Energy Efficient Network-on-Chips with Opportunistic Circuit-Switching for MPSoCs

Masaaki Kondo and Yuan He

Graduate School of Information Science and Technology,
The University of Tokyo

Introduction

- NoCs are becoming the communication backbones for Manycore Processor SoCs
- Needs for scalable / efficient NoCs
  - Size and complexity of NoCs grows as the number of cores increases
  - Performance and power consumption of NoCs are critical to the system
    - It is reported that NoC consumes large portion of total chip power
- Up to date, most NoC designs employ Virtual Channel (VC) flow control for better utilization of link bandwidth
  - Each physical link is bundled with several VCs while each VC is a set of buffers used to store network traffic
  - NoC power is dominated by such VCs (buffers)
  - Using VCs also incurs longer per-hop latency

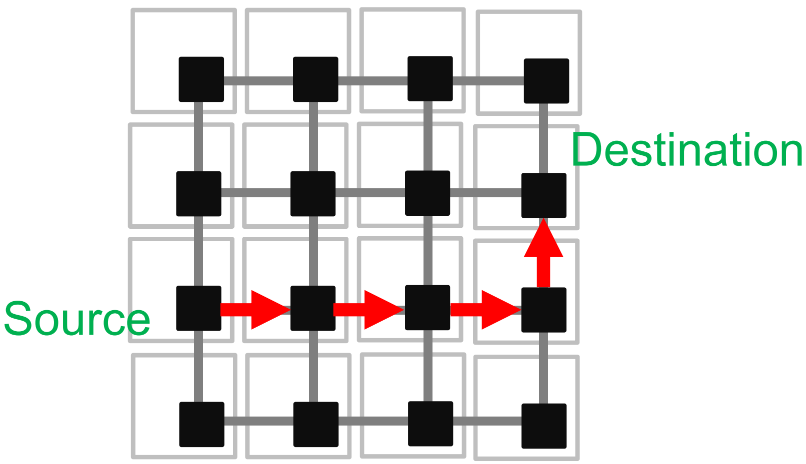
Background and Motivation (1/2)

- Existing solutions to reduce power for VC flow control
  - ☺ DVFS (for dynamic power) and Power-Gating (for static power)
  - ☹ Usually with performance cost
  - ☹ c.f.) low latency router designs: more power hungry
- Modern NoCs with VC is relatively bandwidth plentiful
  - Full link width with VC flow control can be an overkill
- Objective of this research
  - Propose to utilize both VC and circuit-switching (CS)
    - CS requires explicit channel acquisition which deteriorates NoC bandwidth
  - Take the advantages of both VC and CS while removing CS setup
  - Reduce both dynamic and static power consumption while maintaining low latency operation

Background and Motivation (2/2)

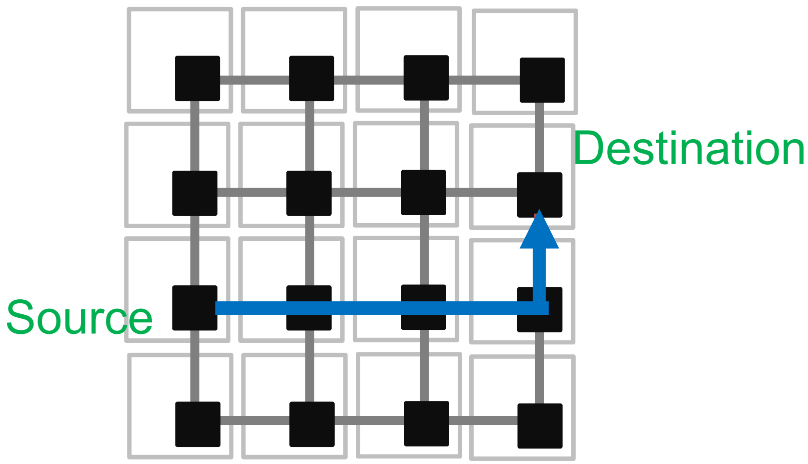
- Virtual Channel vs Circuit Switching

<u>Virtual Channel (VC)</u>



- Buffered
- Deep pipeline
- Per-hop based routing
- ☺Rich bandwidth
- ☹Power hungry

