

Lecture 16

Sidebuffered Deflection Routers



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

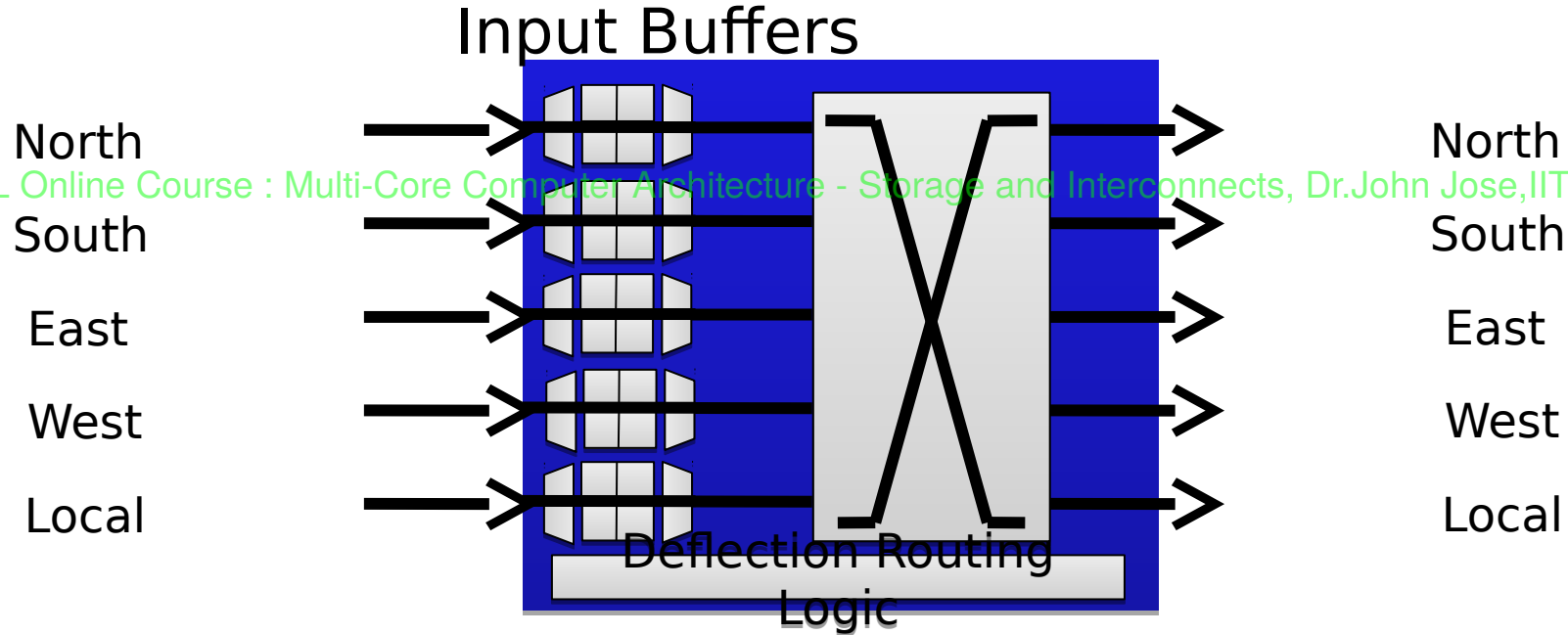
Dr. John Jose

Assistant Professor

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

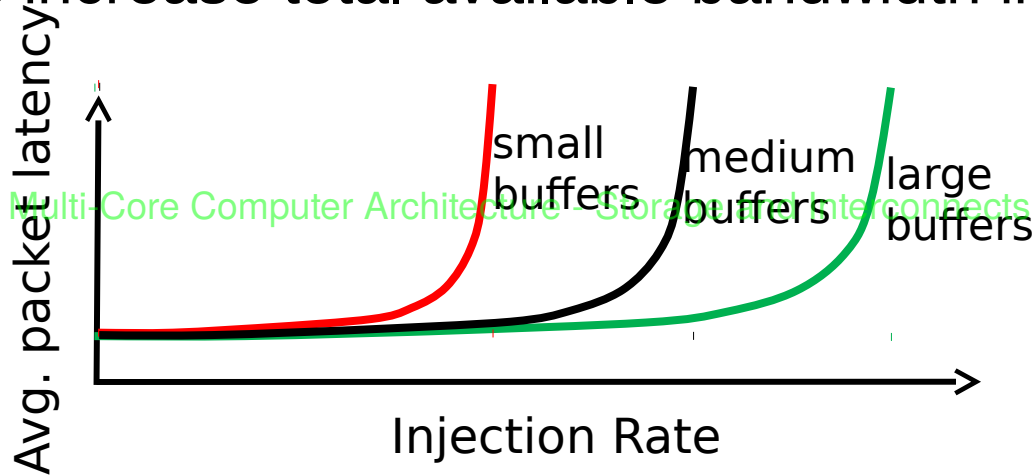
Bufferless Deflection Routing

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

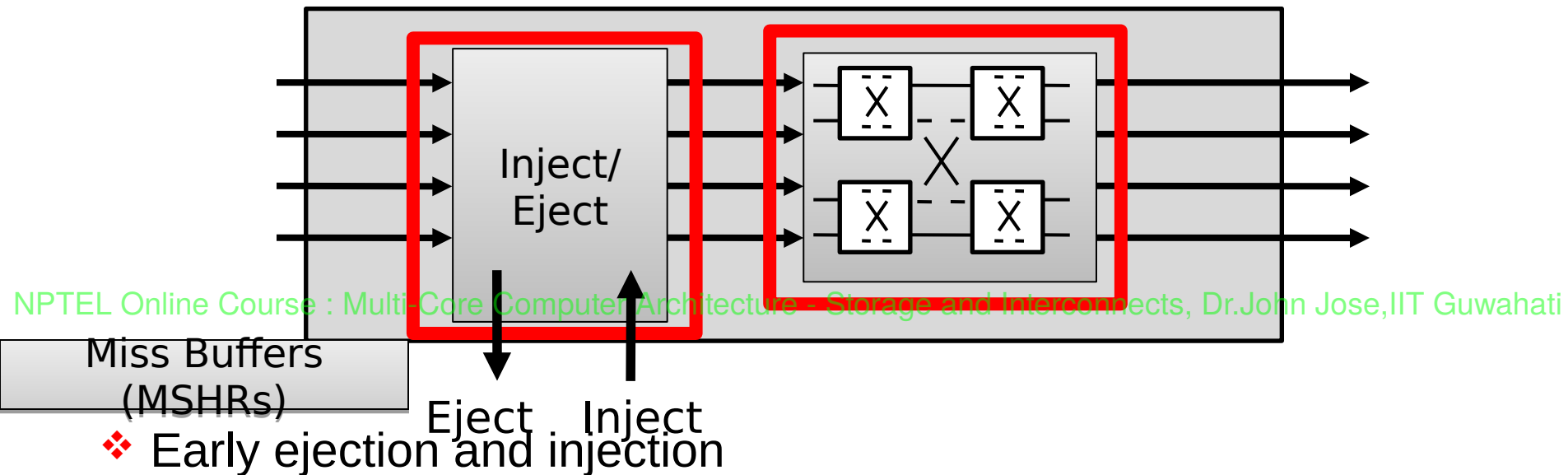


Role of buffers in NoC

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



Parallel Deflection Logic - CHIPPER*



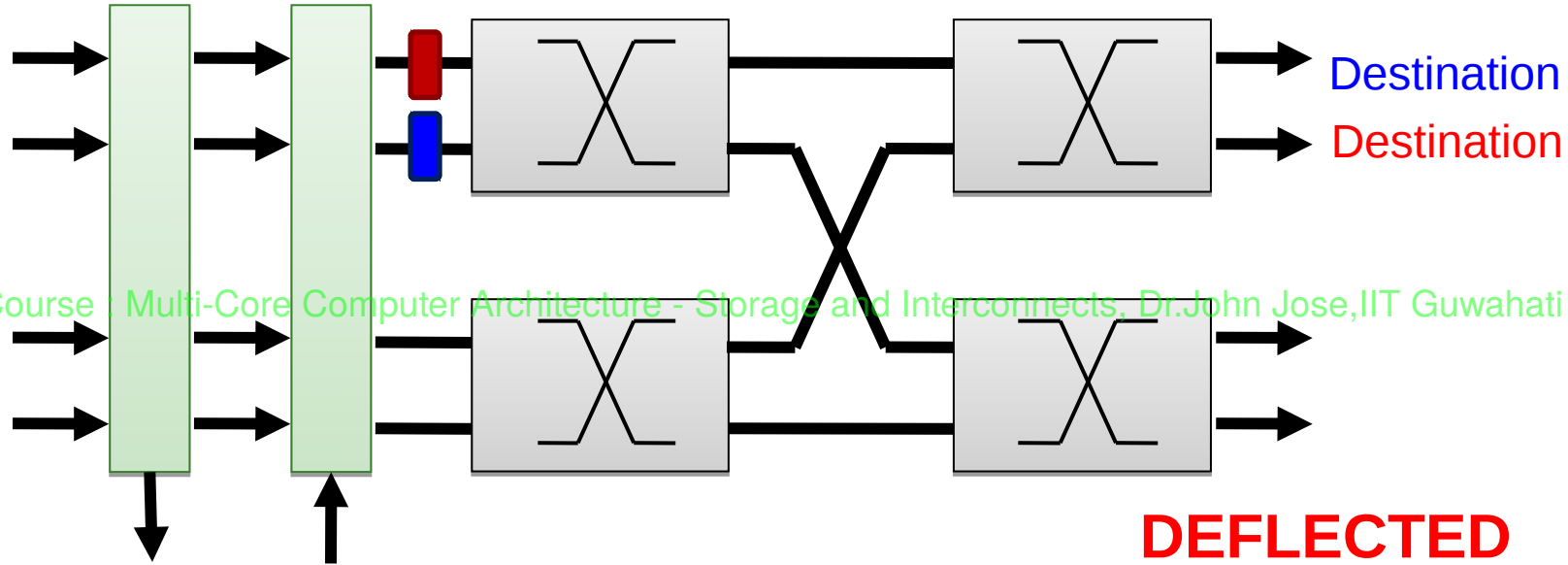
- ❖ Early ejection and injection
- ❖ Injection suppression during busy inputs (starvation)
- ❖ Permutation Deflection Network (PDN) (faster)

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

Key Performance Issues in CHIPPER

- ❖ **Link contention:** no buffers to hold traffic ✉
any link contention causes a deflection ✉ use side buffers
- ❖ **Ejection bottleneck:** only one flit can eject per router per cycle ✉
simultaneous arrival causes deflection ✉ eject up to 2 flits/cycle
- ❖ **Deflection arbitration:** practical (fast) deflection arbiters deflect unnecessarily ✉ new priority scheme (silver flit)

