Design of Bandwidth Aware and Congestion Avoiding Efficient Routing Algorithms for NoCs Platforms

M. Palesi[1], G. Longo[1], S. Signorino[1], R. Holsmark[2], S. Kumar[2], V. Catania[1]

[1]DIIT, University of Catania, Italy
{mpalesi, vcatania}@diit.unict.it

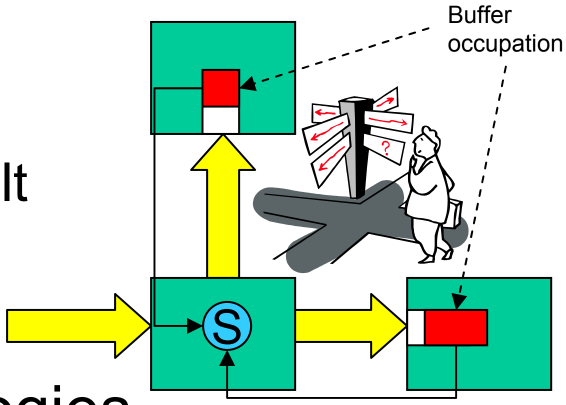
[2]Jőnkőping University, Sweden
{Rickard.Holsmark, Shashi.Kumar}@jth.hj.se

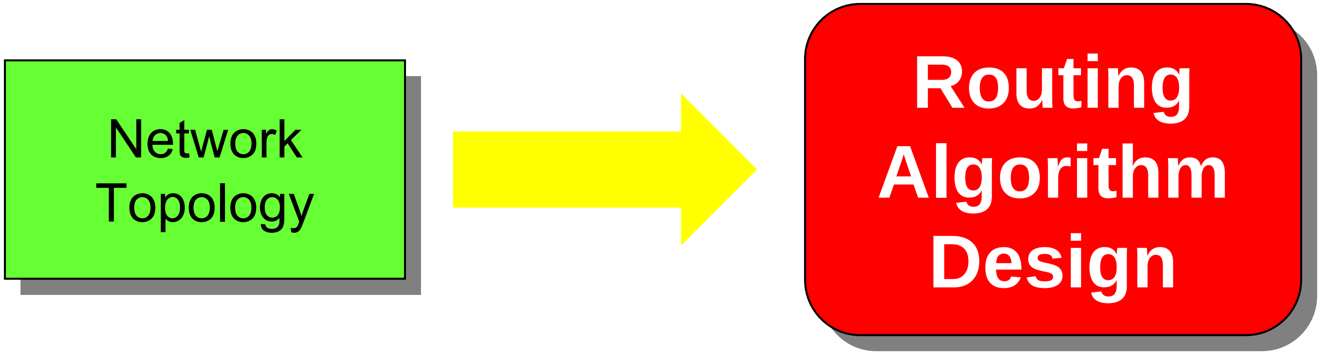
Outline

- Motivation
- Application specific scenario
- Bandwidth aware routing algorithm
- Experiments and Results
- Architectural implications
- Conclusions

Limitations of Current Routing Algorithms

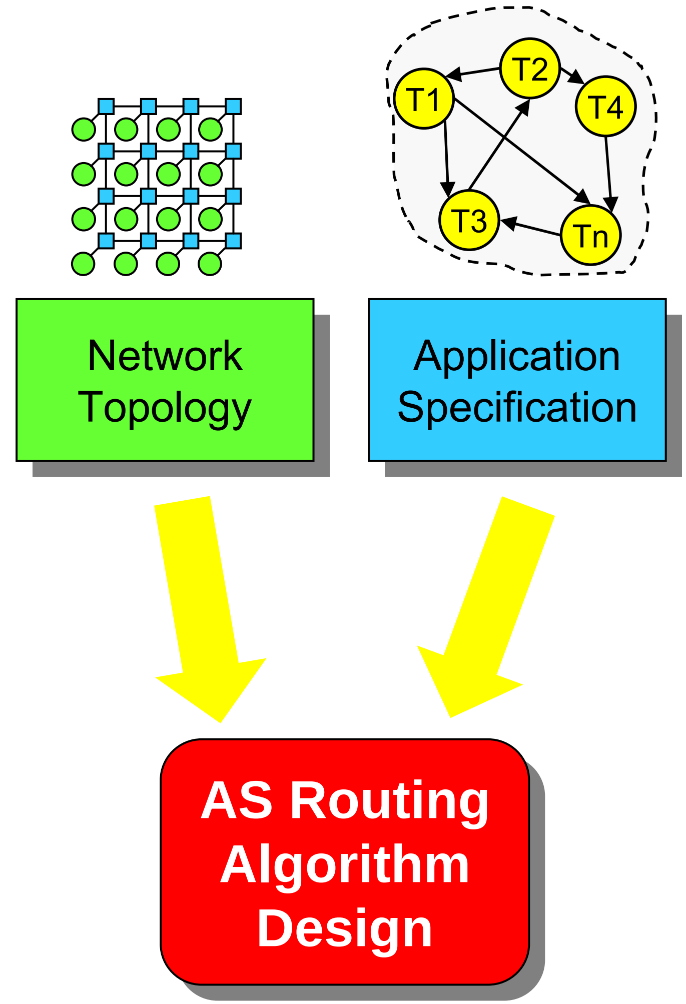
- Efforts biased toward performance
- Side effects like congestion ignored
  - ➔ Estimation and control of congestion is difficult in general
  - ➔ Partially tackled by the selection function
- Designed only for specific network topologies





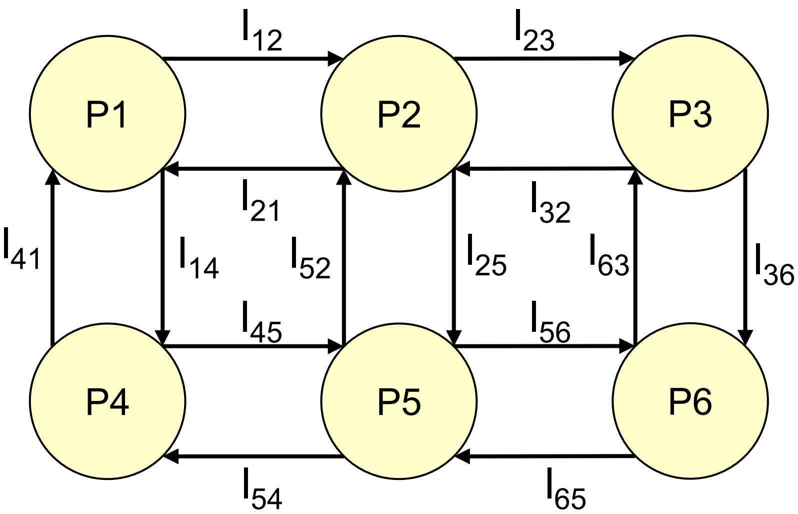
Application Specific Scenario

- Information available about
  - Tasks which communicate and tasks which do never communicate
    - After task mapping ➔ Information about network nodes which communicate
  - Concurrent/non concurrent communications
  - Communication bandwidth requirement for different pais
- Many opportunities
  - Improving performance (e.g., maximize routing adaptivity)
  - Simplify the estimation/control of congestion
  - Design more effective selection policies