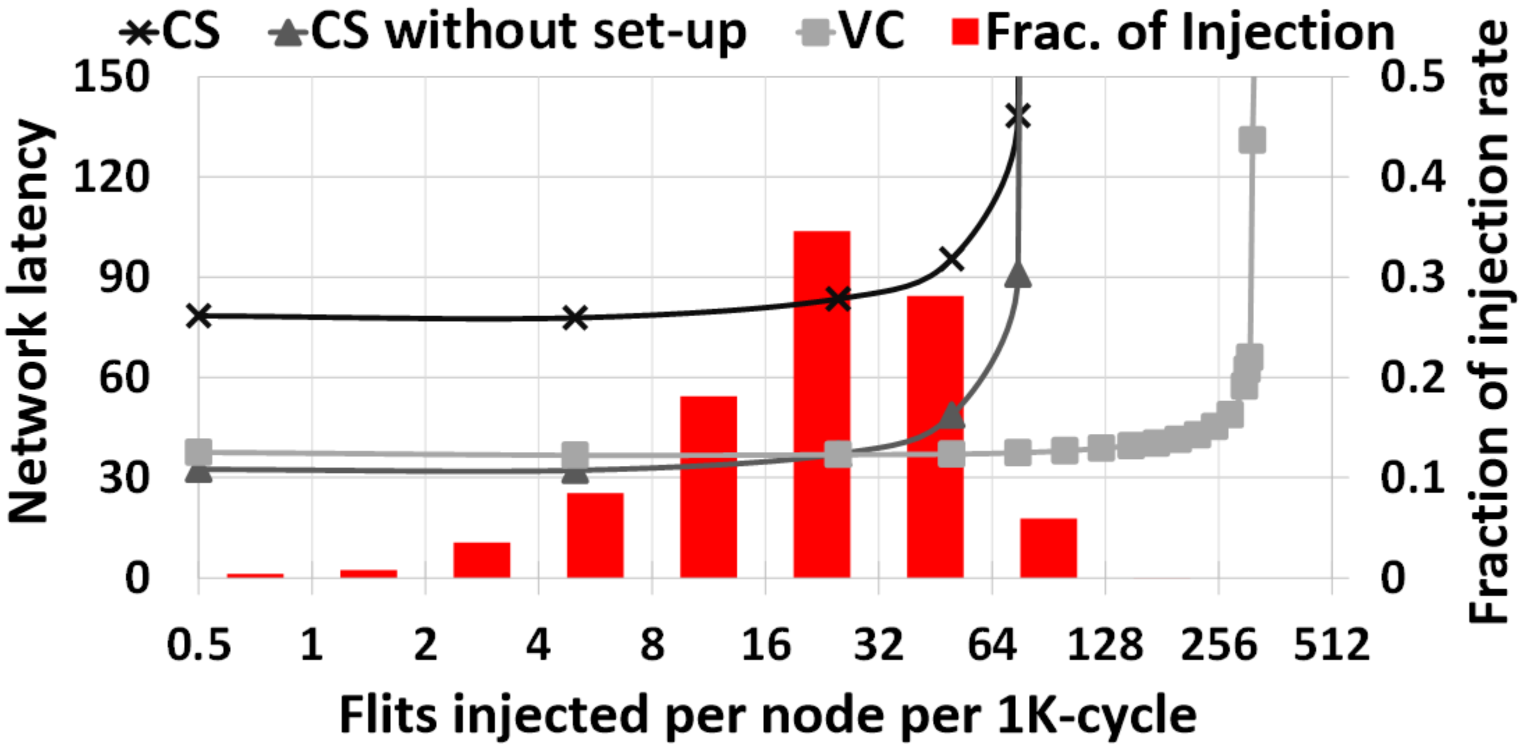
<u>Circuit Switching (CS)</u>



- Bufferless
- One cycle per router
- Explicit route set-up
- ☹Poor bandwidth
- ☺Power efficient