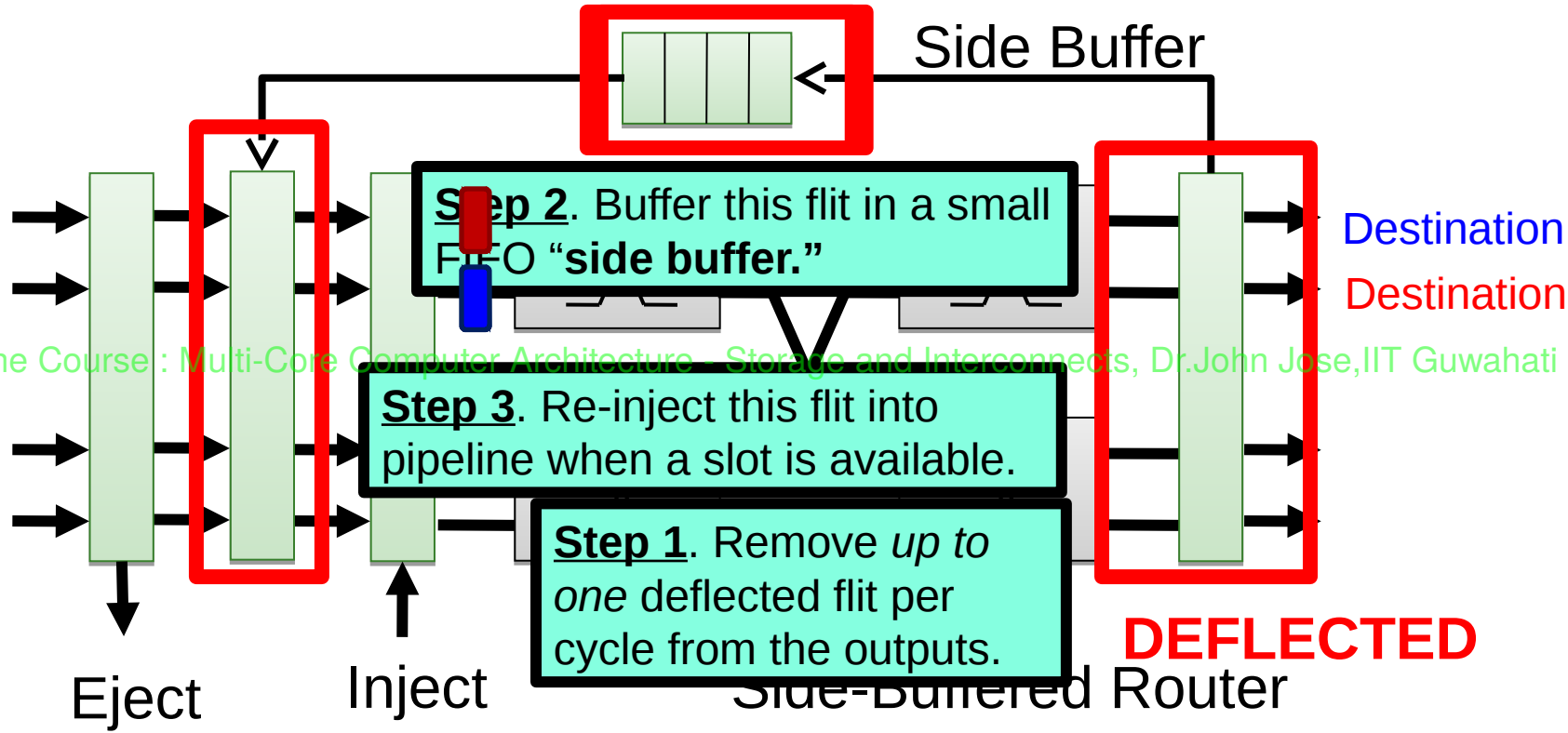
How to Buffer Deflected Flits



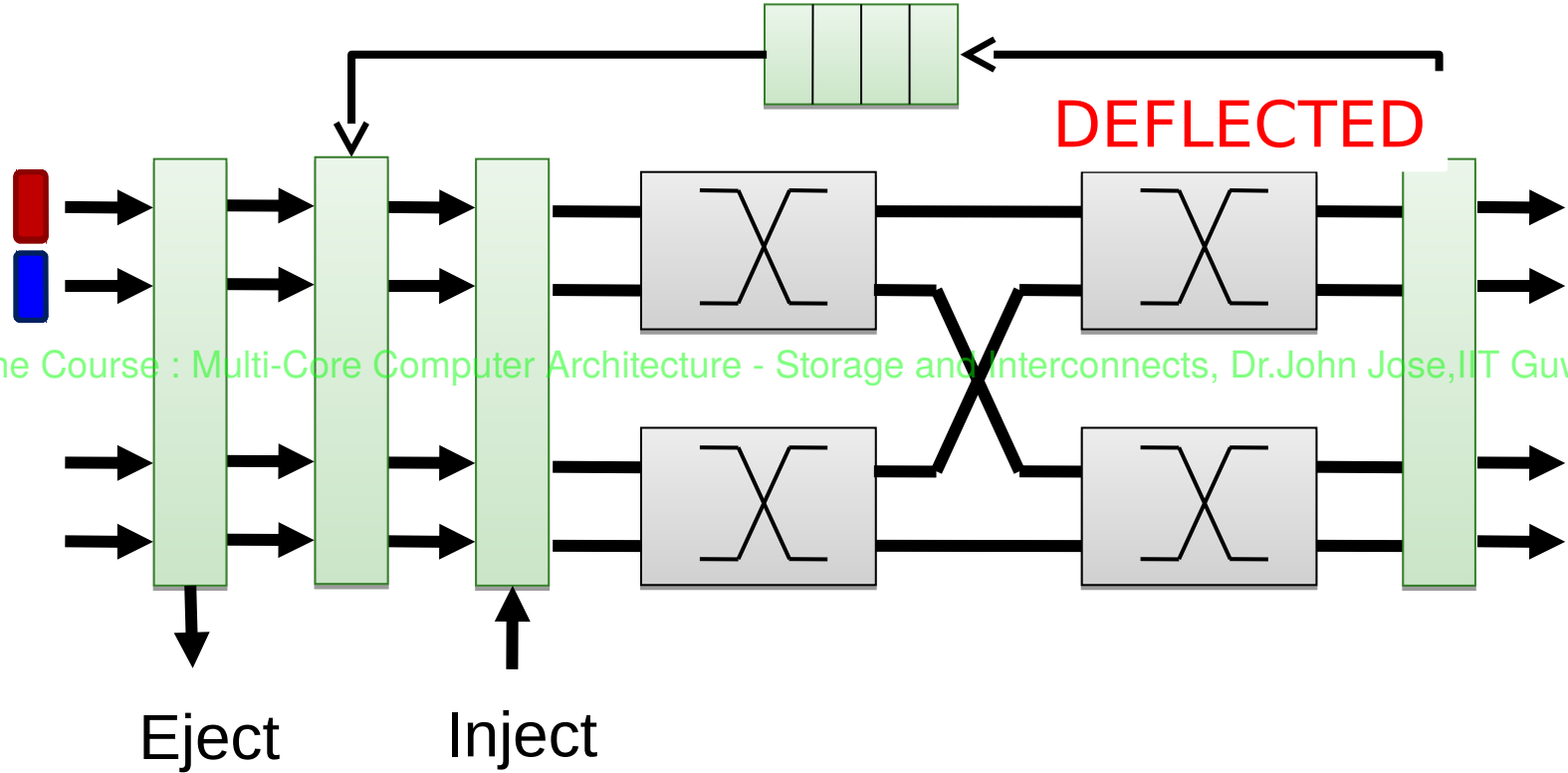
DEFLECTED

CHIPPER

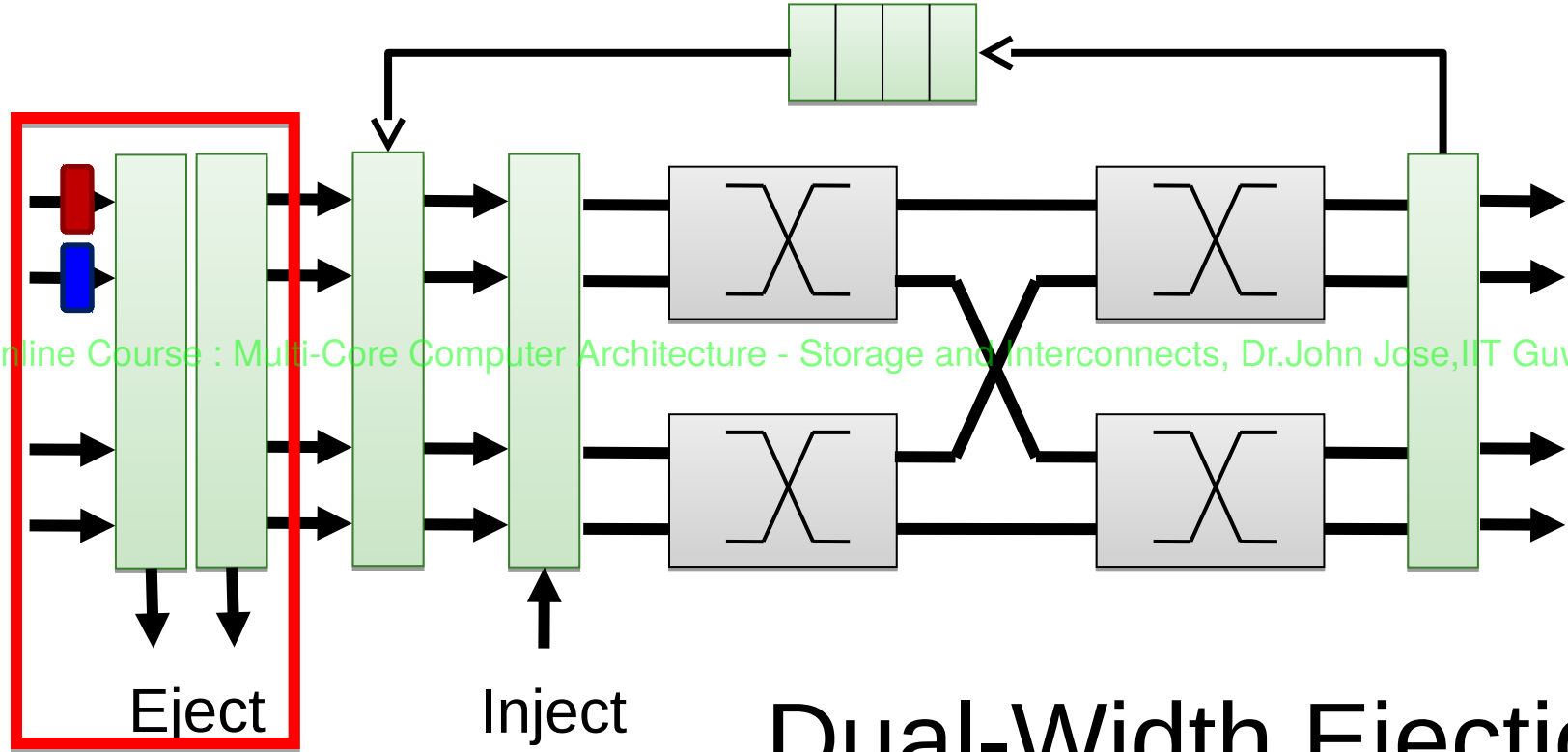
How to Buffer Deflected Flits



Addressing the Ejection Bottleneck



Addressing the Ejection Bottleneck



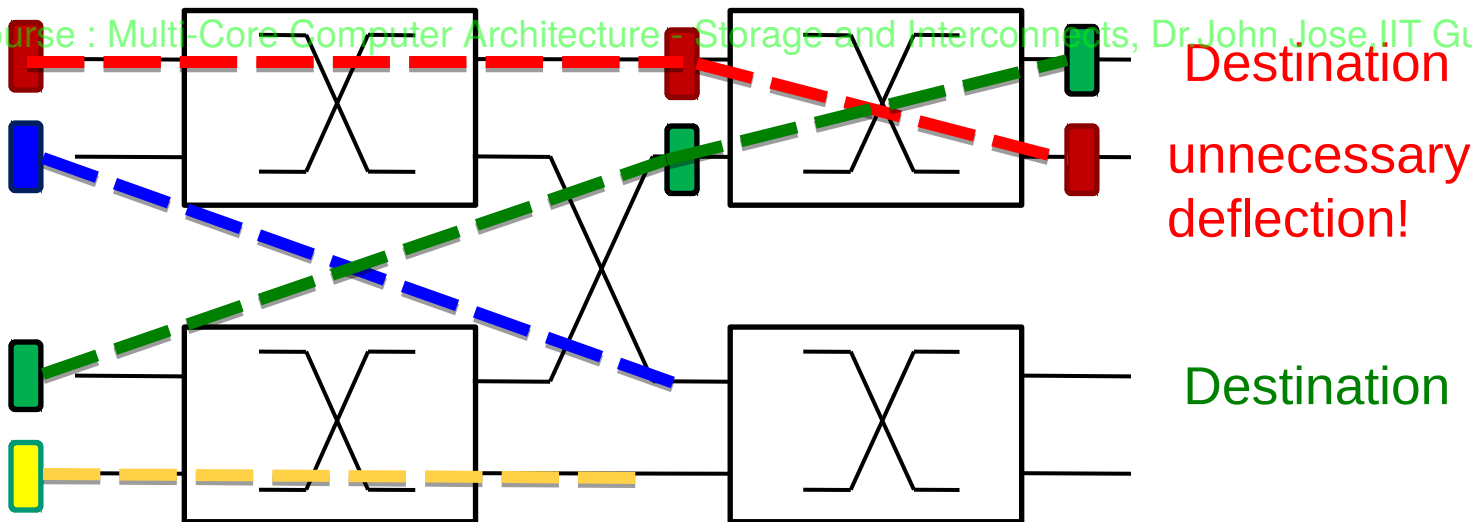
NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

