

Multicore Computer Architecture - Storage and Interconnects

Lecture 15

Energy Efficient Bufferless NoC Routers



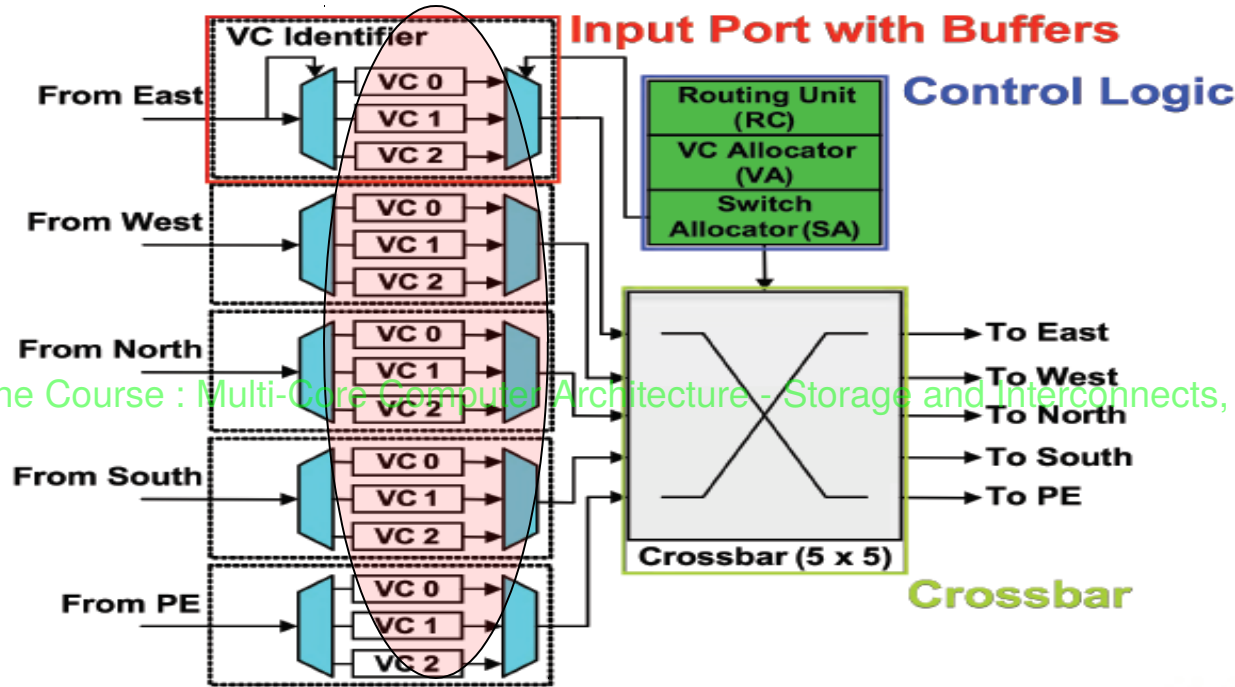
NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

Dr. John Jose

Assistant Professor

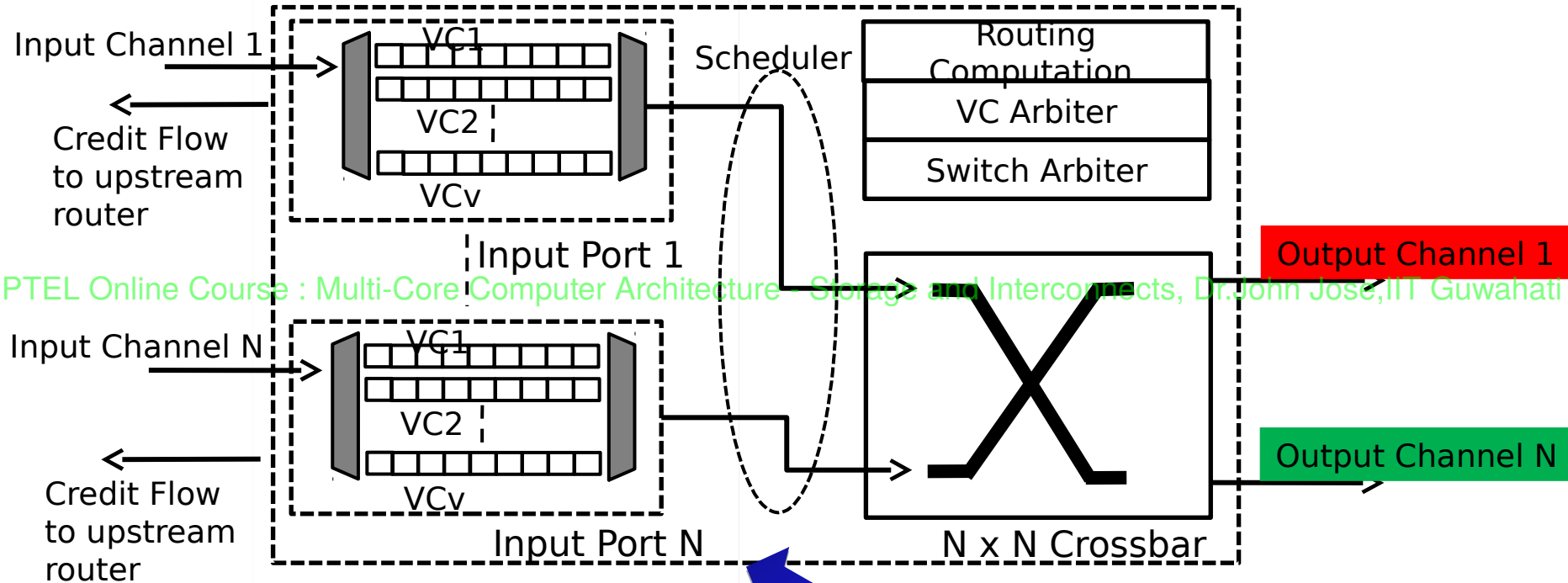
**Department of Computer Science & Engineering
Indian Institute of Technology Guwahati, Assam.**

Virtual Channel NoC router



- ❖ Flits stay in VC buffers during port conflicts
- ❖ Buffers consume much power

