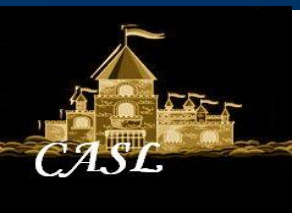


Centralized Buffer Router - A Low Latency, Low Power Router for High Radix NOCs

Syed Minhaj Hassan and Sudhakar Yalamanchili

Center for Research on Experimental Computer Systems
School of Electrical and Computer Engineering
Georgia Institute of Technology

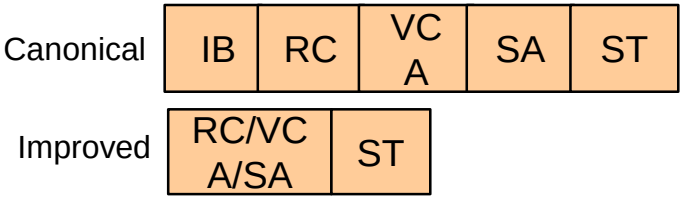
Sponsors: National Science Foundation, Sandia National Laboratories

Overview

- Motivation
- Router Design
- Optimizations
- Deadlock Avoidance
- Results

Router Design Space

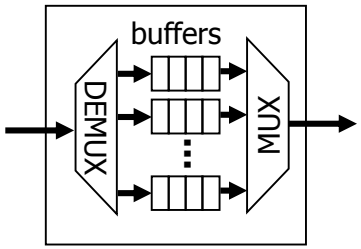
Pipeline Stages



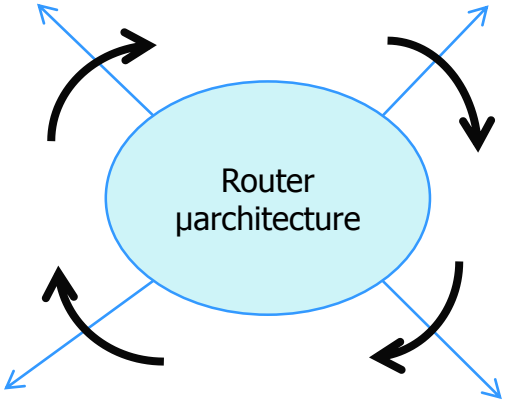
- High throughput
- Reduce Latency
- Ideal – Single cycle

Buffering

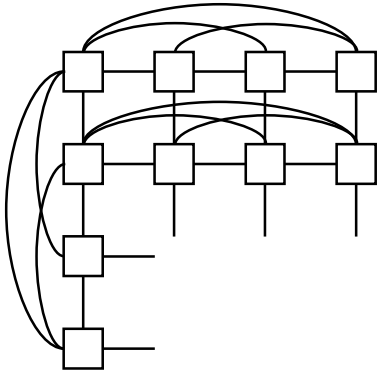
Multiple VCs, Multiple Virtual Net.



- 64 node mesh: (100 – 400KB)
- Power & area consumers
- Ideal – Bufferless

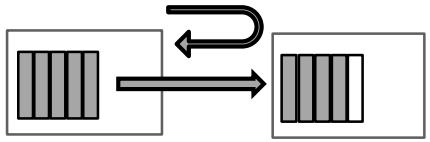


High Radix



- Reduced hop count
- Long wires → ↑buffers
- Ideal – buffer space decoupled from radix

