

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

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
 {rahulboyapati,jyhuang,khkim,yum,ejkim}@cse.tamu.edu *Google Inc. nywang@google.com

ABSTRACT

Reducing static NoC power consumption is becoming critical for energy-efficient computing as technology scales down since NoCs are devouring a large fraction of the on-chip power budget. We propose *Fly-Over (FLOV)*, a light-weight distributed mechanism for power-gating routers. With simple modifications to the baseline router architecture, FLOV links are facilitated over power-gated routers. A Handshake protocol that allows seamless router power-gating in addition to a dynamic routing algorithm, that provides best-effort minimal path without the necessity for global network information, maintain normal NoC functionality. We evaluate our schemes using synthetic workloads as well as real workloads from PARSEC 2.1 benchmark suite. The results show that FLOV can achieve on average 19.2% latency reduction and 15.9% total energy savings.

1. INTRODUCTION

Static power consumption of the on-chip circuitry is increasing at an alarming rate with the scaling down of feature sizes and chip operating voltages towards near-threshold levels. Power-gating, cutting off supply current to idle chip components, is an effective circuit-level technique that can be used to mitigate the worsening impact of on-chip static power consumption. Router Parking (RP) [2] power-gates routers whose attached cores are power-gated, but requires a centralized fabric manager for network reconfiguration, which creates a huge synchronization overhead, and the whole network has to stall until the reconfiguration is completed.

We propose Fly-Over (FLOV), a light-weight distributed power-gating mechanism that eliminates the need for centralized control to power-gate routers. FLOV tries to power-gate routers as soon as the attached cores are powered down by the OS, in a distributed manner. Since such a distributed power-gating mechanism may create interconnect partitions without communication paths, FLOV links in power-gated routers are provided to enable incoming packets to travel straight through for network connectivity. Specifically, FLOV comprises of the FLOV router microarchitecture, handshake protocol, and a partition-based dynamic routing algorithm.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

PACT '16 September 11-15, 2016, Haifa, Israel

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4121-9/16/09.

DOI: <http://dx.doi.org/10.1145/2967938.2974058>

2. FLOV MECHANISM

In this section we explain the three major components of the FLOV mechanism, namely, the router microarchitecture, handshake protocol and the dynamic routing algorithm.

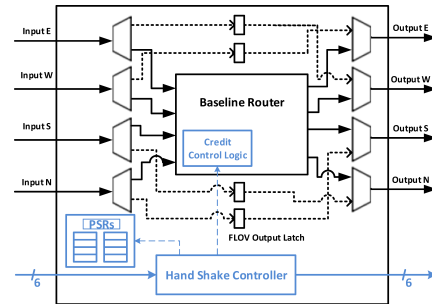


Figure 1: FLOV Router Architecture.

2.1 FLOV Router Architecture

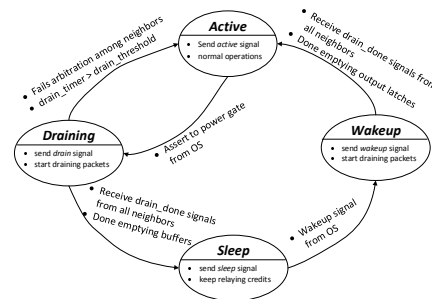
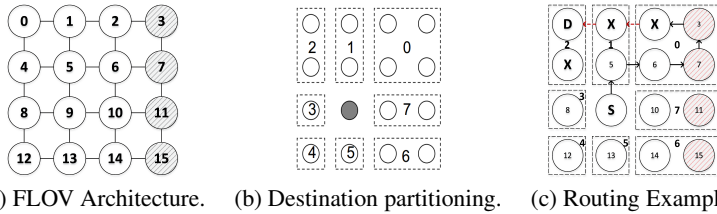


Figure 2: Router Power State Transition Diagram.

As shown in Figure 1, the FLOV router architecture has multiplexers and demultiplexers added to input/output links, in addition to a latch in each direction. When a FLOV router is powered-on, it functions like the baseline 3-stage virtual-channel router, and the muxes/demuxes are set to 0 as well as the latches are power-gated. When the router is power-gated, all the components of the baseline router are power-gated and the muxes/demuxes are set to 1 to activate the FLOV links.

2.2 HandShake Protocol

The HandShake Control logic (HSC) block shown in Figure 1 connects to all the neighboring routers and implements the hand-



(a) FLOV Architecture. (b) Destination partitioning. (c) Routing Example.

Figure 3: FLOV Architecture, Destination partitioning, Routing Algorithm Example.

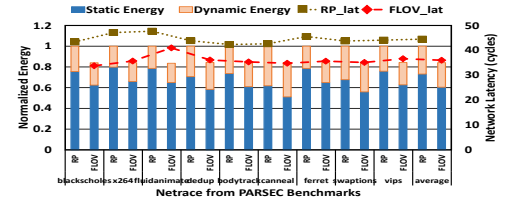


Figure 4: Normalized Energy Consumption and Network Latency.