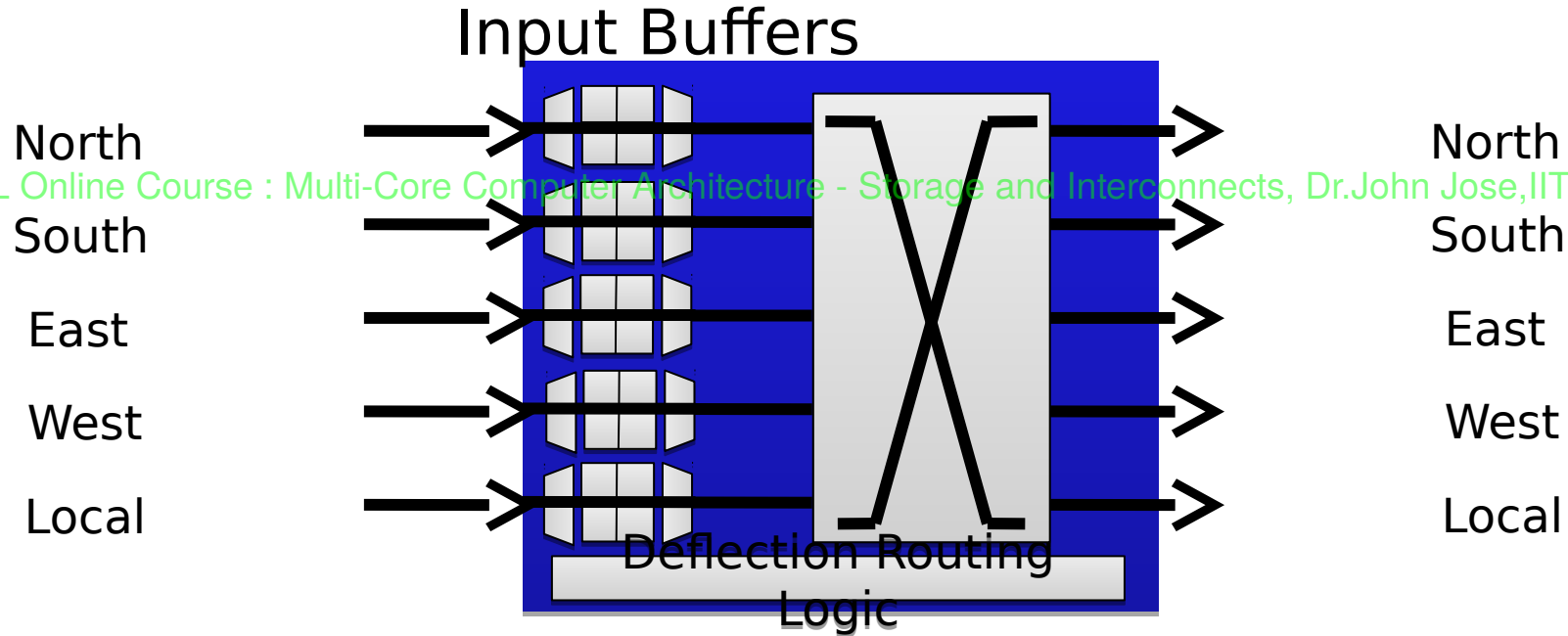
Virtual Channel NoC router



Buffers are integral part of existing NoC Routers

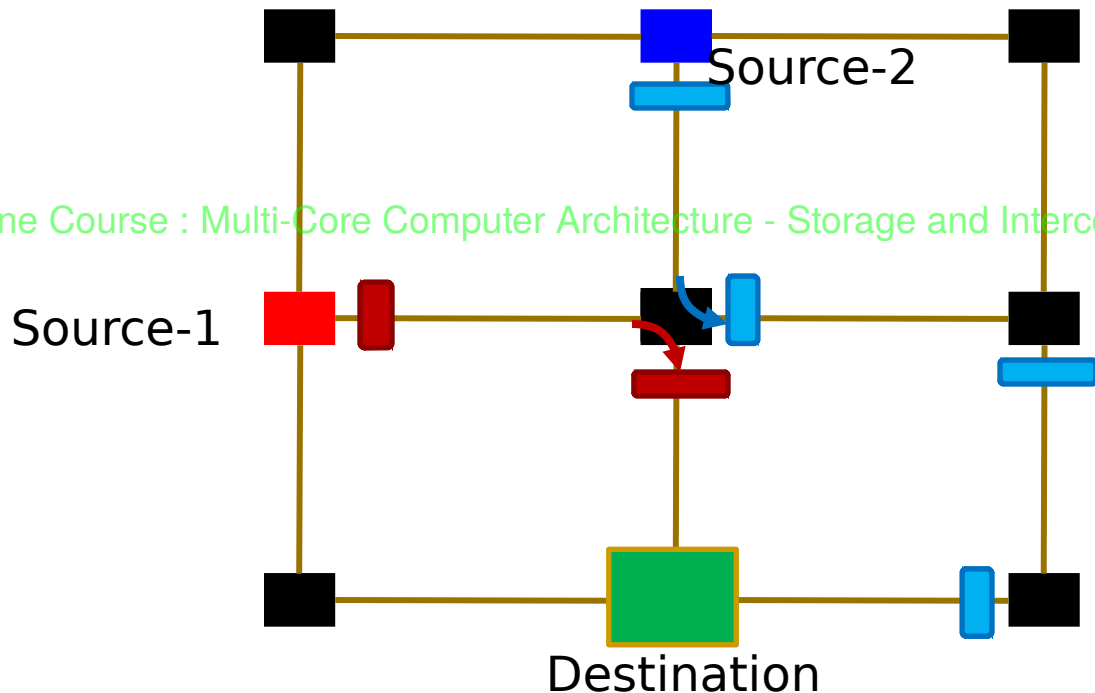
Bufferless Deflection Routing

- ❖ Input buffers are eliminated: flits are buffered in pipeline latches and on network links



Bufferless Deflection Routing

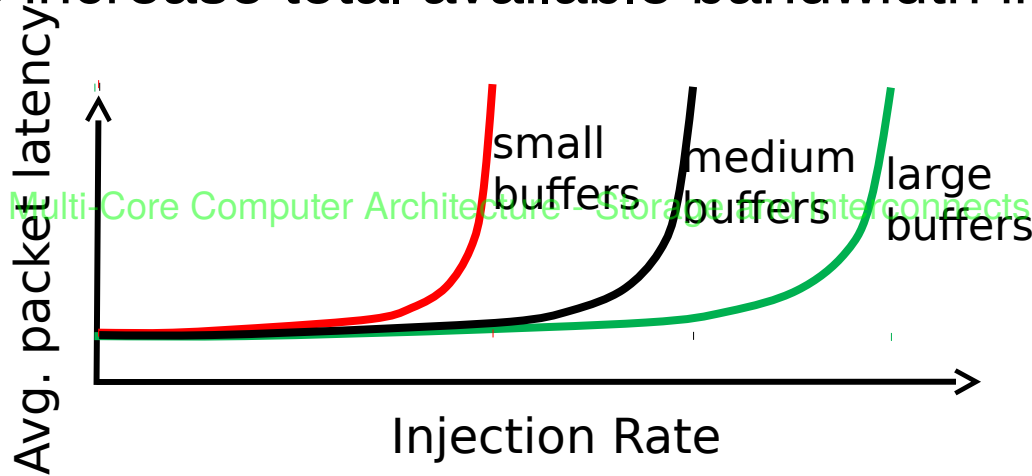
When two packets contend for the same link, one is **given the desired link** and other is **deflected**.



