Fly-Over: A Light-Weight Distributed Power-Gating Mechanism for Energy-Efficient Networks-on-Chip

COMPUTER SCIENCE & ENGINEERING
TEXAS A&M UNIVERSITY

Rahul Boyapati*, Jiayi Huang*, Ningyuan Wang†, Kyung Hoon Kim, Ki Hwan Yum, Eun Jung Kim
Department of Computer Science and Engineering, Texas A&M University, USA †Google Inc.

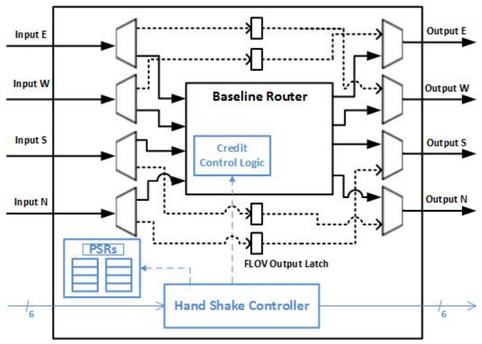
Texas A&M University HPCL Computer Science

Introduction

- Networks-on-Chip (NoCs) are devouring a large fraction of the on-chip power budget as technology scales down.
- In addition, static NoC power consumption is becoming the dominant component.
- **Fly-Over (FLOV)**: a lightweight distributed router power-gating mechanism to reduce NoC static power consumption.
  - ✓ FLOV router: FLOV links for network connectivity and low-latency route over power-gated routers.
  - ✓ Handshake Protocol: allows for seamless power-gating between neighboring routers.
  - ✓ Dynamic Routing Algorithm: best-effort minimal path routing without knowledge of global network status.

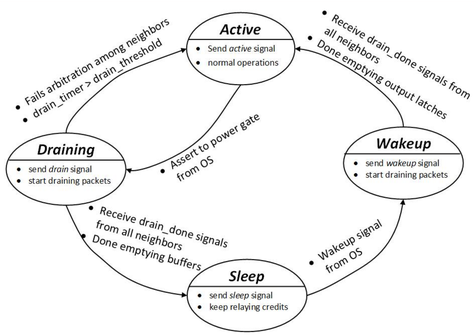
FLOV Router Microarchitecture

- ***ON (Active/Draining)*** mode: Packets are directed through the baseline router.
- ***GATED (Sleep/Wakeup)*** mode: Baseline router portion is power gated/waking up, packets fly over the baseline router through FLOV links.



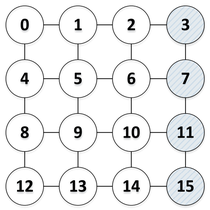
- Handshake Controller: handshaking with neighbors to facilitate power-gating/wakeup.
- Power State Register (PSRs): keeps the power state of physical/logical neighbor routers.
- Credit Control Logic, Augmented to relay credits while router core is gated.

Handshake Protocol

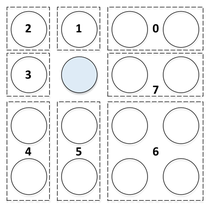


- Power gating and power on procedures are controlled by the handshake protocol and implemented in handshake controller.
- Power Gating: Active – Draining (finish intermittent transmission) – Sleep.
- Power On: Sleep – Wakeup (finish intermittent transmission) – Active.

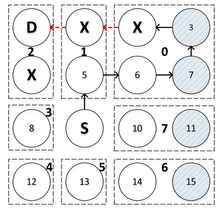
Dynamic Routing Algorithm



(a) FLOV Architecture.



(b) Destination Partitioning.



(c) Routing Example.

- Partition-based dynamic routing algorithm based on YX routing – best effort minimal routing.
- The right-most column always-active routers maintains network connectivity with FLOV links.