Dual-Width Ejection

Unnecessary Deflections in PDN

- ❖ Lack of coordination cause unnecessary deflections of non-golden flits
- ❖ Red flit wins at first stage
- ❖ Green flit loses at first stage (must be deflected now)
- ❖ Red flit loses at second stage; Red and Green are deflected

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

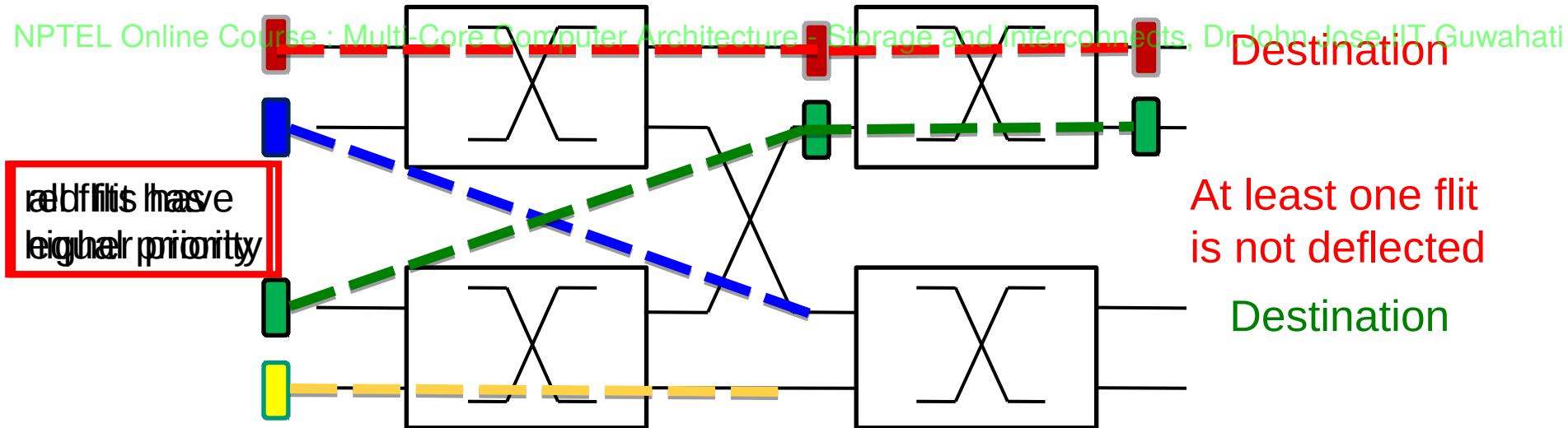


Improving Deflection Arbitration

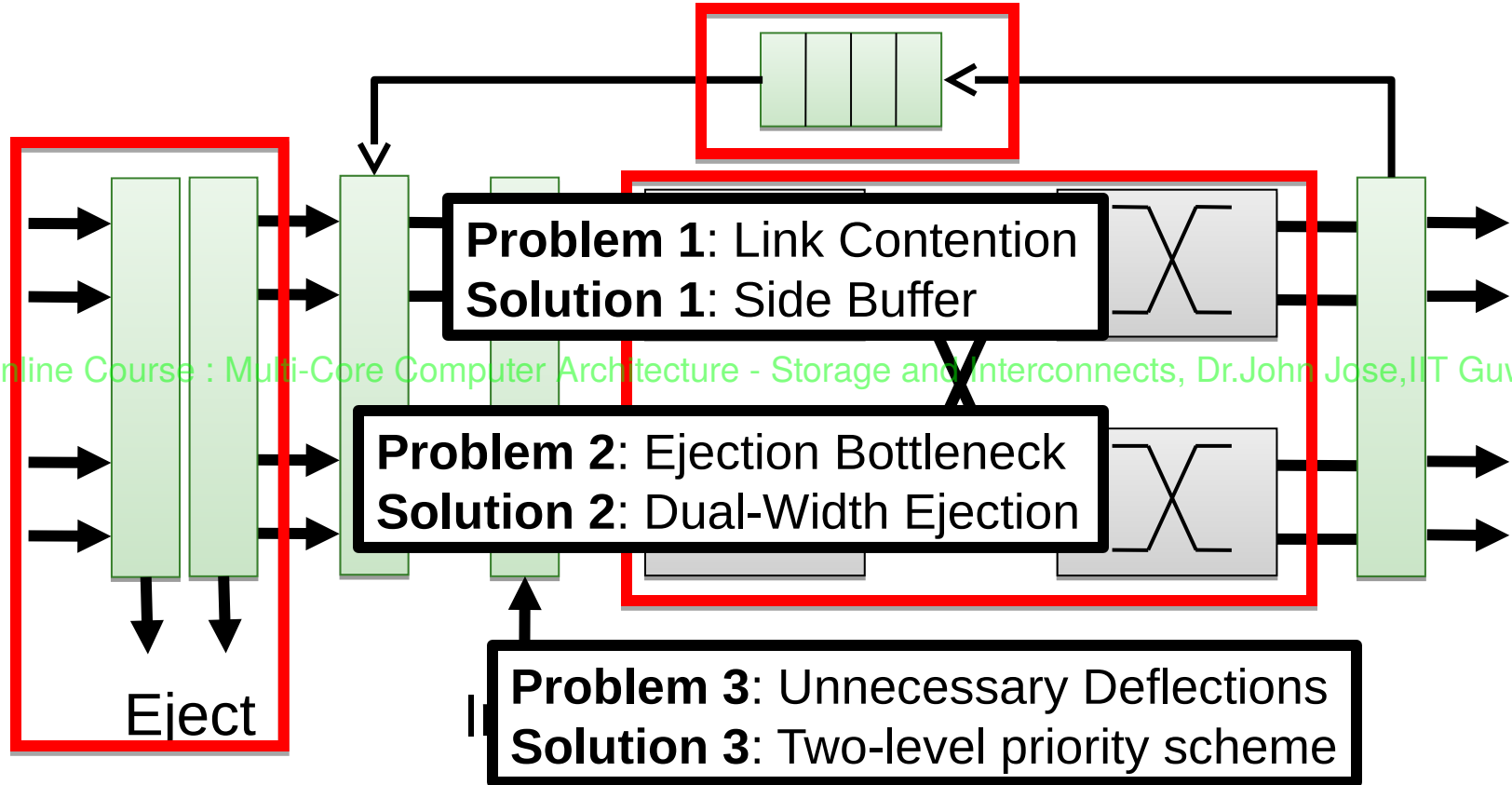
- ❖ **Add a priority level** and prioritize one flit to ensure at least **one flit** is not deflected in each cycle
- ❖ **Highest priority:** one **Golden Packet** in network
 - ❖ Chosen in static round-robin schedule
 - ❖ **Ensures correctness**
- ❖ **Next-highest priority:** one **silver flit** per router per cycle
 - ❖ Chosen pseudo-randomly & local to one router
 - ❖ **Enhances performance**

Adding A Silver Flit

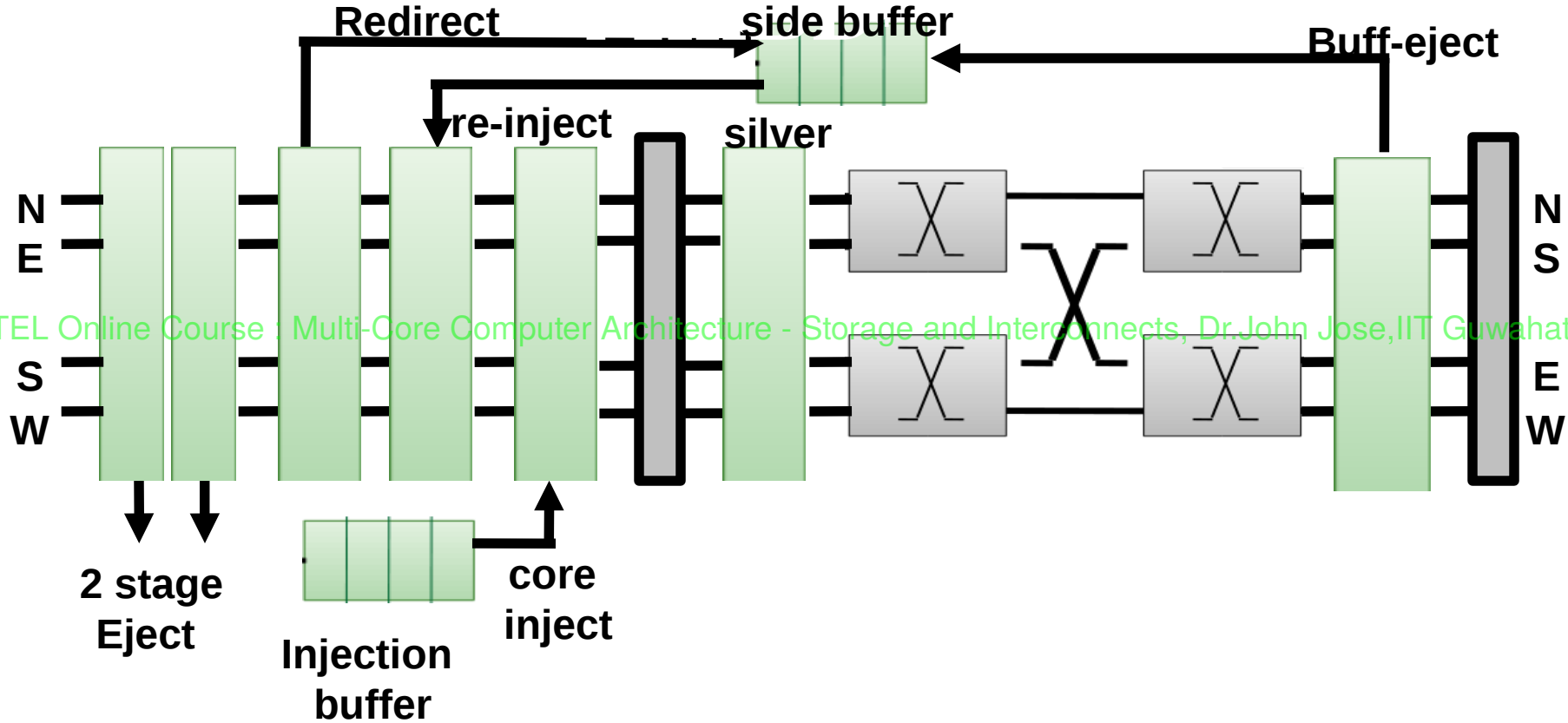
- ❖ No flit is golden but **Red** flit is silver
- ❖ **Red** flit wins at first stage (silver)
- ❖ **Green** flit is deflected at first stage
- ❖ **Red** flit wins at second stage (silver); not deflected