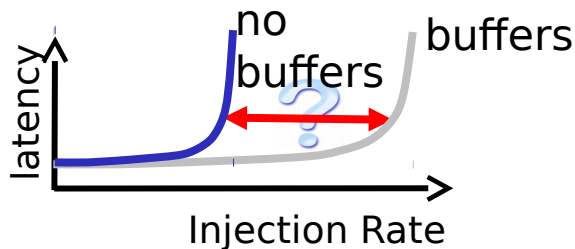
NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

Role of buffers in NoC

- ❖ Buffers are necessary for high network throughput
- ❖ Buffers increase total available bandwidth in network



Bufferless NoC



❖ How much throughput do we lose?

NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

❖ How is latency affected?

❖ Up to what injection rates can we use bufferless routing?

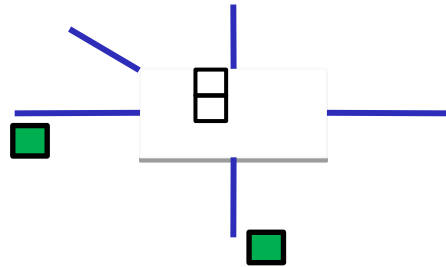
❖ Are there realistic scenarios in which NoC is operated at injection rates below the threshold?

❖ Can we achieve energy, area reduction?

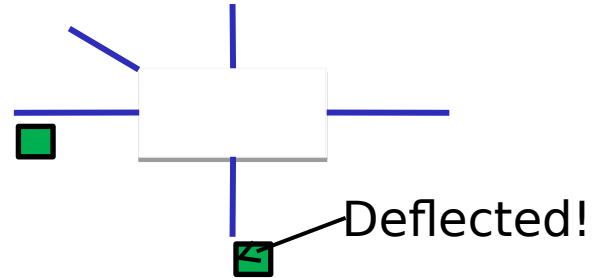
BLESS: Bufferless Routing

- ❖ Always forward all incoming flits to some output port
- ❖ If no productive direction is available, send to another direction  packet is deflected
- ❖ Hot-potato routing

NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

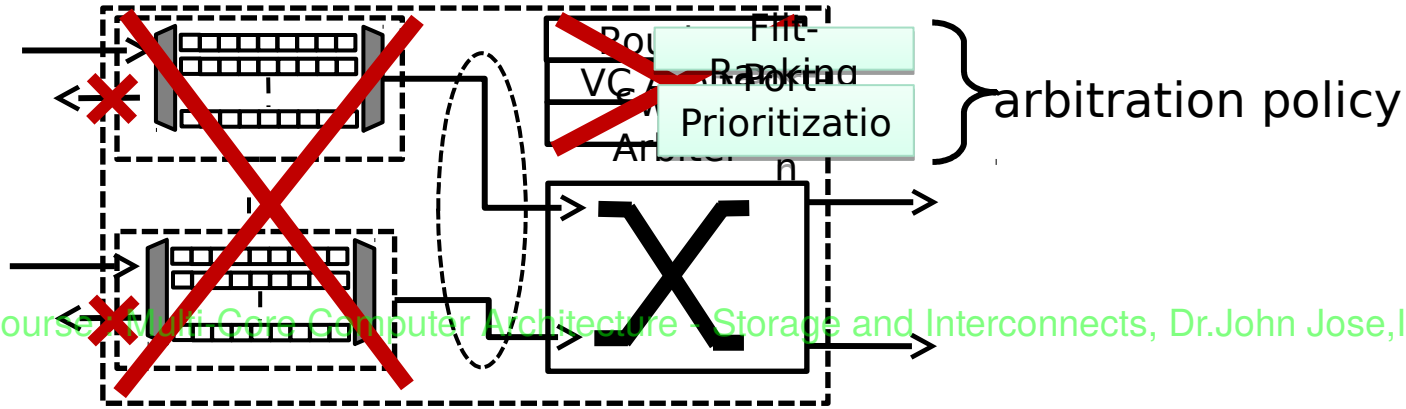


Buffered



BLESS

BLESS: Bufferless Routing



- ❖ **Flit Ranking-** Create a ranking over all incoming flits
- ❖ **Port Prioritization:** For a given flit in this ranking, find the best free output-port
- ❖ Apply to each flit in order of ranking

BLESS: Bufferless Routing

- ❖ Each flit is routed independently.
- ❖ Oldest-first ranking
- ❖ Assign flit to productive port, if possible.
- ❖ Otherwise, assign to non-productive port (deflect).
- ❖ Flow Control Completely local
- ❖ Injection Policy: Inject whenever input port is free
- ❖ Absence of Deadlocks: every flit is always moving
- ❖ Absence of Livelocks: with oldest-first ranking

BLESS: Advantages & Disadvantages

Advantages

- ❖ No buffers and local flow control
- ❖ Simplicity
 - no credit-flows
 - no virtual channels
 - simplified router design
- ❖ No deadlocks, livelocks
- ❖ Adaptivity: packets are deflected around congested areas!
- ❖ Router area reduction

Disadvantages

- ❖ Increased latency
- ❖ Reduced bandwidth
- ❖ Increased buffering at receiver for reassembly
- ❖ Header information at each flit
- ❖ Oldest-first arbitration complex
- ❖ QoS becomes difficult

Livelock Freedom in BLESS

- ❖ What stops a flit from deflecting forever?
- ❖ All flits are **timestamped**
- ❖ **Oldest flits** are assigned their desired ports
- ❖ **Total order among flits**

