Microarchitecture of a High-Radix Router

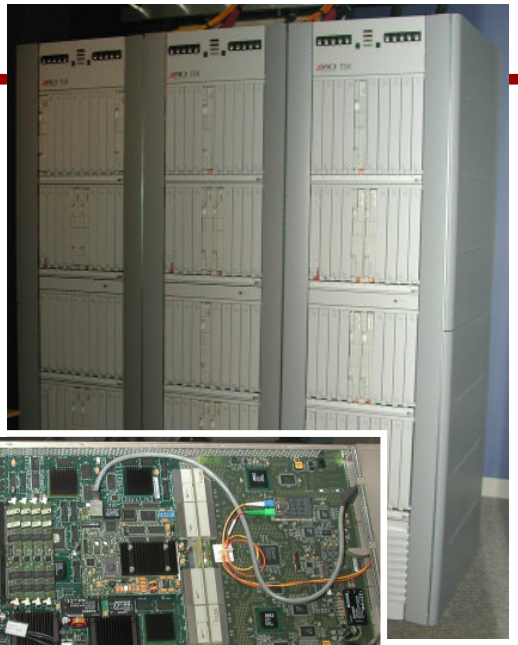
John Kim, William J. Dally, Brian Towles[1], and Amit K. Gupta

Concurrent VLSI Architecture
Stanford University

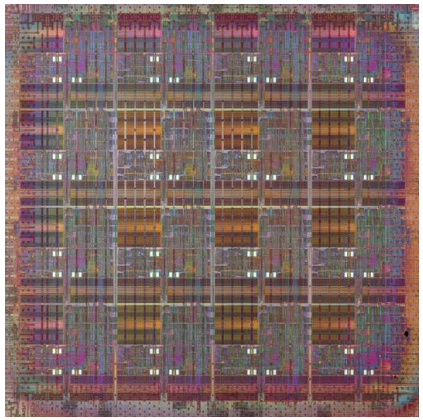
[1]D.E. Shaw
Research & Development

Interconnection Network

Supercomputer networks:
Cray X1

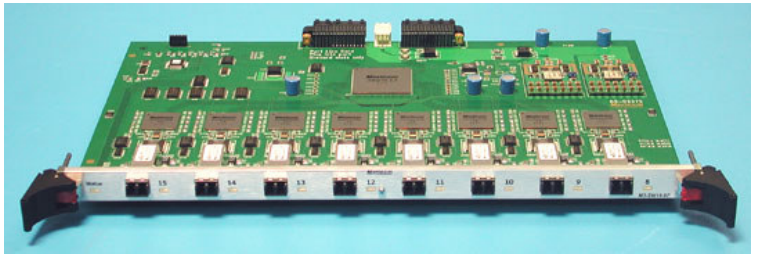


Router Fabrics:
Avici TSR

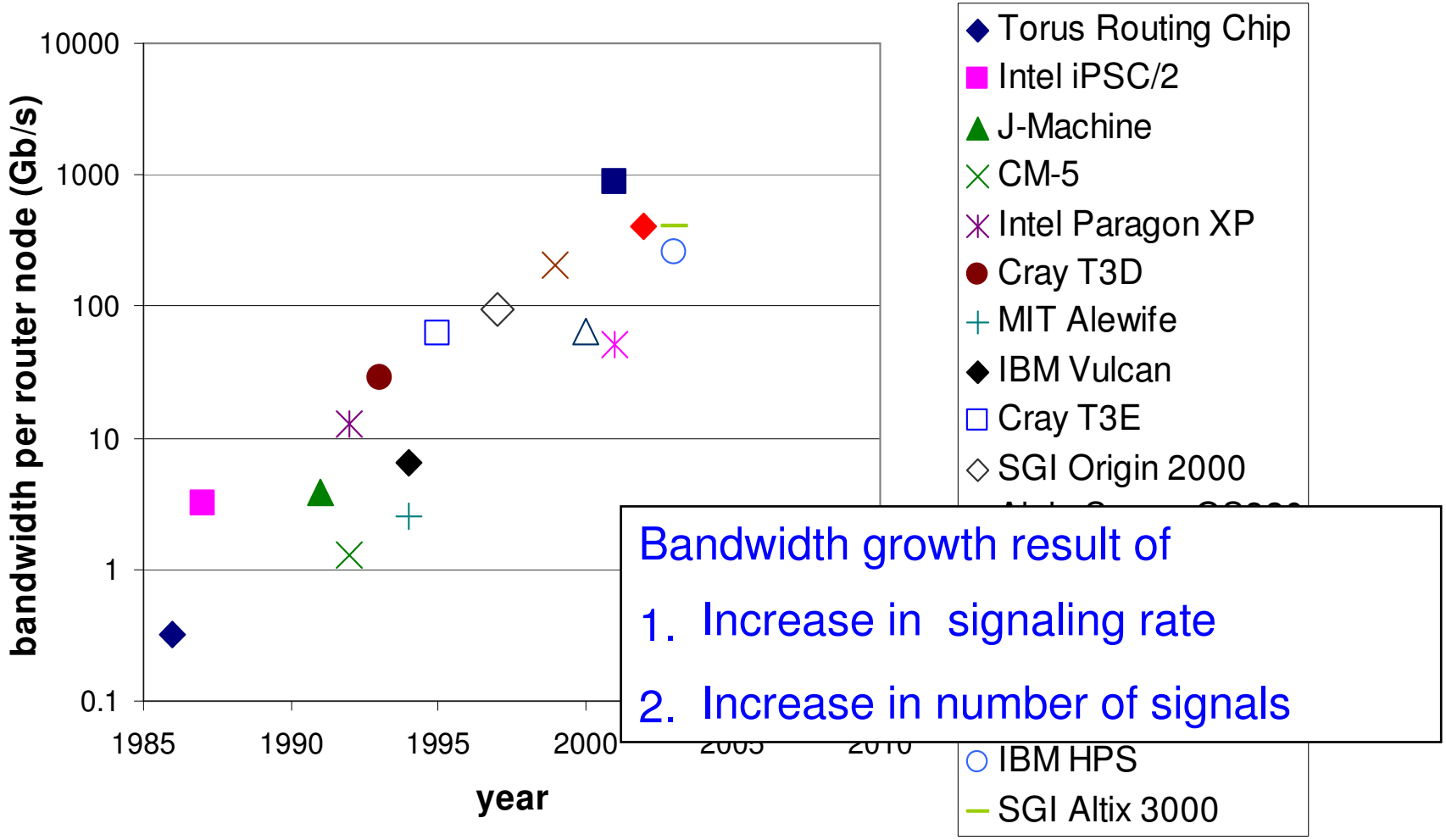


On-chip Networks:
MIT RAW

I/O Interconnects:
Myrinet/Infiniband



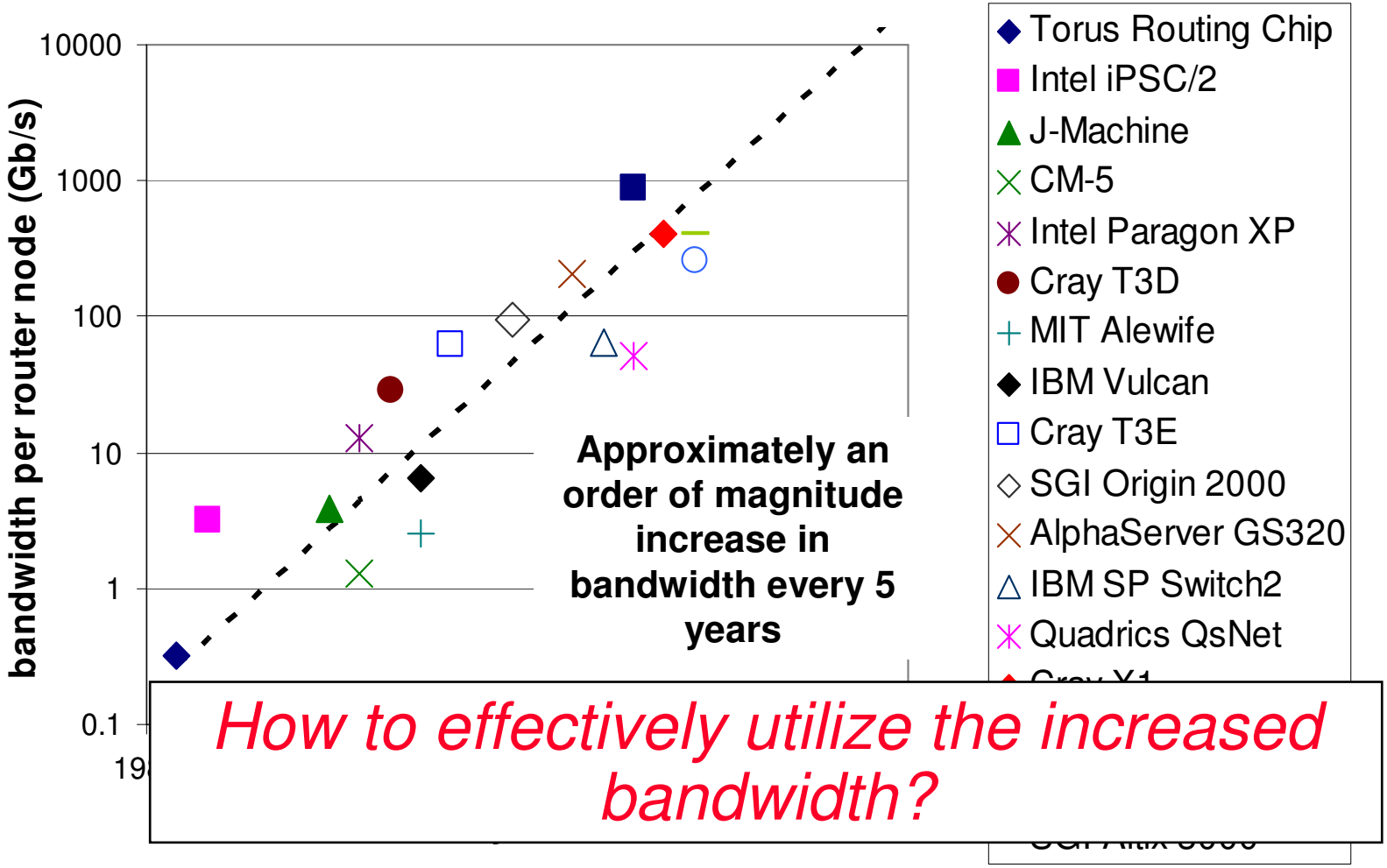
Bandwidth Trend



Bandwidth growth result of

1. Increase in signaling rate
2. Increase in number of signals

Bandwidth Trend



How to effectively utilize the increased bandwidth?

Current Interconnection Networks

- Earlier works [Dally '90, Agarwal '91] suggested "lower-radix" networks
  - Examples: Torus Routing Chip, Cray T3E, SGI Altix 3000
- Current Routers
  - Myrinet : radix-32
  - Quadrics : radix-8
  - IBM SP2 Switch : radix-8

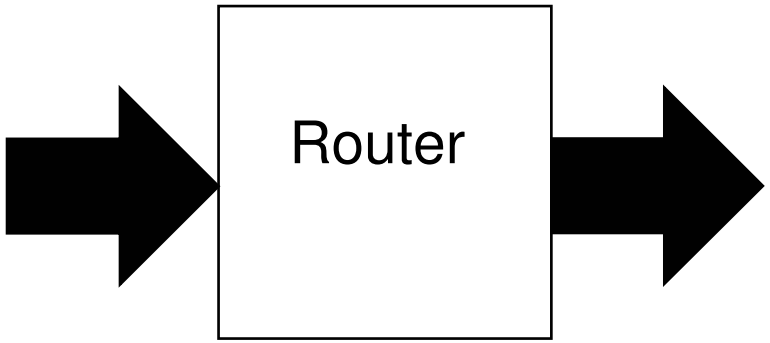
- **Technology trends**
  - **Increasing Bandwidth**
  - Optical signaling

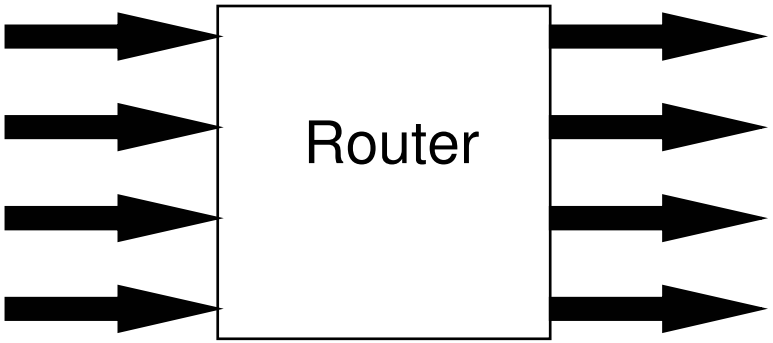
High-Radix Routers *can take better advantage of these technology trends*.

Outline

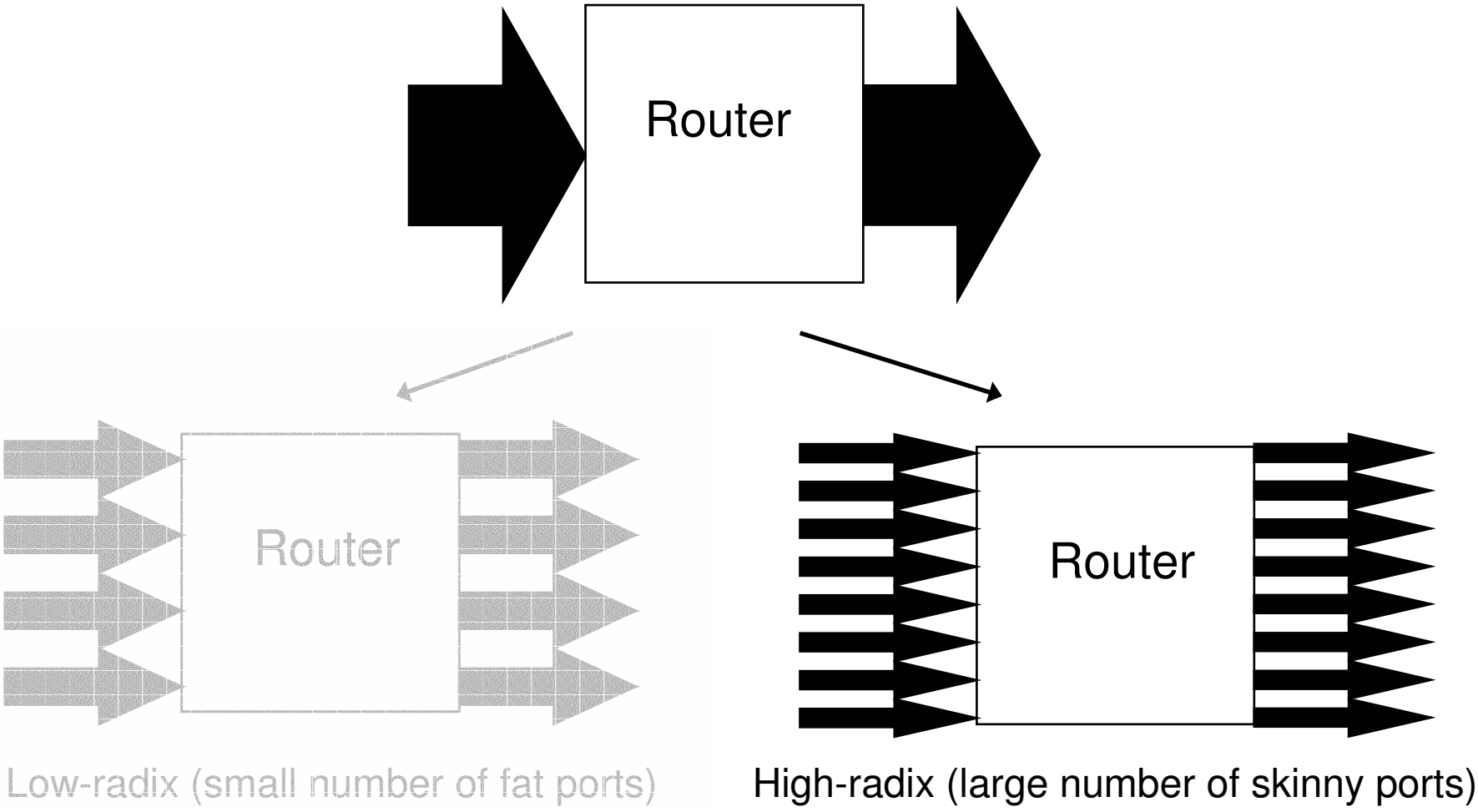
- Technology Trends for High-Radix Router
- Motivation for High-Radix Router
- High-Radix Router Architectures
  - Baseline Architecture
  - Fully Buffered Crossbar
  - Hierarchical Crossbar
- Conclusion & Future Work

High-Radix Router





High-Radix Router



Latency in a Network

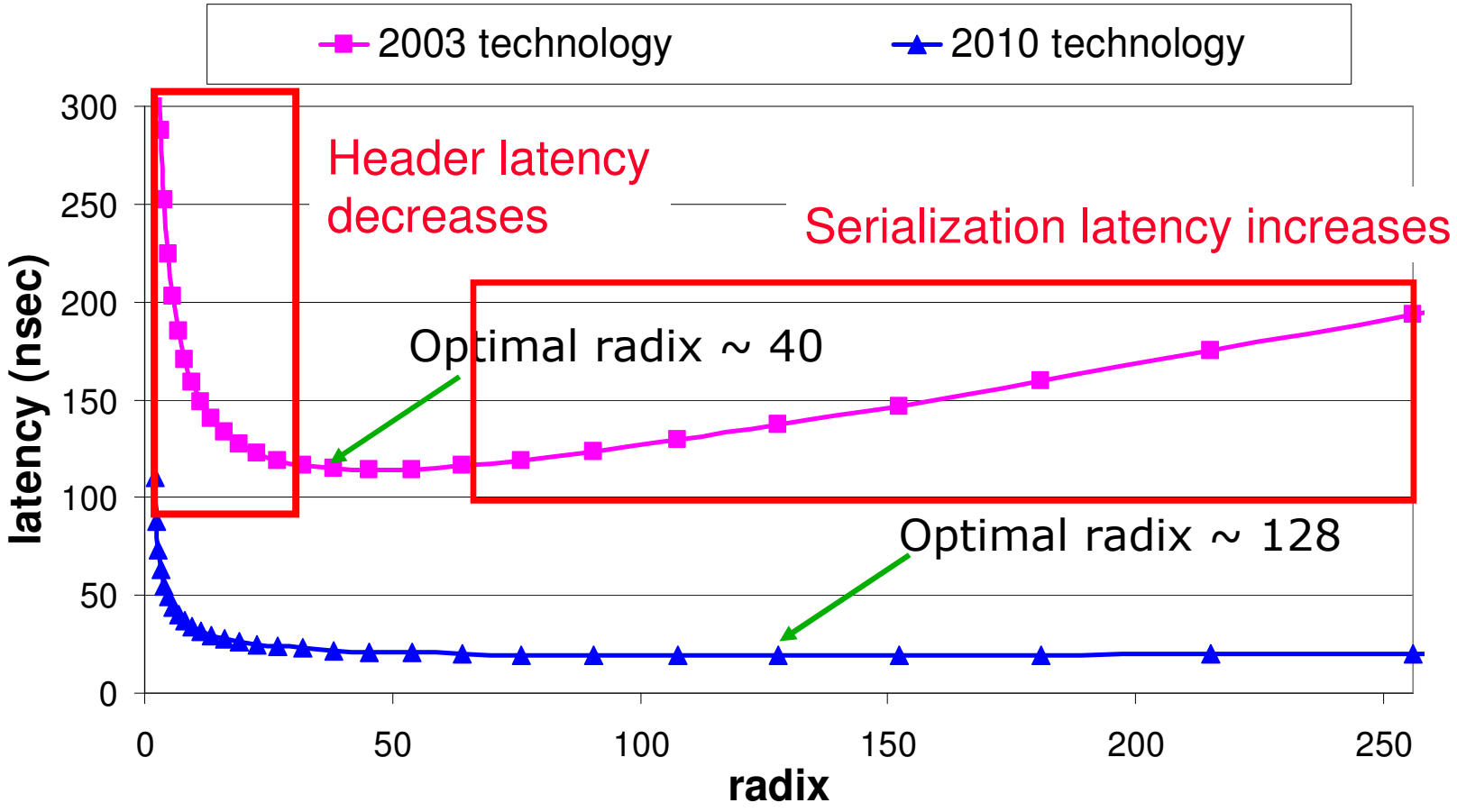
Latency = Header Latency + Serialization Latency

$= Ht_r + L/b$

$= 2t_r \log_k N + 2kL / B$

where k = radix
B = total Bandwidth
N = # of nodes
L = message size

Latency vs. Radix



Determining Optimal Radix

Latency = Header Latency + Serialization Latency

$= H\,t_r + L / b$

$= 2t_r \log_k N + 2kL / B$

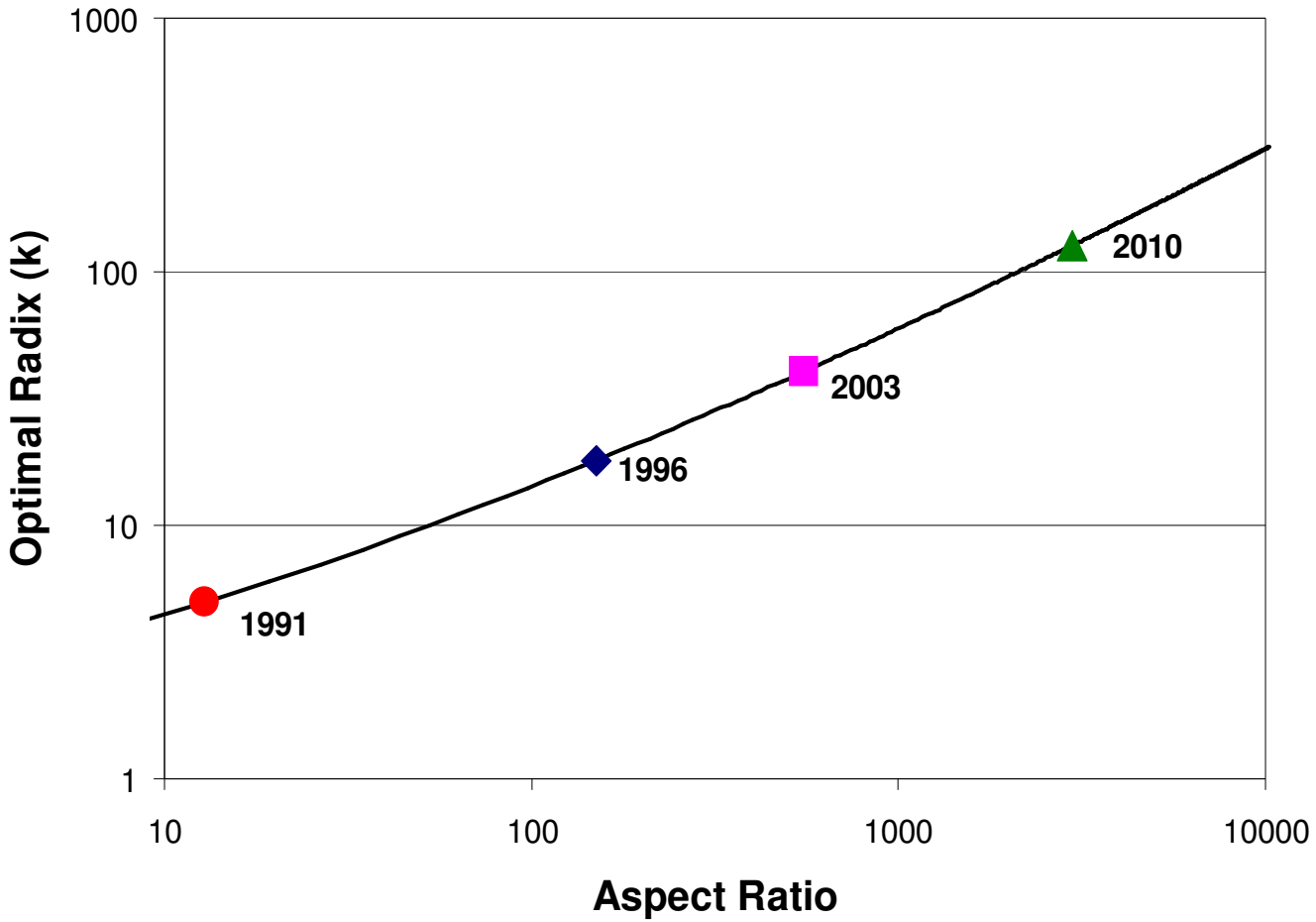
where k = radix
B = total Bandwidth
N = # of nodes
L = message size

Optimal radix

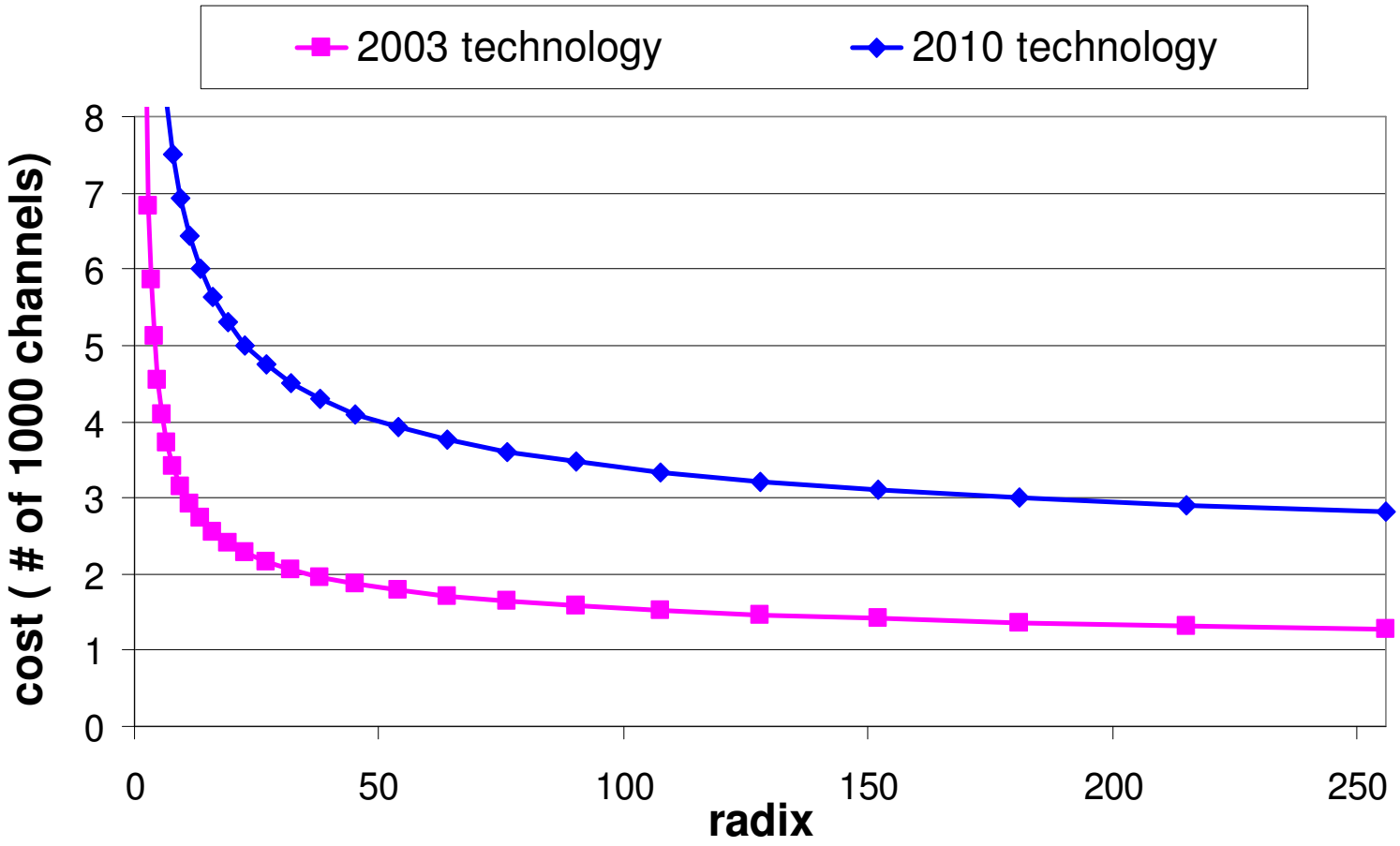
➔ $k \log_2 k = (B\,t_r \log N) / L$

$=$ ***Aspect Ratio***

Higher Aspect Ratio, Higher Optimal Radix



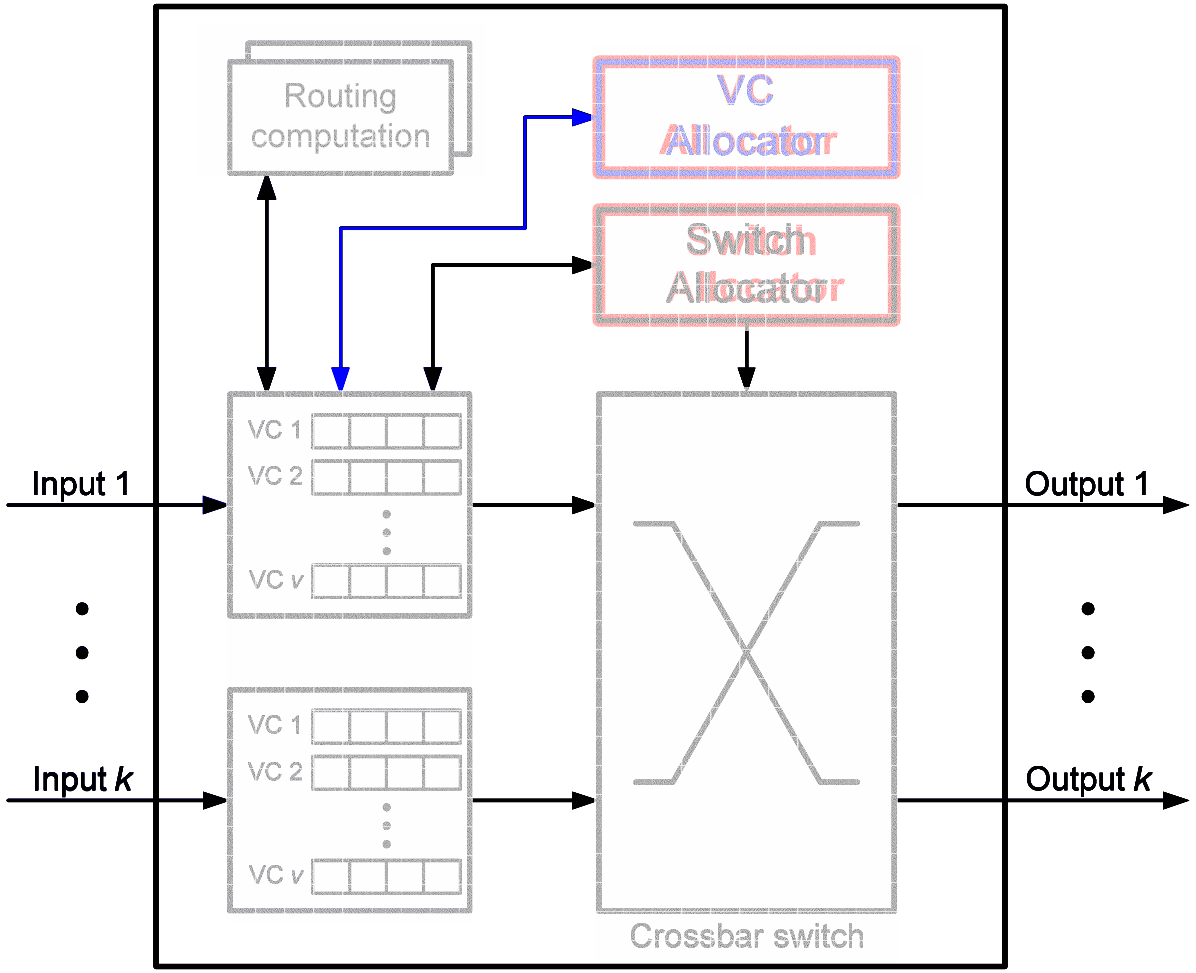
Higher Radix, Lower Cost



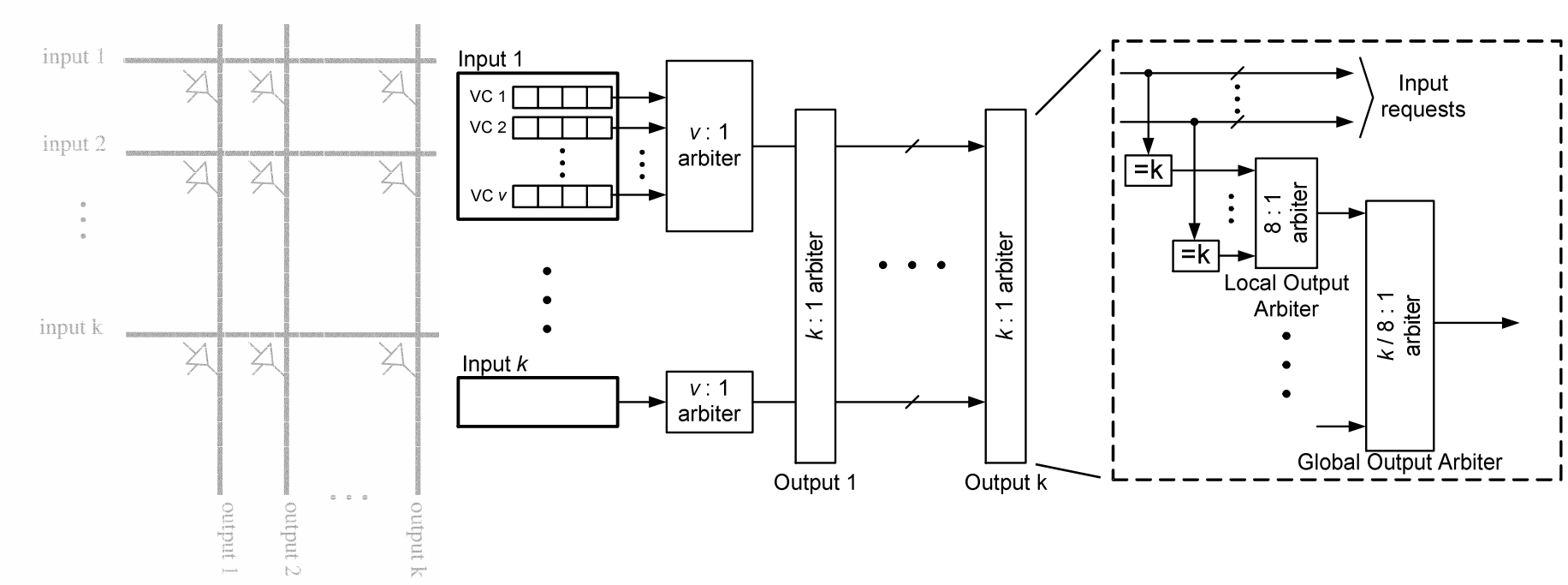
Outline

- Technology Trend for High-Radix Router
- Motivation for High-Radix Router
- High-Radix Router Architectures
  - Baseline Architecture
  - Fully Buffered Crossbar
  - Hierarchical Crossbar
- Conclusion & Future Work

Virtual Channel Router Architecture



High-Radix Switch Architectures (I)



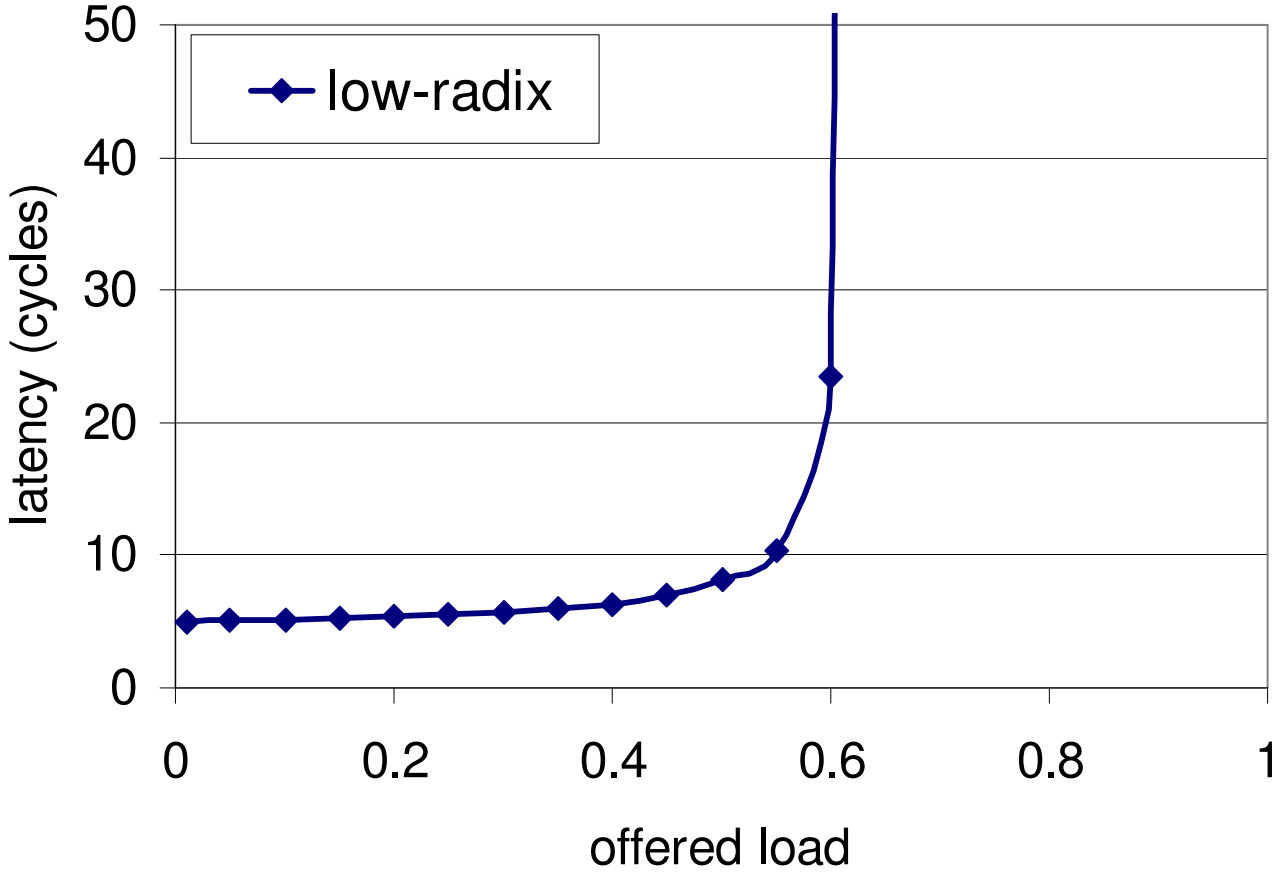
(a) Baseline design

Simulation Methodology

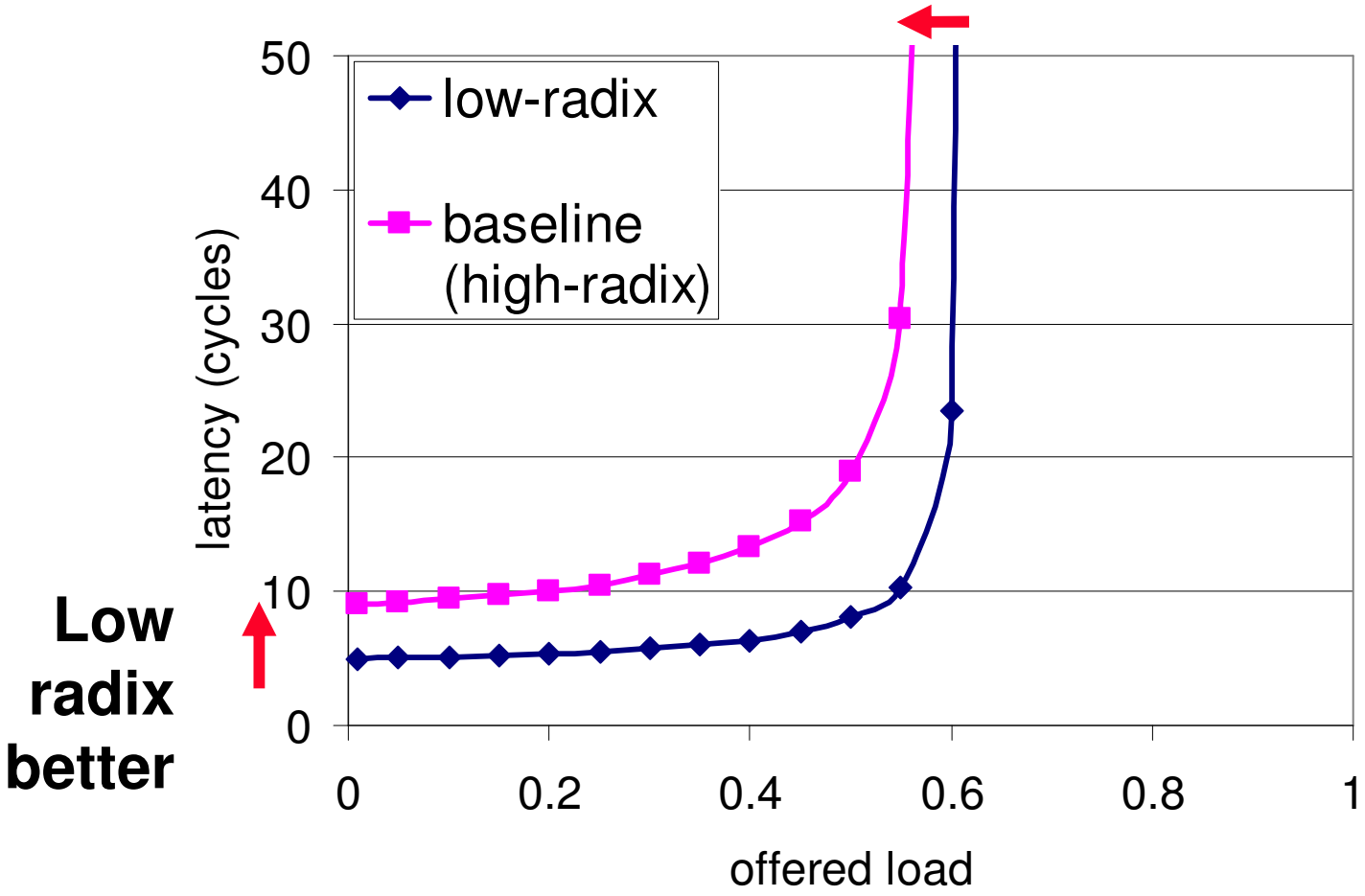
- Open-loop simulation done with steady-state measurement
- Output switch arbitration required two cycles
- Wire delay included in the pipeline
- Uniform random traffic pattern
- Each packet was assumed to be a single flit
- Radix=64 router evaluated with 4 VCs per input

- Metrics – latency & throughput
- Cost – area

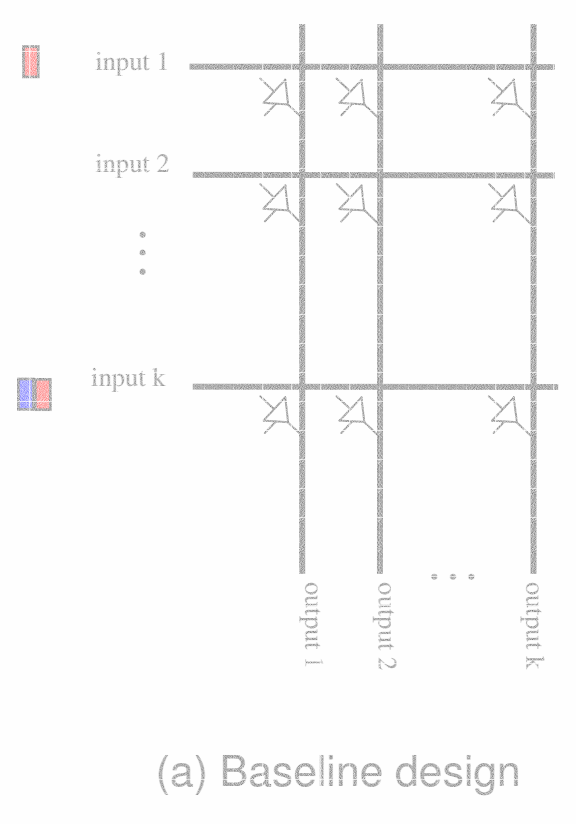
Baseline Performance Evaluation



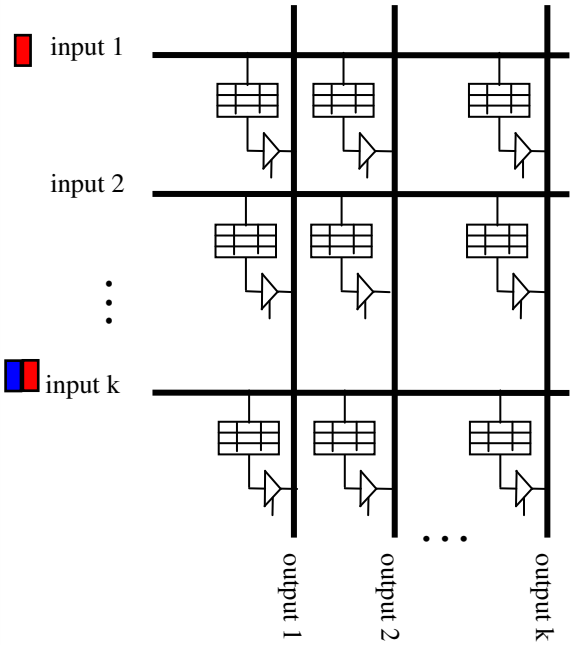
Baseline Performance Evaluation



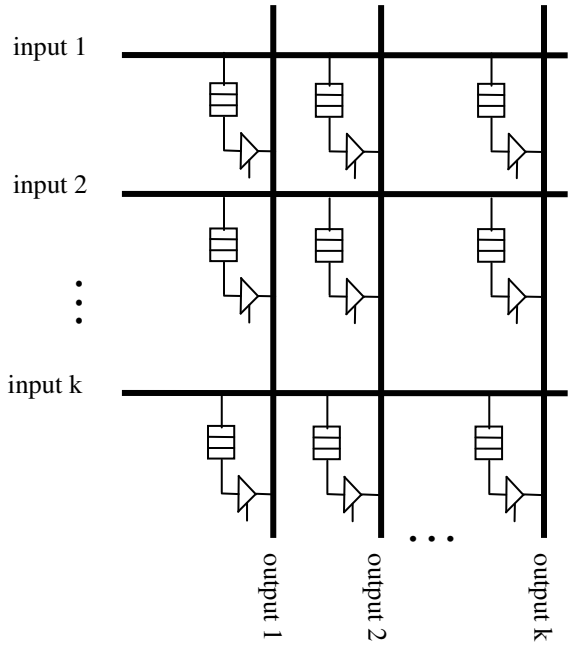
High-Radix Switch Architectures (II)



(a) Baseline design

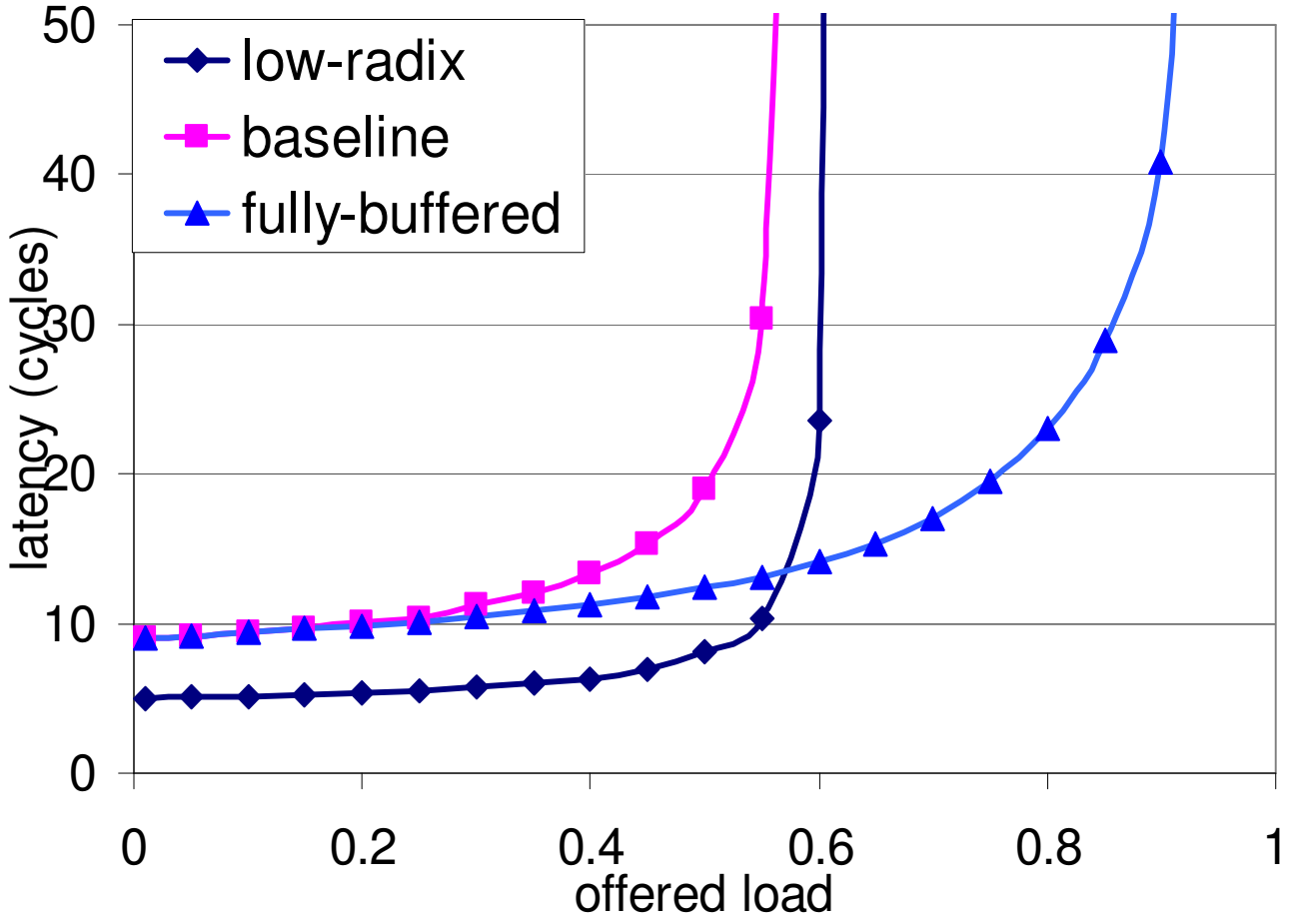


(b) Fully buffered crossbar

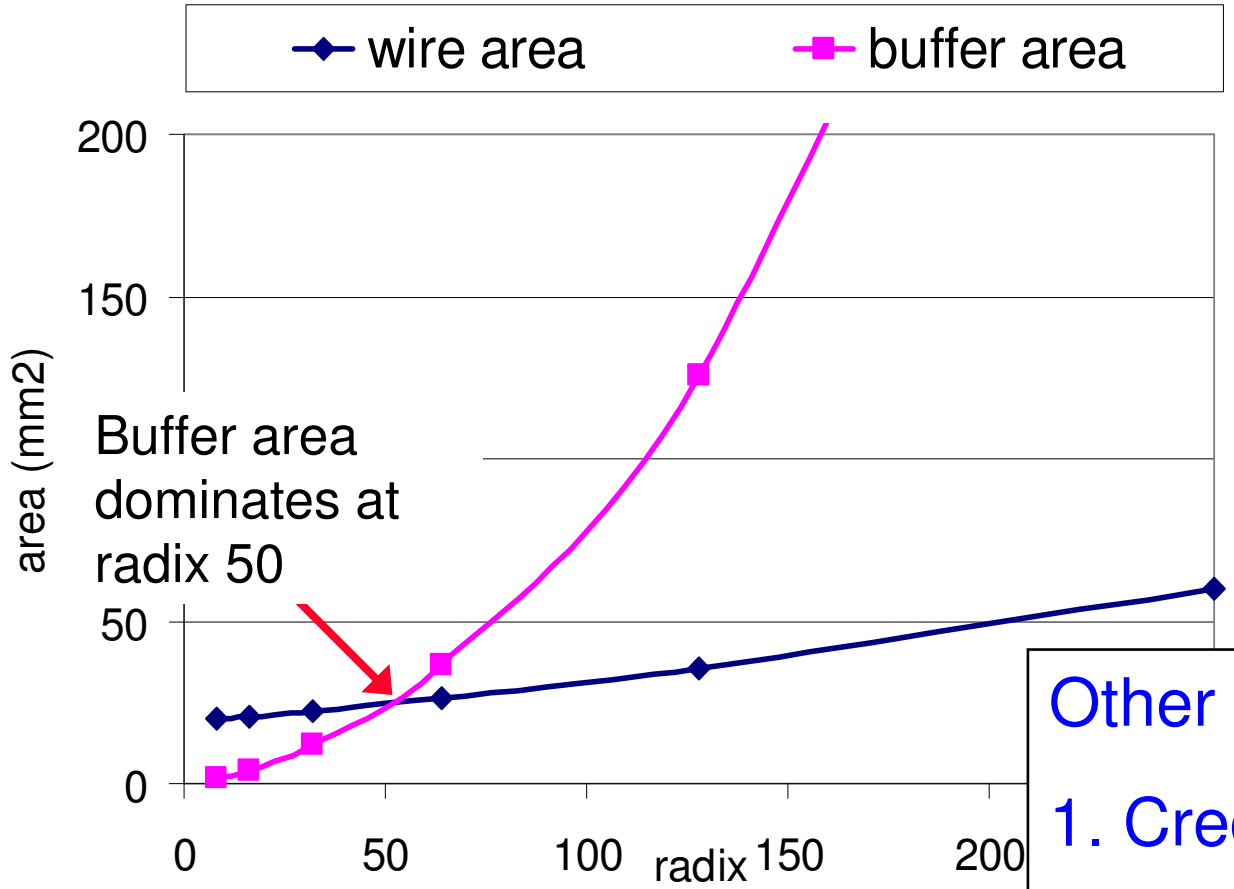


Fully buffered crossbar with shared buffers

Fully Buffered Crossbar Provides High Performance but ....



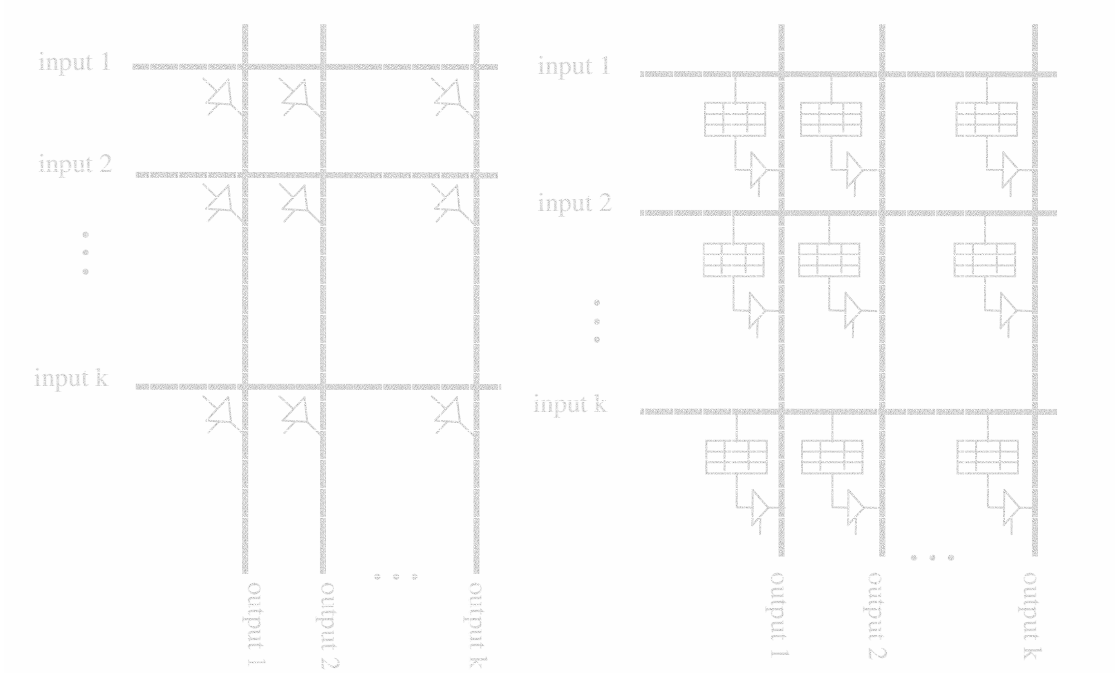
... Becomes Costly to build



Other issues:

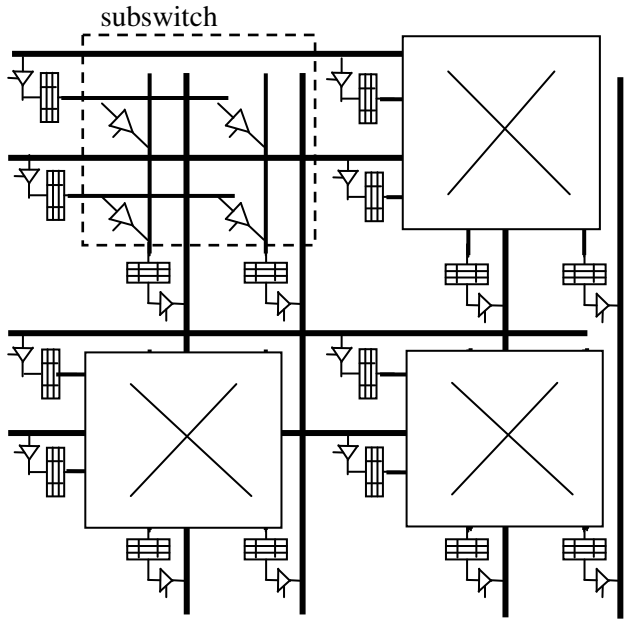
1. Credit Information
2. Adaptive Routing

High-Radix Switch Architectures (III)



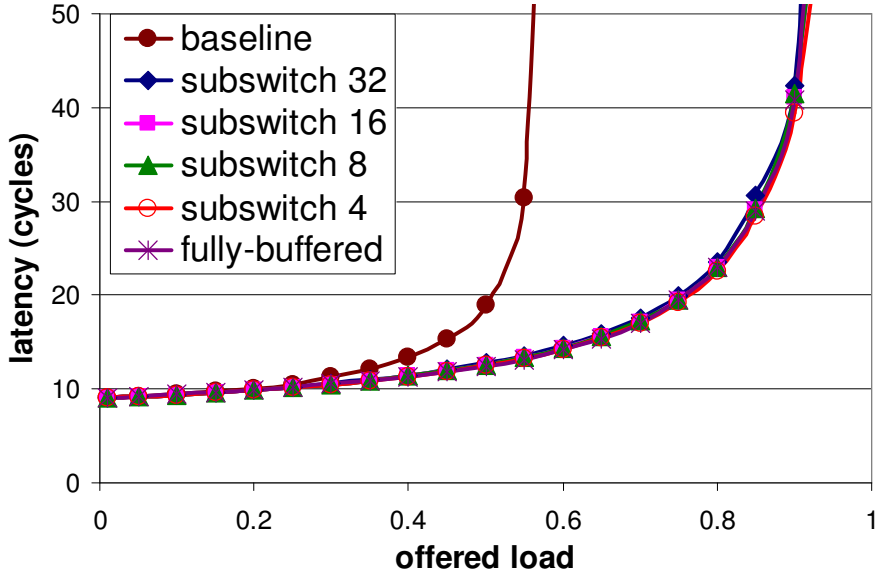
(a) Baseline design

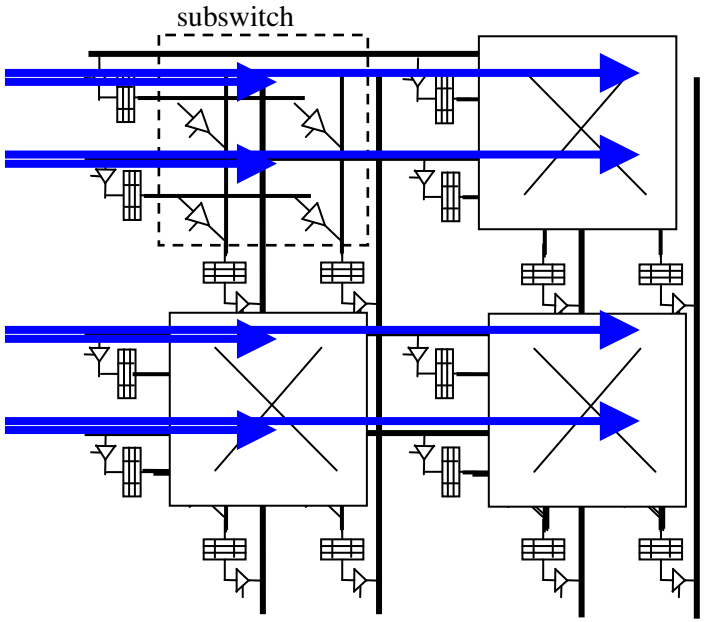
(b) Fully buffered crossbar



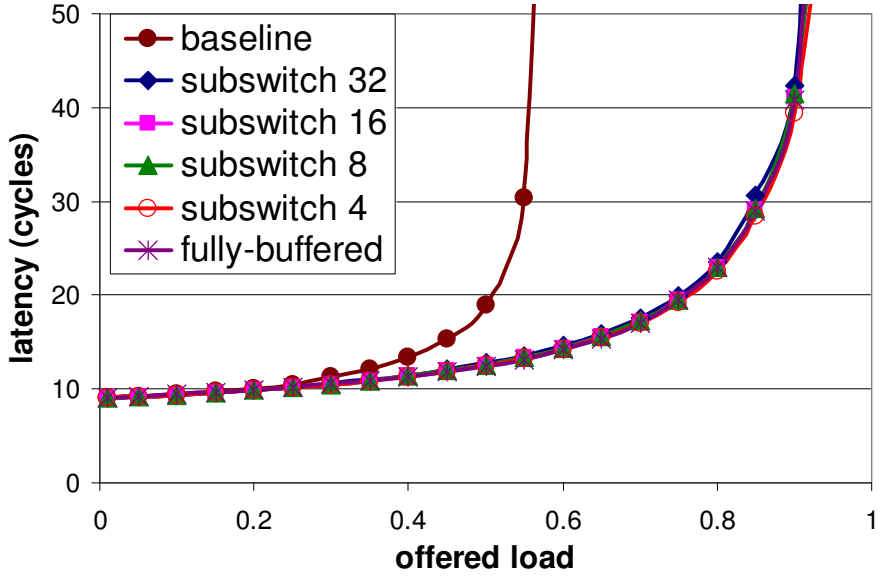
(c) Hierarchical crossbar

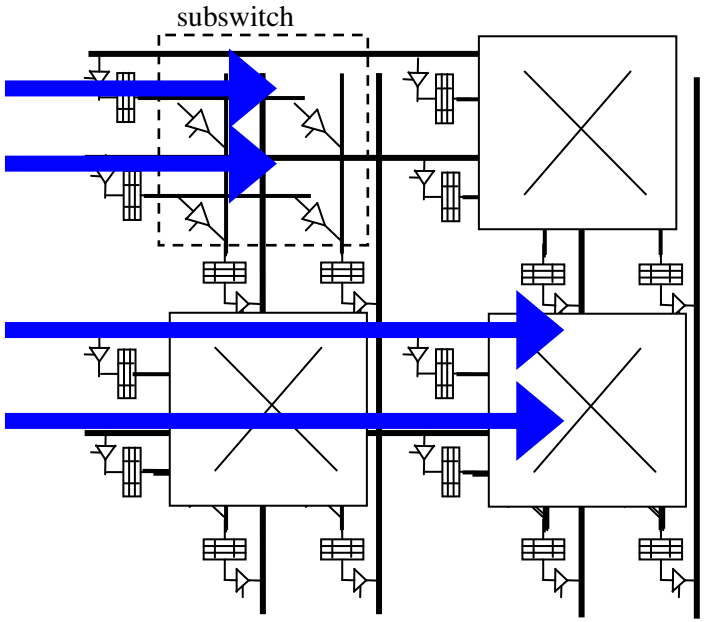
Hierarchical Crossbar Performance on Uniform Random Traffic



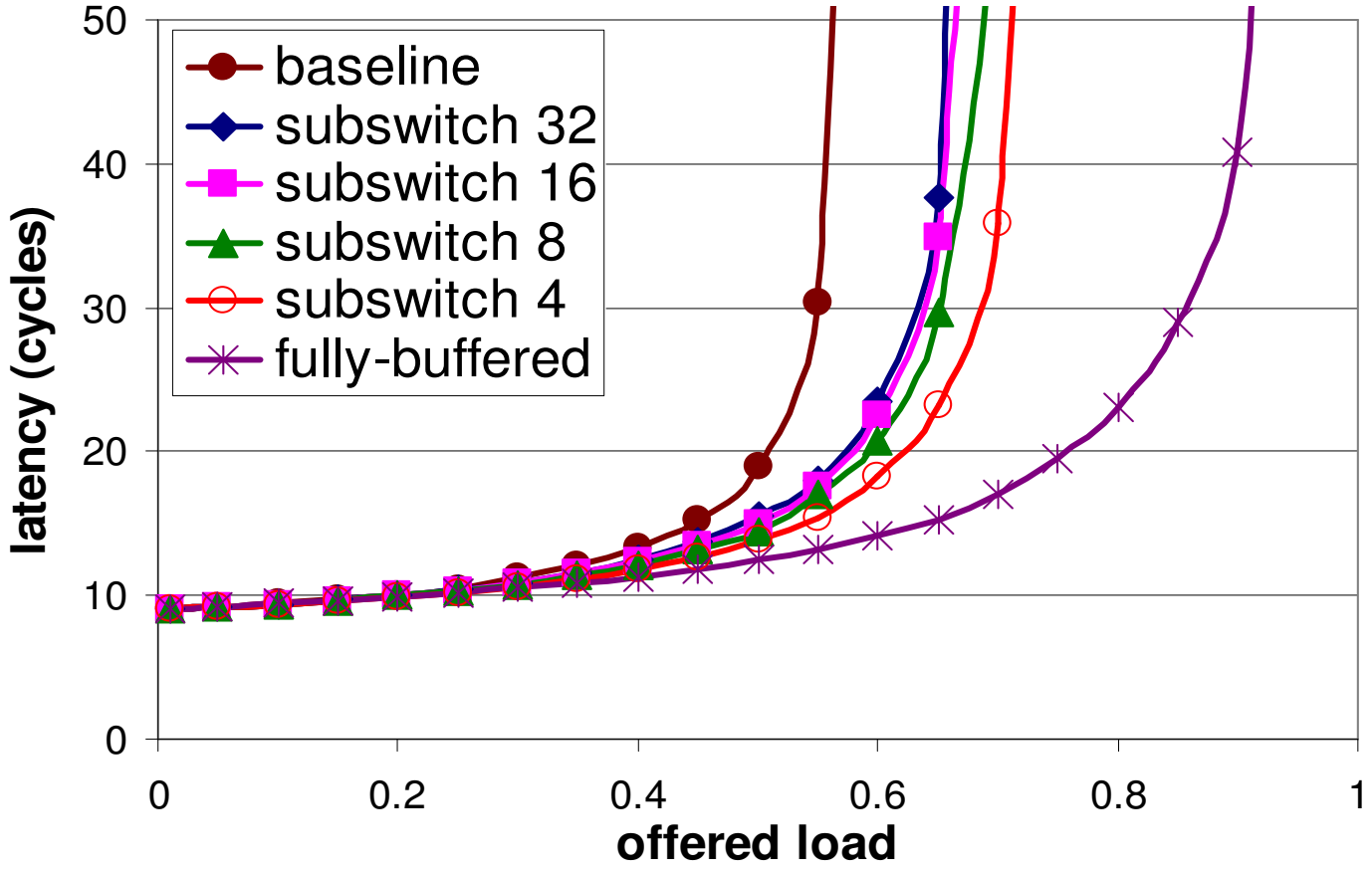


Worst-case Traffic Pattern

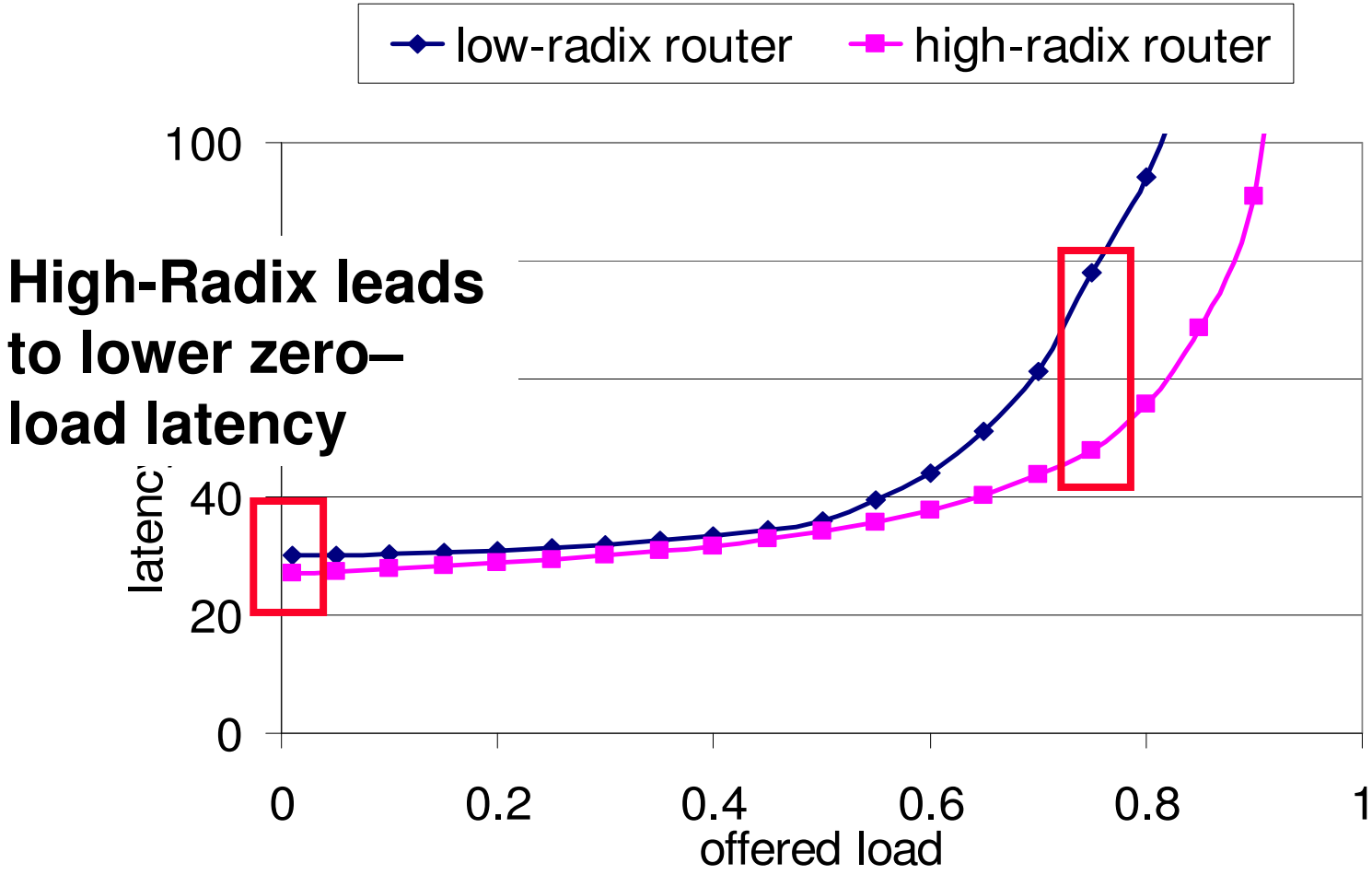




Worst Case Performance Comparison



4k Node Network Performance



High-Radix Router Microarchitecture

- In building high-radix router, there are scaling issues as the number of ports increase
- Baseline design provides poor performance
- Fully buffered crossbar leads to high performance but a costly design
- Hierarchical crossbar provides a feasible design with minimal loss in performance

Conclusion

- Performance of digital systems is often limited by the interconnection network
- Increasing on-chip bandwidth can be more efficiently used by a ***high-radix*** routers.
- High-radix lead to lower cost and lower latency networks
- A high-radix router requires a different architecture
- A hierarchical organization makes high-radix routers feasible

Future Work

- Optimal topology for high-radix network
  - Clos topology suited for high-radix routers
- Routing- How to adaptively route in a high-radix router?
- Flow control with high-radix router
- Simulating a larger network
- Microarchitecture - Alternative to crossbars : multistage switch organizations, network on-chip (torus, mesh, etc.)

Thank you

Questions?