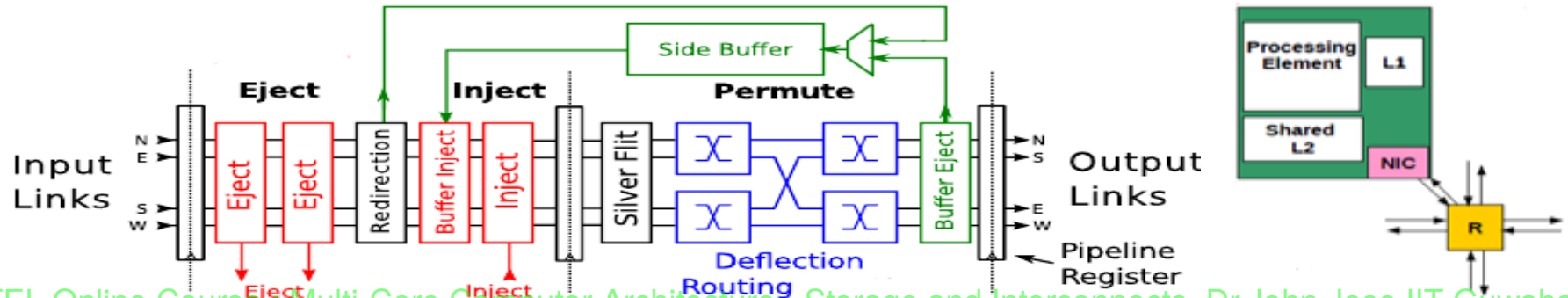
Minimally-Buffered Deflection Router



Idea of Side Buffer & Dual Eject -



MinBD* - Limitations



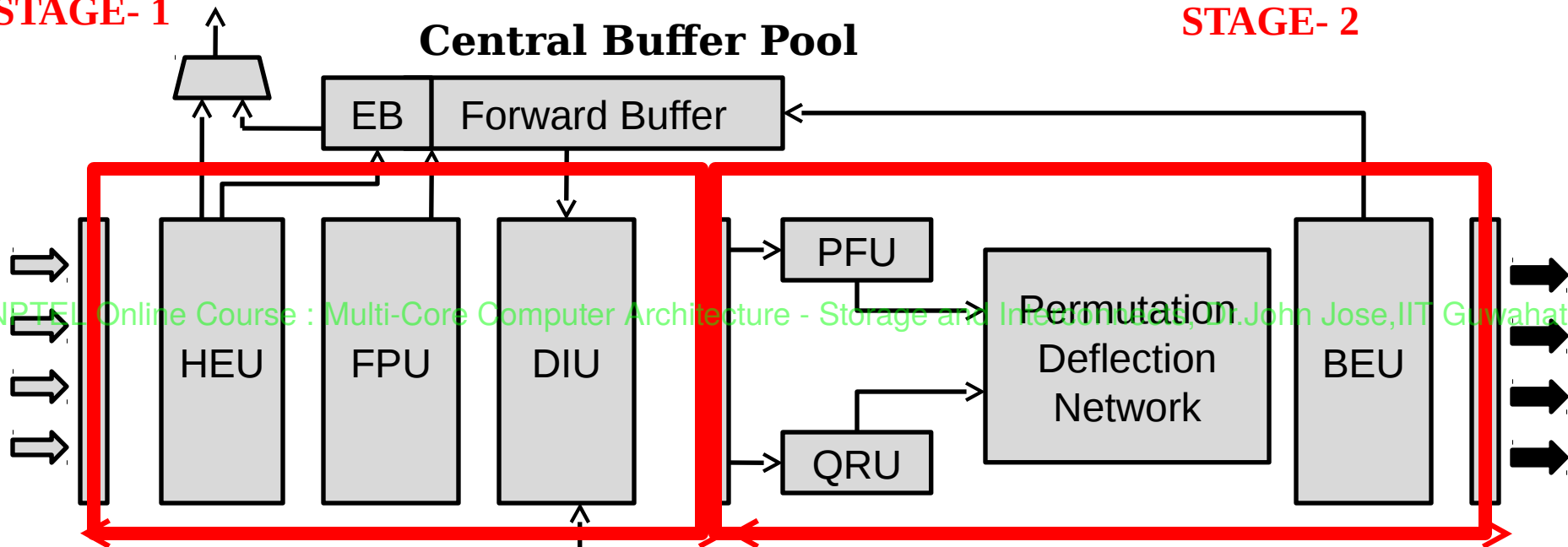
NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati

- ❖ Two separate stages of ejection and dual ejection port.
- ❖ Preemption will take care of starvation from side buffer only.
- ❖ Core buffer starvation is not addressed.
- ❖ Random selection for silver flit – cannot ensure progress.