Latency and Bandwidth Comparison

- Network latency vs. packet injection rate



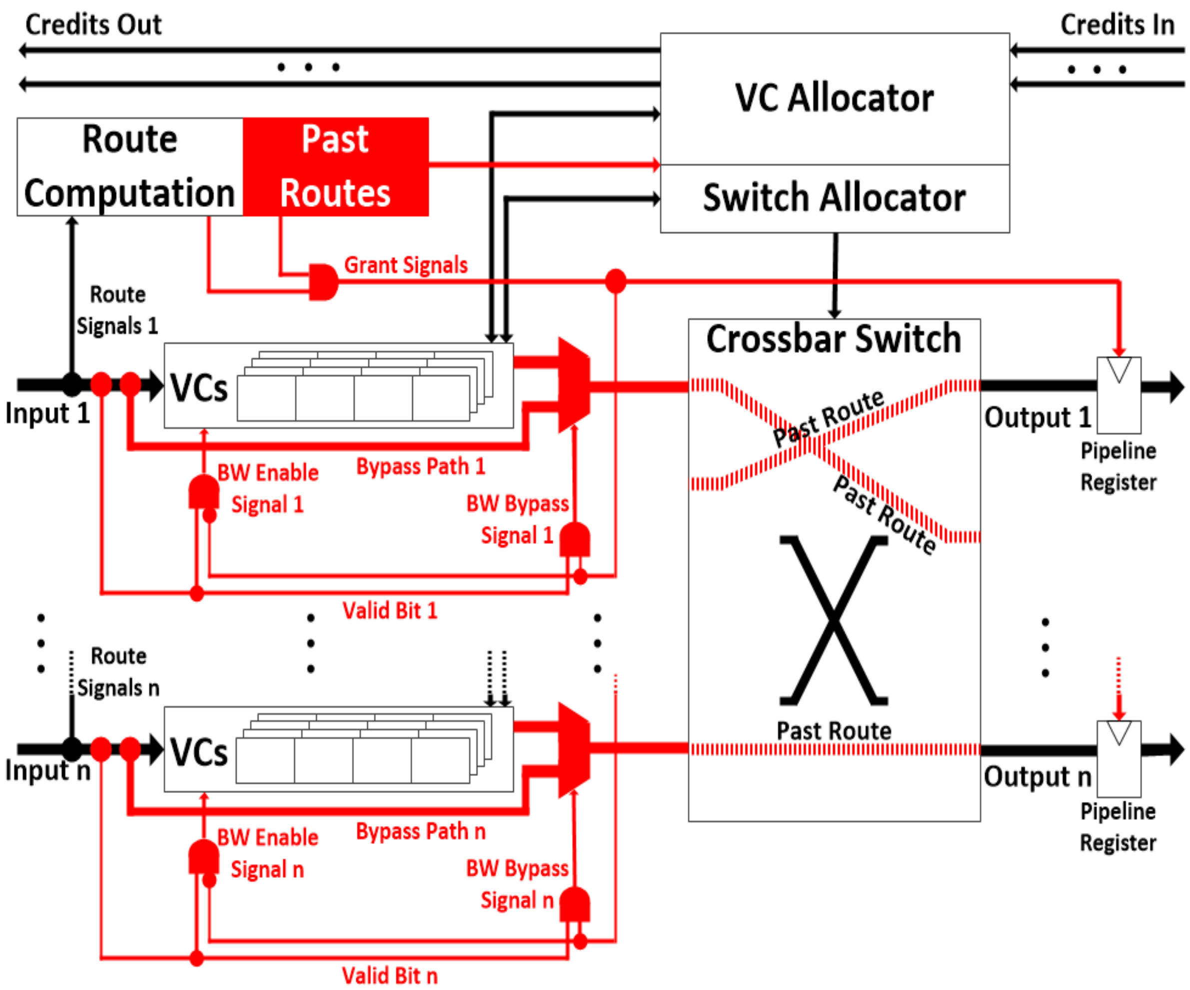
  - VC flow control is rich in bandwidth
  - Traditional CS has poor bandwidth and network latency
  - CS without set-up can potentially achieve the lowest latency

Opportunistic Circuit-Switching (OCS)

- Basic Concept
  - Preset the route using predictions to form circuits in each router
    - No explicit setup to acquire a channel
  - Verified with actual routes at packet traversal in the router
  - Prediction hit: a flit traverses like CS without buffer write
  - Prediction miss: use traditional VC flow control with buffer write
- Benefit of OCS
  - Dynamic power saving at the buffers
  - Static power saving due to longer power-gating interval for buffers
  - Improved network latency by bypassing router pipeline