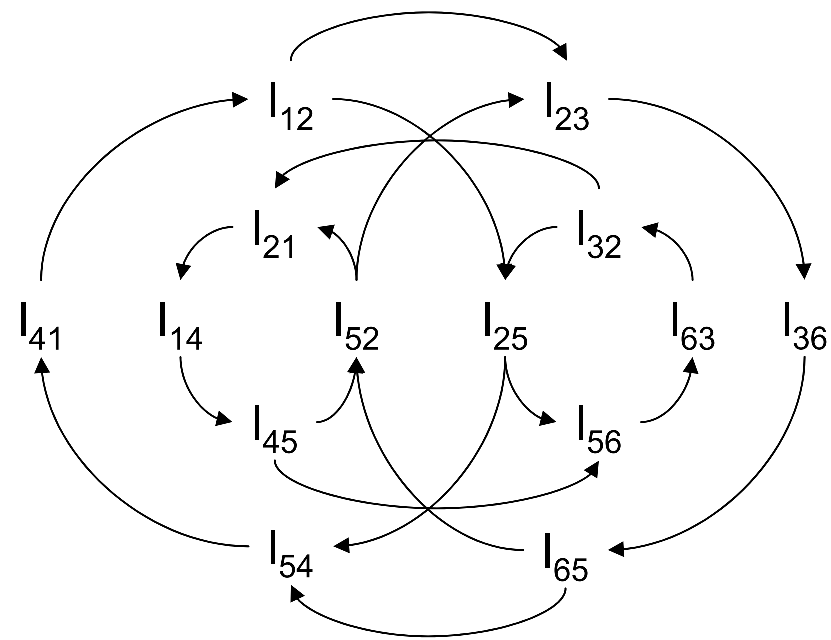


Channel Dependency Graph

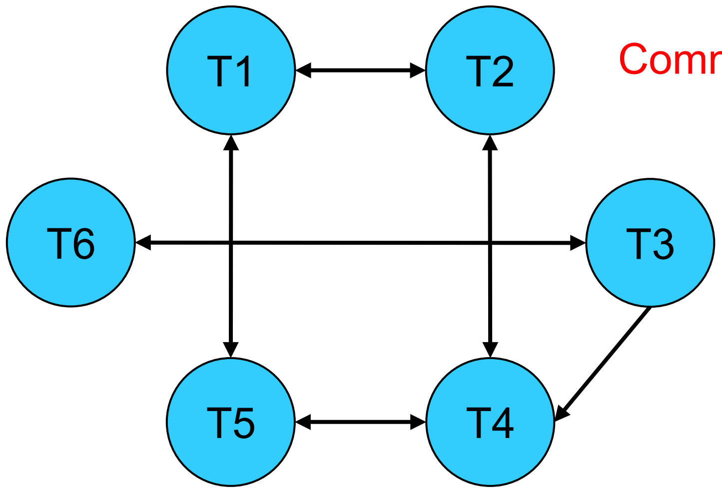
Topology Graph



Channel Dependency Graph

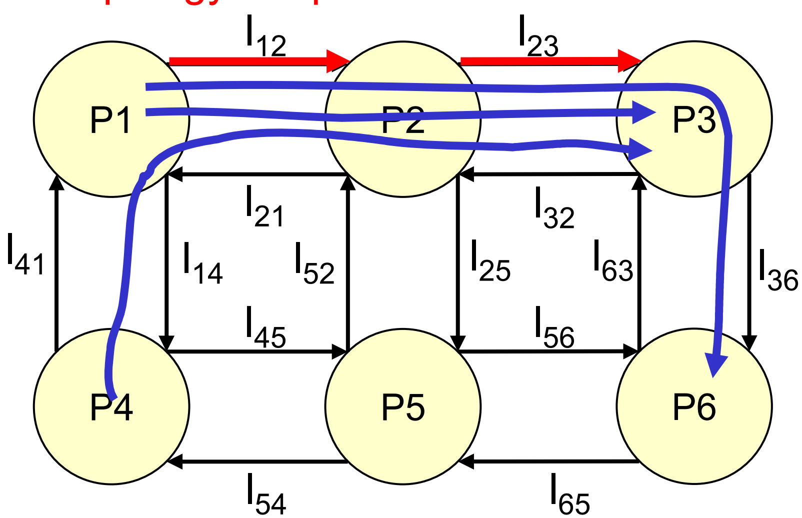


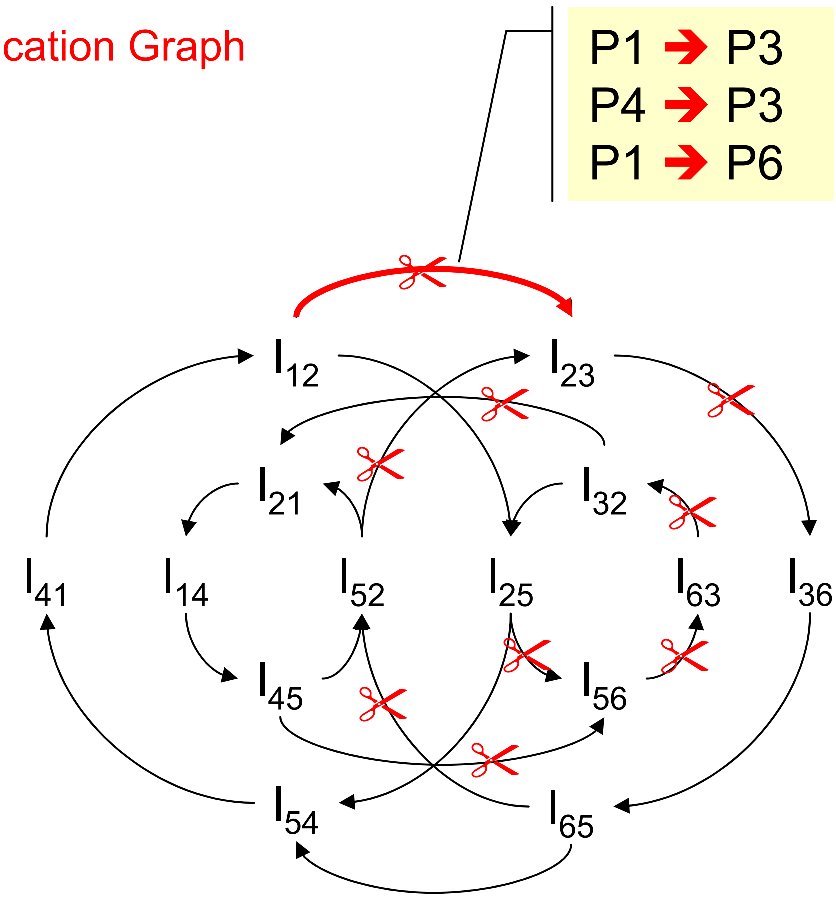
Application Specific CDG



Communication Graph

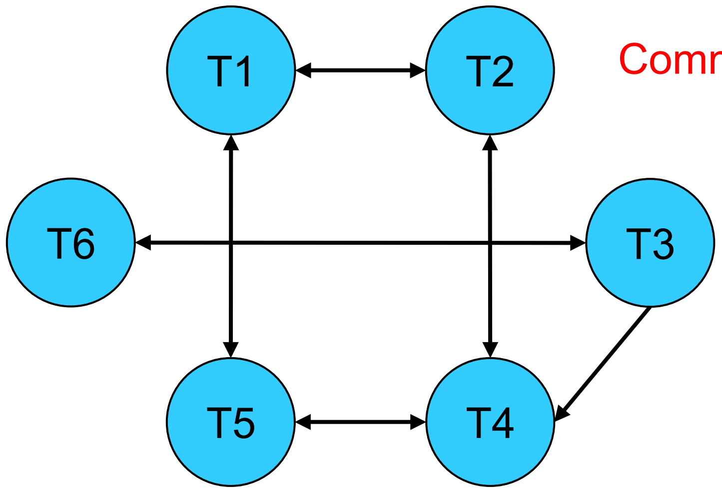
Topology Graph





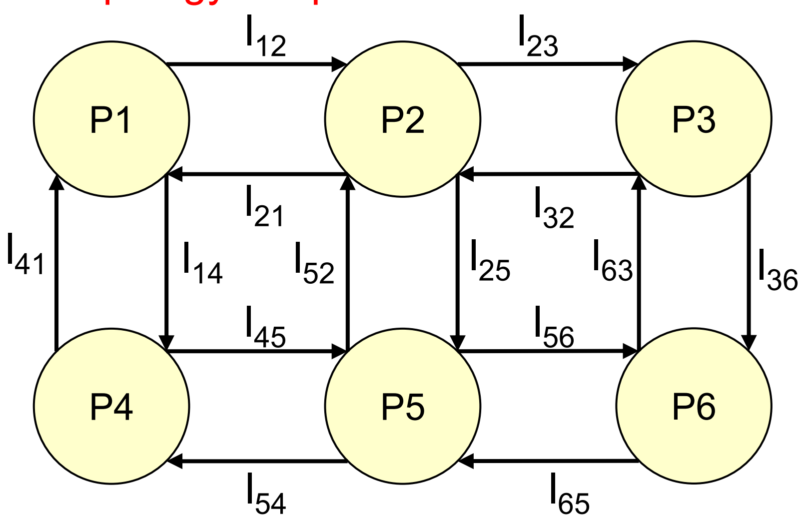
Channel Dependency Graph

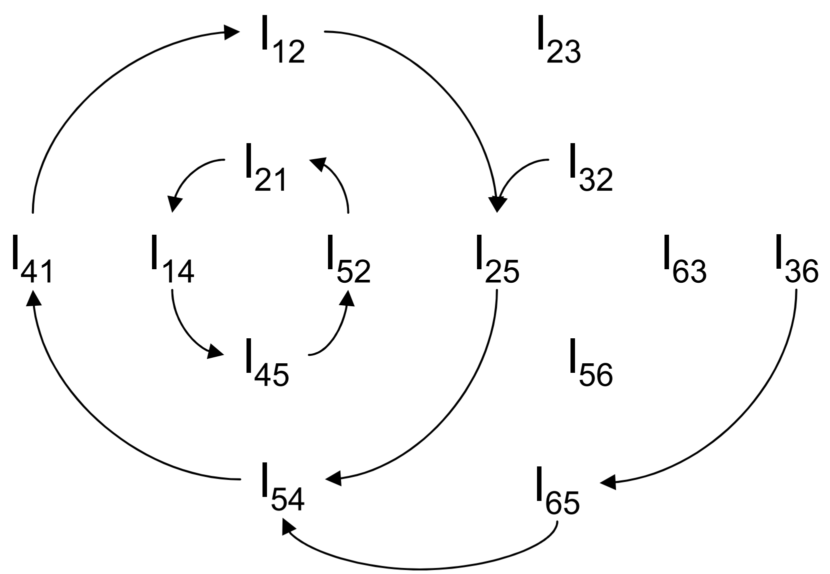
Application Specific CDG



Communication Graph

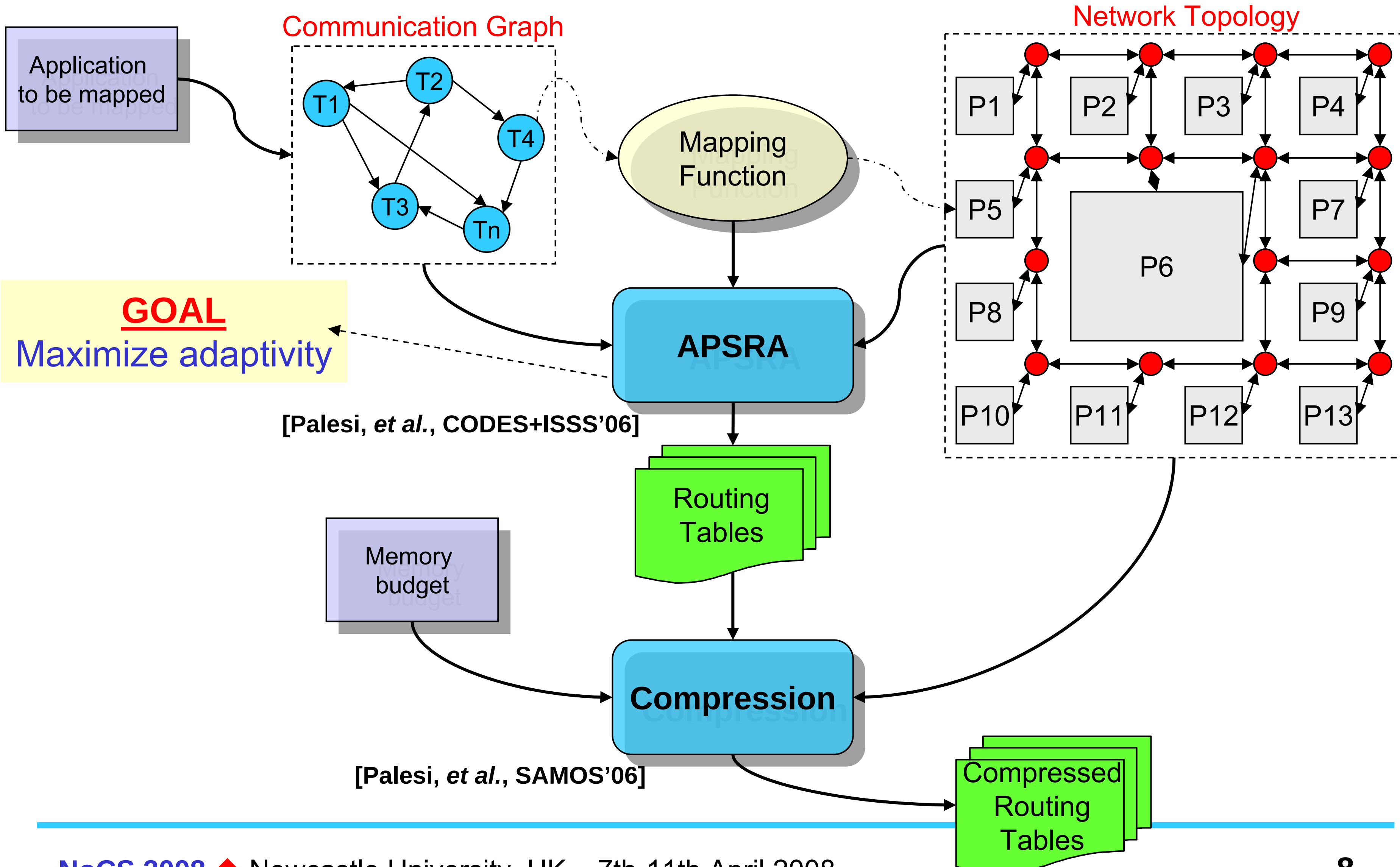
Topology Graph



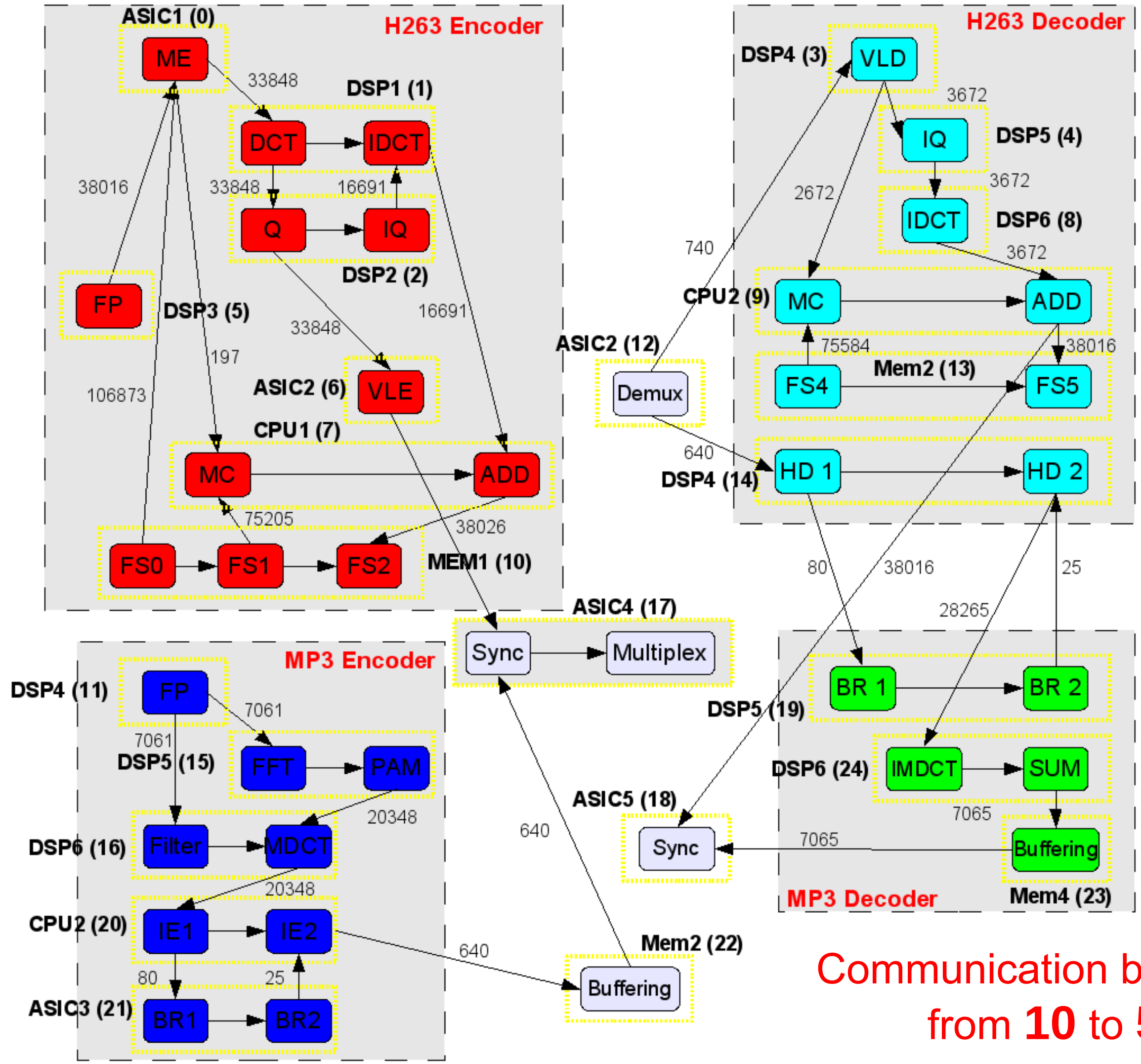


Application Specific
Channel Dependency Graph

APSRA Design Methodology



Bandwidth Variation: Multimedia Example



Source: Hu and Marculescu, TCAD 24(4), 2005

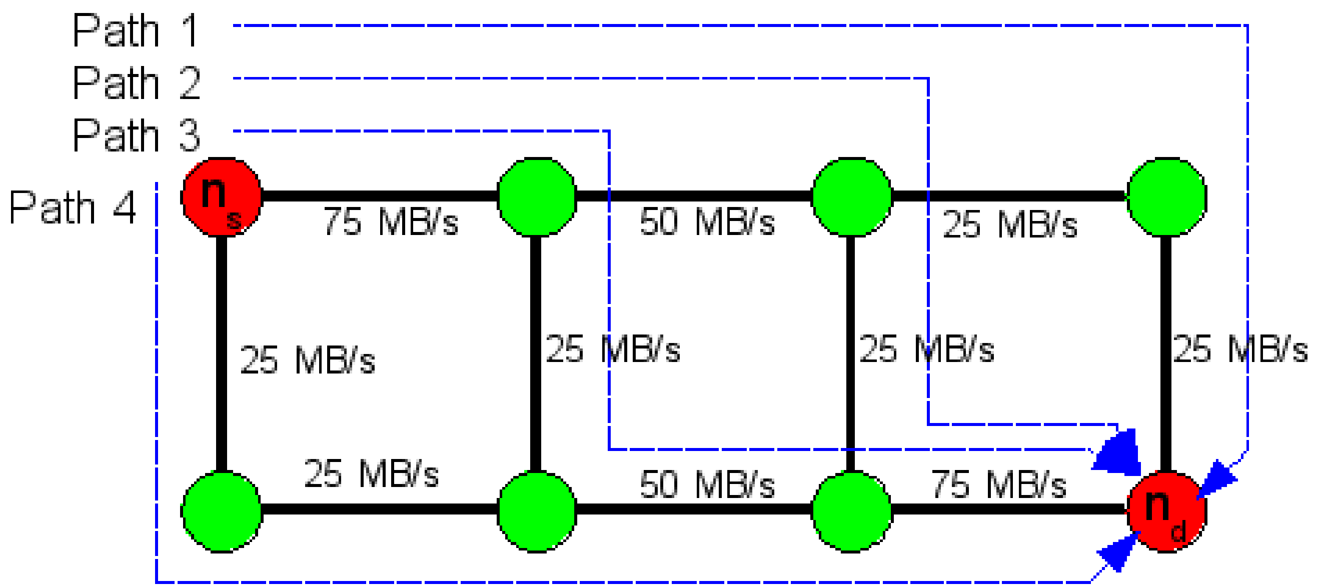
Communication bandwidth ranges from **10** to **500 MB/s**

Contributions

- Design APSRAs which
  - ➔ are highly adaptive
    - ✓ Translates into high performance, in general
  - ➔ uniformly distribute traffic over the network
  - ➔ allow maintenance of load of links under a given bandwidth threshold

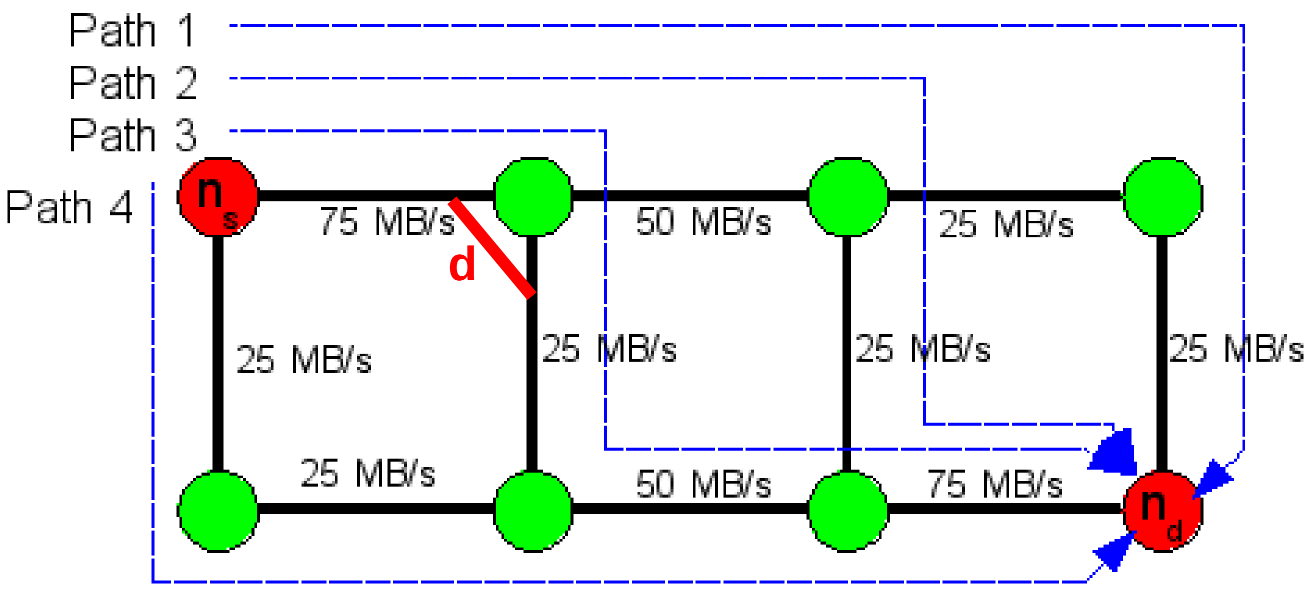
1st Phase

- Removing a dependency ***d***
  - ➔ Removing all the paths which use ***d***
- As soon as a path is removed
  - ➔ The fraction of bandwidth it transports must be redistributed between the remaining paths



1st Phase

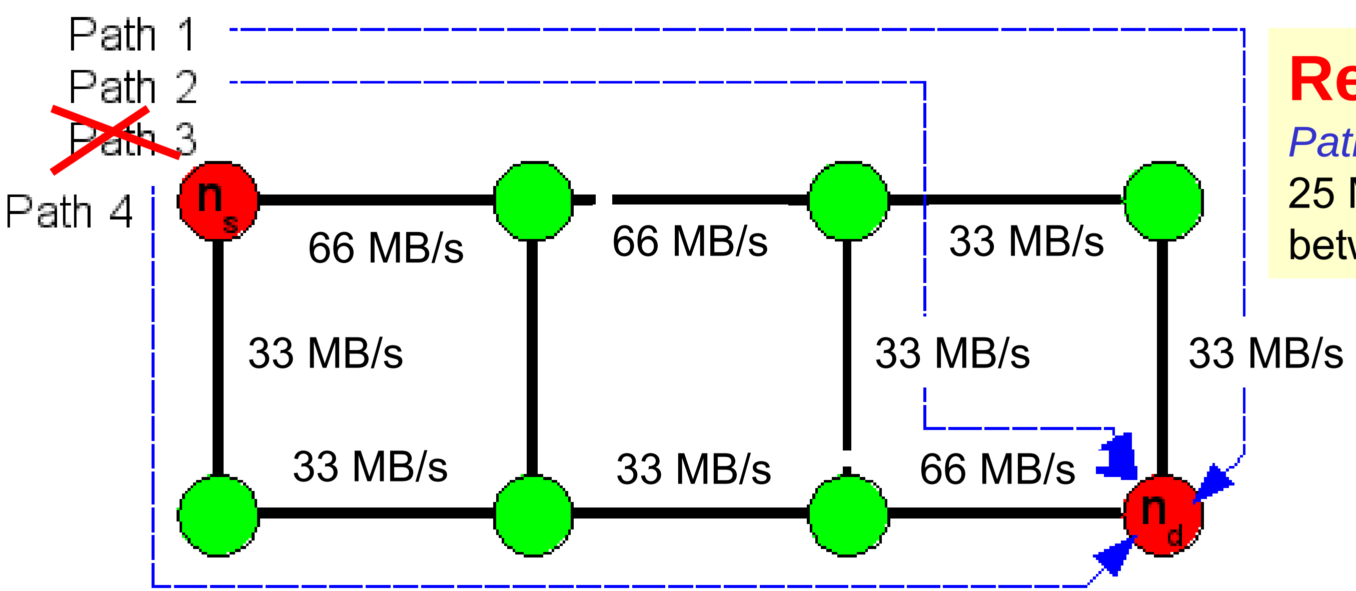
- Removing a dependency ***d***
  - ➔ Removing all the paths which use ***d***
- As soon as a path is removed
  - ➔ The fraction of bandwidth it transports must be redistributed between the remaining paths



d due to *Path 3*

1st Phase

- Removing a dependency ***d***
  - ➔ Removing all the paths which use ***d***
- As soon as a path is removed
  - ➔ The fraction of bandwidth it transports must be redistributed between the remaining paths

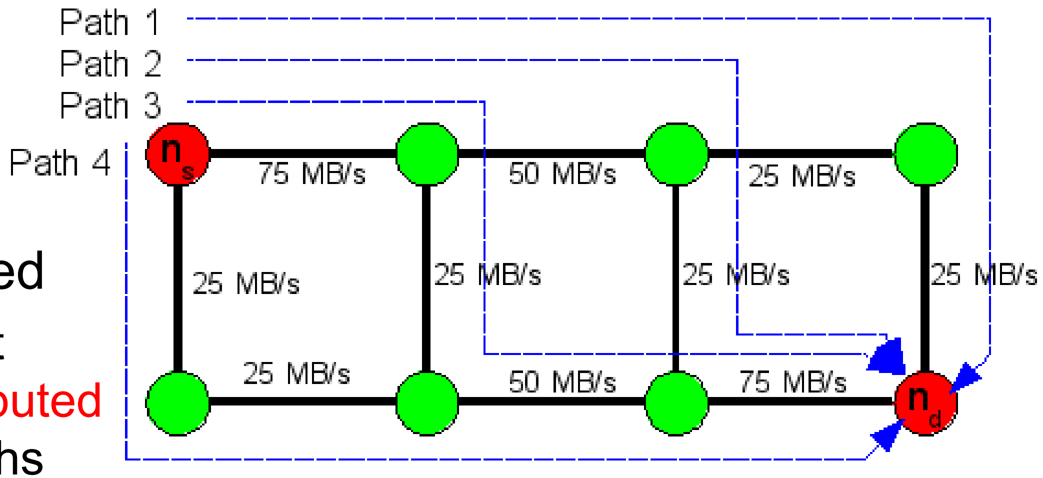


Remove d

Path 3 must be removed
25 MB/s must be redistributed between the remaining paths

1st Phase

- Removing a dependency ***d***
  - ➔ Removing all the paths which use ***d***
- As soon as a path is removed
  - ➔ The fraction of bandwidth it transports must be redistributed between the remaining paths