Flow Control

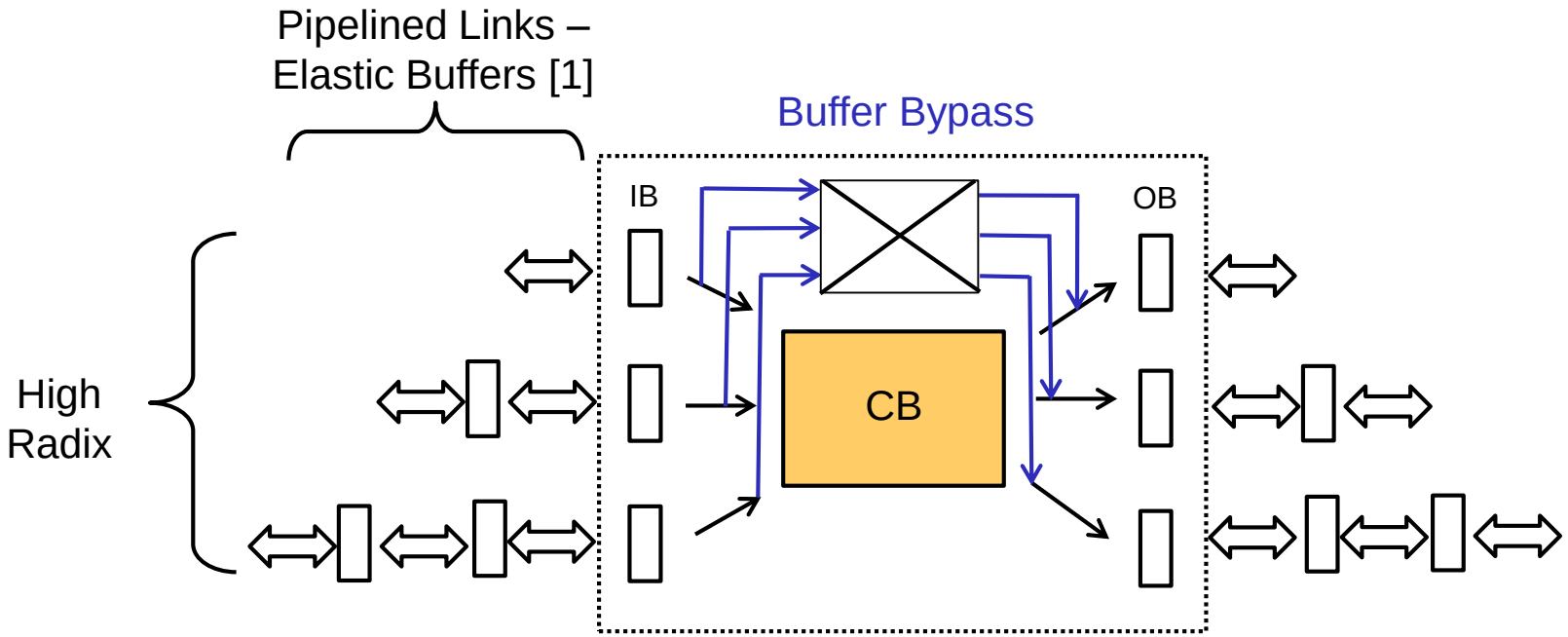


- Remove pipeline bubbles & high link utilization
- Buffer size = $\mathcal{F}$(RTT latency)
- Ideal – buffers size decoupled from wire length

Overview

- Motivation
- Router Design
- Optimizations
- Deadlock Avoidance
- Results

Centralized Buffer Router: High Level Overview



- Central buffers reduce buffer space dependency on radix.
- Elastic Buffer (EB) links to decouple buffer size from wire length.
- Buffer bypass to reduce latency at low load
- Bubble flow control using central buffers for deadlock avoidance.

[1] Michelogiannakis, G. Elastic Buffer Flow Control for On-Chip Networks, HPCA 2009

Centralized-Elastic-Bubble (CEB) Router (An Instance of Central Buffer Routers)

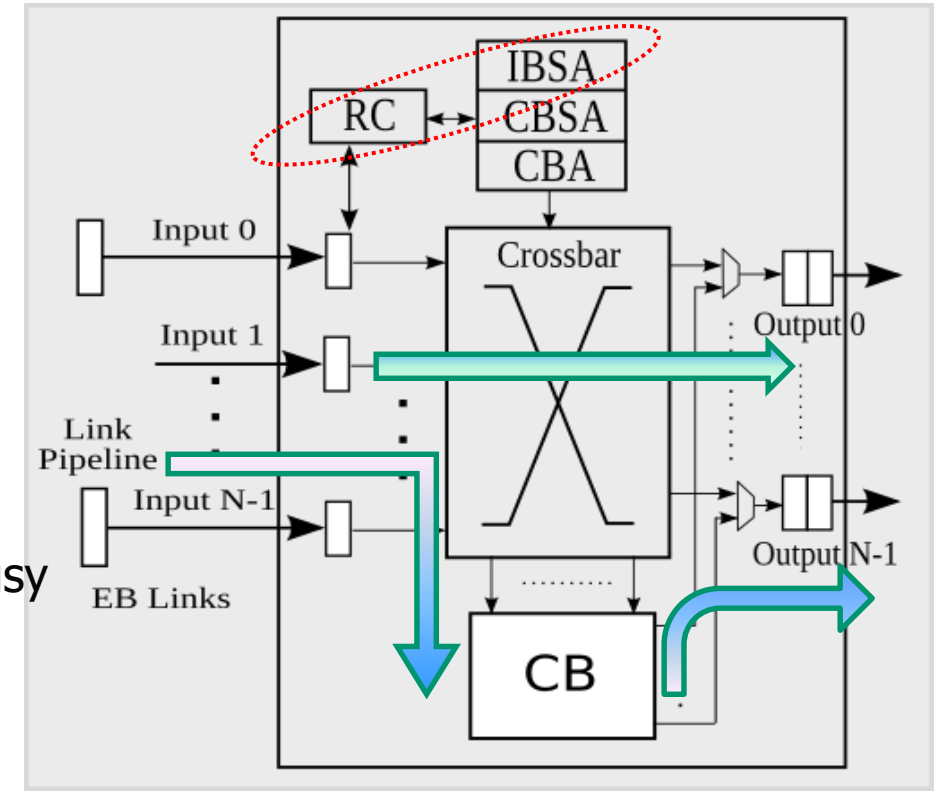
- Two Paths
  - Bypass
    - Used at low loads