Schematic Design of the OCS Router

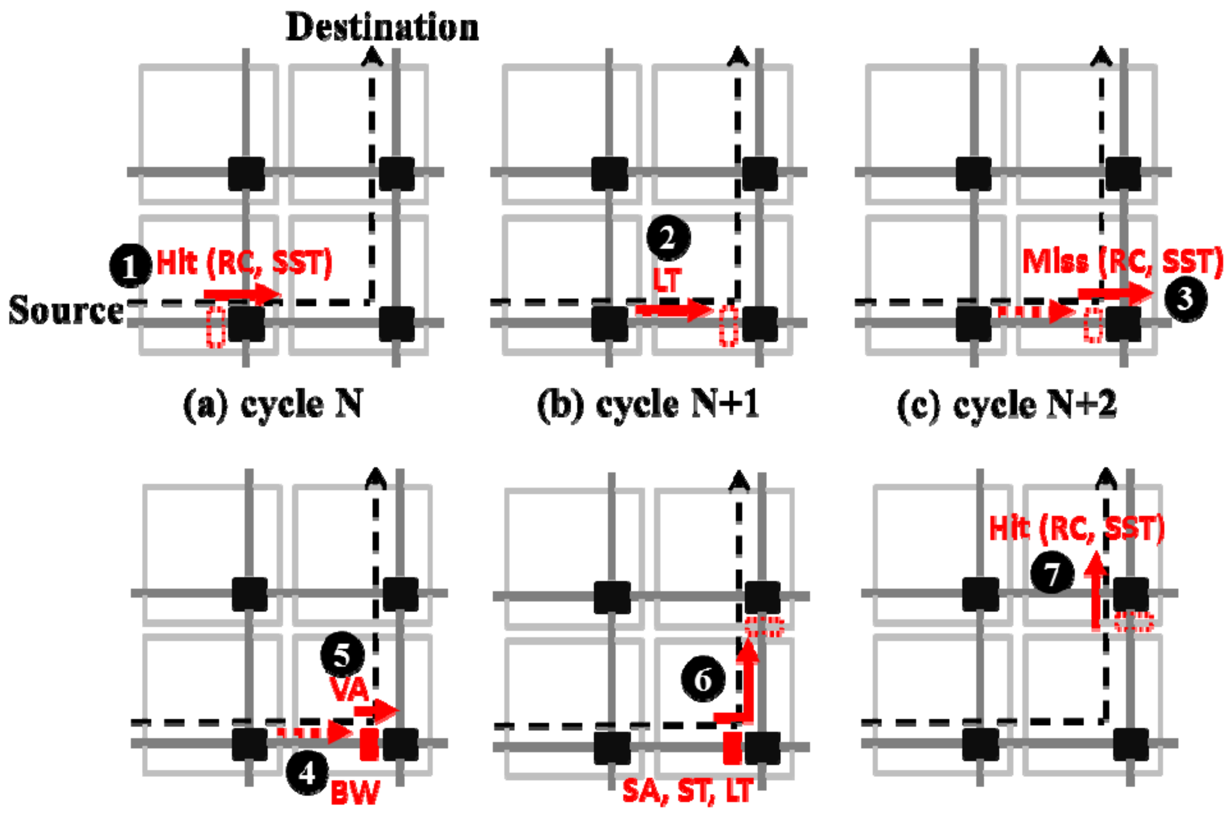
- Required extension to routers for OCS



- History buffer to store the latest past routes
- Extra wires and multiplexers for packets to bypass the VC
- Extra wires to issue grant signals when OCS hit
- OCS is cancelled if the downstream router has no credit (packet is stored in the buffer on the current router)

