



Dynamic Power Management for Power Optimization of Interconnection Networks Using On/Off Links

Vassos Soteriou Li-Shiuan Peh
{soteriou, peh}@ee.princeton.edu

Princeton University

August 20, 2003



Outline

- On/off link architecture and preliminaries
- Motivation for low-power interconnection networks
- Challenges
- Methodology:
 - Power-performance connectivity graph
 - On/off routing algorithm
 - On/off decision mechanism
- Router architecture incorporating on/off links
- Experimental results
- Conclusions

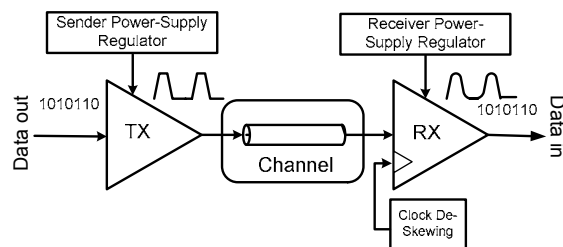
August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

2

On/off link architecture

- Link behavior:
 - Link on:
 - Active, operates at design frequency, traffic traversal, full power consumption
 - Link off:
 - Inactive, no traffic traversal, no power consumption
- Link performance metrics:
 - On to off transition delay, $t_{off}(\text{link})$
 - Off to on transition delay, $t_{on}(\text{link})$
 - Transition energy, E_{tr}



August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

3

Benefits of on/off links

- Industry products available:
 - Examples: Motorola WarpLink, TI, IBM Blue Logic SerDes
 - Backplane and board-to-board communication
 - Smaller design overhead, quick solution
 - No further design validation
- More robust than DVS links (still in research):
 - L. Shang et. al. [HPCA '02], G. Y. Wei et. al. [ISSCC '00]
 - Smaller CMOS area overhead
 - Lower hardware complexity
 - No idle power
- Electrical and optical
- Spawns many intellectual research challenges
 - Motivation for circuit designers to design more efficient links

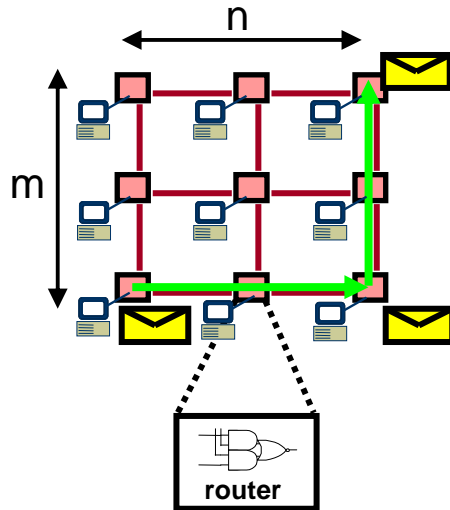
August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

4

What are interconnection networks?

- Fabric connecting subsystems within a digital system:
 - Multiprocessor servers
 - Chip multiprocessors
 - Clusters of workstations
 - Ports of IP routers
 - I/O devices



August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

5

What makes low power interconnection networks important?

- Heavily performance driven
 - Low latency and high throughput
 - Little work on power-aware designs
- Power consumption increases as network bandwidth and frequencies upscale
- Significant consumer of power:
 - Thermal problems
 - Cooling and power supply cost and space
 - Circuit reliability issues
 - Portability, battery life in mobile systems

August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

6

Why target links for low power interconnection networks?