IBSA / RC | ST

    - Look-ahead routing
  - Buffered
    - If corresponding output port is busy

CBA / RC | CBT | CBSA | CBOT

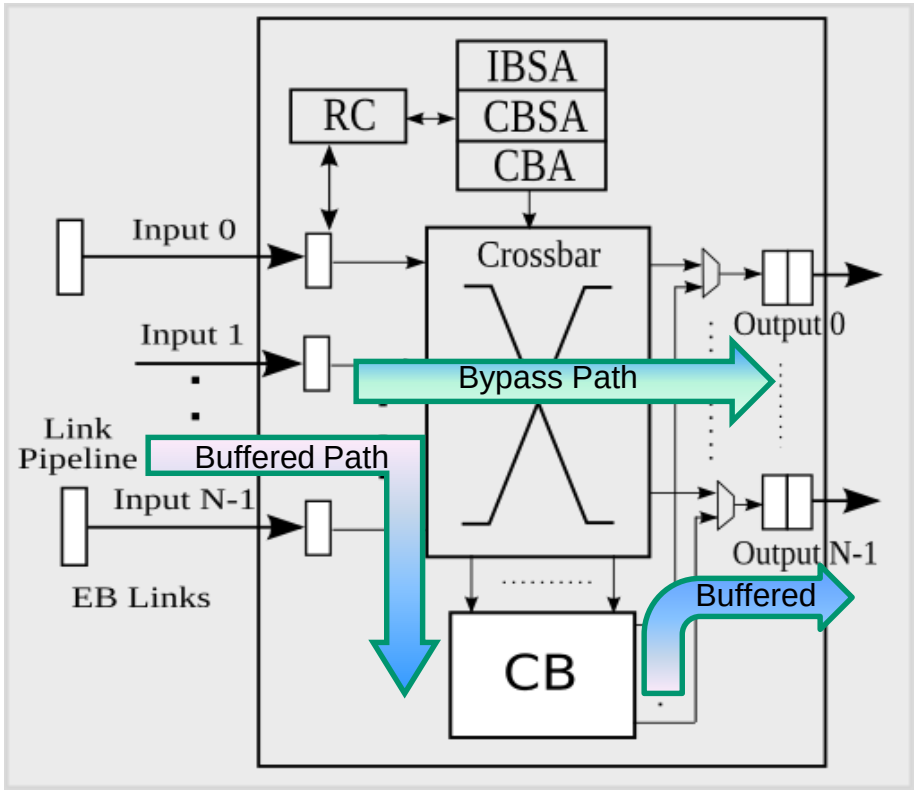
- CB is shared among different output ports
  - 1 head/tail ptr for each output port.
  - Flits have to move through crossbar to enter corresponding port of CB.



IBSA – Allocation of Input Buffers (IB) to OB
CBSA – Allocation of Central Buffers (CB) to OB
CBA – Allocation of IB to CB
ST – Traversal from IB-OB
CBOT – Traversal from CB-OB
CBT – Traversal from IB-CB

Centralized-Elastic-Bubble (CEB) Router

- 3 simultaneous allocation operations
  - CBSA given higher priority than IBSA
  - IBSA given higher priority than CBA
- Once a flit is allocated an OB or CB, the whole packet follows the same path.