DeBAR

STAGE- 1

STAGE- 2



HEU: Hybrid Ejection Unit

FPU: Flit Preemption Unit

DIU: Dual Injection Unit

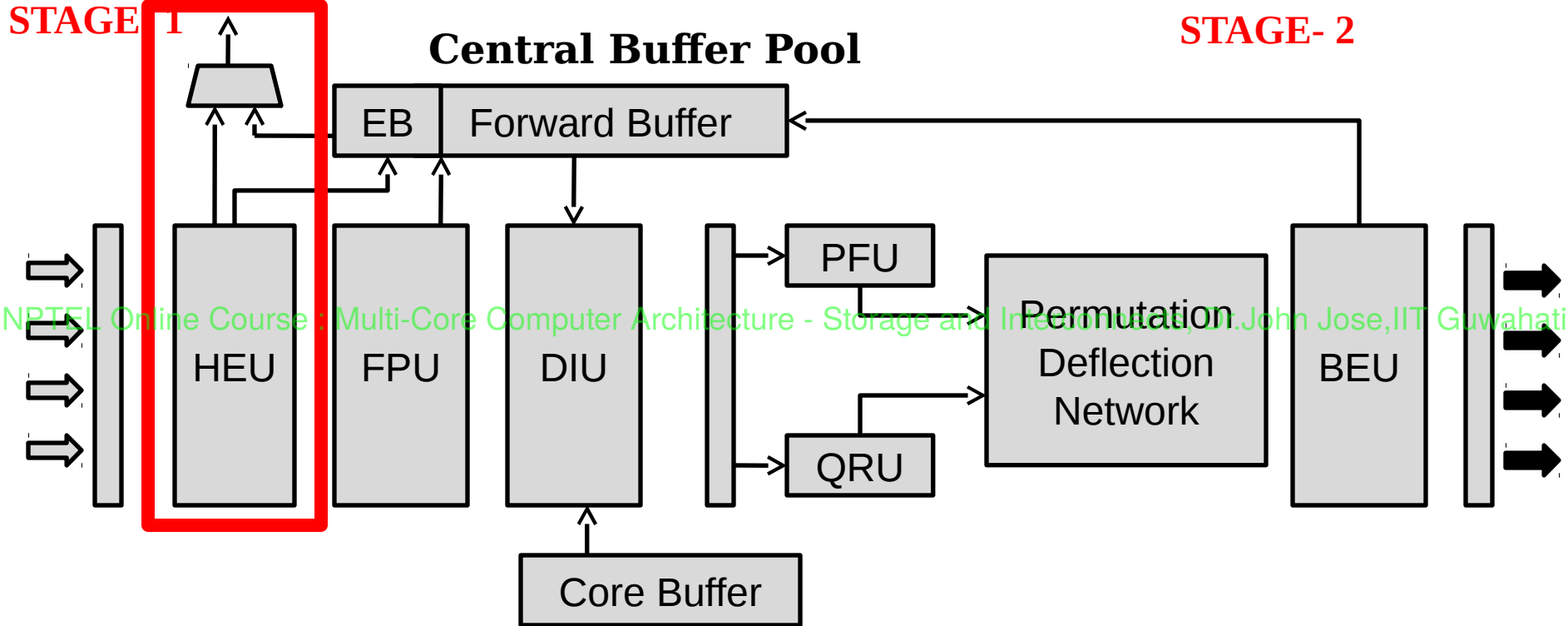
Core Buffer

PFU: Priority Fixer Unit

QRU: Quadrant Routing Unit

BEU: Buffer Ejection Unit

DeBAR Pipeline



The Hybrid Ejection Unit

Ejection Flits



Pass through flits



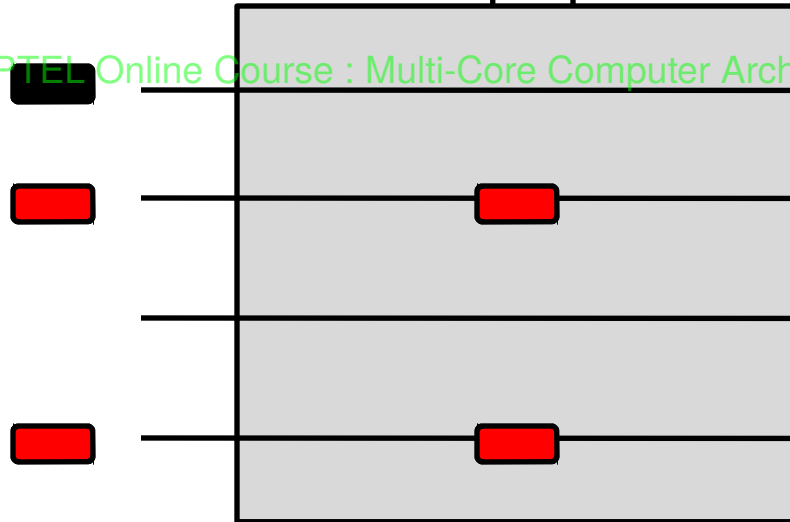
Common Buffer Pool

FB

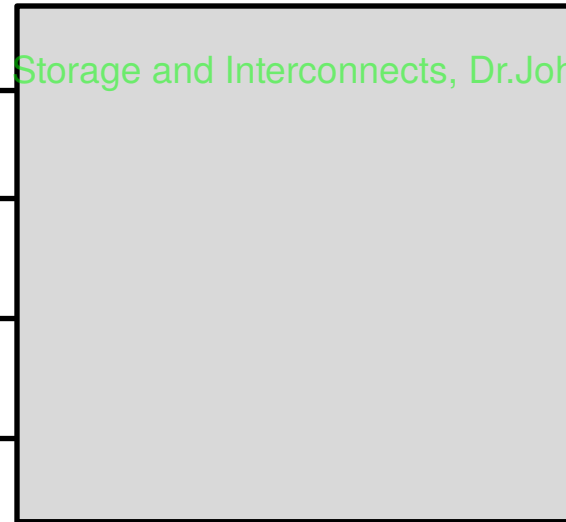
Forward Bank



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



Hybrid Ejection Unit



Flit Preemption Unit

The Hybrid Ejection Unit

Ejection Flits



Pass through flits



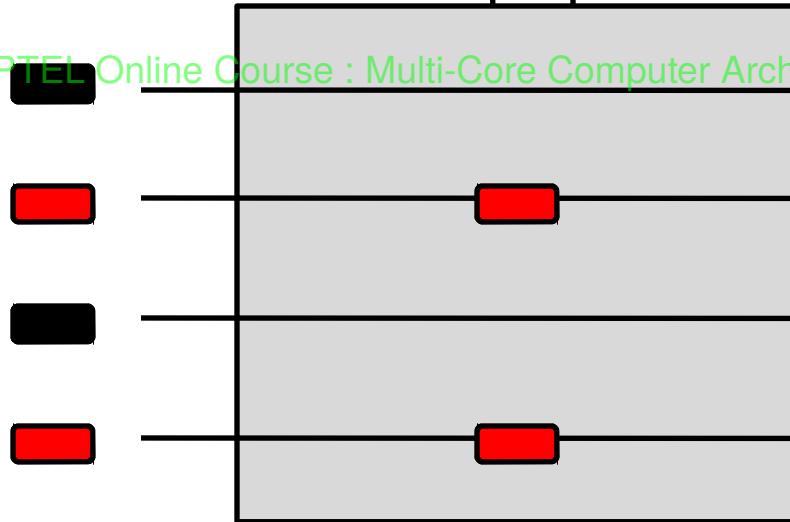
Common Buffer Pool

FB

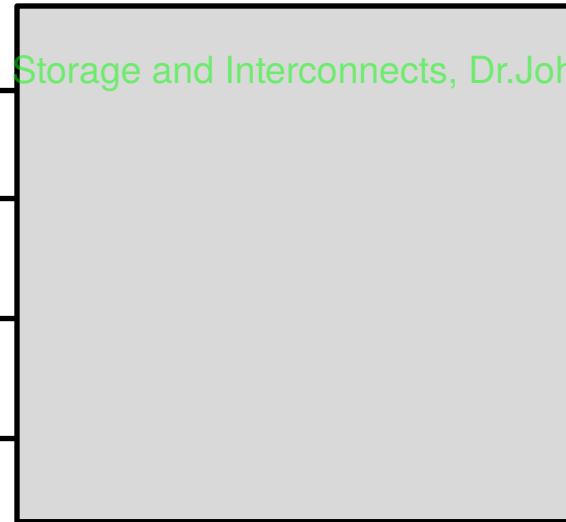
Forward Bank



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



Hybrid Ejection Unit



Flit Preemption Unit

The Hybrid Ejection Unit

Ejection Flits



Pass through flits



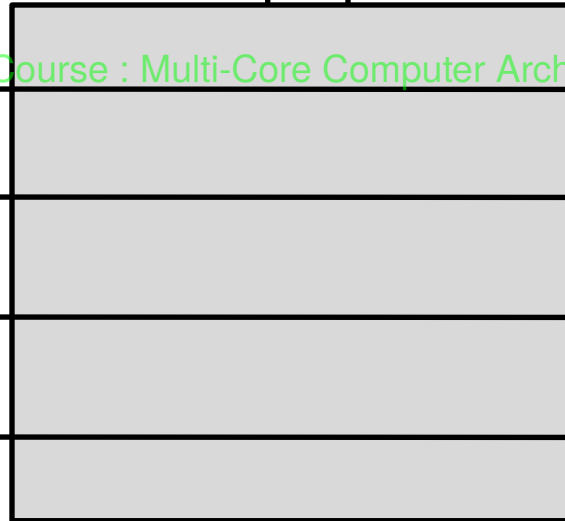
Common Buffer Pool

FB

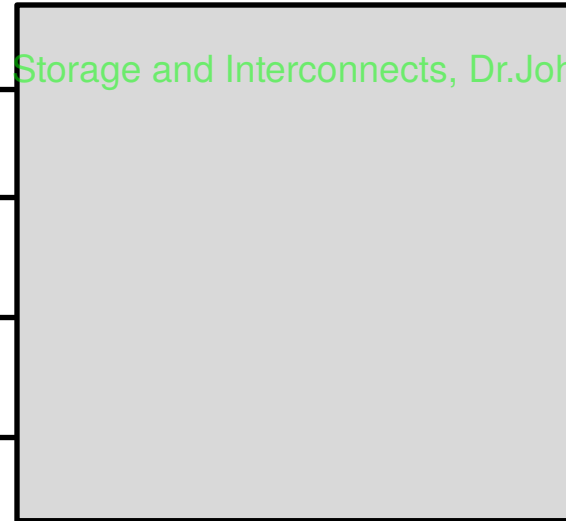