New traffic is lowest
priority



Flit age forms total order

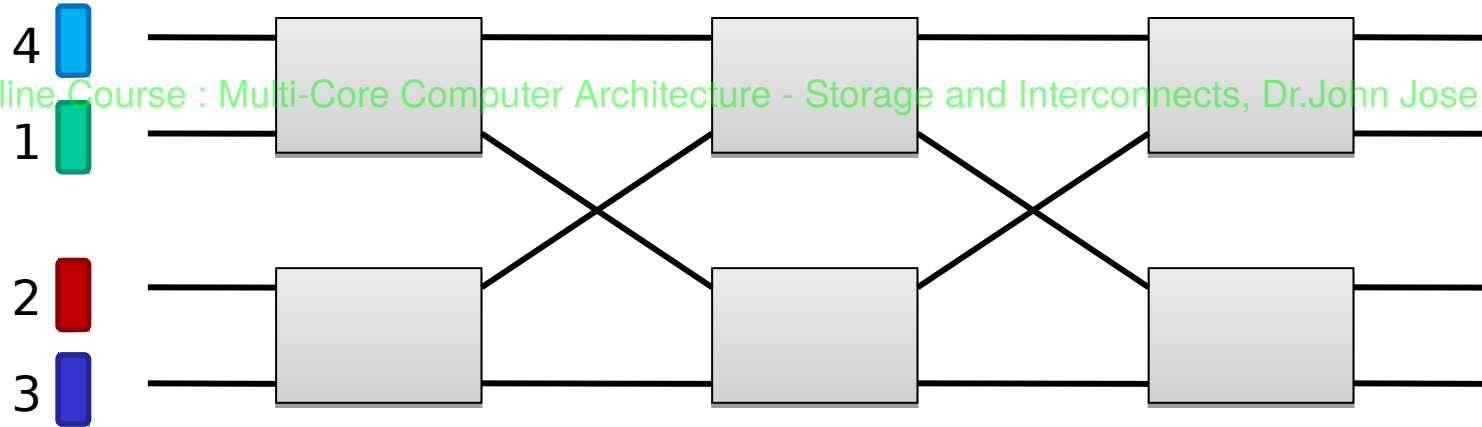
Guaranteed
progress!

- ❖ But what is the **cost** of this?

Expensive Age-Based Priorities

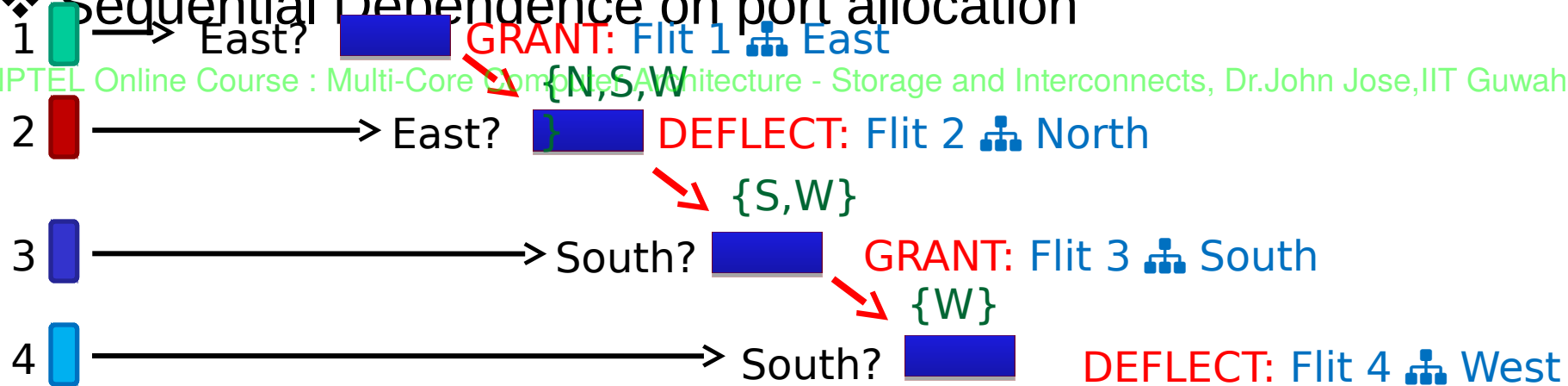
❖ Router must **sort flits by age**: long-latency sort network

– **Three comparator stages** for 4 flits



Expensive Age-Based Priorities

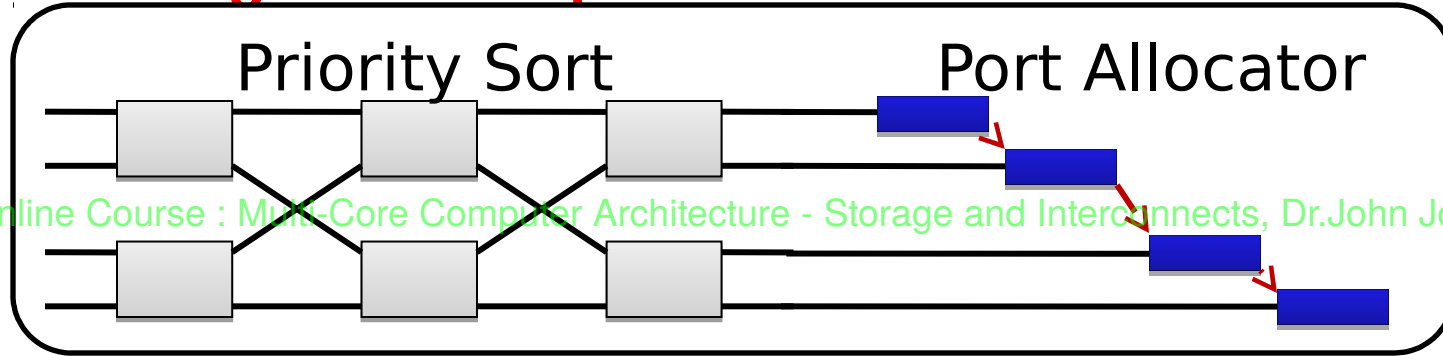
- ❖ After sorting, flits assigned to output ports in priority order
- ❖ Port assignment of younger flits depends on that of older flits
- ❖ Sequential Dependence on port allocation



Age-Ordered Flits

Expensive Age-Based Priorities

- ❖ Overall, **deflection routing logic** based on **Oldest-First** has a **longer critical path** than a buffered router



- ❖ Do we have a cheaper way to route while guaranteeing livelock-freedom?