- Major consumer of overall system power
 - Example: the link circuitry of the Alpha 21364 μ P consumes **58%** of the 23W allocated router power
- The off-chip link's power consumption does **not vary considerably** with **traffic fluctuations**
 - Power consumption in links will become more dominant as compared to the average system power budget as operating frequencies keep increasing

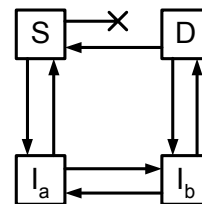
August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

7

Challenges to be met

- Reduction in path diversity
 - Impact on performance
 - Packet dropping and loss
- Routing
 - Deadlocks and Livelocks
 - Similarities with fault-tolerant networks
- Global information sharing among routers to determine active paths
 - Problems with timing and currency of information
 - Power overhead
- Hardware complexity

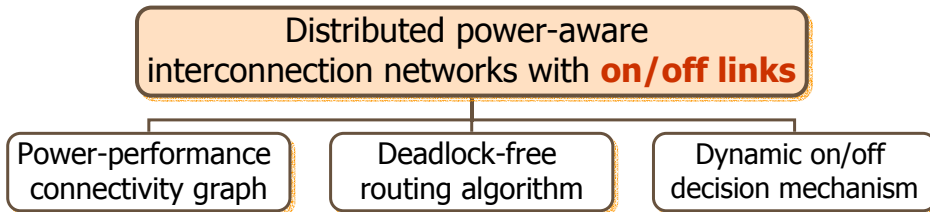


August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

8

DPM methodology: 3-phase systematic approach

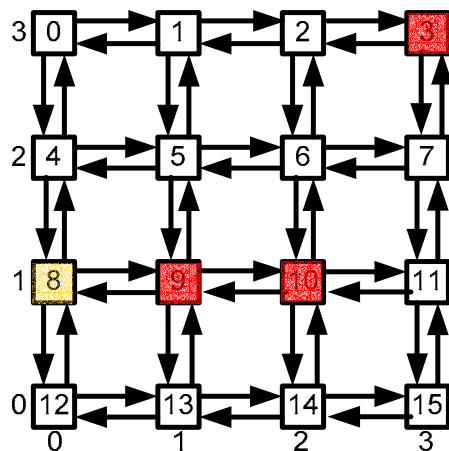


August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

9

A fully random on/off link approach can cause network disconnectivity



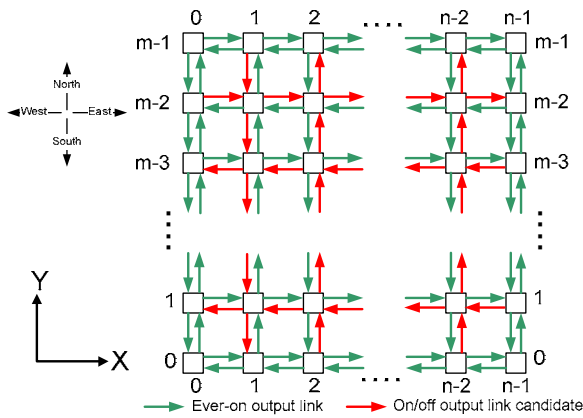
- A core router:
 - $2n$ ($n=2$) off links per router can cause network disconnection
- One link per router allowed to turn off:
 - Routers 3, 9 and 10:
 - Directly and completely disconnected
 - Router 8:
 - Indirectly disconnected

August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

10

Power-performance connectivity graph



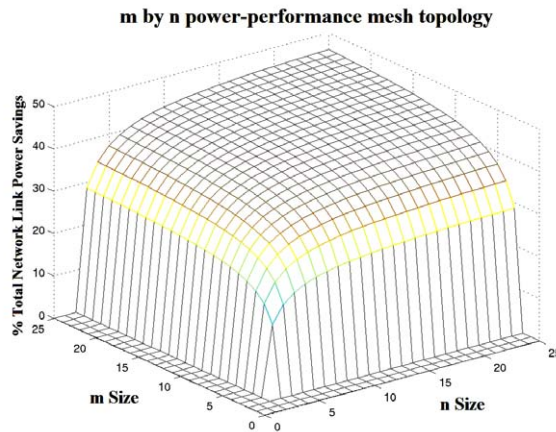
- Generalized model
 - Scalable
- Alternating on/off rows and columns
 - Small impact on hop count when rerouting
- Up to 2 links off per router
 - Up to 50% power savings

