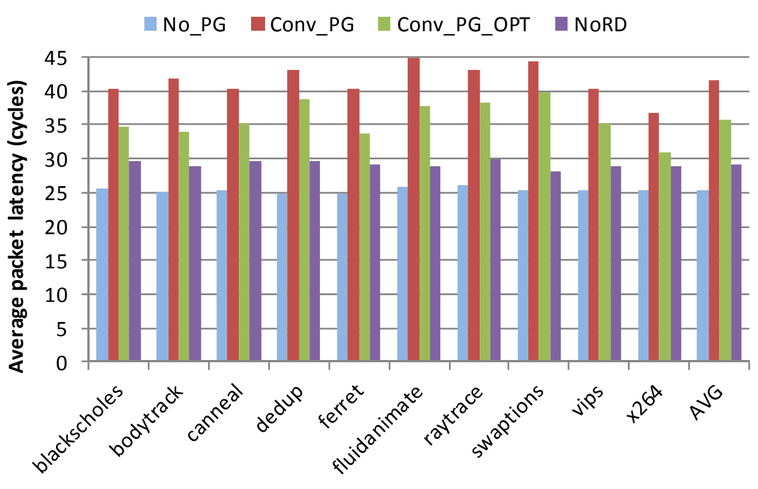
Static Energy Savings

- NoRD Improvement: 23.9% and 29.9%



Average Packet Latency Penalty

- NoRD Improvement: 48.6% and 26.3%



L. Chen and T. M. Pinkston, "NoRD: Node-Router Decoupling for Effective Power-gating of On-Chip Routers," in *45th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2012. {lizhongc, tpink}@usc.edu

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering