HiRy: An Advanced Theory on Design of Deadlock-free Adaptive Routing for Arbitrary Topologies

2017/12/17

Ryuta Kawano (Keio Univ., Japan)
Ryota Yasudo (Keio Univ., Japan)
Hiroki Matsutani (Keio Univ., Japan)
Michihiro Koibuchi (NII, Japan)
Hideharu Amano (Keio Univ., Japan)

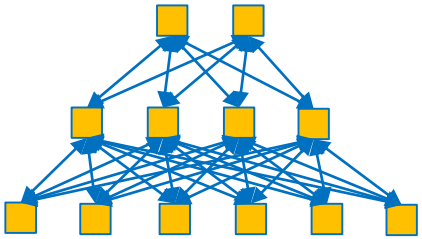
Outline

- Low-latency Network Topologies for HPC systems
- Conventional Deadlock-free Routing Methods
- EbDa – A Generalized Theorem to Design Adaptive Routing for *Mesh and Torus*
- **HiRy** - An Advanced Theorem to Design Adaptive Routing for *Arbitrary Topologies*
- Evaluation by Network Simulation
- Conclusion

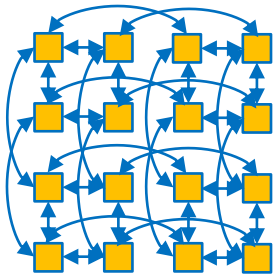
Subject: Inter-switch Networks for HPC Systems

- **Network topologies** are determined based on the required performance and scalability.

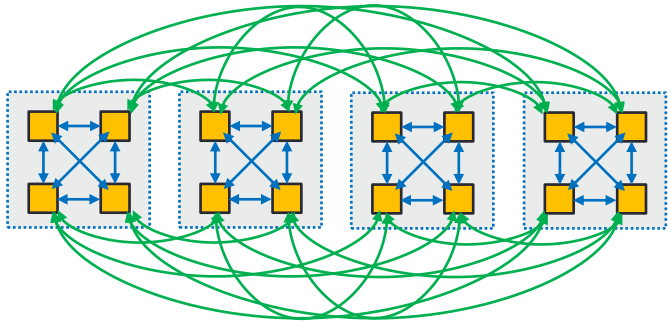
- Fat-tree, Torus, Dragonfly [1] are widely used for HPC systems.



Fat-tree



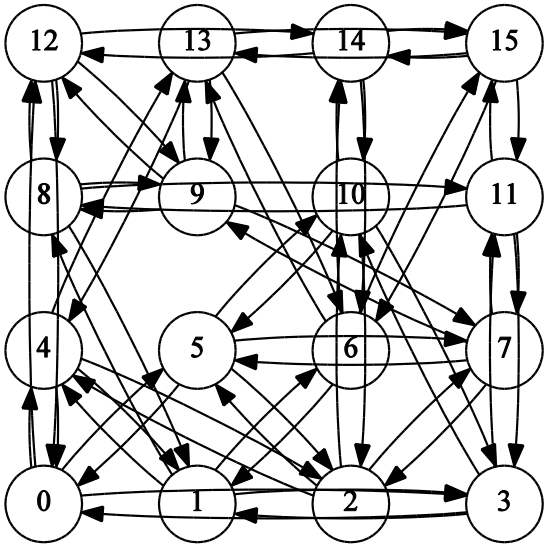
Torus



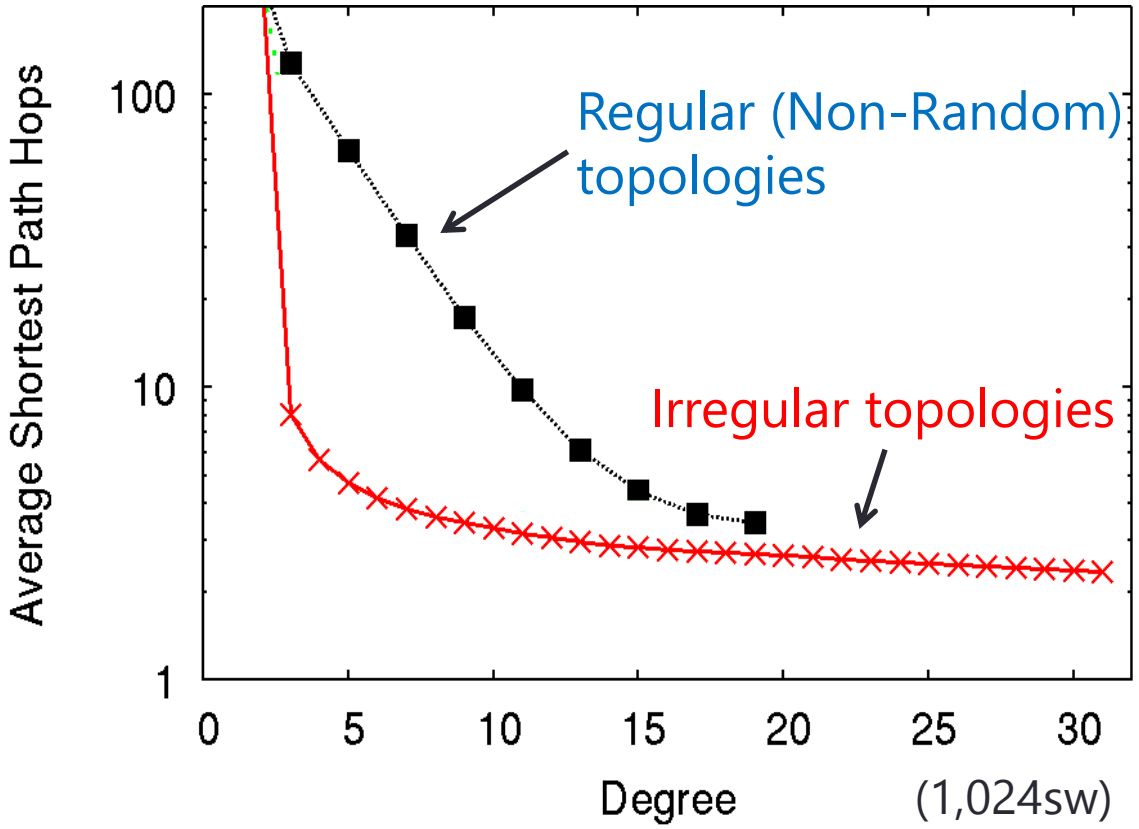
Dragonfly [1]

[1] J. Kim, W. J. Dally, S. Scott and D. Abts: “Technology-Driven, Highly-Scalable Dragony Topology", ISCA’08.

Low-latency Irregular Topologies [2,3] for HPC systems



Inter-Switch
Irregular Topology



Reduction of # of hops with randomized links

[2] M. Koibuchi et al.: "A Case for Random Shortcut Topologies for HPC Interconnects", ISCA'12.
[3] H. Yang et al.: "Dodec: Random-Link, Low-Radix On-Chip Networks", MICRO'14.

Outline

- Low-latency Network Topologies for HPC systems
- Conventional Deadlock-free Routing Methods
- EbDa – A Generalized Theorem to Design Adaptive Routing for *Mesh and Torus*
- **HiRy** - An Advanced Theorem to Design Adaptive Routing for *Arbitrary Topologies*
- Evaluation by Network Simulation
- Conclusion

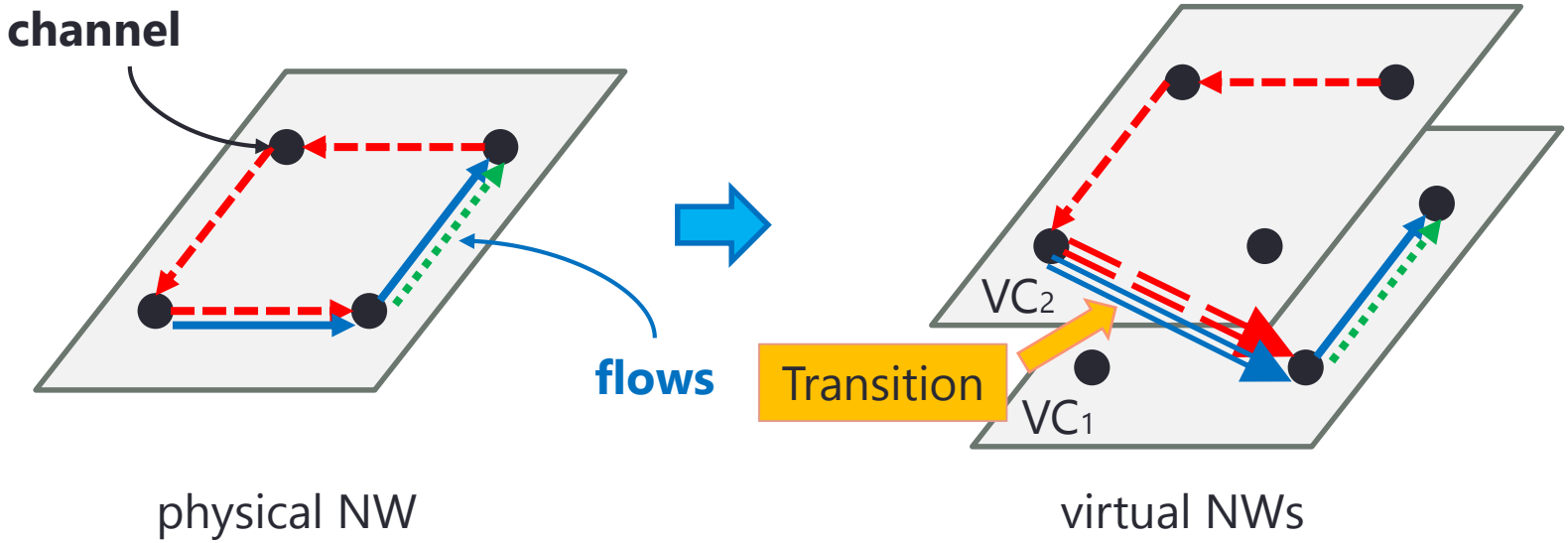
Challenge: Deadlock-free Routing

- Routing methods for irregular topologies have to support **deadlock-freedom** while
  - **reducing the # of hops** to achieve the low latency.
  - **making alternative paths available** to avoid the congestion.

- Conventional topology-independent routing methods for irregular topologies
  - LASH-TOR
  - Duato's protocol

LASH-TOR [4]

- Layered virtual networks generated with multiple Virtual Channels (VCs)
  - Permitting ***transitions*** to achieve minimal routing
- ◯: **Minimal paths,**
  ×: Alternative paths



[4] T. Skeie, O. Lysne, J. Flich, P. Lopez, A. Robles and J. Duato: "LASH-TOR: A Generic Transition-Oriented Routing Algorithm", ICPADS'04.

Duato's Protocol [5]

- Layered virtual networks generated with multiple Virtual Channels (VCs) as LASH-TOR
- Minimal routing on a virtual network and non-minimal and deadlock-free routing on another virtual network
- △: **Minimal paths,**
  ◯: **Alternative paths**
  - Non-minimal routing on high load

[5] F. Silla and J. Duato: "Improving the Efficiency of Adaptive Routing in Networks with Irregular Topology", HiPC'97.

Comparison of Topology-independent Routing Methods

	LASH-TOR	Duato's
Minimal Paths	○	△
Alternative Paths	×	○

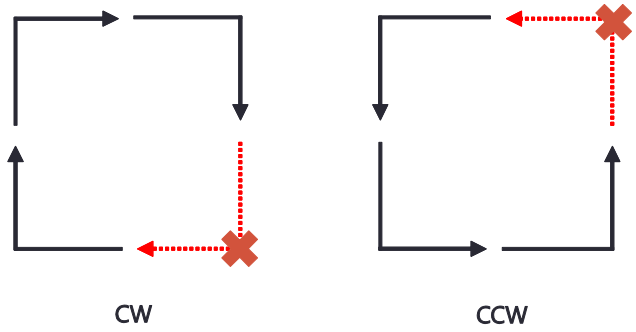
- Challenge: Designing routing methods achieving **minimal paths** and **alternative paths** for irregular networks

Outline

- Low-latency Network Topologies for HPC systems
- Conventional Deadlock-free Routing Methods
- EbDa – A Generalized Theorem to Design Adaptive Routing for *Mesh and Torus*
- **HiRy** - An Advanced Theorem to Design Adaptive Routing for *Arbitrary Topologies*
- Evaluation by Network Simulation
- Conclusion

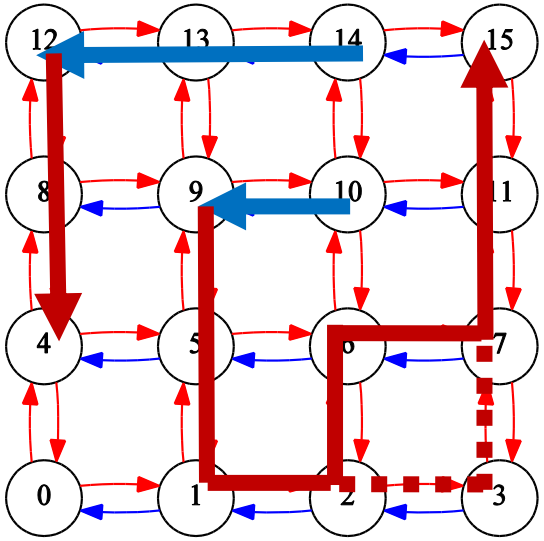
Turn Model

- Routing theorem for Mesh and Torus
  - prohibiting a part of turns to avoid loops



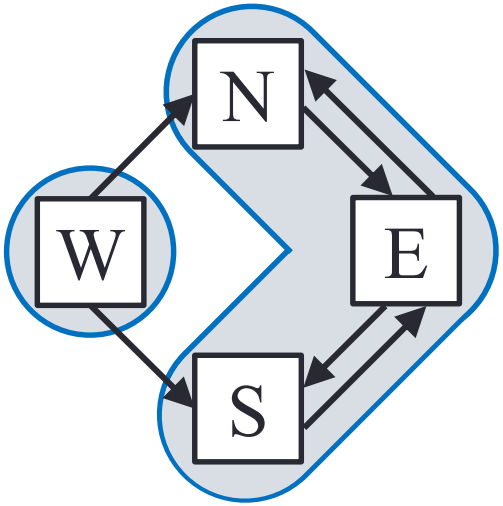
- Example: West-first routing
  – West channels are available before using {North East, South} channels.

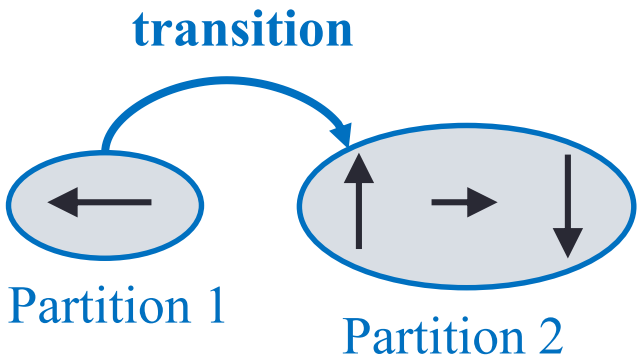
- ○: Minimal paths,
  ○: Alternative paths



EbDa [6] - Generalized Theorems of the Turn Model

- Available turns on West-first routing are illustrated by arrows in the left figure.
  - The directions available arbitrarily and repeatedly can be arranged into a group called a **partition** in EbDa.
- A **transition** between partitions can be illustrated in the right figure.

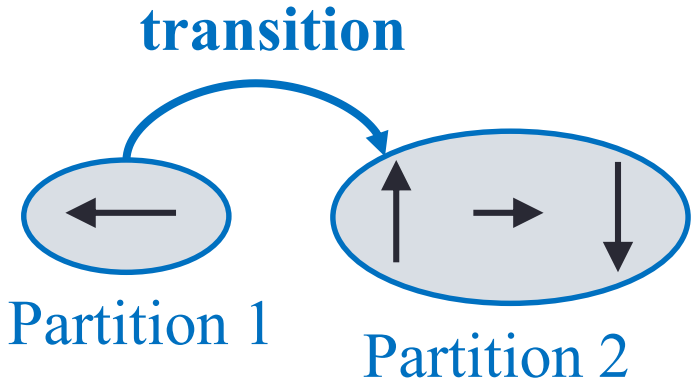


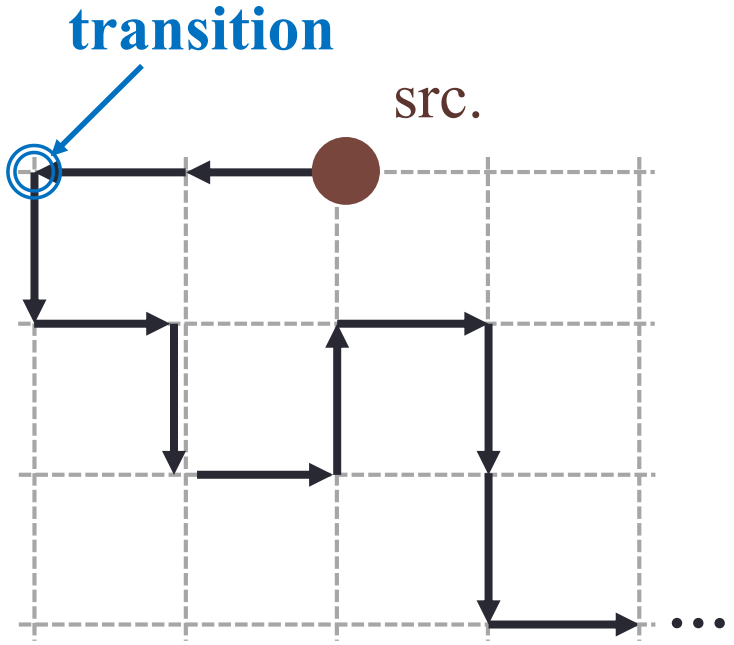


[6] M. Ebrahimi et al: " EbDa: A New Theory on Design and Verification of Deadlock-free Interconnection Networks", ISCA'17.

Deadlock-free Routing in EbDa

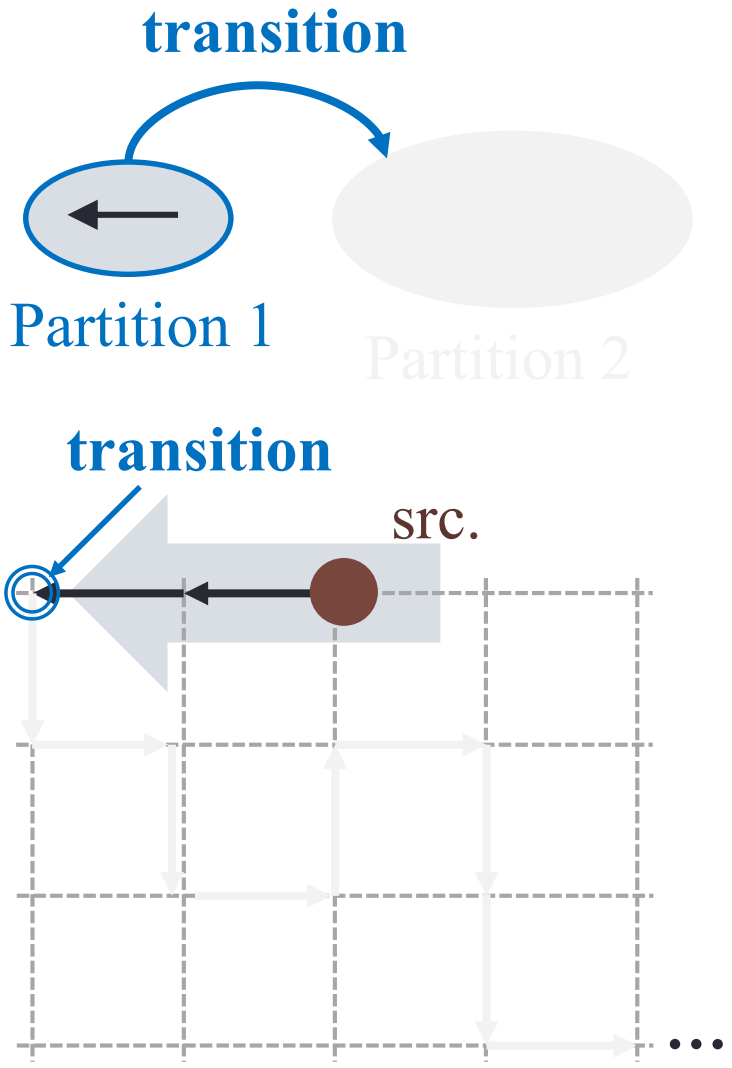
- An intuitive proof for deadlock-freedom
  - An example of a routed path in the bottom-right figure





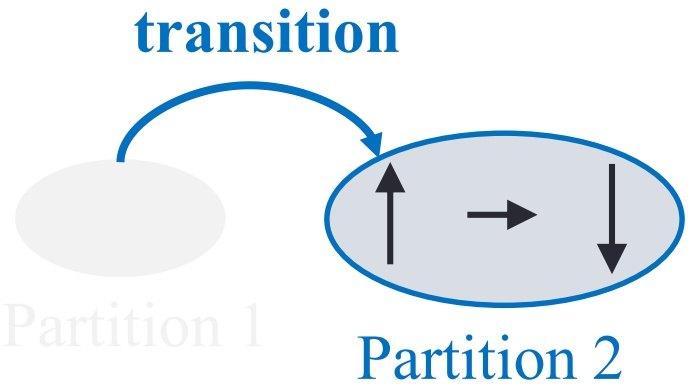
Deadlock-free Routing in EbDa

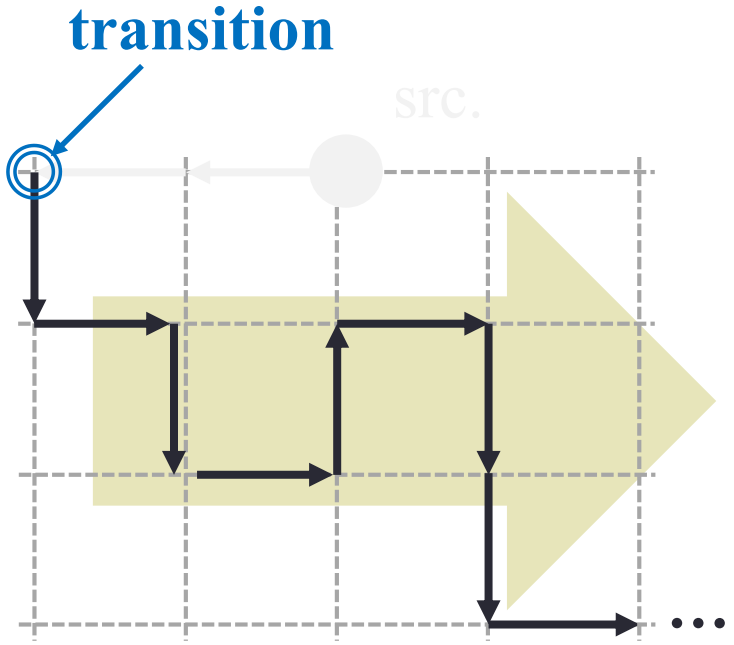
- An intuitive proof for deadlock-freedom
  - An example of a routed path in the bottom-right figure
  - West channels available before the transition
  - The uni-directional transition can avoid loops among partitions.



Deadlock-free Routing in EbDa

- An intuitive proof for deadlock-freedom
  - An example of a routed path in the bottom-right figure
  - West channels available before the transition
  - The uni-directional transition can avoid loops among partitions.
  - After the transition, {North, East, South} channels are available.
  - Packets cannot cause loops because they have to move along the eastern direction **monotonically**.



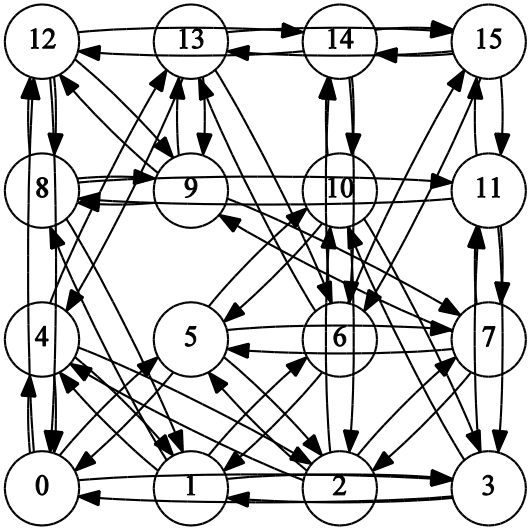


Outline

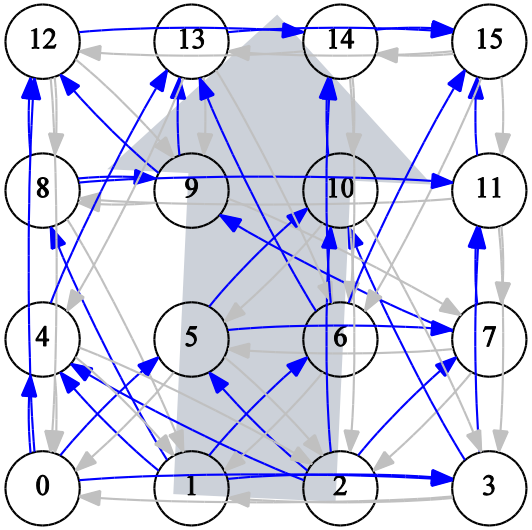
- Low-latency Network Topologies for HPC systems
- Conventional Deadlock-free Routing Methods
- EbDa – A Generalized Theorem to Design Adaptive Routing for *Mesh and Torus*
- **HiRy** - An Advanced Theorem to Design Adaptive Routing for *Arbitrary Topologies*
- Evaluation by Network Simulation
- Conclusion

Proposal：Extention of the EbDa Theorems for *Arbitrary Networks* (≒ *Irregular NWs*)

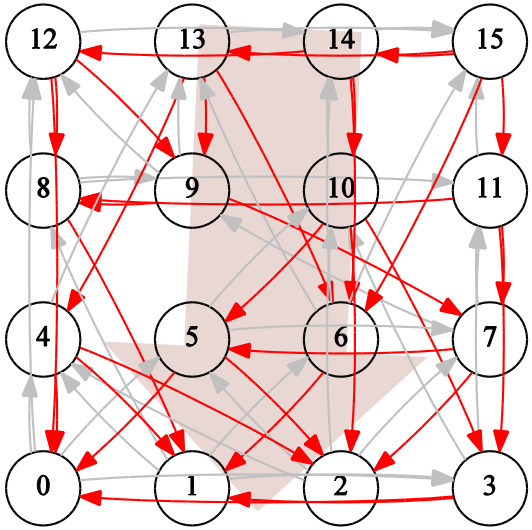
- Grouping channels based on their monotonic directions including diagonal ones
  - An example in the bottom figures
    - Partition1: North channels
    - Partition2: South channels



4×4 Random Topology



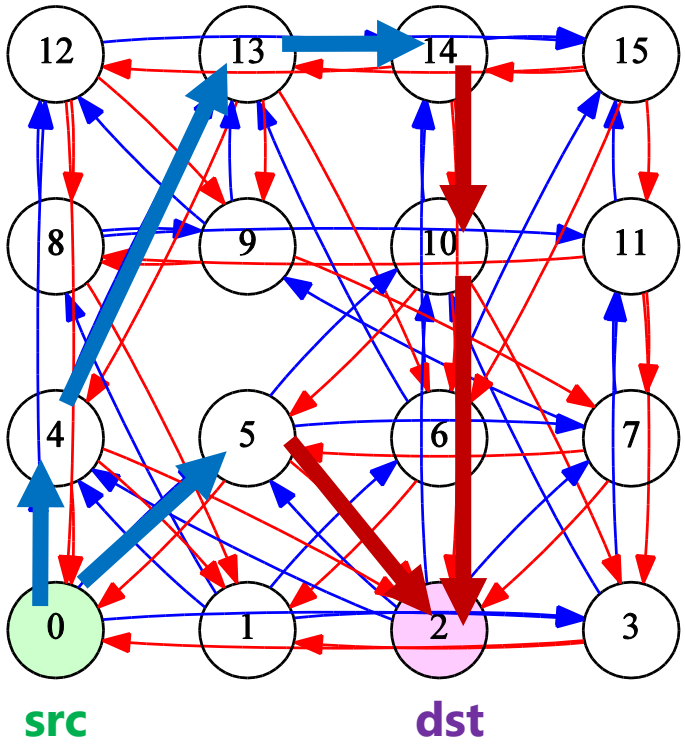
Partition 1



Partition 2

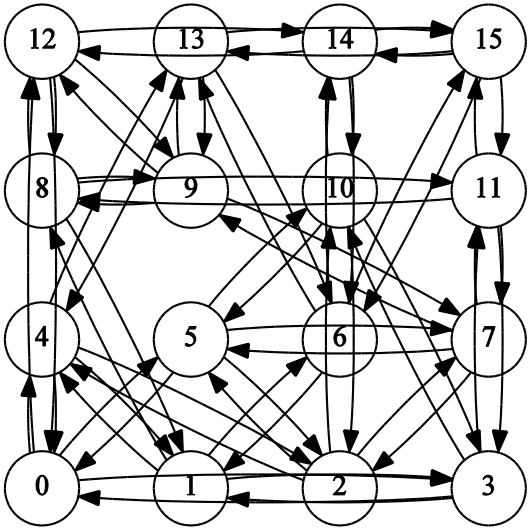
Design of Routing based on the Proposed Theory

- An example of routed paths （the right figure)
  - The channels in Partition 1 available before those in Partition 2
    - Packets can avoid loops because they have to move monotonically in each partition.
  - As the turn model, congestion can be avoided by alternative paths.

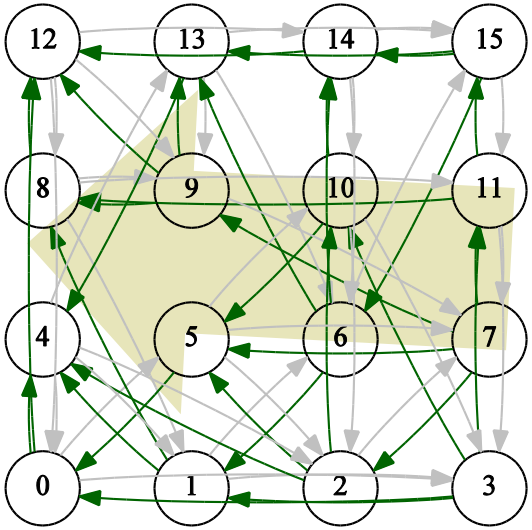


Other Partitions Derived from the Different Monotonic Directions

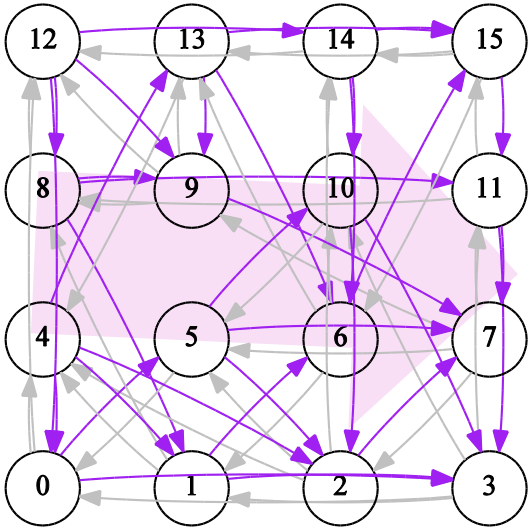
- Partitions can be generated for arbitrary monotonic directions.
  - An example in the bottom figures
    - Partition1: West channels
    - Partition2: East channels



4×4 Random Topology



Partition 1

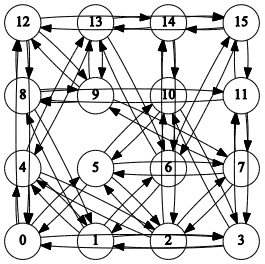


Partition 2

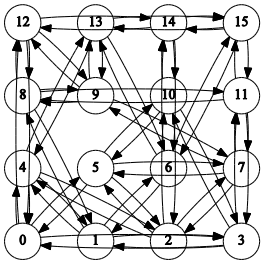
An Implementation of Deadlock-free Routing based on the proposed theory

- Virtual networks generated with multiple Virtual Channels (VCs) as LASH-TOR and Duato's protocol

(# of VC = 2)



Virtual NW 1

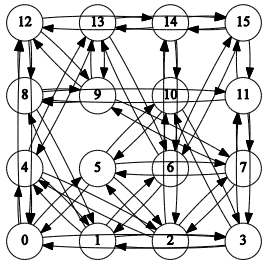


Virtual NW 2

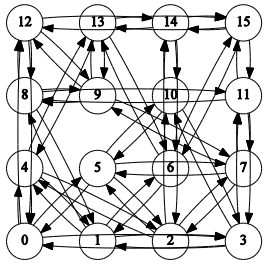
An Implementation of Deadlock-free Routing based on the proposed theory

- Virtual networks generated with multiple Virtual Channels (VCs) as LASH-TOR and Duato's protocol
  - Partitions generated in each virtual Network

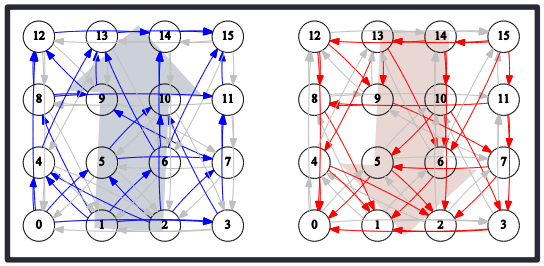
(# of VC = 2)

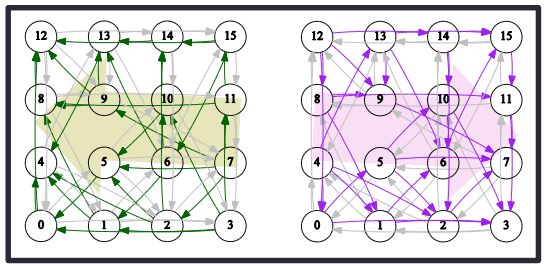


Virtual NW 1



Virtual NW 2

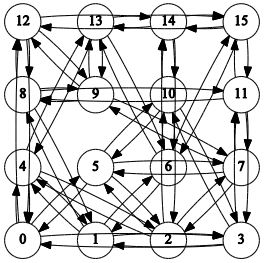




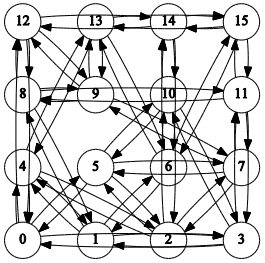
An Implementation of Deadlock-free Routing based on the proposed theory

- Virtual networks generated with multiple Virtual Channels (VCs) as LASH-TOR and Duato's protocol

(# of VC = 2)



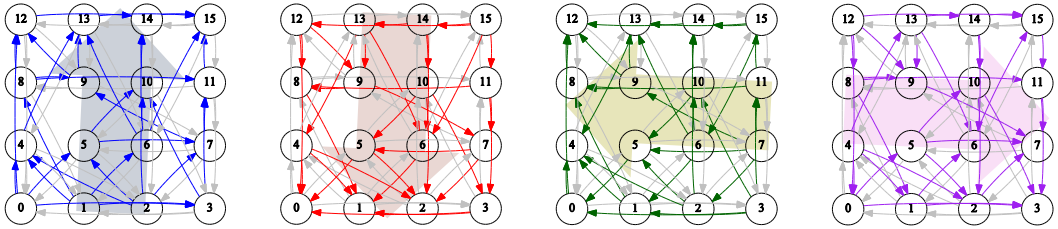
Virtual NW 1



Virtual NW 2

- Partitions generated in each virtual Network

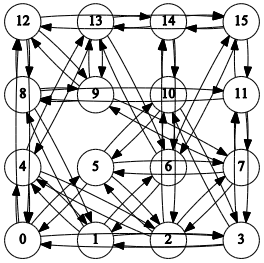
- The order of the partitions are sorted to reduce the average path hops.



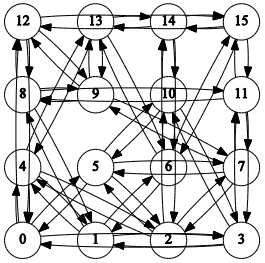
An Implementation of Deadlock-free Routing based on the proposed theory

- Virtual networks generated with multiple Virtual Channels (VCs) as LASH-TOR and Duato's protocol

(# of VC = 2)

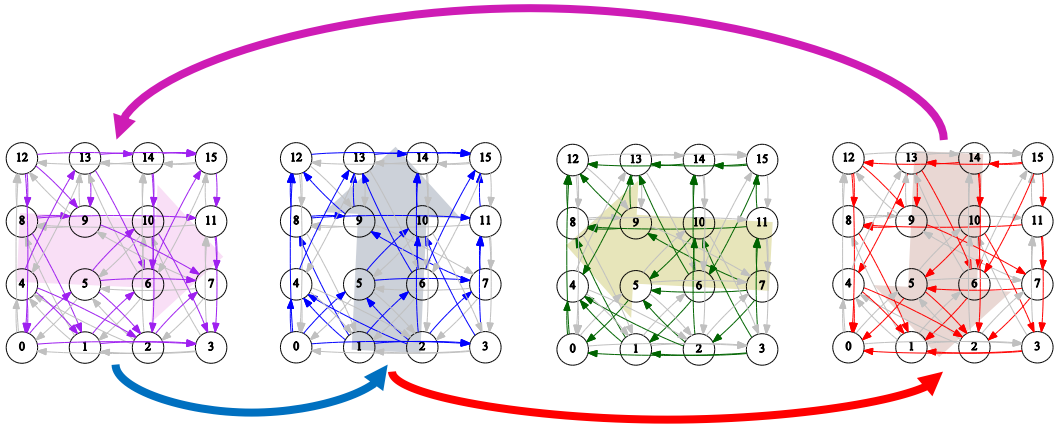


Virtual NW 1



Virtual NW 2

- Partitions generated in each virtual Network
- The order of the partitions are sorted to reduce the average path hops.



Partition 1 Partition 2 Partition 3 Partition 4

Outline

- Low-latency Network Topologies for HPC systems
- Conventional Deadlock-free Routing Methods
- EbDa – A Generalized Theorem to Design Adaptive Routing for *Mesh and Torus*
- **HiRy** - An Advanced Theorem to Design Adaptive Routing for *Arbitrary Topologies*
- Evaluation by Network Simulation
- Conclusion

Network Simulation Environment

- Booksim simulator [7]
- Evaluating
  - LASH-TOR
  - Duato's protocol
    - up*/down* routing for non-minimal deadlock-free paths
  - **HiRy**-based implementation
    - # of dimensions =2, 3, 4
- Applying 4 traffics
  - Uniform, Transpose, Reverse, Shuffle

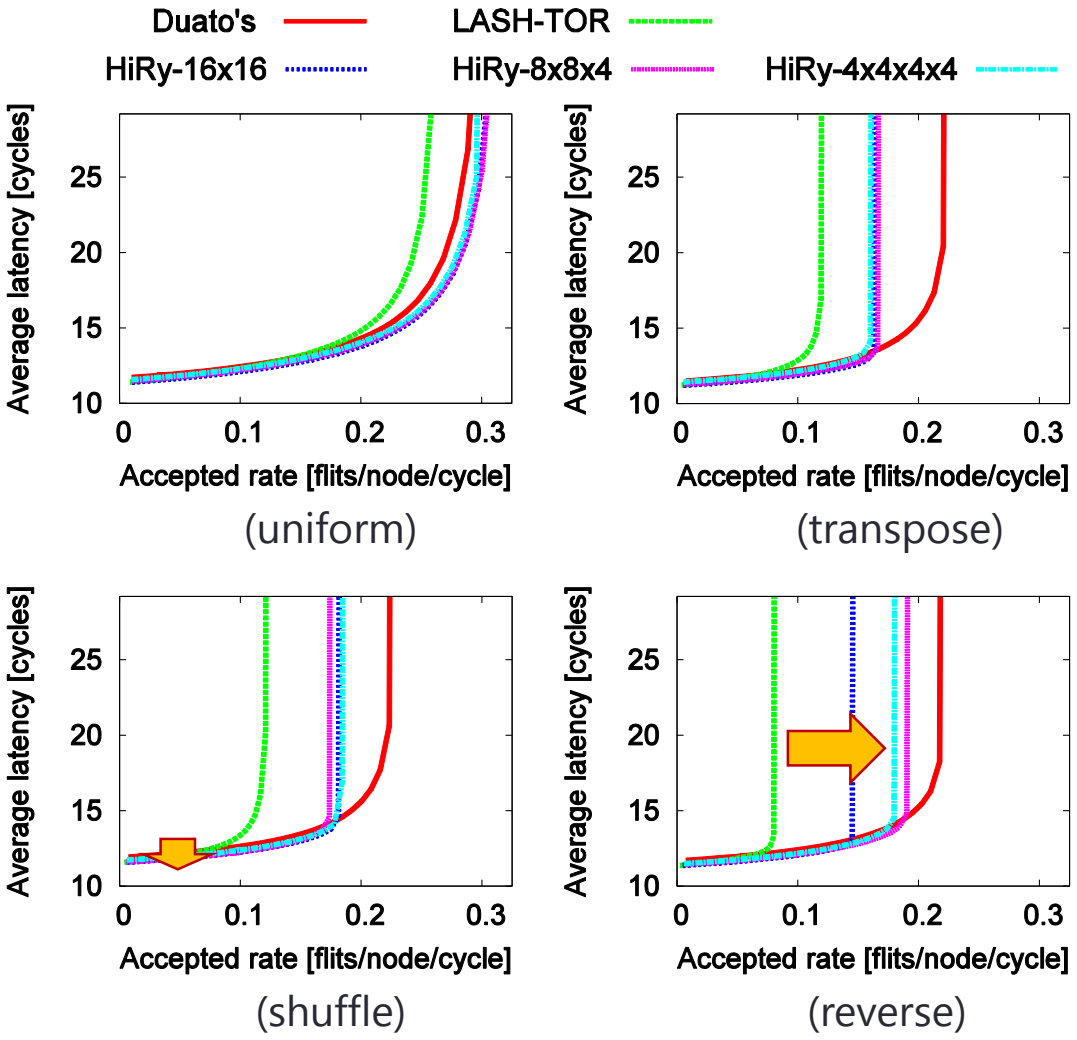
Topology and simulation parameters

NW topology	Random regular topology
# of nodes (SWs)	256
Degree (# of ports)	13 (required for LASH-TOR)
Simulation period	100,000 cycles
Packet size	1 flit
# of VCs	2
Buffer size / VC	8 flits
# of pipeline stages	4

[7] N. Jiang et al. : "A Detailed and Flexible Cycle-Accurate Network-on-Chip Simulator," ISPASS'13.

NW Simulation Results (256 nodes)

- Improving the throughput with alternative paths by up to **138%** compared with LASH-TOR
- Reducing the latency with minimal paths by up to **2.9%** compared with Duato's protocol



Conclusions

- **HiRy**, a theory to design deadlock-free routing with the low latency and the high throughput for irregular networks
  - Extention of the EbDa theorems, generalization of the turn model
- An Implementation of the routing method based on **HiRy**
  - Improving the throughput by up to **138%** compared with LASH-TOR
  - Reducing the latency by up to **2.9%** compared with Duato's protocol