Forward Bank



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



Hybrid Ejection Unit



Flit Preemption Unit

The Hybrid Ejection Unit

Ejection Flits



Pass through flits



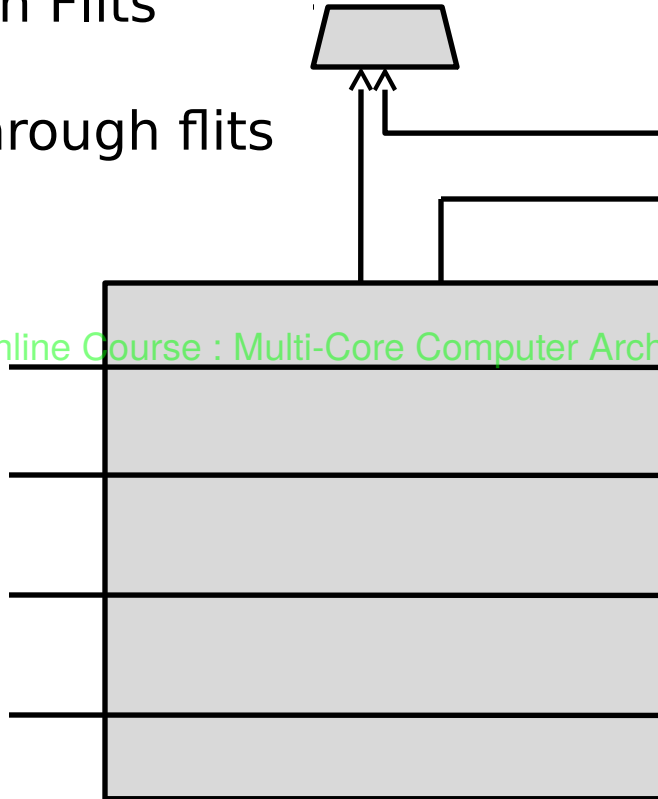
Common Buffer Pool

FB

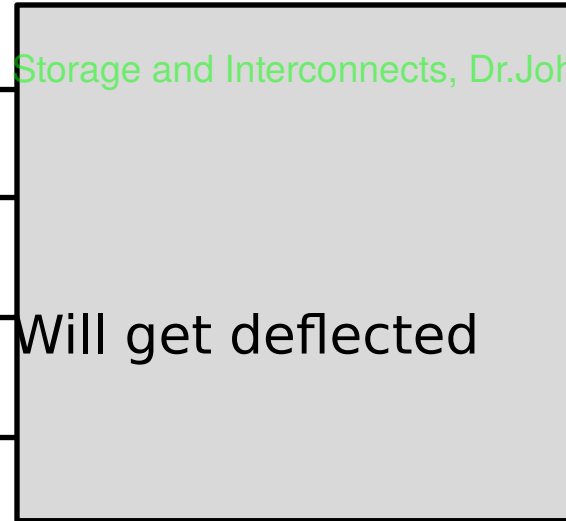
Forward Bank



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



Hybrid Ejection Unit

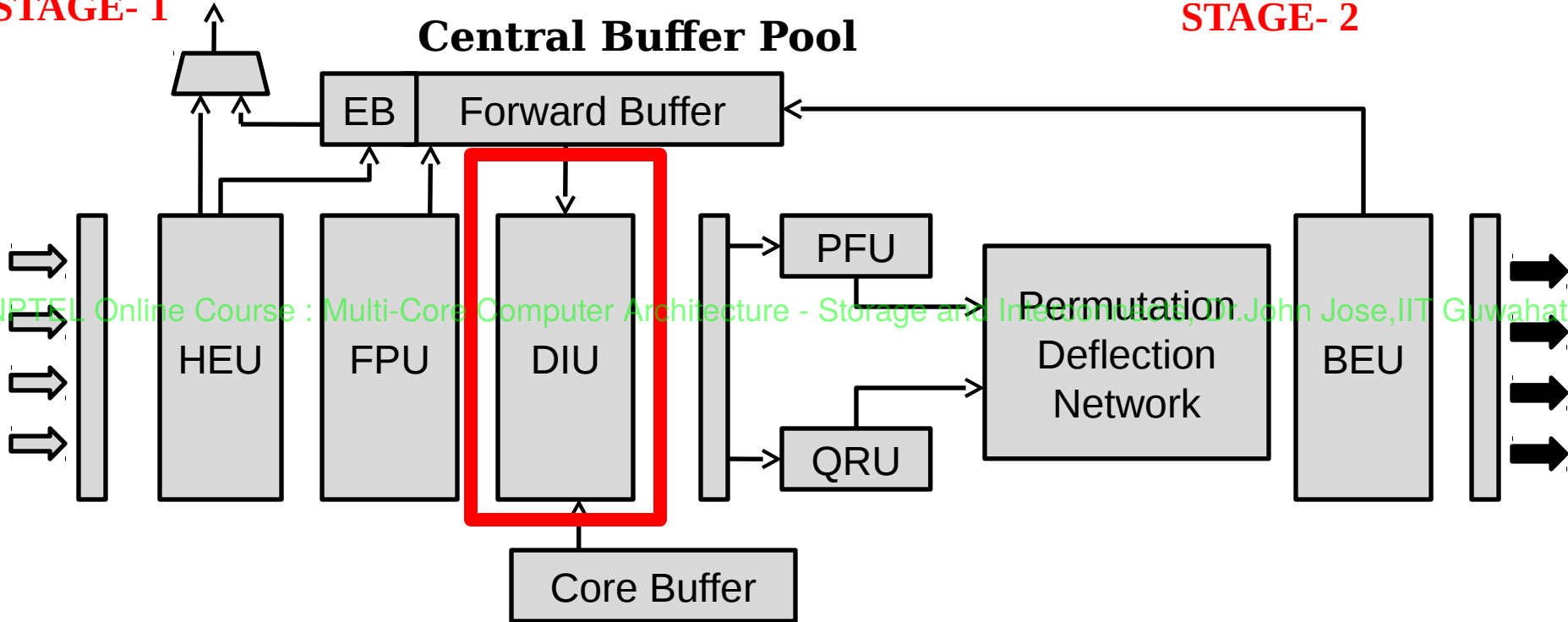


Flit Preemption Unit

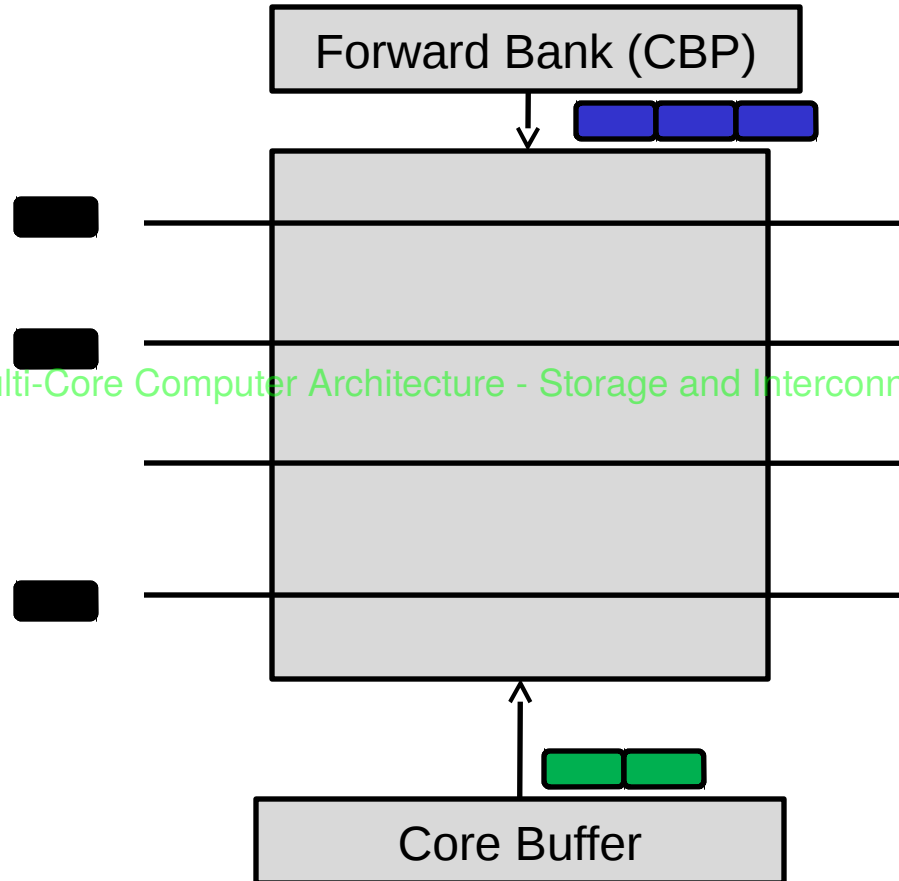
DeBAR Pipeline

STAGE- 1

STAGE- 2

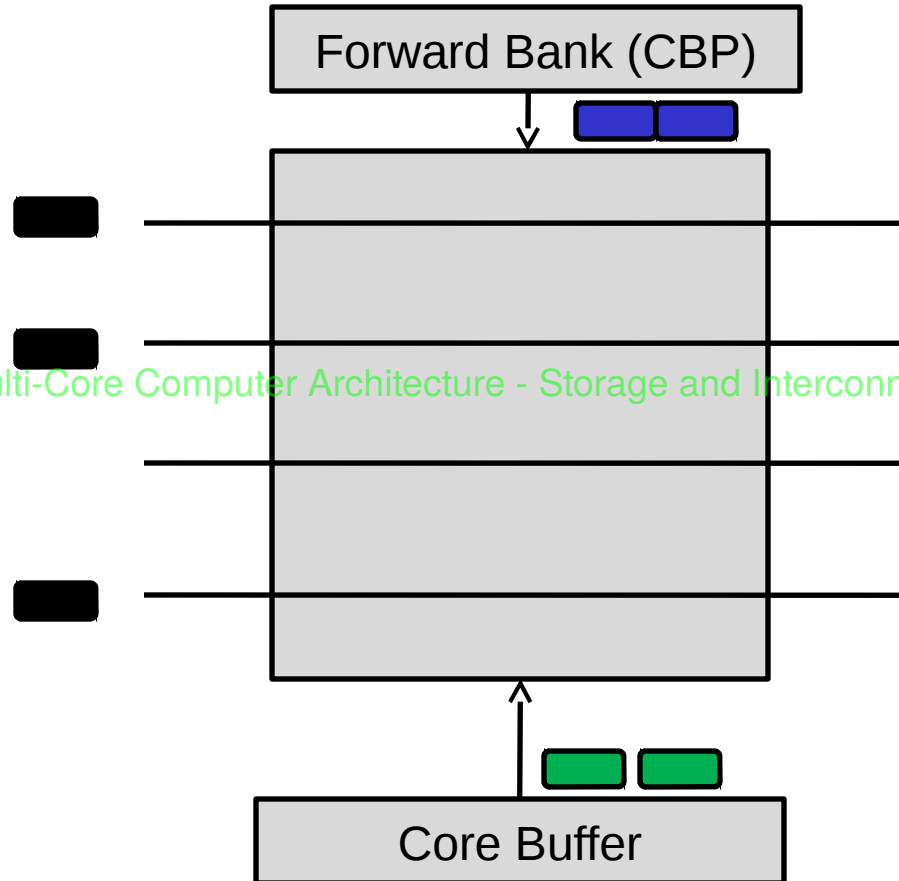


The Dual Injection Unit



ODD CYCLE

The Dual Injection Unit

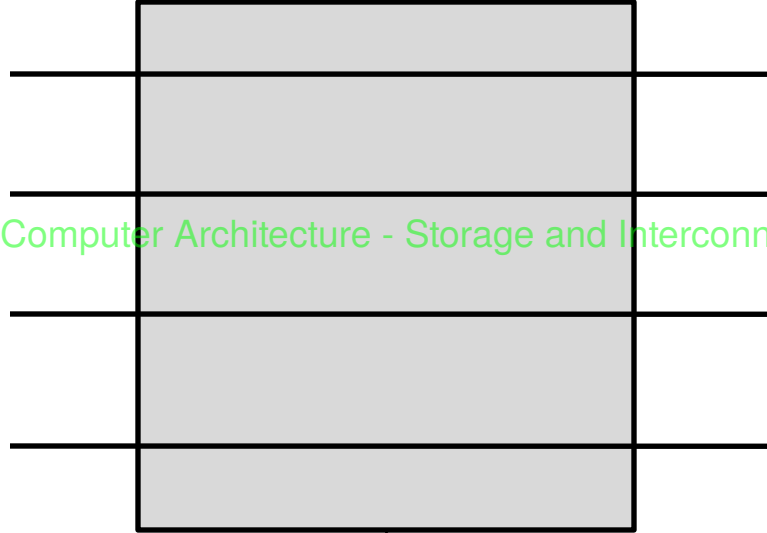


EVEN CYCLE

The Dual Injection Unit

Forward Bank (CBP)

ODD / EVEN CYCLE



Core Buffer

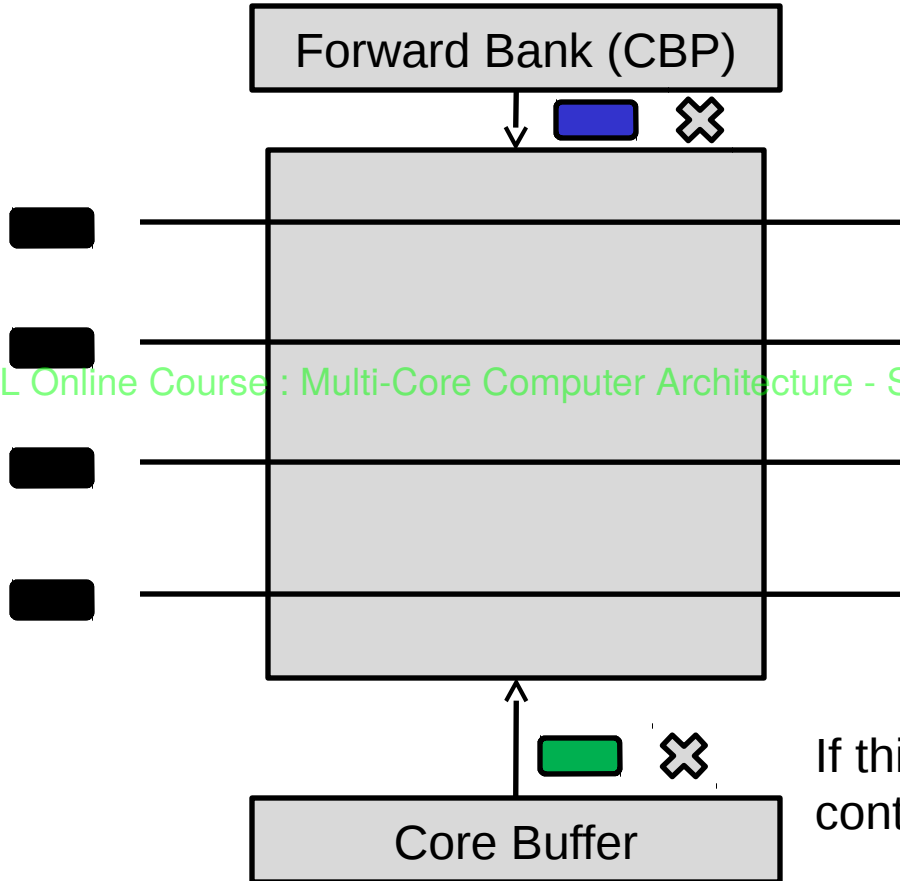
The Dual Injection Unit

ODD / EVEN CYCLE

RII – Re-Inject Interval

CII – Core Inject Interval

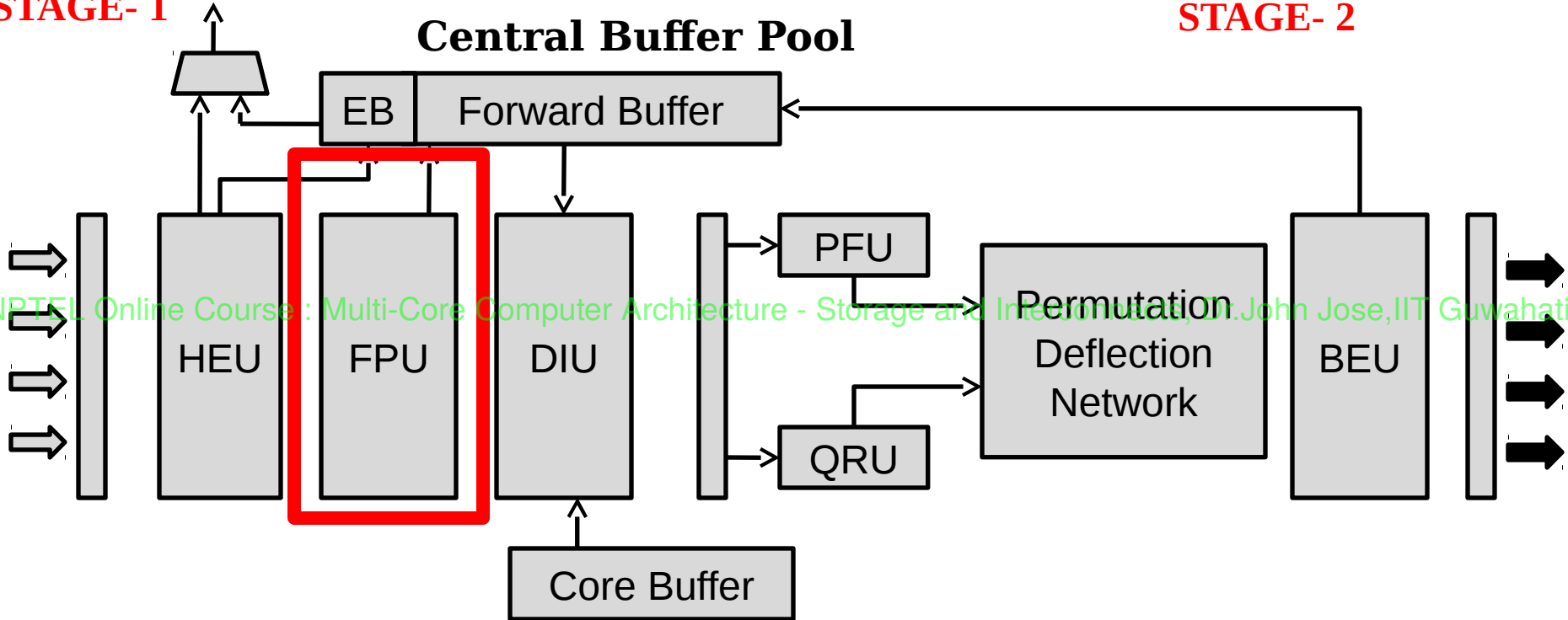
If this scenario continues for multiple continuous cycles then FPU helps