August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

11

Potential maximal power savings in an m x n mesh topology



$$P_s(m, n), m \geq 3, n \geq 3$$

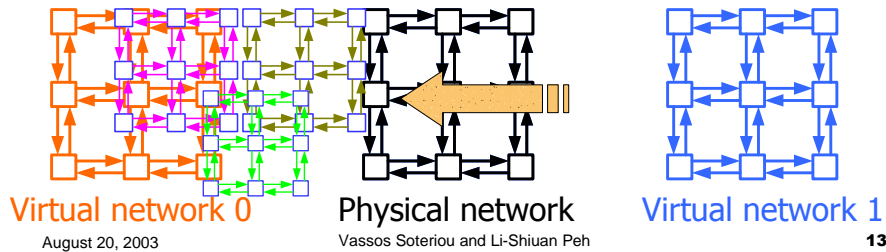
August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

12

On/off routing algorithm

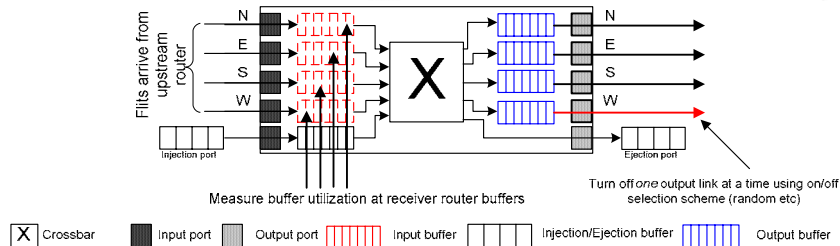
- Physical network split into two virtual networks
- Combination of two non-minimal, fully-adaptive, deadlock-free, distributed routing algorithms
 - East-Last and West-Last (WLEL, Turn Model)
 - Strongly coupled to the connectivity graph
- Each protocol runs on a separate virtual network
 - **WL** protocol (VC0) in eastward direction
 - **EL** protocol (VC1) in westward direction
- One or more virtual channels per virtual network



Statistics acting as dynamic on/off decision mechanism drivers

- Statistics or counters obtained from the network reflect the level of network traffic
- We use *aggregate* input buffer utilization

$$U_{buffer} = \frac{\sum_{p=1}^P \sum_{t=1}^{t_{sw}} (F(t, p)/B)}{t_{sw} \times P}, 0 \leq U_{buffer} \leq 1$$



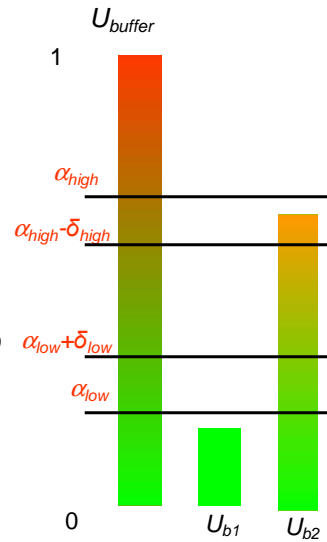
Thresholds as link state switch deciders

- Simple thresholds used as link on/off deciders:
 - Low level utilization ratio α_{low}
 - Low level utilization change ratio δ_{low}
 - High level utilization ratio α_{high}
 - High level utilization change ratio δ_{high}
- U_{buffer} compared against these ratios to decide link state