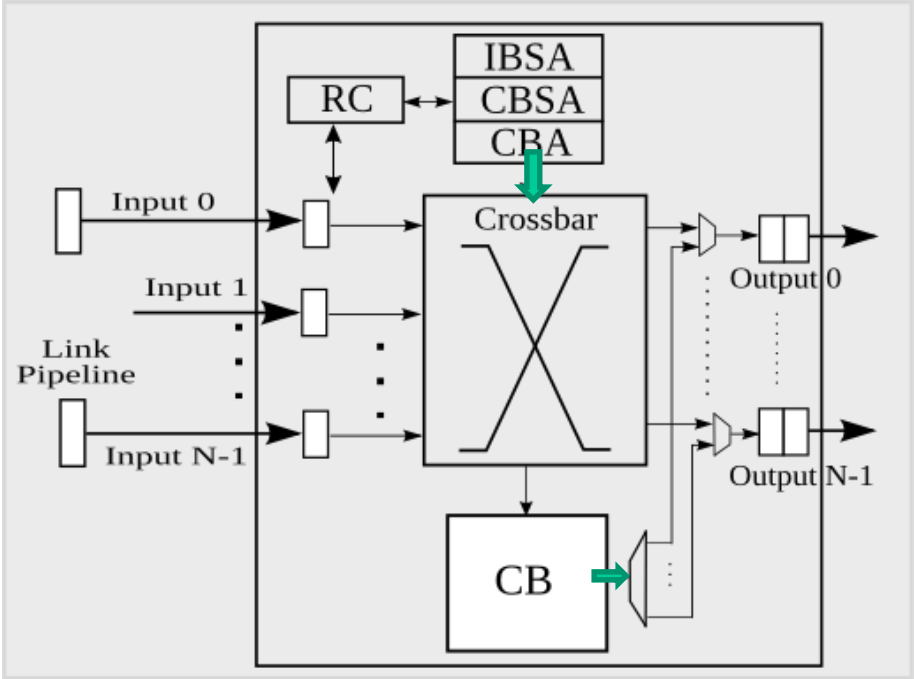
IBSA – Allocation of IB -OB
CBSA – Allocation of CB -OB
CBA – Allocation of IB -CB

Overview

- Motivation
- Router Design
- Optimizations
- Deadlock Avoidance
- Results

Optimizations

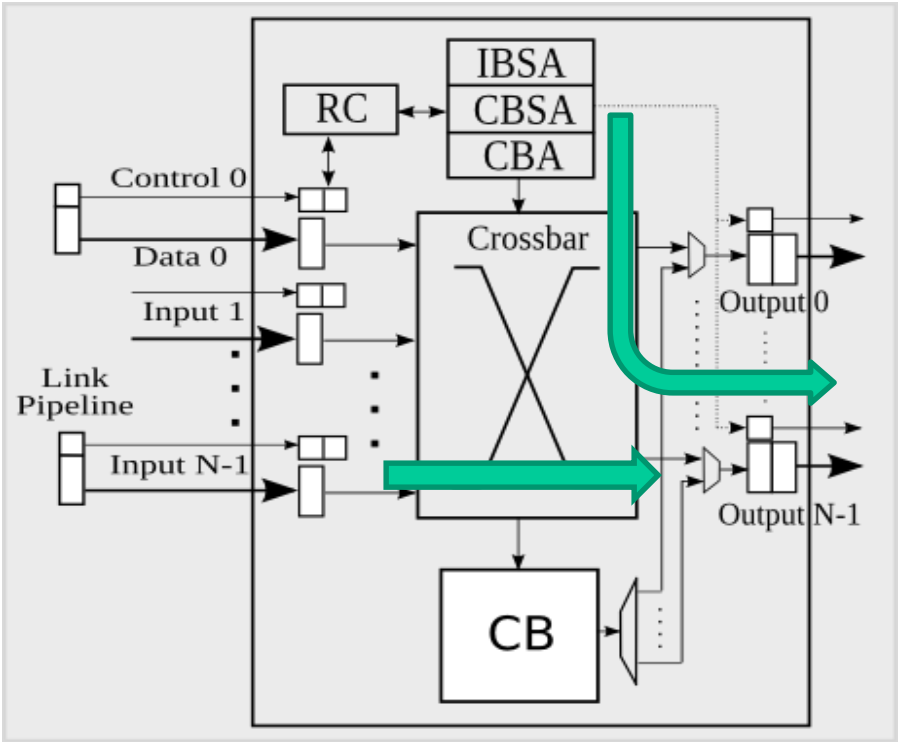
1. Reduce #CB ports
   - Serialize the input and output of the CB.
   - CBSA & IBSA has to perform this serialization
   - Reduces crossbar size to *port* x *port+1*.

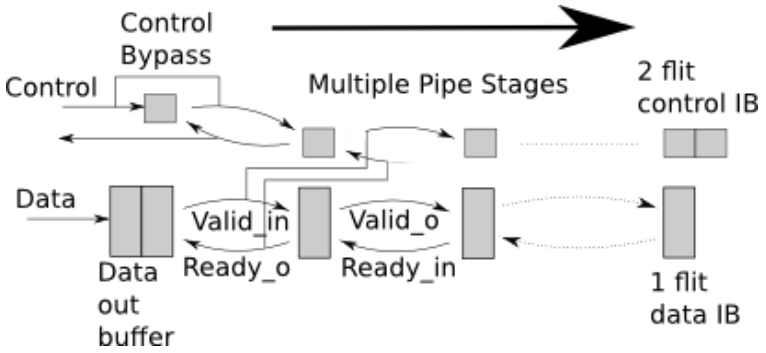


2. Power Gating of CB (Coarse grained)
   - Turned on when a head flit waits for more than X (=500) cycles.
   - Turned off when empty and minimum wait time has passed.