DeBAR Pipeline

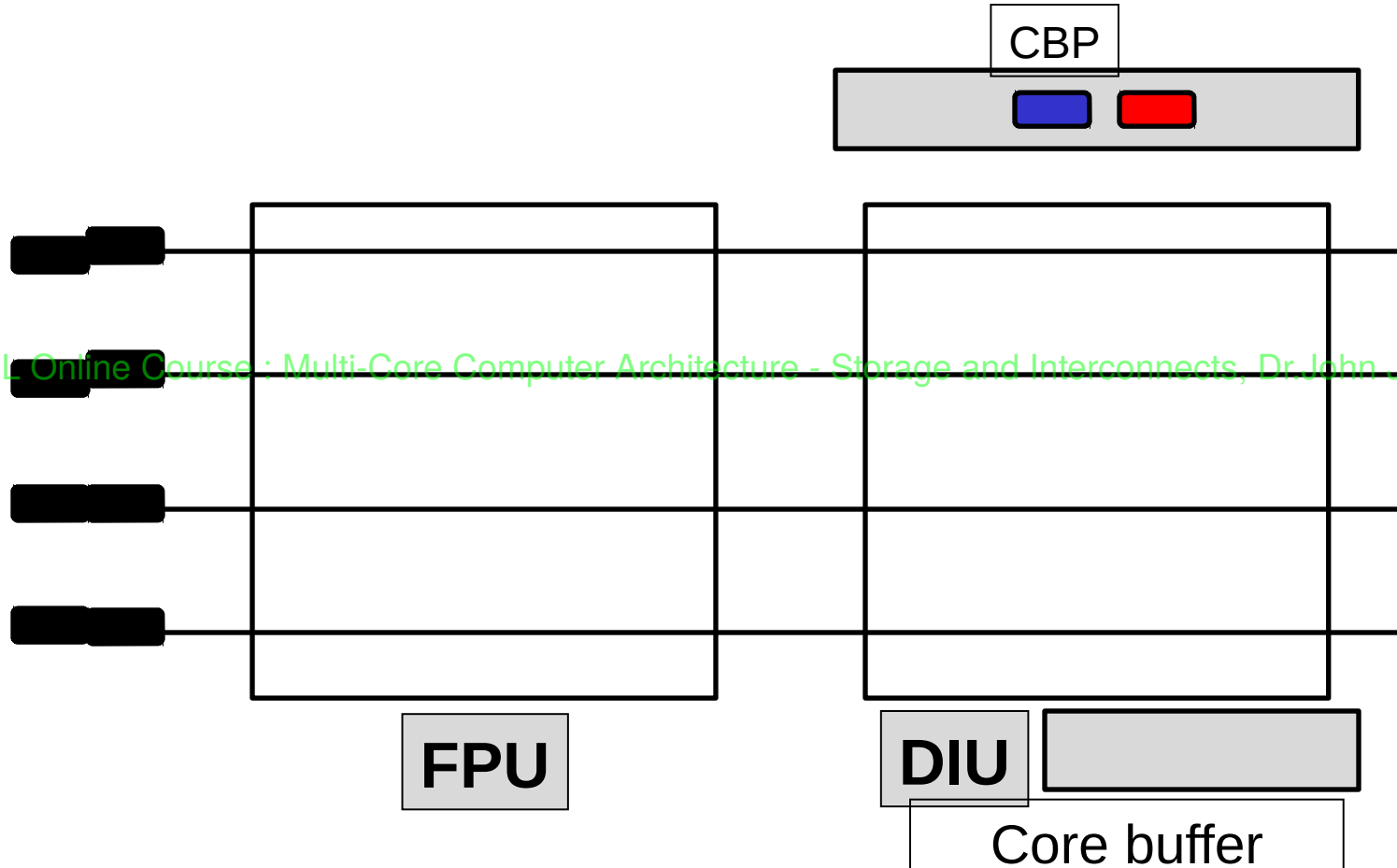
STAGE- 1

STAGE- 2



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

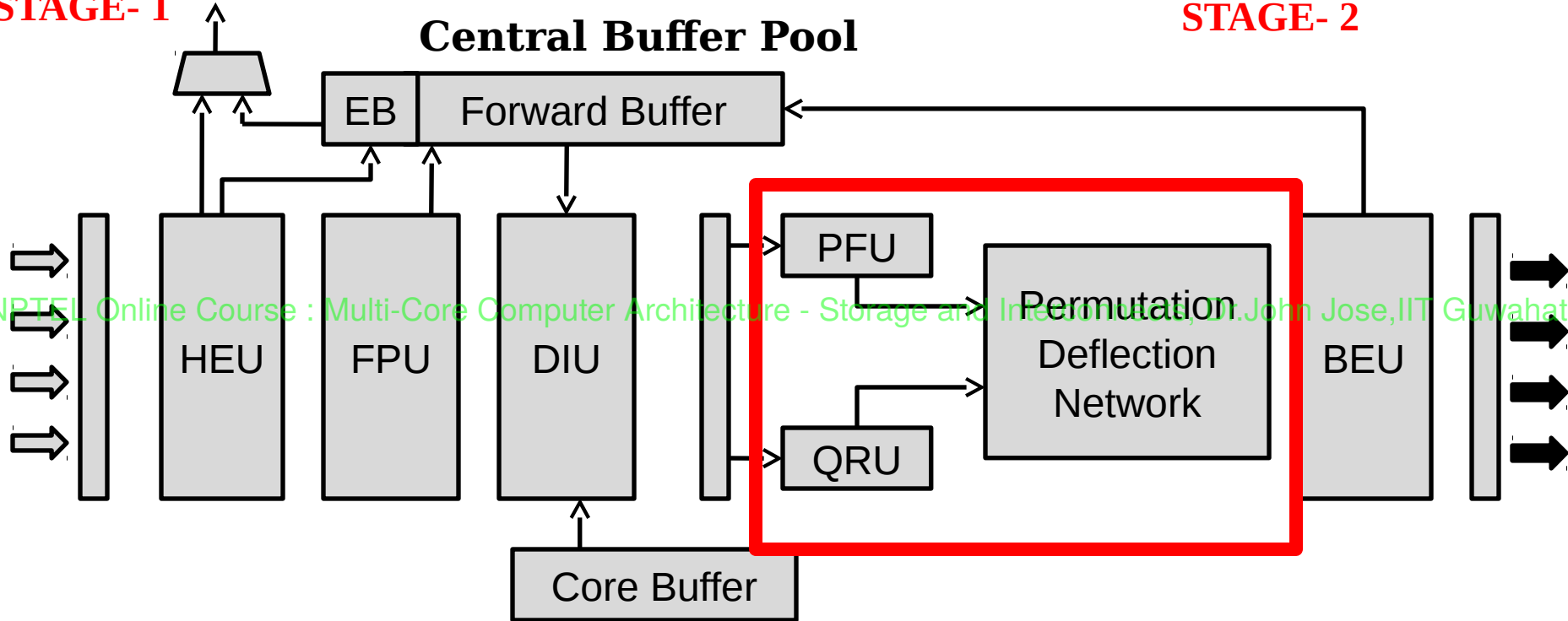
Flit Preemption Unit



DeBAR Pipeline

STAGE- 1

STAGE- 2

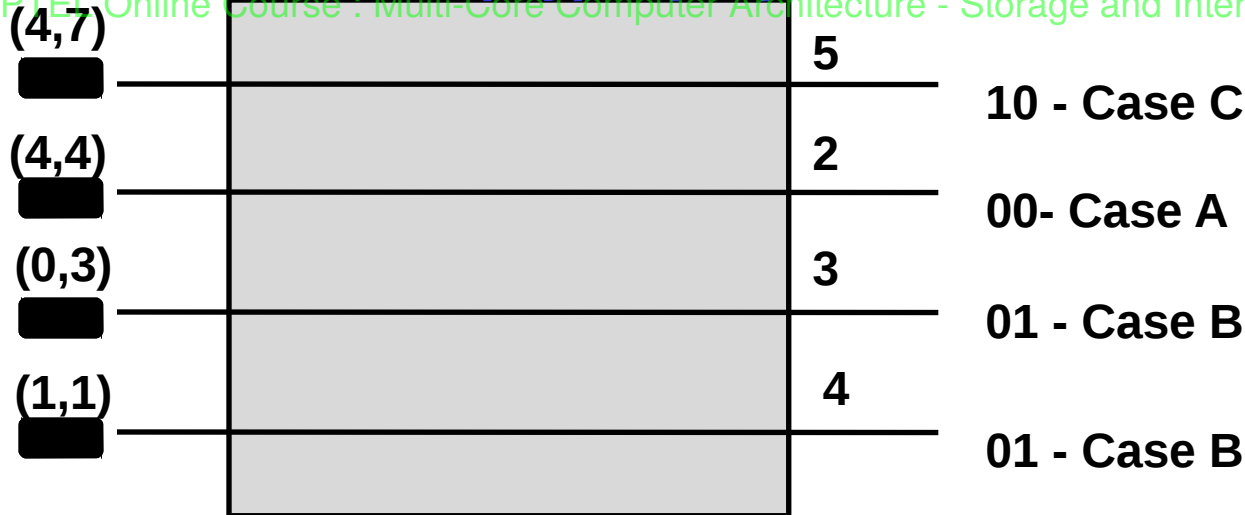


Priority Fixing Unit

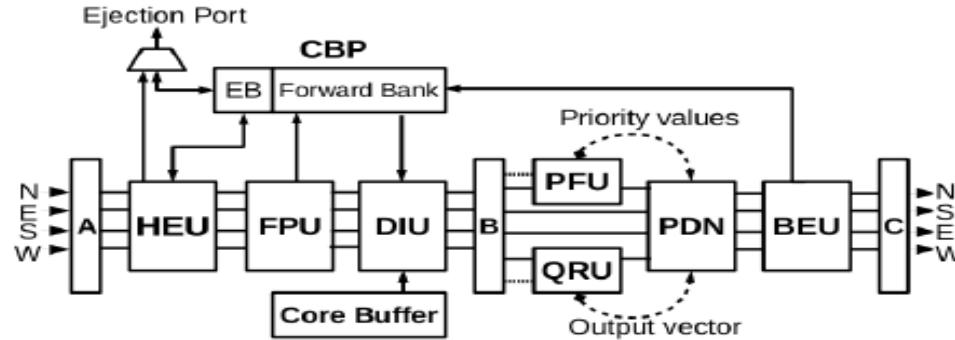
The hops to destination (D) is computed and encoded into one among the 3 cases.