- Strategy
  - ➔ Removing the dependency ***d*** which minimizes the overhead of bandwidth that should be allocated to the remaining paths that do not use ***d***

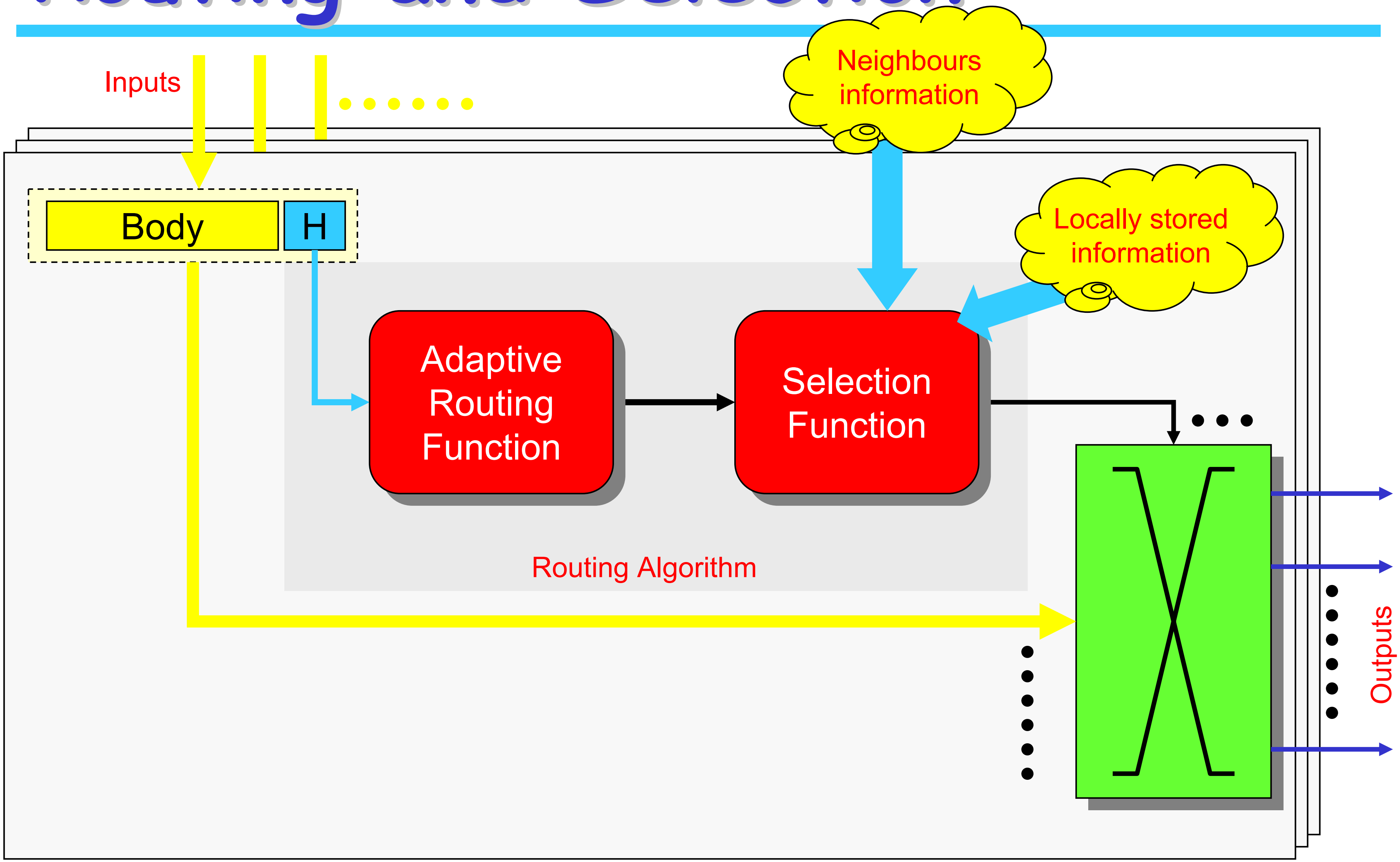
$$\mathrm{cost}(d) = \sum_{c \in C} \frac{B(c) \times \left|PT^2(c,d)\right|}{\left|P(c)\right| \times \left(\left|P(c)\right| - \left|PT^2(c,d)\right|\right)}$$

2nd Phase

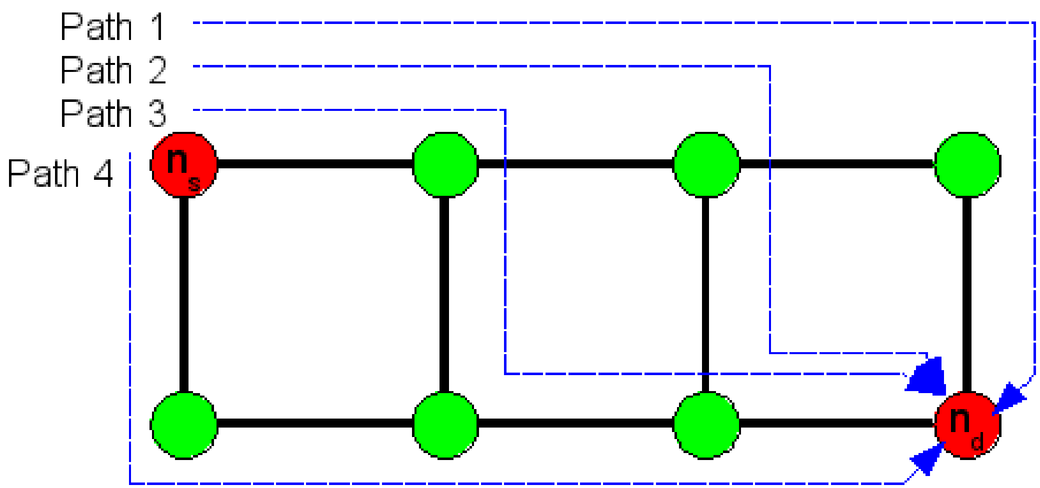
- After *Phase 1* we have a routing function which is
  - ➔ Deadlock free
  - ➔ Provides more adaptivity to communications characterized by higher communication bandwidth
- But…
  - ➔ It is possible that the agregate bandwidth (***AB***) on some links exceeds the capacity of that link
- **"Some"** routing paths passing on that link, must be removed to reduce the ***AB*** on that link down to the link capacity

$$\forall \text{ link } l \rightarrow AB(l) \leq Cap(l)$$

Routing and Selection



Load Balancing Selection Function



Dst	Admissible Out		Selection Prob	
…	…	…	…	…
…	…	…	…	…
n_d	E	S	0,75	0,25
…	…	…	…	…
…	…	…	…	…

Overhead

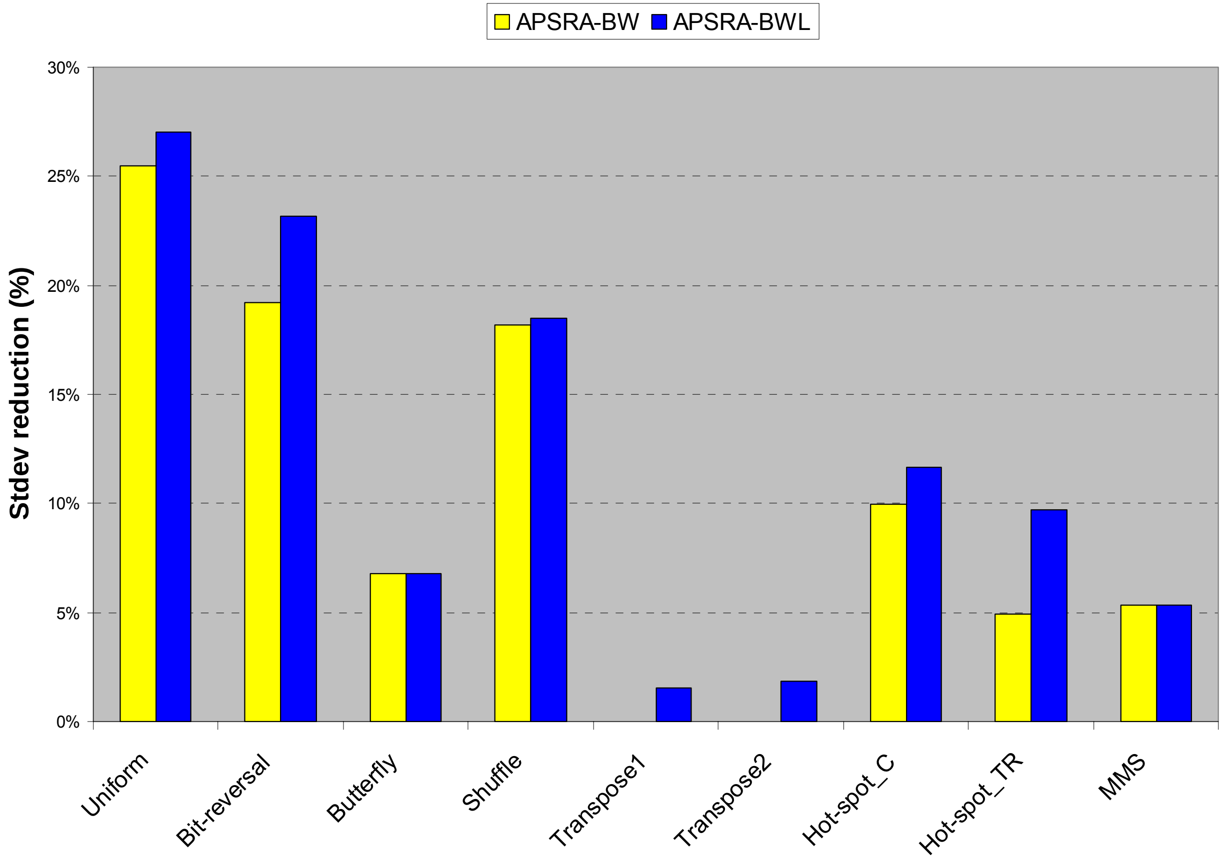
- The probability to select output channel ***l*** is proportional to the number of admissible paths starting from ***l*** and that can be used to reach the destination

Experimental Setup

- 8x8 mesh based NoC
- Buffer size 4-flits
- Simulation time 100,000 cycles
- Warmup time 20,000 cycles
- Traffic injection distribution
  - Poisson (for synthetic traffic scenarios)
  - Self-similar (for MMS traffic)

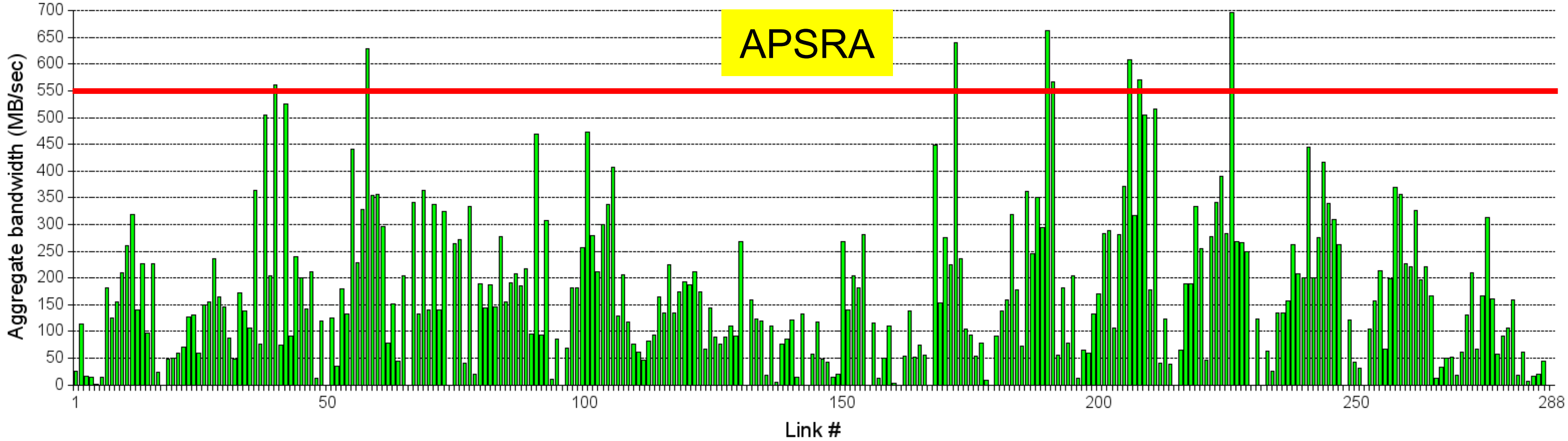
Stdev Reduction

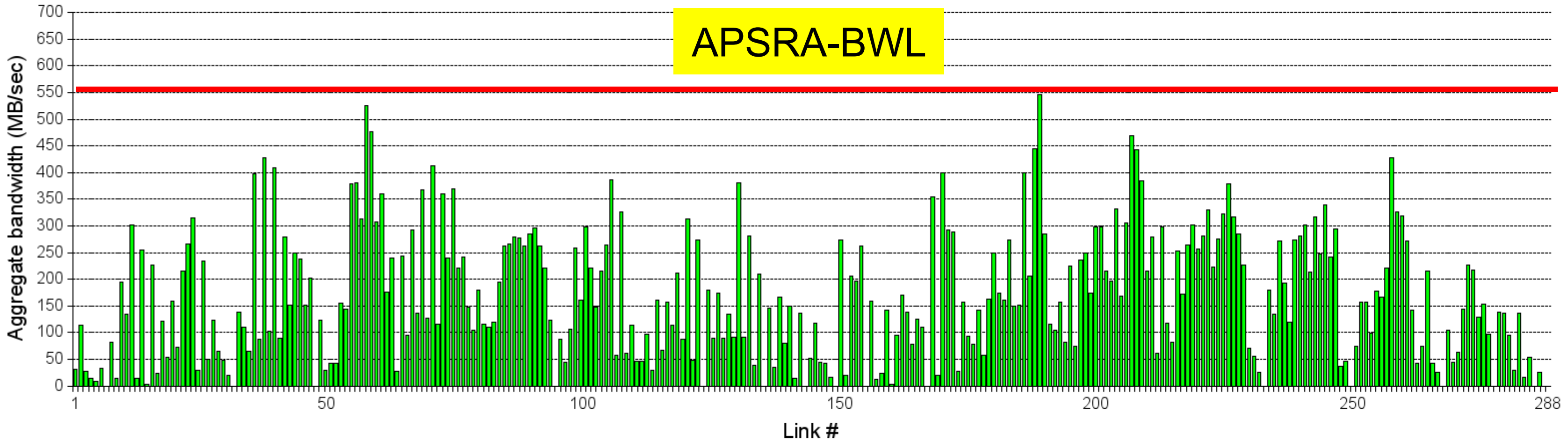
- Percentage reduction of standard deviation of the aggregated bandwidth in network links



Aggregate bandwidth

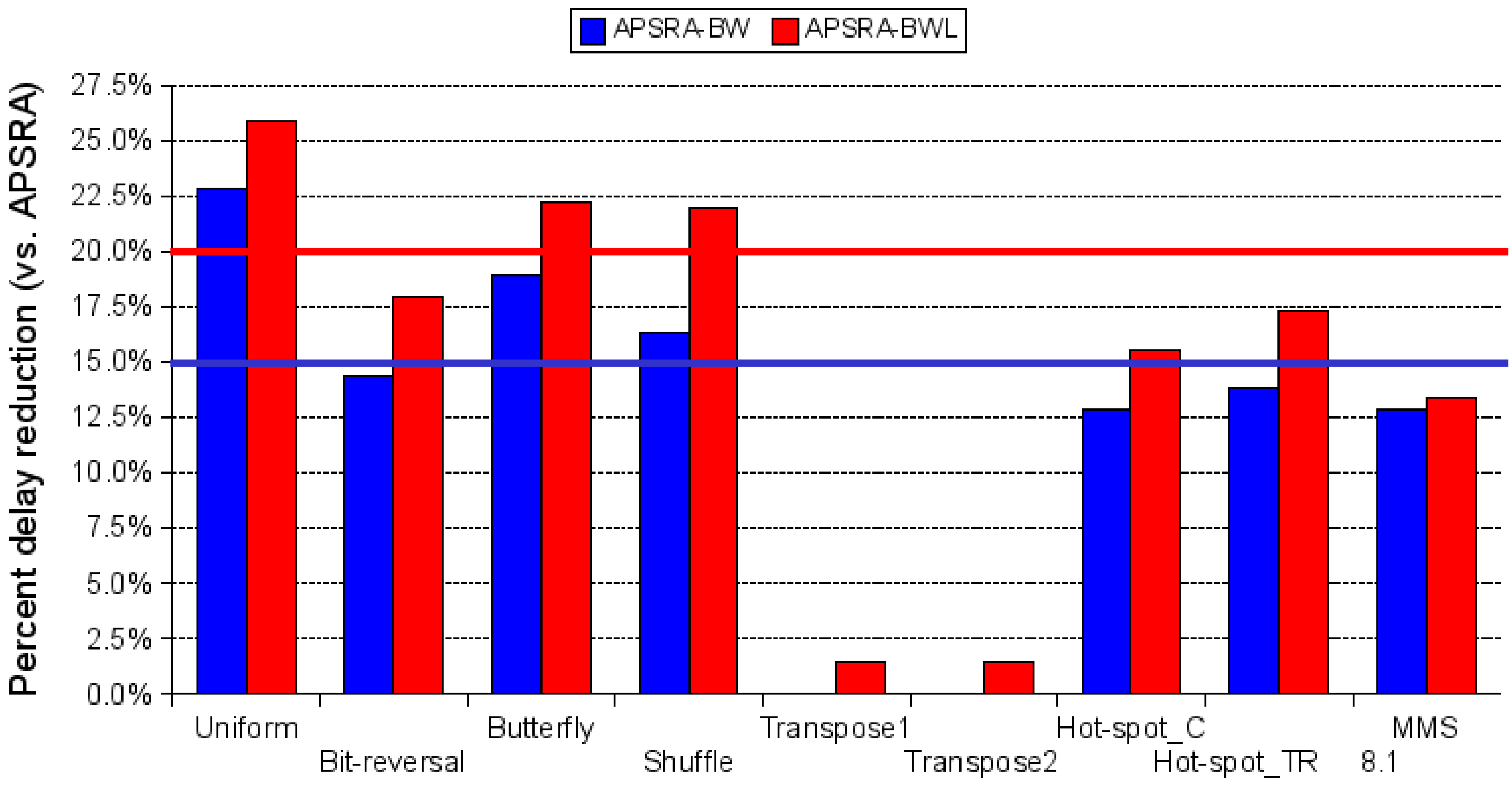
- Aggregate bandwidth per link for a 9x9 mesh-based NoC under *uniform* traffic