Optimizations (cont.)

3. Lookahead Switch Allocation (SA)
   - Reduce bypass path latency to 1 cycle (buffered to 3 cycles)
   - Perform IBSA/CBA one cycle ahead in parallel with IB.
   - Requires control information 1 cycle ahead of data.
   - Split the control and data path in the EB link.
   - Send control 1 cycle early during the ST cycle.
   - Need 1-cycle ahead guarantee.



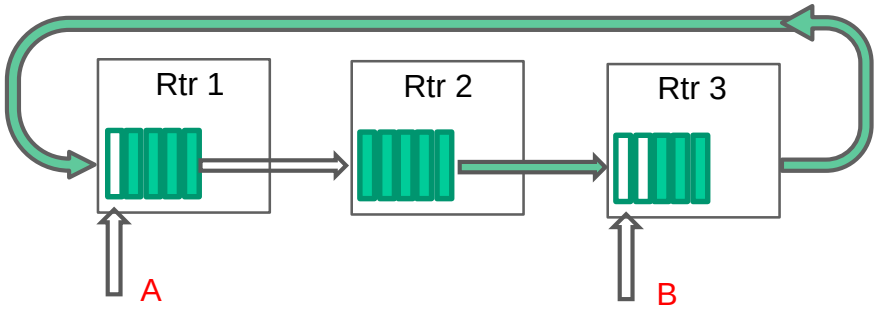


Overview

- Motivation
- Router Design
- Optimizations
- Deadlock Avoidance
- Results

Deadlock Avoidance (DA)

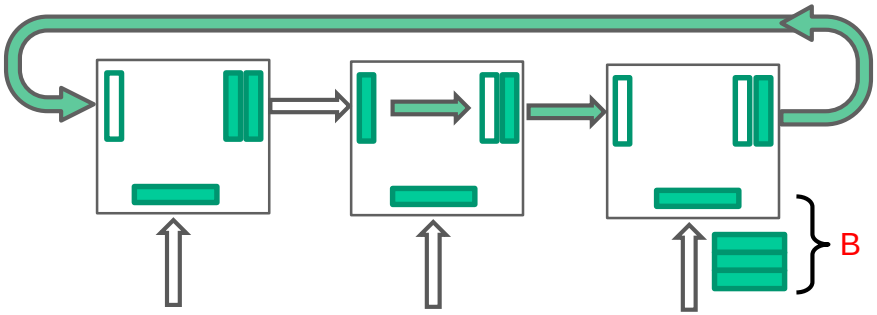
- EB links create dependency among VCs with the link
- Bubble flow control avoid deadlocks with single VC
  - ensures at least one packet free buffer (bubble) remains in each ring.

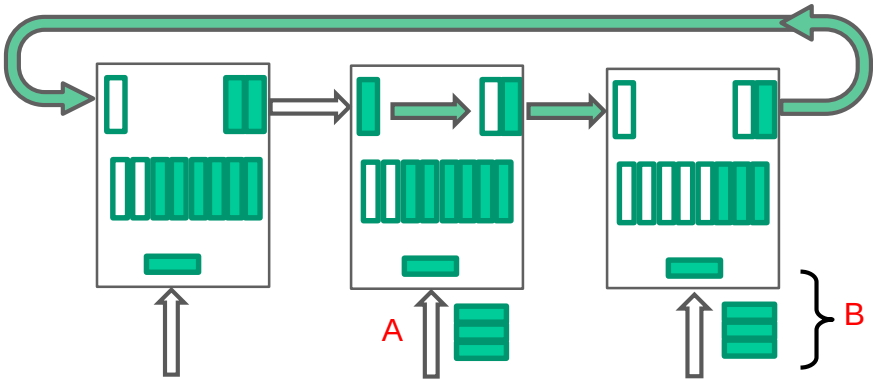


- DA Conditions
  - Every ring must have a bubble
  - External packets entering the ring cannot destroy the bubble.
  - If a ring has a bubble, packets cannot wait indefinitely on any other condition.
- However, this only works at packet level
  - Output channel is allocated for the whole packet.

Deadlock Avoidance (DA) (cont.)

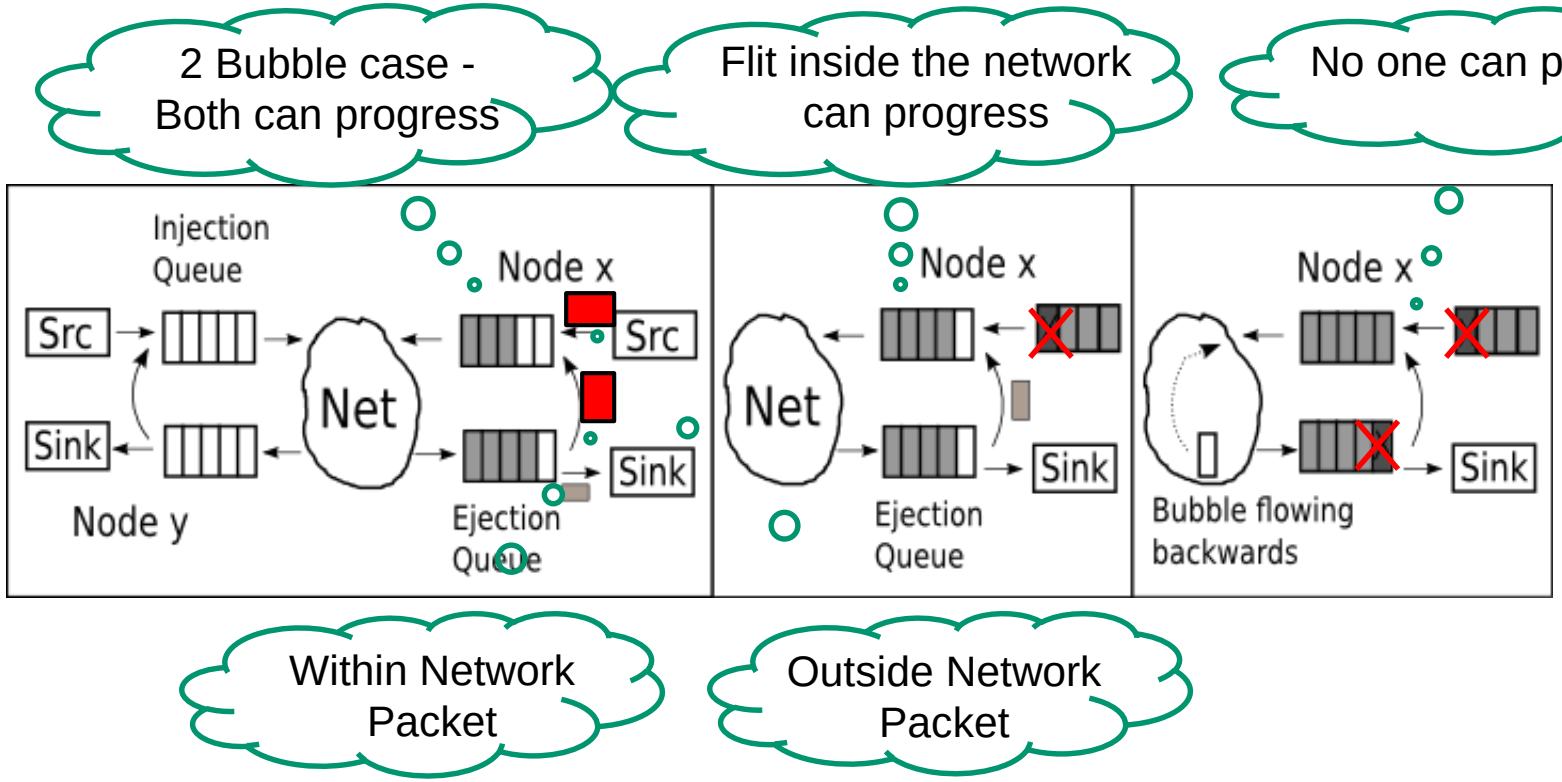
- Single flit bubble is enough for flit level ring.
- Packets that begin to enter the ring must be guaranteed entry
  - A PktLength+1 bubble can be provided by CBs.
- CB can be bypassed.
  - Flits in the output ring will eventually drain to the empty space of CB.
- Suitable for EB links (no downstream router information required)
- CB buffer space can be shared among rings (min. space req.).





Message dependent Deadlock

- Standard solution: separate virtual networks.
- CB Solution: A reply can be treated as a 180 degree turn
- Bubble flow control can be used to avoid deadlocks in a cycle.



- Ensure that a flit inside the ring makes progress if a single bubble is available.