$$0 \leq \alpha_{low} \ll \alpha_{high} \leq 1$$

$$0 < \delta_{low} < \alpha_{low}$$

$$0 < \delta_{high} < \alpha_{high}$$



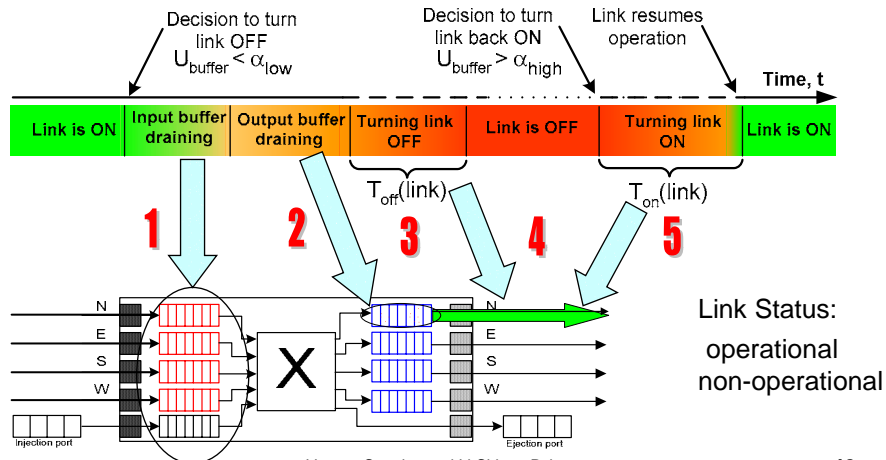
August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

15

Process of switching link states

- 5 stages between on (active) and off (inactive) link states

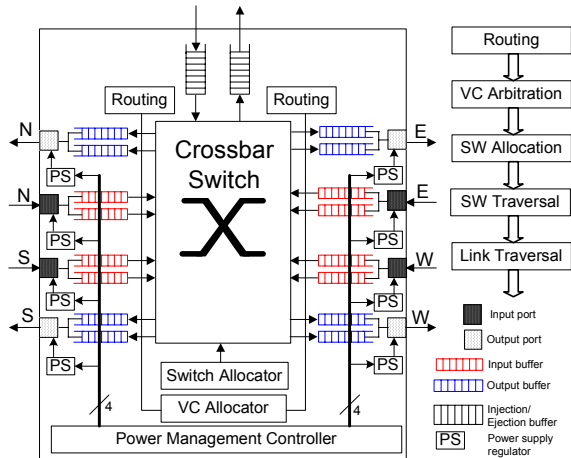


August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

16

Virtual-channel router architecture with on/off links



- Five stage pipelined VC routers
- Separate on/off mechanism for every link individually
- Centralized VC and switch arbitration and allocation

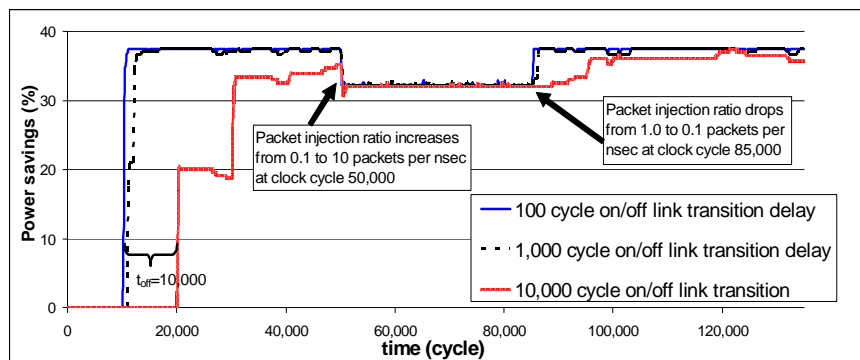
August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

17

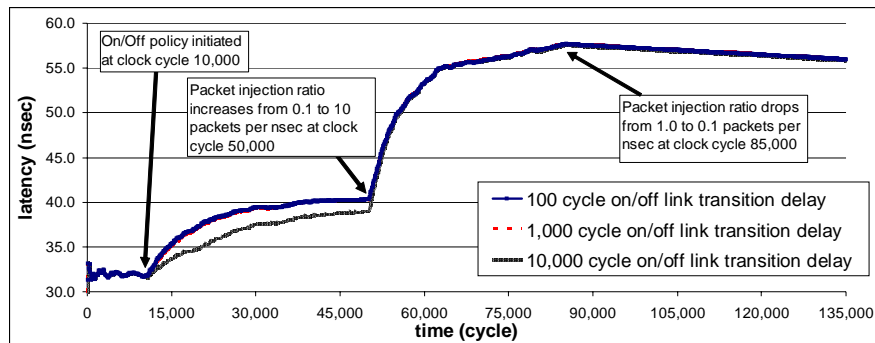
Uniform random traffic for an 8x8 mesh with varying packet injection rate (1 of 2)