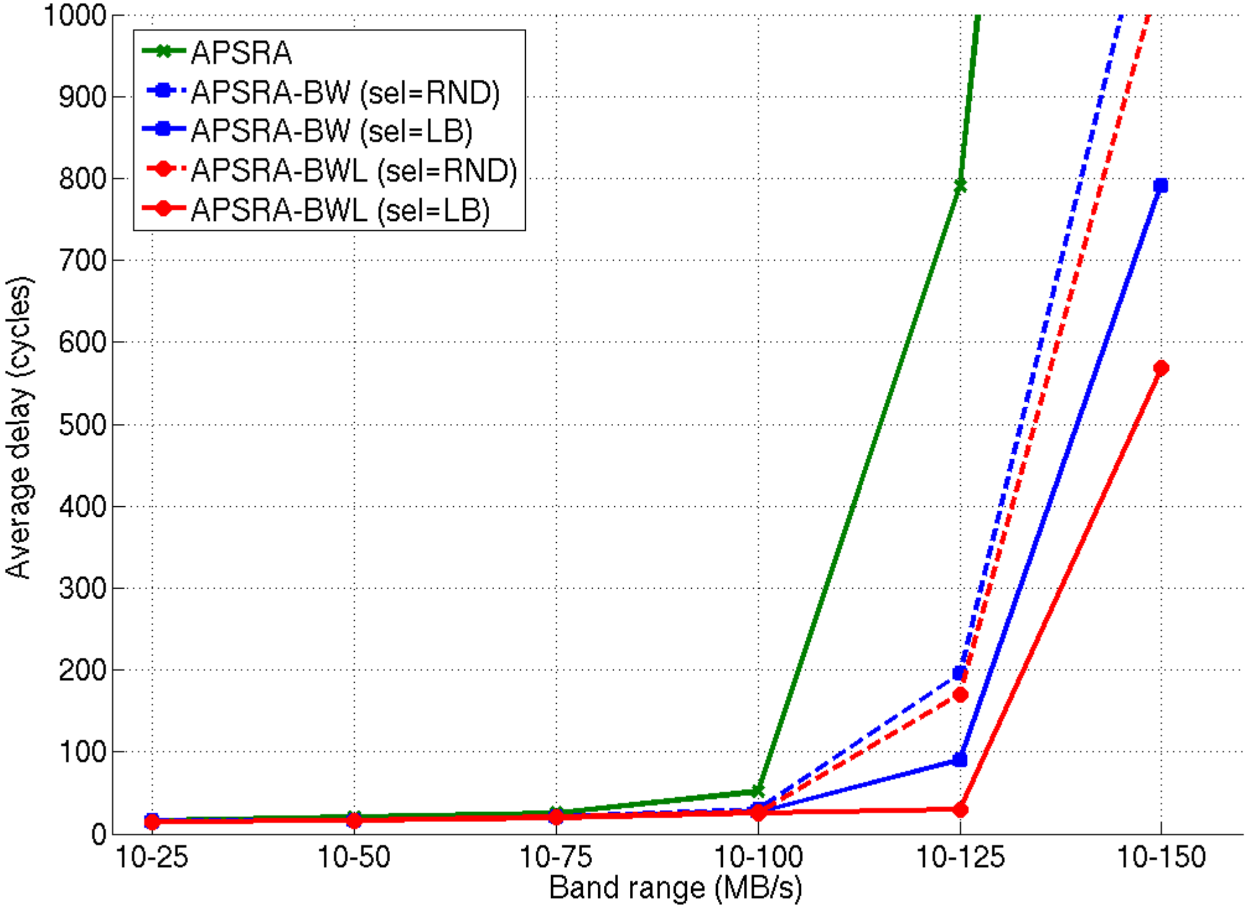
Delay Reduction

- Average delay reduction obtained when APSRA-BW and APSRA-BWL are used taking APSRA as a baseline



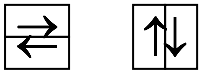
Average Delay Variation

- Average delay variation under uniform traffic for different ranges of communication bandwidth

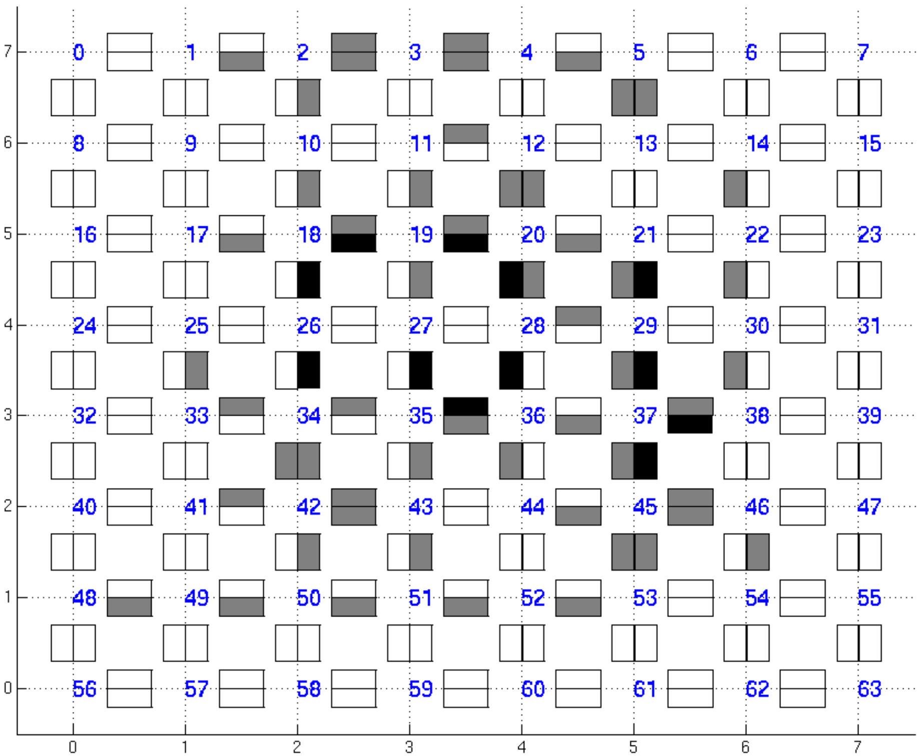


Links Utilization

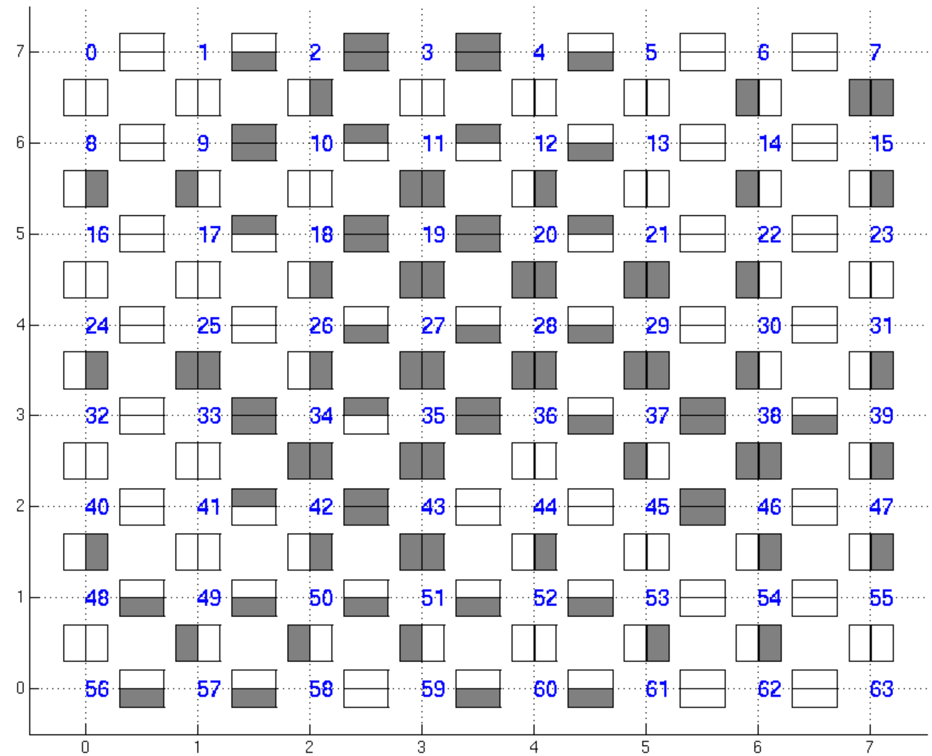
- Links utilization under *uniform* traffic for APSRA and APSRA-BWL