Overview

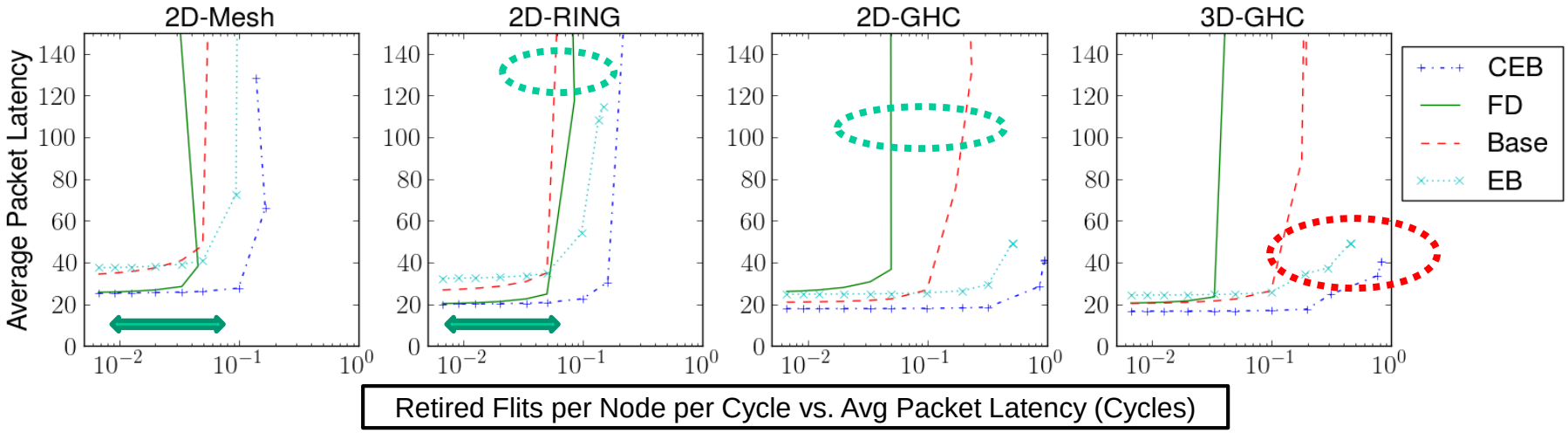
- Motivation
- Router Design
- Optimizations
- Deadlock Avoidance
- Results

Simulation Methodology

- 4 different routers
  - **Baseline:** Standard 2 stage, 2VC per virtual network, 5-flit IB.
  - **Flit Deflection (FD)**: Similar to flit BLESS.
    - Age based priority
    - Retire as soon as the tail arrives (no wait due to rearranging out of order arrival of head/body)
  - **EB**: 2 Stage EB, half wide links, twice number of flits
- **Network**: 64 cores 8x8 or 4x4x4 (Mesh, Torus, GHC)
  - GHC has link delay equal to the number of hops between the routers
  - Others has single cycle link delay
- **Results**: Average of 4 synthetic traffic patterns.

Performance (Other Routers)

- Low Loads
  - CEB has low-load latency equal to deflection (look-ahead SA)
  - Baseline suffers due to constant 2 cycle latency within the router.
  - EB suffers due to narrower channels

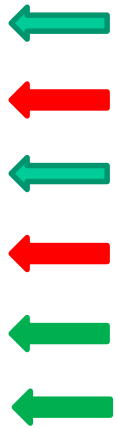


Retired Flits per Node per Cycle vs. Avg Packet Latency (Cycles)

- High loads
  - CEB and EB has higher saturation throughput due to EB links.
  - Credit flow control cause bubbles in the pipeline of baseline router and reduces throughput.
  - FD increases its deflection with higher radix & injection rate.
- Improvement in performance gets higher in large radix routers.

Buffer Space Analysis

	2D Torus (KB)	3D Torus (KB)	3D GHC (KB)	2D GHC (KB)
Baseline-M1	100	124	110	145
EB-M1	60	68	60	70
Baseline-M4	200	376	320	460
EB-M4	120	152	120	160
FD-P4	55	61	70	85
FD-P20	135	141	150	165
CEB-Nogate	73	79	88	103
CEB-Gate	55	61	70	85



Single flit IB, 2 flit OB, 18 flit CB. 128 bit flits. 20 flit inj. & Ej. Torus has 2-VC, GHC 1-VC

Baseline requires large buffer space 

EB requires low, (mostly in NI) but increases significantly with higher msg. classes 