Behavior of NoC with Opportunistic Circuit-Switching

- Case for Hit → Miss → Hit



(a) cycle N (b) cycle N+1 (c) cycle N+2

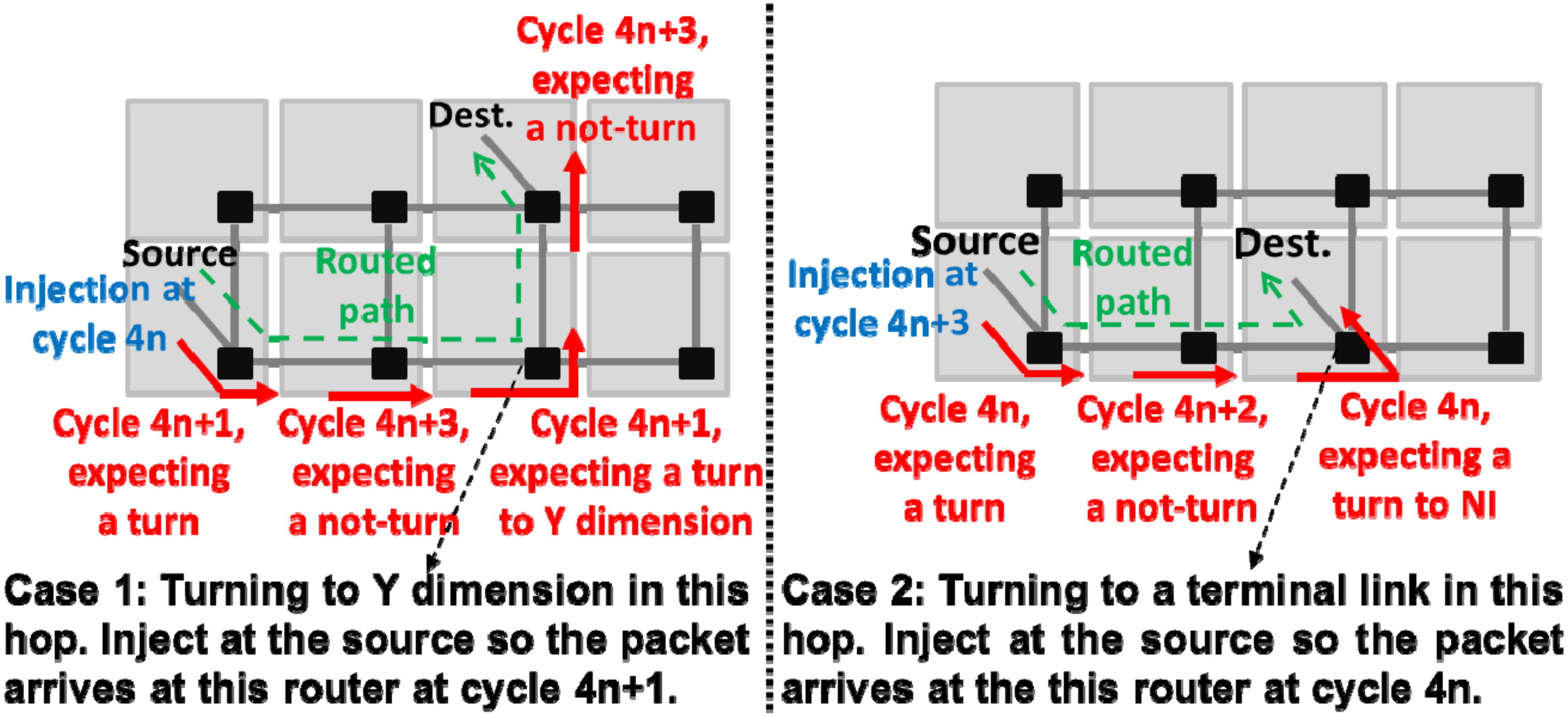
(d) cycle N+3 (e) cycle N+4 to N+6 (f) cycle N+7

RC: route computation	VA: Virtual channel Allocation
SA: Switch Allocation	ST: Switch Traversal, LT: Link Transfer
BW: Buffer Write	SST: Speculative Switch Traversal

① OCS Hit: a flit traverses a router within one cycle without buffering
② it traverses the link to the next router
③ OCS Miss: it is transmitted under VC flow control
④ It is written to the buffer in the current router
⑤ VA is taken for the next hop
⑥ It goes through the three remaining pipeline stages
⑦ OCS Hit again

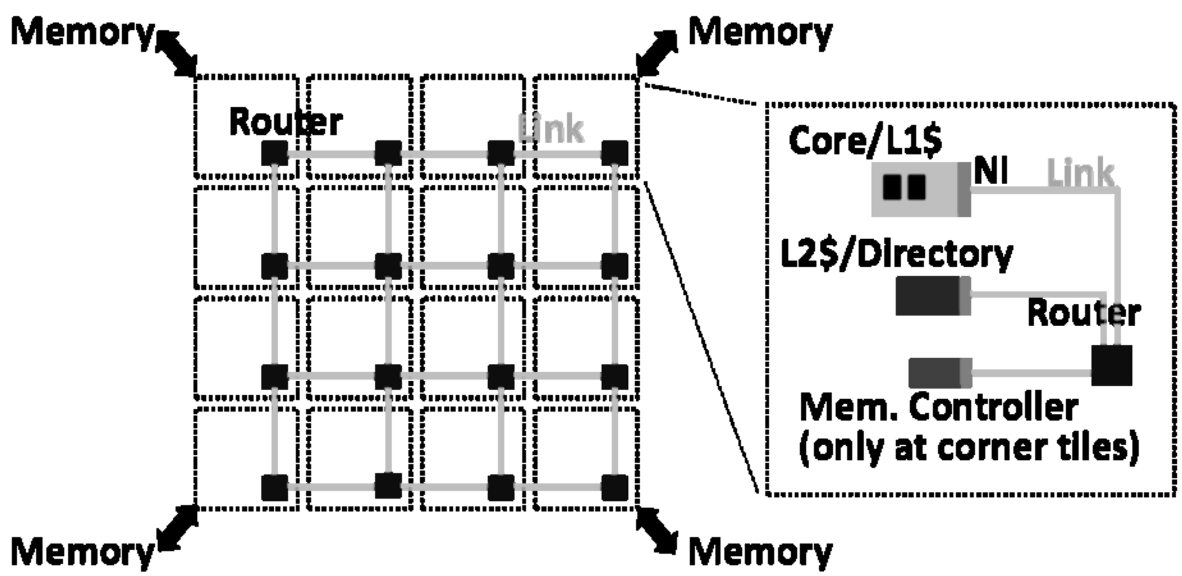
Scheduled Injection and Cycle-Aware Route-Reuse (SICR)