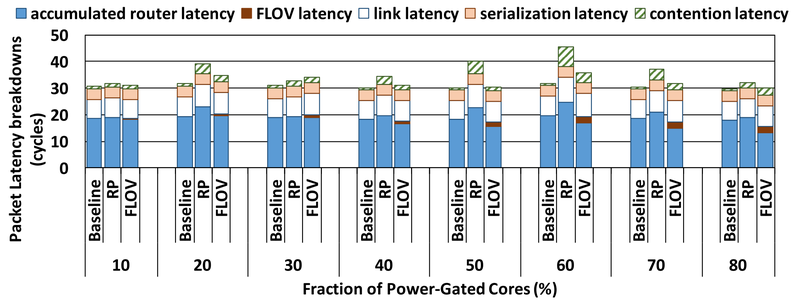
Conclusions

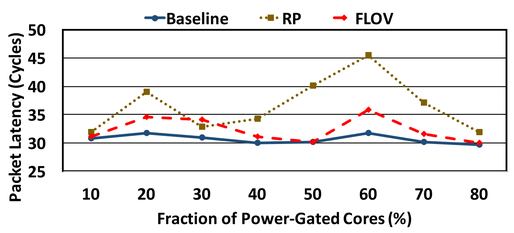
- FLOV power gates more routers thus achieving better NoC energy saving.
- Low latency FLOV links and the best-effort minimal routing avoids aggregated traffic rerouting, compensate for detour overhead.
- Compared with Router Parking (RP), FLOV achieves static energy and dynamic energy reduction by 17.3% and 11.9%, respectively.
- FLOV reduces average latency by 19.2% with respect to RP.

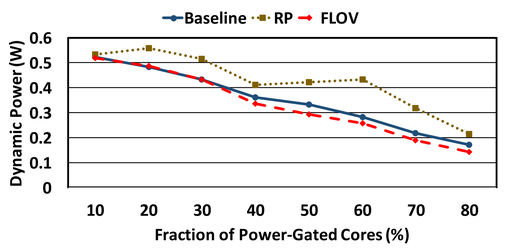
Evaluation Methodology & Results

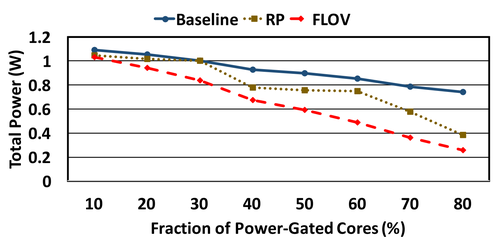
Simulation Testbed Parameters	
Network Topology	8x8 Mesh
Router	3-stage (3 cycle) router
Virtual Channel	3 regular VCs and 1 escape VC per vnet, 3 vnets, 6-flit depth
Packet Size	4 flits/packet for synthetic workload
Memory Hierarchy	32KB L1 I/D$, 8MB L2$, MESI, 4 MCs at 4 corners
Technology	32 nm
Clock Frequency	2 GHz
Link	1 mm, 1 cycle, 16B width
Power-Gating	Overhead = 17.7 pJ; Wakeup latency = 10 cycles
Baseline Routing	YX routing

Synthetic workload: 0.08 flits/core/cycle Uniform random traffic

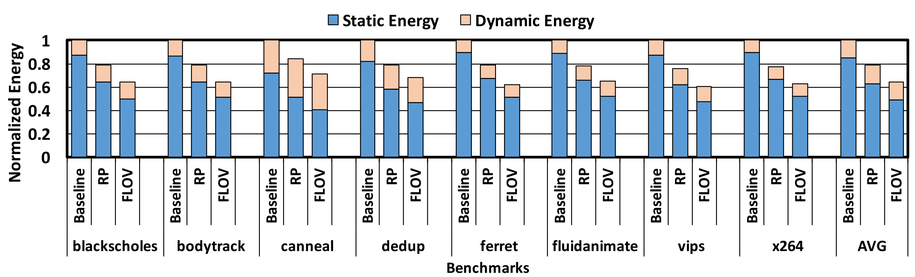






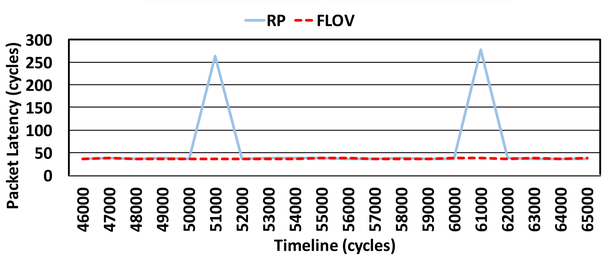


PARSEC Benchmarks Evaluation



- Average Energy reduction w.r.t. RP: 22% static energy saving, and 18% saving in total.
- Performance degradation w.r.t Baseline: <1% on average.

Network Reconfiguration Overhead



- FLOV power-gating is light-weight in terms of latency, while RP's centralized power gating control incurs high network latency, reconfiguration phase I is more than 700 cycles.

Reference:

A. Samih, R. Wang, A. Krishna, C. Maciocco, C. Tai and Y. Solihin, "Energy-Efficient Interconnect via Router Parking," in *International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE, 2013, pp. 508-519.