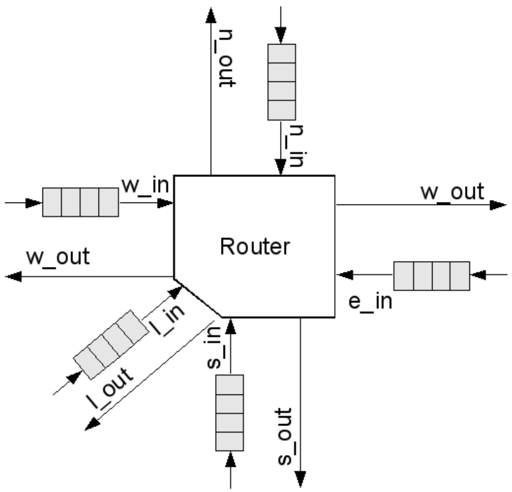
APSRA

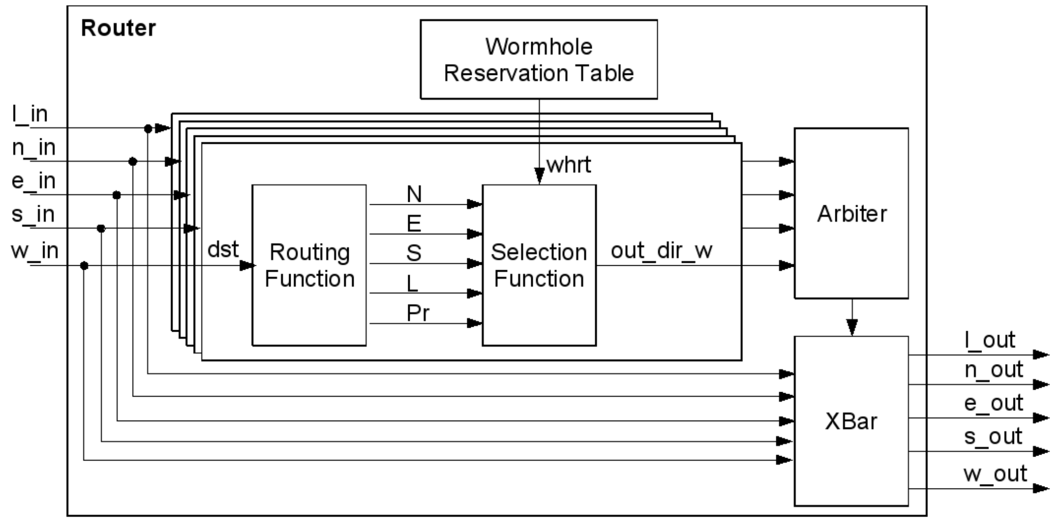


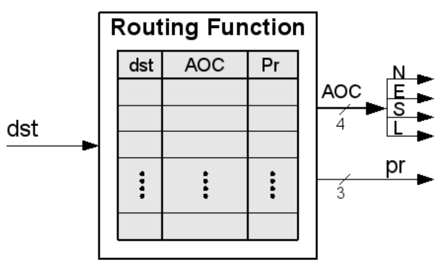
APSRA-BWL

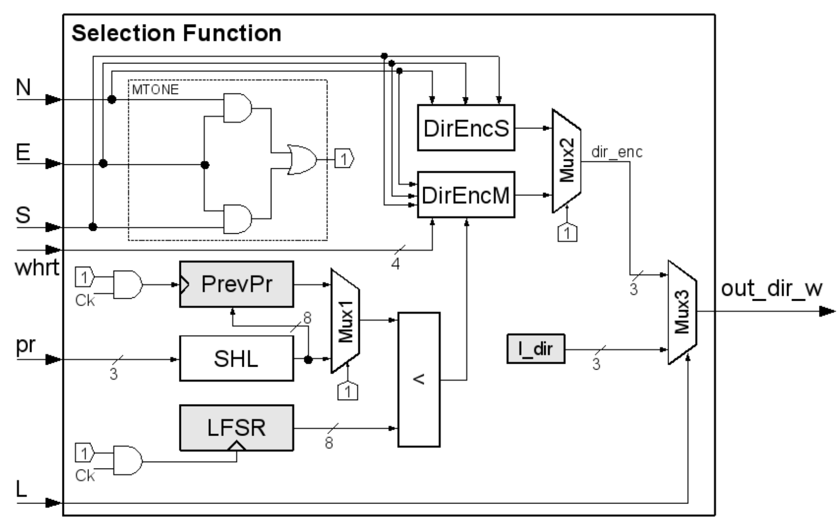


Router Architecture

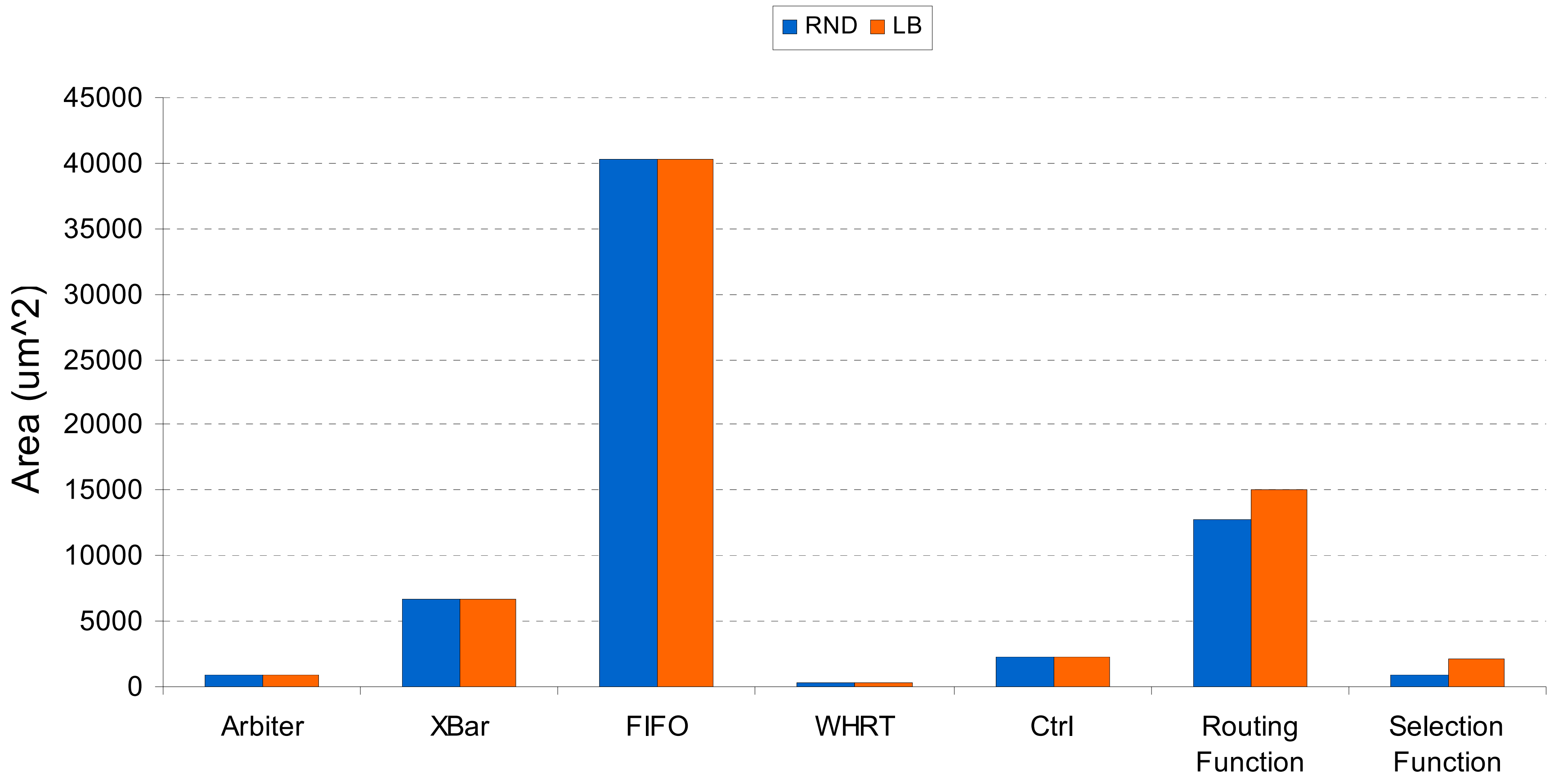








Area Overhead



+5% overall area overhead

Conclusions

- Bandwidth aware routing algorithm
  - ➔ Highly adaptive
  - ➔ Reduces the variation of load in the network links
  - ➔ Ensures that the link bandwidth is not violated
- Evaluate the idea for irregular mesh topology