- Hit rate for reusing past routes is the key for OCS
- Improving the hit rate by SICR
  - Use multiple predicted/predefined routes according to cycle number
  - Control packet injection timing expecting cycle-specific predicted or predefined routes are used at routers coming at different cycles



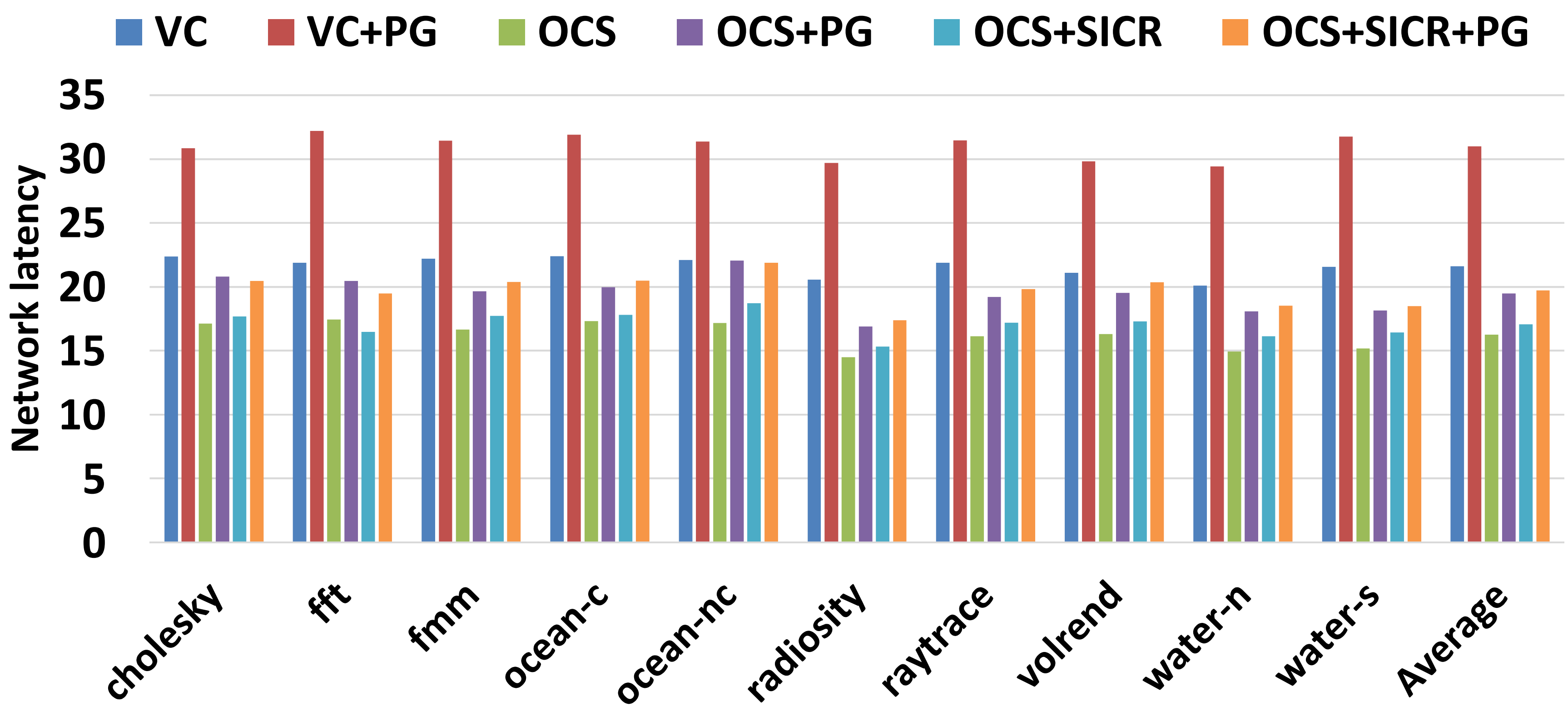
Evaluation Methodology

- Simulated on a 16-core system
  - GEMS/Simics with GARNET (network) and McPAT (power)
- Network under 2D mesh topology
  - 128-bit link width
  - X-Y routing
  - 4-cycle pipeline for VC
- Process parameter
  - 32nm technology
  - Three device types
    - High performance (HP)
    - Low standby power (LSP)
    - Low operating power (LOP)
- Workload from SPLASH-2 and NPB 3