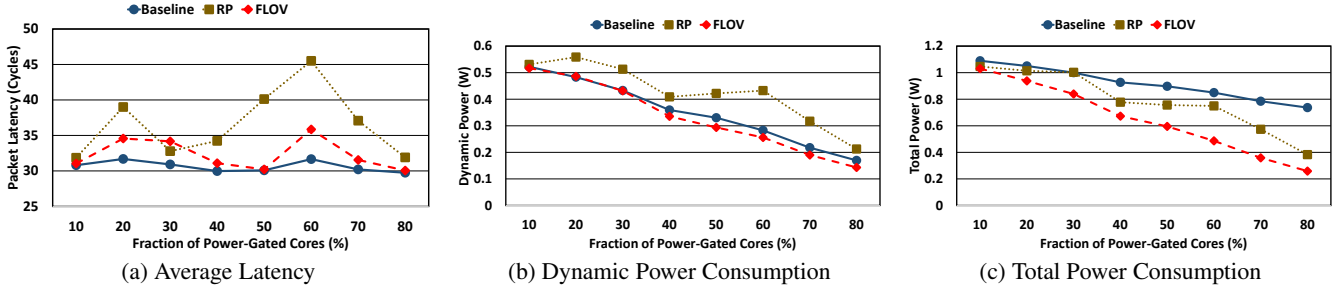


Figure 5: Latency, Dynamic and Total Power Comparison for Injection Rate 0.08 flits/node/cycle with Uniform Random Traffic.

shake protocol between logical adjacent routers required before power-gating a router.

The state transition diagram in Figure 2 depicts the power states a router can be in. The router initially goes into *Draining* from *Active* when it wants to be power-gated, once the attached core is power-gated. The router immediately sends a drain signal to its neighbors and starts to drain the packets residing in its input buffers. Routers that want to drain at the same time but fail to win arbitration with their neighbors come back into *Active*. Once all the router's neighbors finish any intermittent transmissions destined to it and the packet draining is finished, the router can go into *Sleep*. In the *Sleep* state the router sends a sleep signal to all its neighbors after turning off the baseline router operation and starting the FLOV operation. The router starts relaying credits between its powered-on neighbors using the Credit Control Logic(CCL). A router goes to *Wakeup* from *Sleep* when its attached core is powered on.

2.3 Dynamic Routing Algorithm

Figure 3(a) shows a (4×4) 2D mesh network with the proposed FLOV routers. The pattern-shaded routers (3, 7, 11, and 15) are connected to memory controller (MC) nodes that should be never power-gated. We propose a partitioned-based dynamic routing algorithm based on YX routing. Each router divides the network into partitions as shown in Figure 3(b). The routing decision is made based on two variables, the partition which the destination falls into and the power states of neighboring routers.

In the example show in figure 3(c), the packet is rerouted through the last column since routers 1 and 0 are power-gated. The main idea behind the algorithm is that in the worst case the packets are directed to the EAST direction so we can guarantee that the packet will be able to make a turn toward the destination in the always powered-on edge router of the corresponding row.

3. EXPERIMENTAL EVALUATION

In this section we evaluate the FLOV mechanism by comparing the power consumption and NoC latency with Router Parking [2]. We use a cycle-accurate network simulator, BookSim [1], that models all the router pipeline stages and link latencies. The simulation testbed is an 8×8 2D mesh network, with 4 VCs (6-flit depth) in

each input channel and a packet size of 4 flits. We simulate a 2 GHz clock frequency with 32nm technology.

Figure 5 shows the performance and power consumption of the different mechanisms as the percentage of power-gated cores in increased. As shown in Figure 5 (a), FLOV performs better than RP since the RP mechanism always reroutes packets through powered-on routers and links connecting them leading to increased path length whereas FLOV uses the FLOV links to achieve near minimal path. We observe from Figures 5 (b) and (c) that the total and dynamic power consumption of FLOV is less than RP. The dynamic power consumption of FLOV is smaller since in RP when reroutes happen, every hop in the elongated traversal path will execute the total router pipeline whereas with FLOV we can use low cost FLOV links even with detour. As shown in Figure 4 for PARSEC network traces evaluation, FLOV achieves 19.2% lower latency than RP and reduce total energy by 15.9% on average when 29 out of 64 cores are off.

4. CONCLUSIONS

We proposed Fly-Over (FLOV), a light-weight distributed router power-gating mechanism for NoCs. FLOV power-gates routers attached to powered-down cores without global network information, but still ensures network connectivity. Performance evaluations using synthetic and real workloads show that FLOV not only achieves better NoC power savings due to power-gating more routers but avoids aggregated traffic rerouting in the network unlike Router Parking.

5. REFERENCES

- [1] N. Jiang, D. U. Becker, G. Michelogiannakis, J. Balfour, B. Towles, D. E. Shaw, J.-H. Kim, and W. J. Dally. A Detailed and Flexible Cycle-Accurate Network-on-Chip Simulator. In *International Symposium on Performance Analysis of Systems and Software (ISPASS)*, pages 86–96. IEEE, 2013.
- [2] 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)*, pages 508–519. IEEE, 2013.