- Demonstrates on/off policy response to temporal variance
- Injection ratio (5-flit packets):
 - 0.1 to 1.0 and back to 0.1 packets per nsec



Uniform random traffic for an 8x8 mesh with varying packet injection rate (2 of 2)

- Latency for various t_{off} and t_{on} is almost the same for all three cases
 - Power savings is more critical



August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

19

Conclusion

- **Systematic** architectural and algorithmic approach to reduce power consumption in interconnection networks using **on/off links**
 - Power-performance connectivity graph for meshes
 - Deadlock-free, distributed routing algorithm
 - Router microarchitecture design of on/off mechanism
- High reduction in dynamic and idle power dissipation
- Met various design challenges
- Provided a motivation for circuit researchers to design links with better on/off and power characteristics

August 20, 2003

Vassos Soteriou and Li-Shiuan Peh

20

References (1 of 2)

- W. J. Dally and B. Towles. Route packets, not wires: On-chip interconnection networks. In Proc. Design Automation Conference, pages 684-689, June 2001.
- W. J. Dally, P. Carvey, and L. Dennison., The Avici Terabit switch/router, in Proc. Hot Interconnects 6, Aug. 1998.
- Hang-Sheng Wang, Li-Shiuan Peh and Sharad Malik, A Power Model for Routers: Modeling Alpha 21364 and InfiniBand Routers., In Proceedings of the 10th Symposium on High Performance Interconnects (Hot Interconnects), Stanford, CA, August 2002, pp. 21-27.
- N. J. Boden, D. Cohen, R. E. Felderman, A. E. Kulawik, C. L. Seitz, J. N. Seizovic, and W.-K. Su. Myrinet – a Gigabit-per second local-area-network. IEEE Micro, 15(1):29-36, Feb. 1995.
- S. Mukherjee et al., The Alpha 21364 network Architecture, In Proc. Hot Interconnects, 9 Aug. 2001.
- Hang-Sheng Wang, Xinping Zhu, Li-Shiuan Peh and Sharad Malik, Orion: A Power-Performance Simulator for Interconnection Networks., In Proceedings of the 35th International Symposium on Microarchitecture (MICRO), Istanbul, Turkey, November 2002.
- Li Shang, Li-Shiuan Peh and Niraj K. Jha, Dynamic Voltage Scaling with Links for Power Optimization of Interconnection Networks., In Proceedings of the 9th International Symposium on High-Performance Computer Architecture (HPCA), Anaheim, CA, January 2003.

References (2 of 2)

- C. Patel, S. Chai, S. Yalamanchili, and D. Schimmel. Power-constrained design of multiprocessor interconnection networks. In Proc. Inter. Conference on Computer Design, pages 408-416, Oct. 1997.
- G. Wei, J. Kim, D. Liu, S. Sidiropoulos, and M. Horowitz, A variable-frequency parallel I/O interface with adaptive power-supply regulation, J. Solid-State Circuits, vol. 35, no. 11, pp.1600-1610, Nov. 2000.
- <http://www.motorola.com/>
- J. Duato, A Theory of Fault-Tolerant Routing in Wormhole Networks, IEEE Transactions on Parallel and Distributed Systems, Vol. 8, NO. 8, August 1997.
- Patrick. T. Gaughan and Sudhakar Yalamanchili, A Family of
- Fault-Tolerant Routing Protocols for Direct Multiprocessor Networks, IEEE Trans. on Parallel and Distributed Systems, Vol. 6, NO. 5, May 1995.
- Christopher J. Glass and Lionel M. Ni, The Turn Model for Adaptive Routing, Proceedings of the 19th International Symposium on Computer Architecture, May 1992.
- <http://www.ee.princeton.edu/~lshang/popnet.html>
- InfiniBand Trade Alliance. The InfiniBand architecture. <http://www.infinibandta.org>.