Evaluation Result – Network Latency

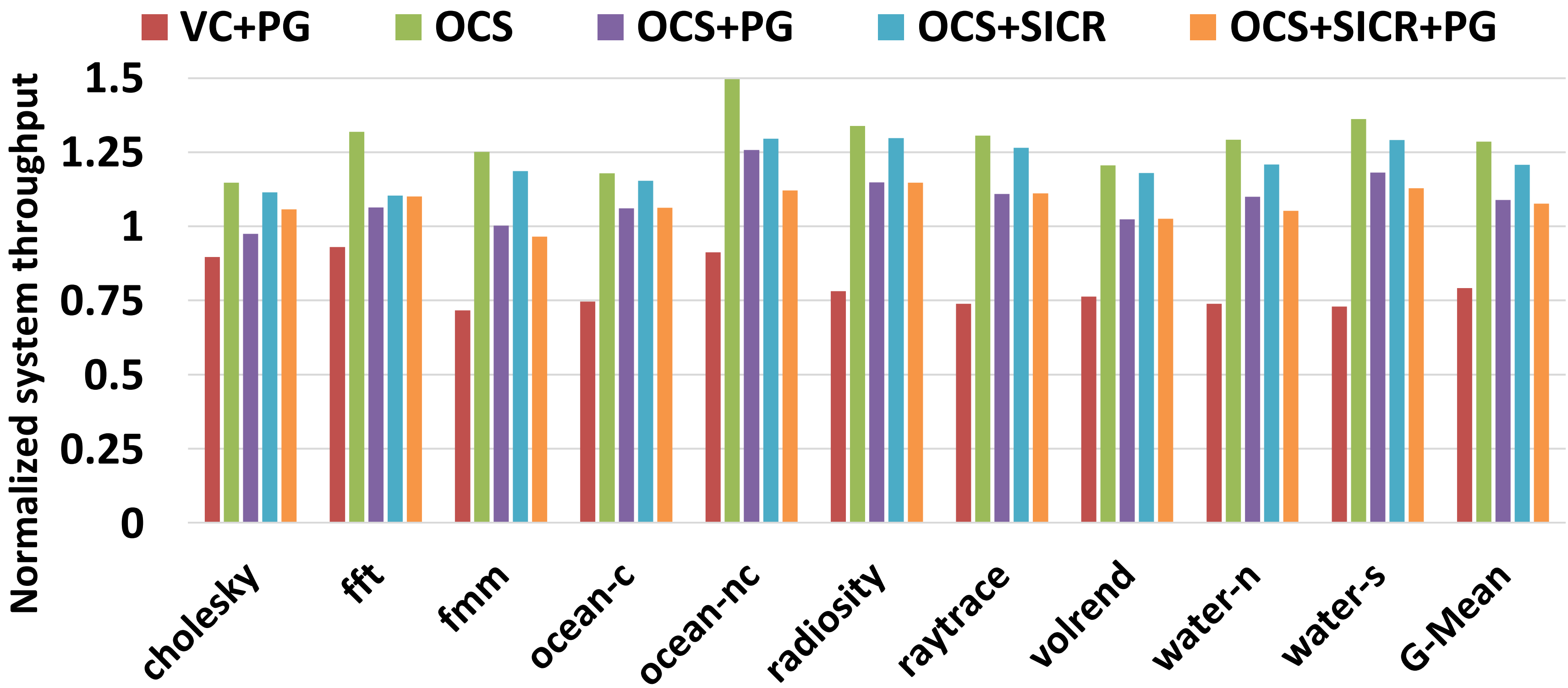
- Network latency per flit under different NoC designs



- VC+PG slows down the network significantly
- Latency reduction by OCS
  - Applying OCS and PG together alleviates the latency issue of PG
- Slight latency increase with PG or SICR due to latency overhead