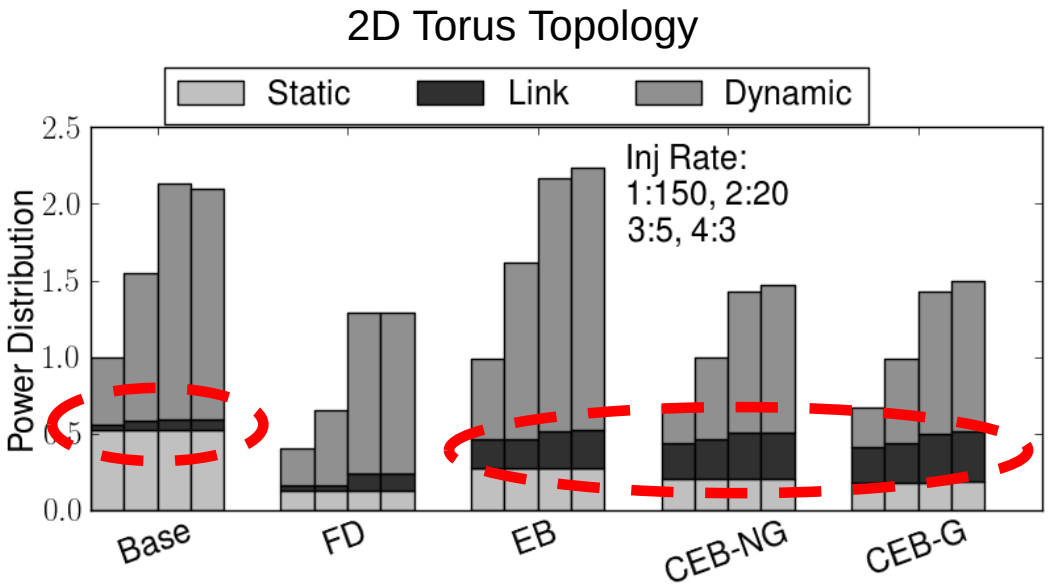
FD has least but reorganizing flits require large buffers in NI. 

CEB has low buffer requirement with any number of message classes.

It has slow increase of buffer space with higher radix (Similar to FD).

Power Comparison

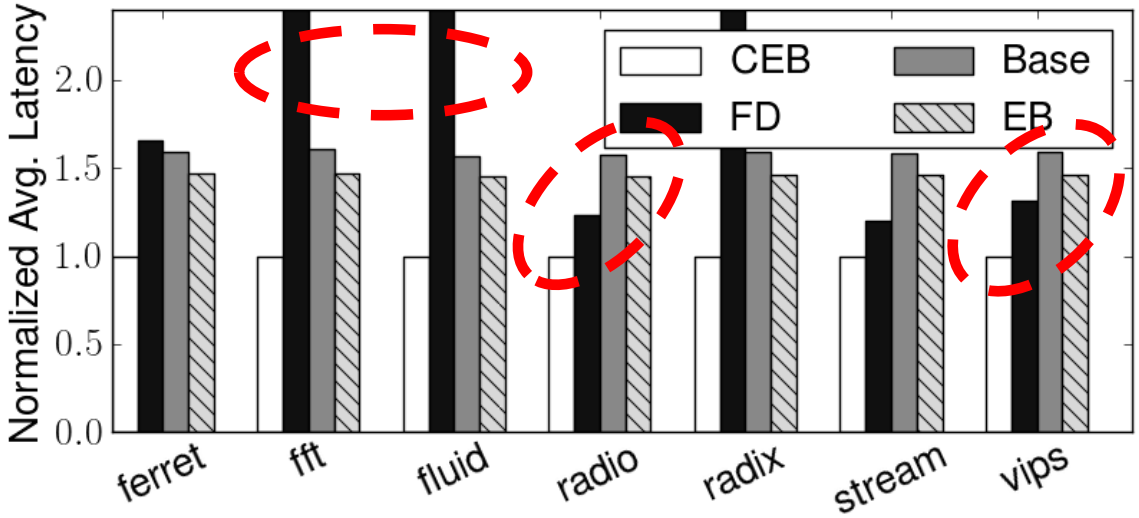
- Orion 2.0 is used
  - Activity estimated using timing simulations and fed to Orion
- Orion Modifications
  - 3X more leakage & device unit power in EB links.
  - 3X more arbiter power in CEB router
- EB & Baseline router
  - 2 Msg classes (2VC per class)



- Baseline has high static power due to large buffer space
- Power in EB links is very high
- FD has lower power than CEB.
  - It increases rapidly with higher number of deflections

Application Traces

- Traces from 64 core, 16 MC simulation at the back side of coherent L1 cache.
- Most benchmarks operate at low loads.
  - Low load latency of CEB & deflection are least.
- However, sometimes higher traffic.
  - FD latency become too high.



Conclusions & Next Step

- Centralized buffer routers are a good candidate for low-latency, high-radix NOCs.
- CEB mix advantages of central buffers, EB links and bubble flow control.
- Next Steps
  - Hardware Implementation
  - Extension with worm-bubble
  - Adaptive Routing
  - QoS Support

THANK YOU !!