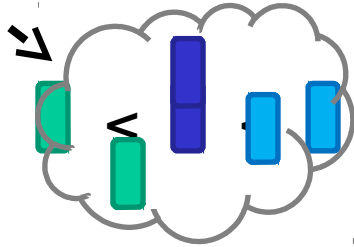
Golden Flit Concept

Key Insight: No total order. it is enough to:

1. Pick one flit to **prioritize until arrival**
2. Ensure any flit is **eventually** picked

NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

New traffic is
lowest-priority



**Guaranteed
progress!**

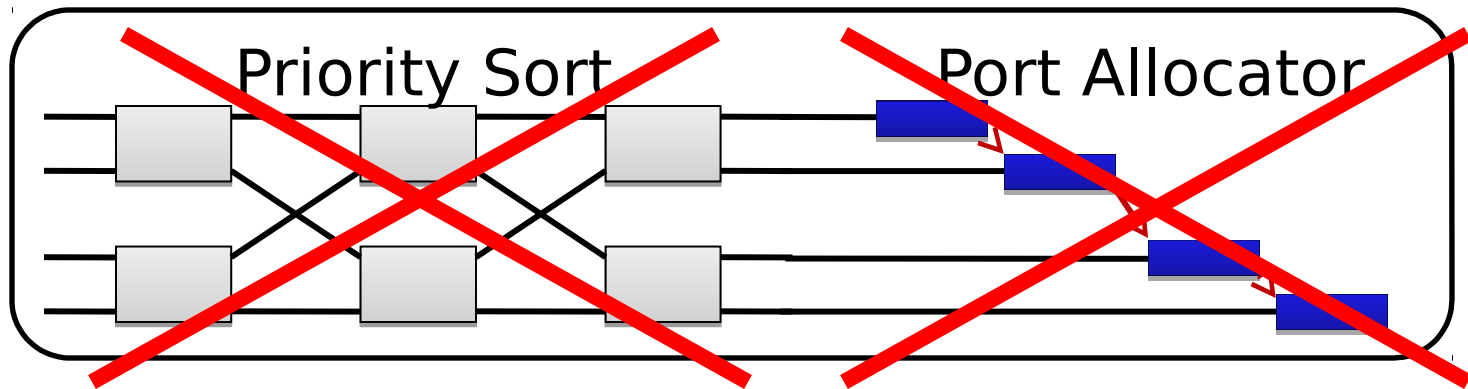
<
“Golden Flit”

Flit age forms total order
partial ordering is sufficient!

Golden Flit Concept

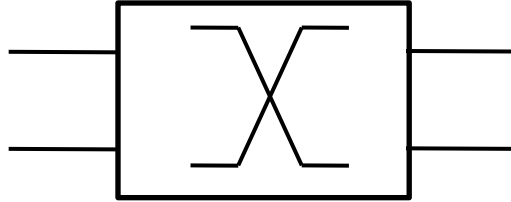
- ❖ Only **need** to properly route the Golden Flit
- ❖ **First optimization:** no need for full sort
- ❖ **Second optimization :** no need for sequential allocation

NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr.John Jose,IIT Guwahati



Golden Packet Concept

- ❖ Routing of the Golden Flit in a two-input router

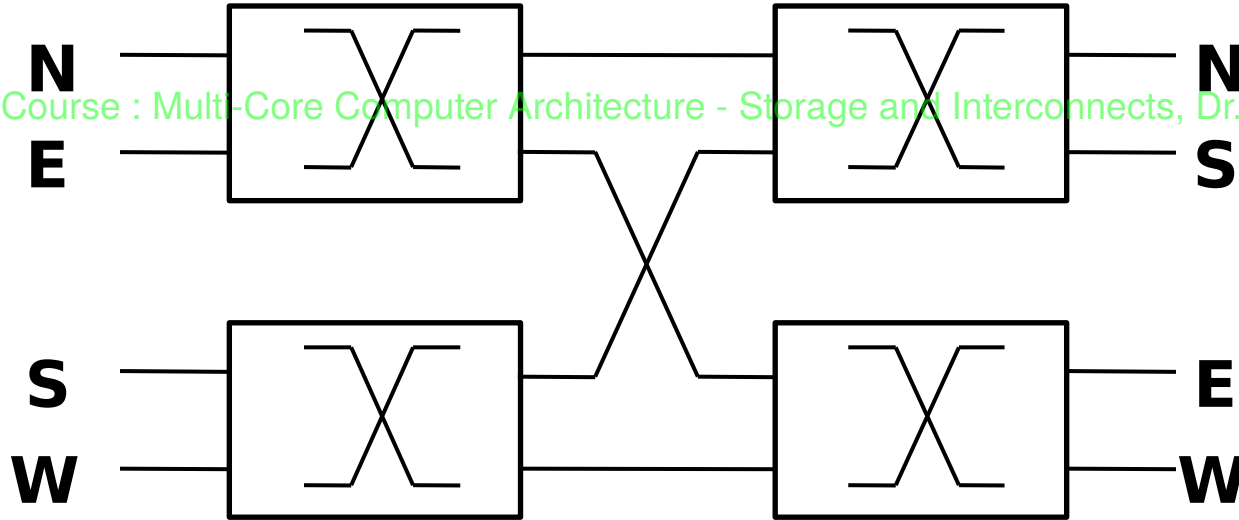


NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

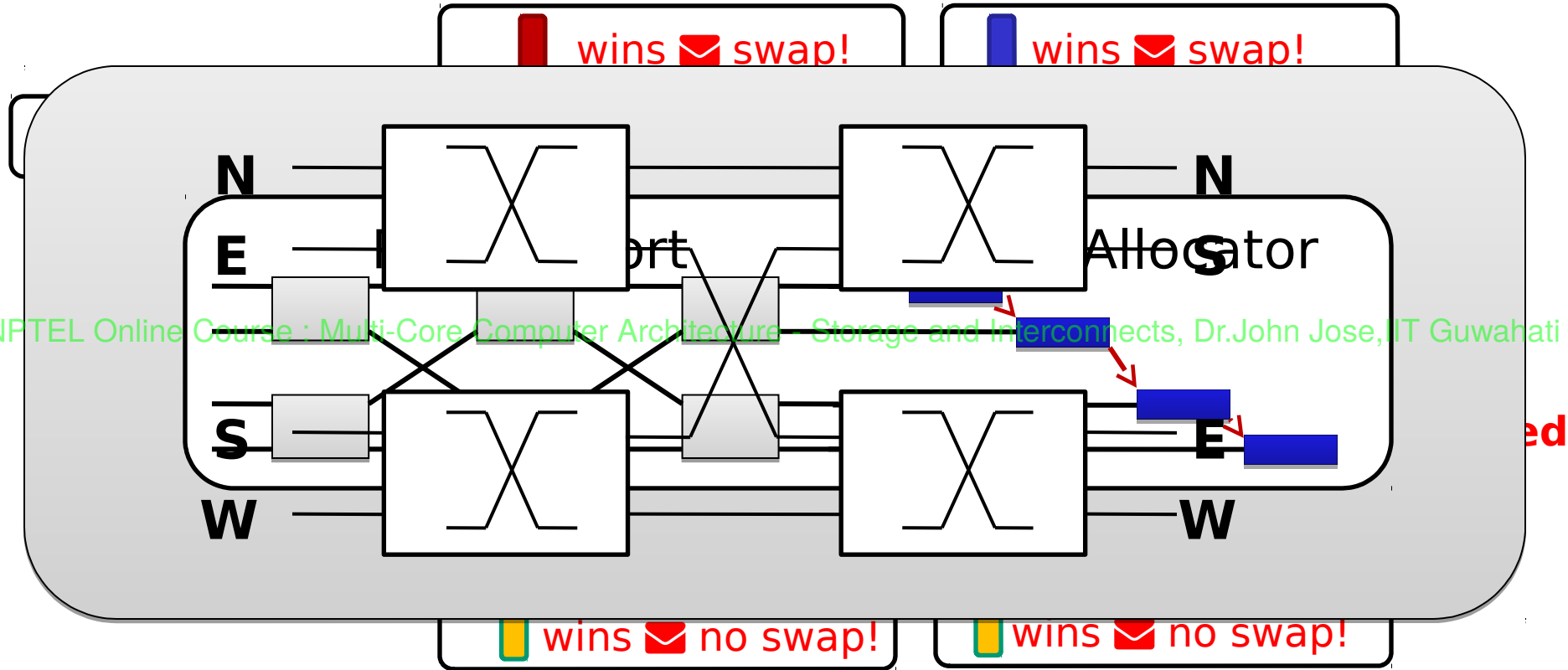
- ❖ **Step 1:** pick a **winning flit**: Golden Flit, else random
- ❖ **Step 2:** steer the winning flit to its **desired output**
- ❖ and **deflect other flit**
- ❖ **Golden Flit always routes toward destination**

Golden Flit Routing

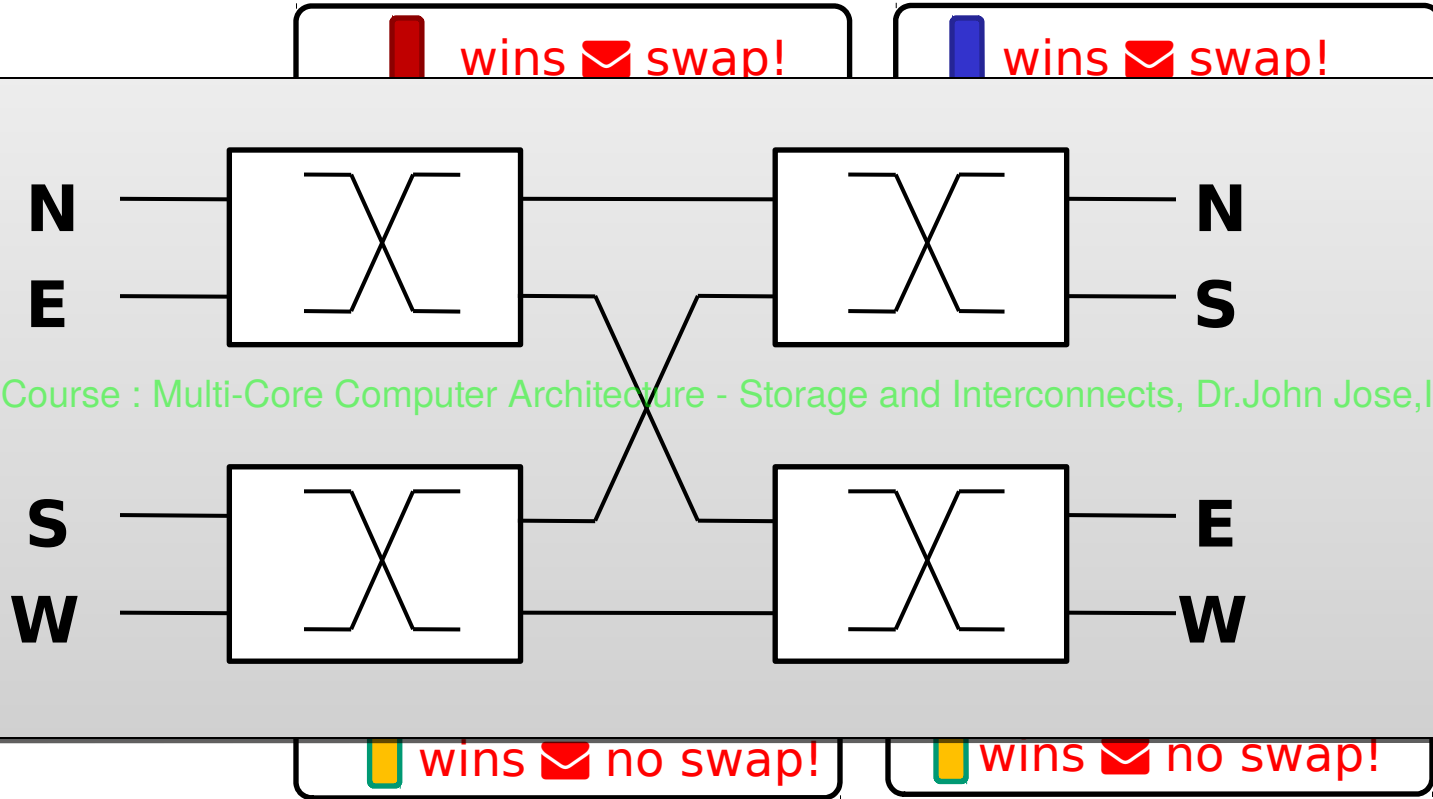
- ❖ Each block makes decisions **independently!**
- ❖ Deflection is a distributed decision