Evaluation Result – Throughput

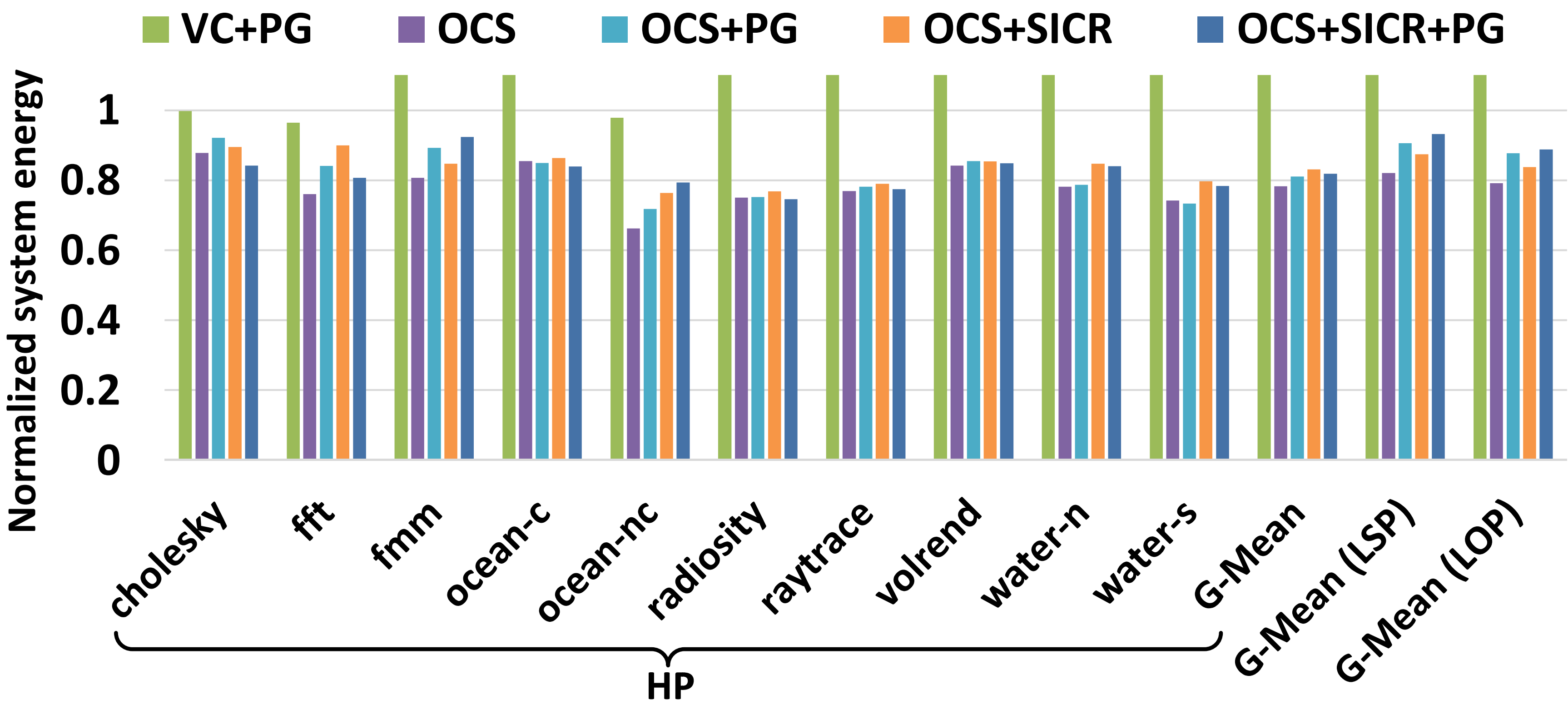
- System throughput normalized to original VC



  - OCS can provide more than 25% of improvement on throughput
  - In OCS+PG, improvement drops but still about 10%
  - VC+PG degrades system throughput by about 20%

Evaluation Result – Energy Consumption

- System energy normalized to original VC



- Energy reduction by around 20% for all three device types
  - Large energy saving even including processor energy consumption
- OCS alone results in the smallest energy footprint
  - PG and/or SICR have performance impact

Summary

- We questioned the necessity of virtual channel flow control
- Opportunistic Circuit-Switching
  - Preset the route using predictions to form circuits in each router based on the past route history
  - A flit traverses like circuit switching when prediction hits
- Evaluation with cycle level simulation
  - About 25% throughput improvement
  - About 20% system energy reduction
- Future work
  - Consider better routes prediction strategies
  - Evaluate OCS under various network configurations