Permutation Network Operation

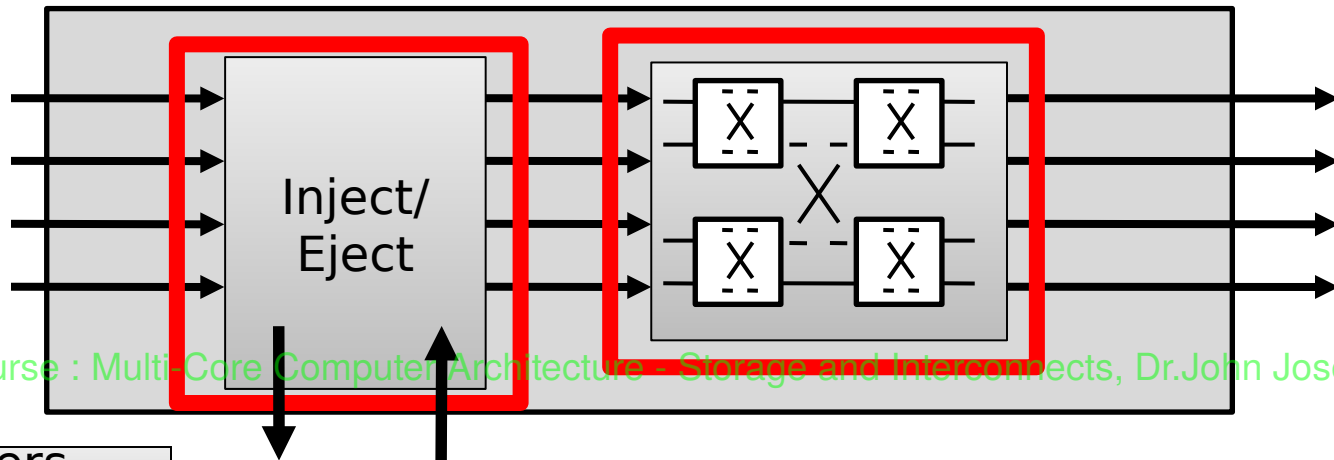


Permutation Network Operation



NPTTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr.John Jose,IIT Guwahati

Parallel Deflection Logic - CHIPPER*



NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

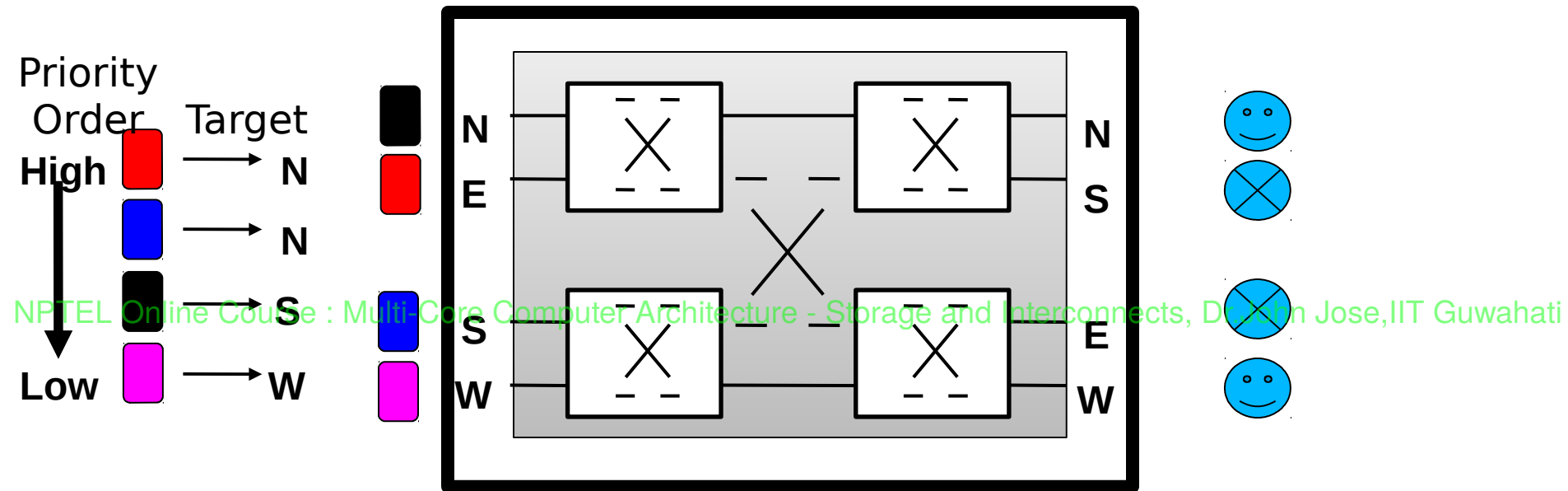
Miss Buffers
(MSHRs)

Eject Inject

- ❖ Early ejection and injection
- ❖ Injection suppression during busy inputs (**starvation**)
- ❖ Permutation Deflection Network (PDN) (**faster**)

* C. Fallin et al., "CHIPPER: A low complexity bufferless deflection router," HPCA- 2011.

Parallel Deflection Logic - CHIPPER*

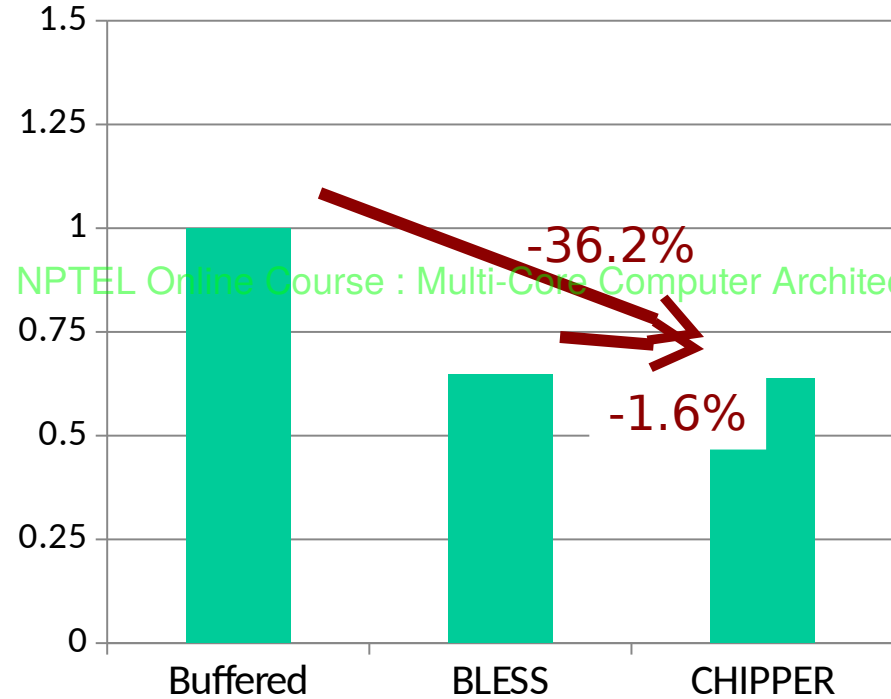


- ❖ Deflection for low priority flits.
- ❖ Poor flit latency at high injection rate.

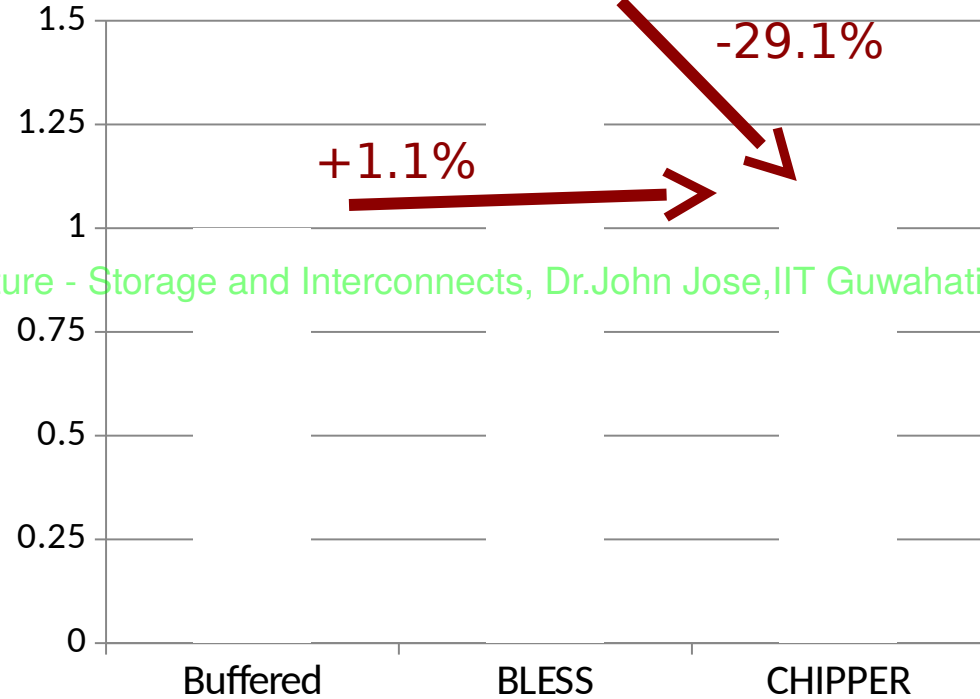
* C. Fallin et al., "CHIPPER: A low complexity bufferless deflection router," HPCA- 2011.

Performance Analysis

Normalized Router Area

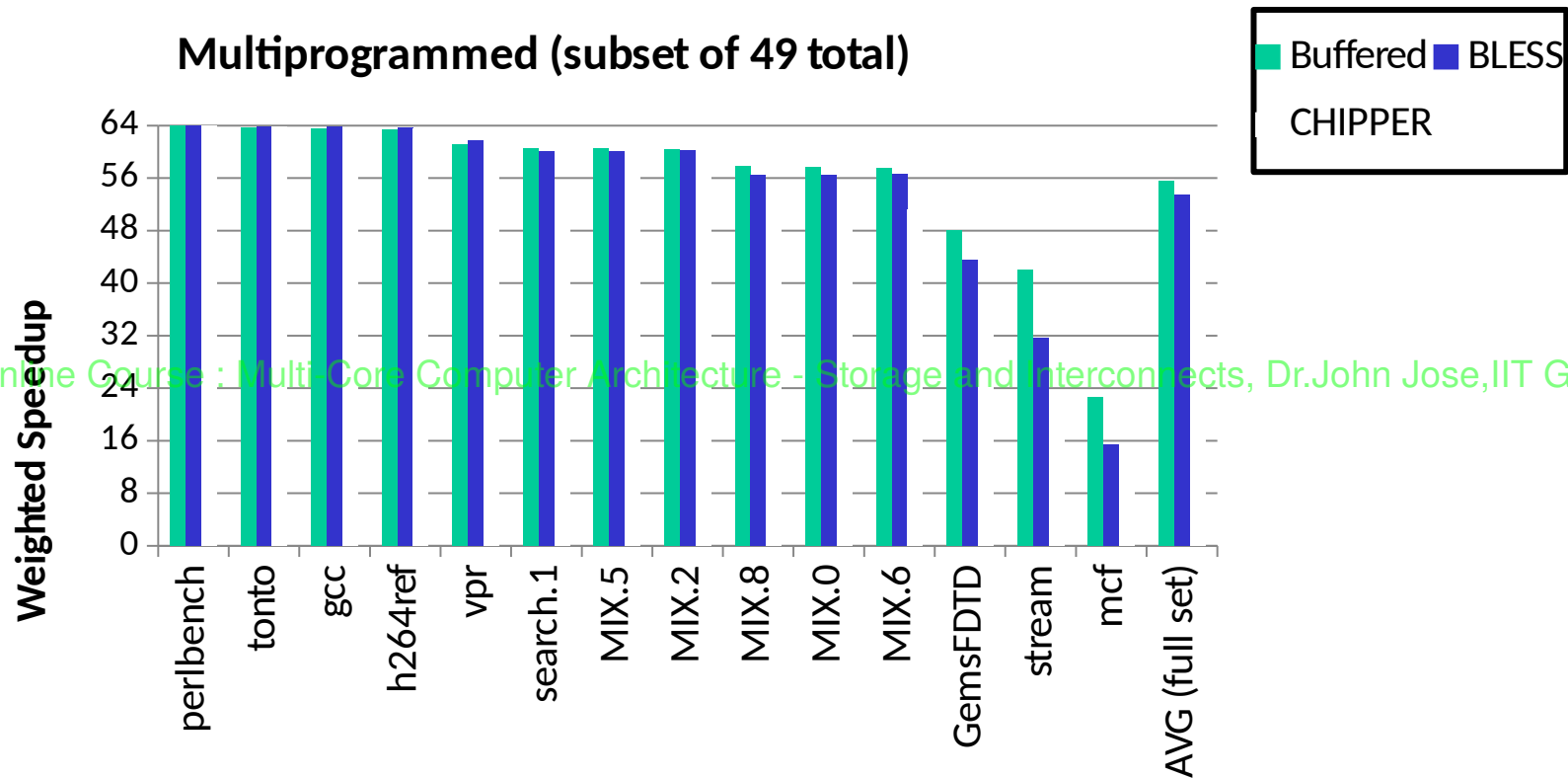


Normalized Critical Path



NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

Performance Analysis



**THANK
YOU!**

FOR

**YOUR
PRESENCE**

johnjose@iitg.ac.in

<http://www.iitg.ac.in/johnjose/>