Case	# hops to D	Priority value
A (high)	≤ 2	00
B (med)	$>2 \ \&\leq 4$	01
C (low)	> 4	10

Router (x,y) = (3,3)

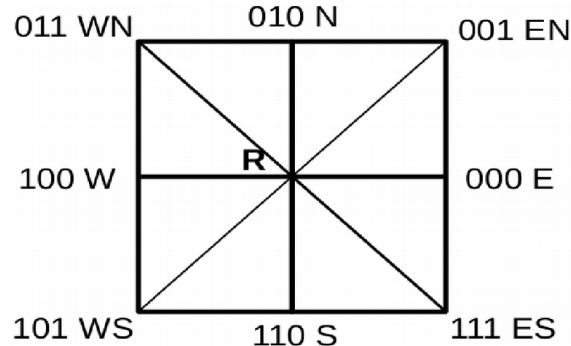


Quadrant Routing and Marking



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

**Output vector generation
using quadrant routing**



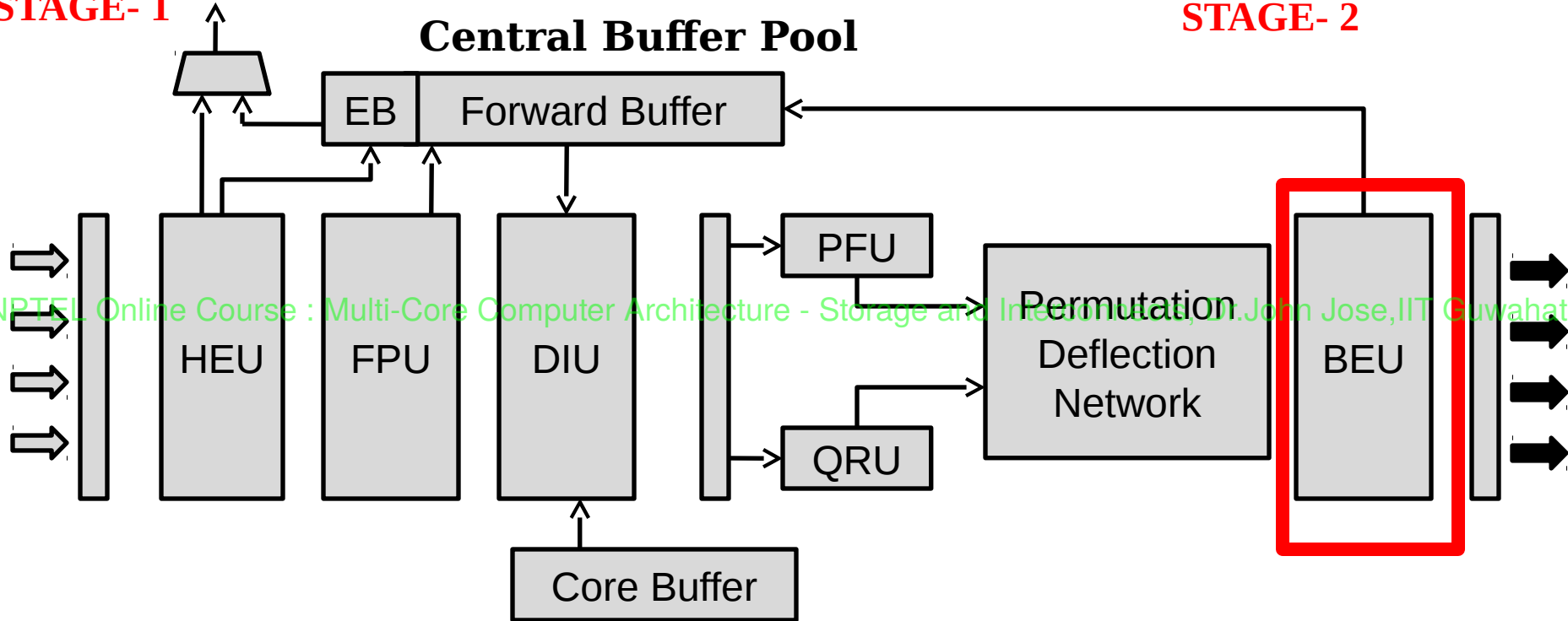
**Marking logic from output
vector and PDV port allocation**

Port Allocation in PDN	Output Vector							
	000 E	001 EN	010 N	011 WN	100 W	101 WS	110 S	111 ES
East (000)	0	0	1	1	1	1	1	0
North (010)	1	0	0	0	1	1	1	1
West (100)	1	1	1	0	0	0	1	1
South (110)	1	1	1	1	1	0	0	0

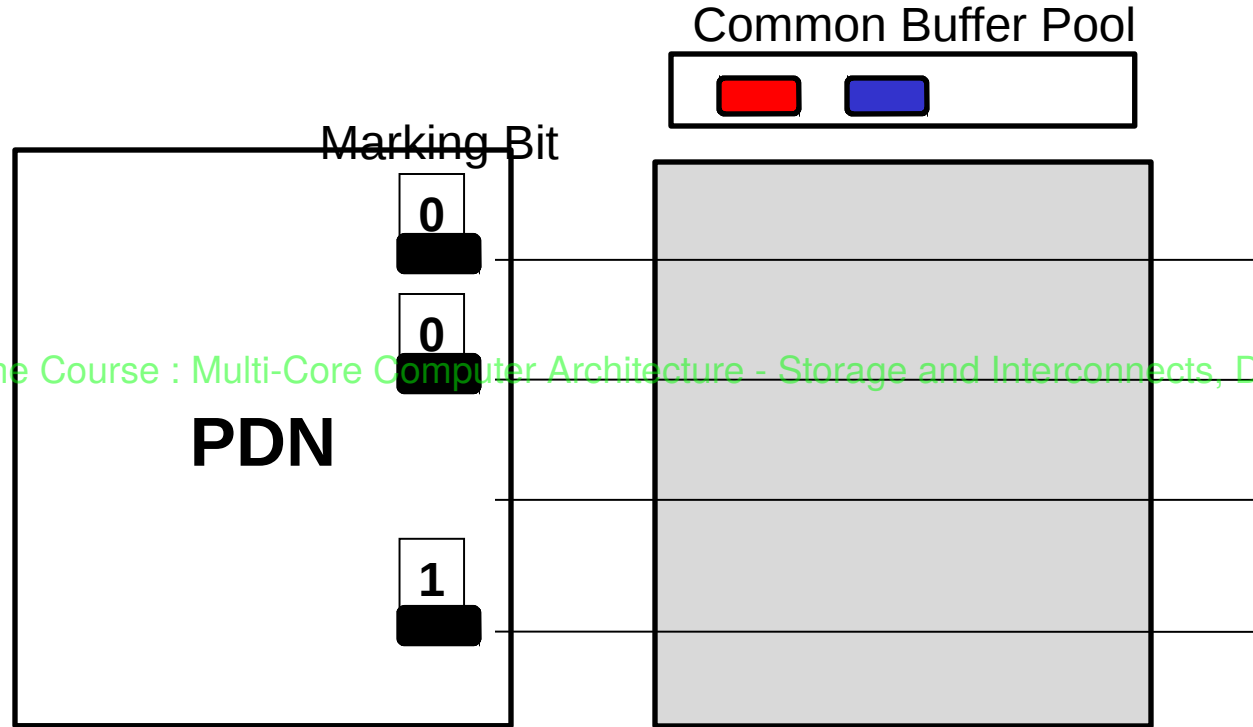
DeBAR Pipeline

STAGE- 1

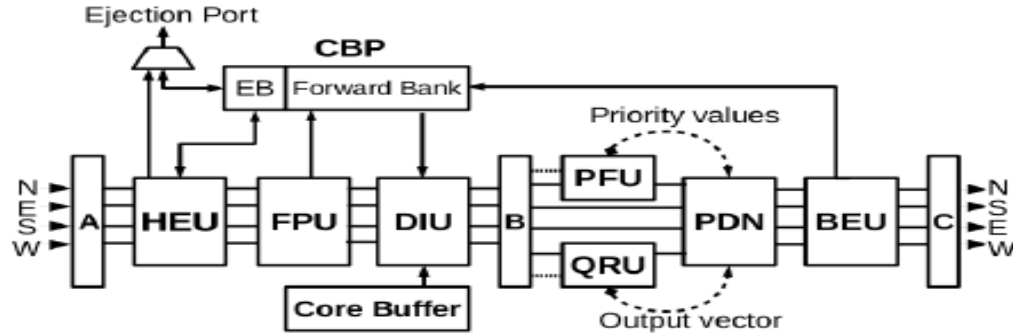
STAGE- 2



Buffer Eject Unit



DeBAR - Salient Features



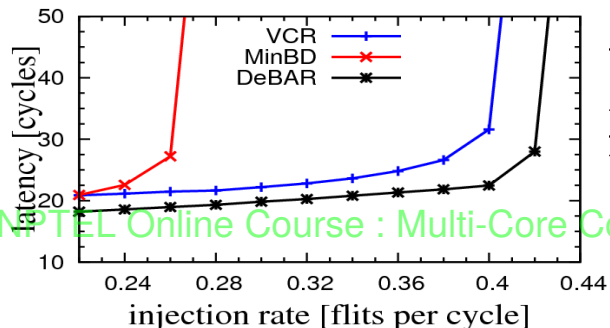
NPTEL Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr.John Jose,IIT Guwahati

- ❖ Dual ejection support without a 2 stage ejection unit.
- ❖ Ensures fairness in progress of flits from CBP and core buffer.
- ❖ Priority computation and output vector generation in parallel .
- ❖ PDN does port allocation and flit marking.

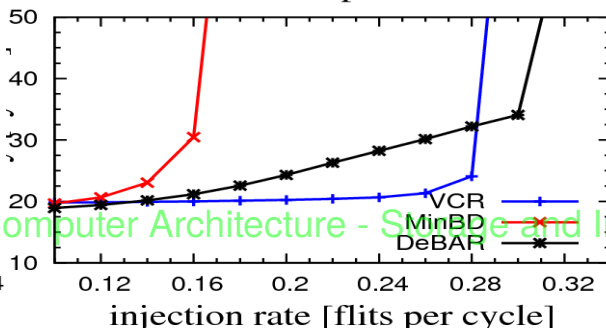
Performance Comparison

Average packet latency & deflection rate in 8x8 mesh using Booksim

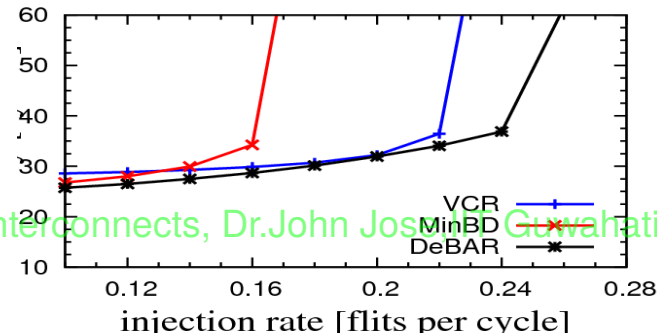
Uniform



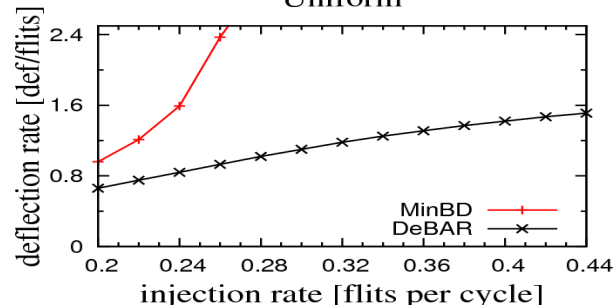
Transpose



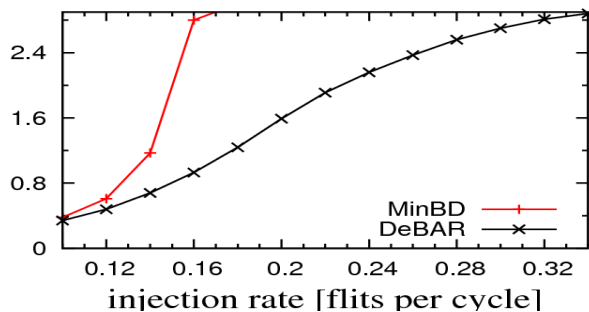
Bit Complement



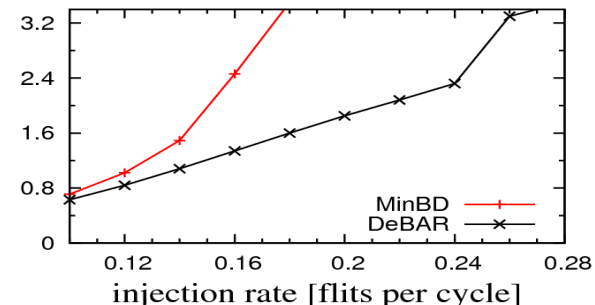
Uniform



Transpose

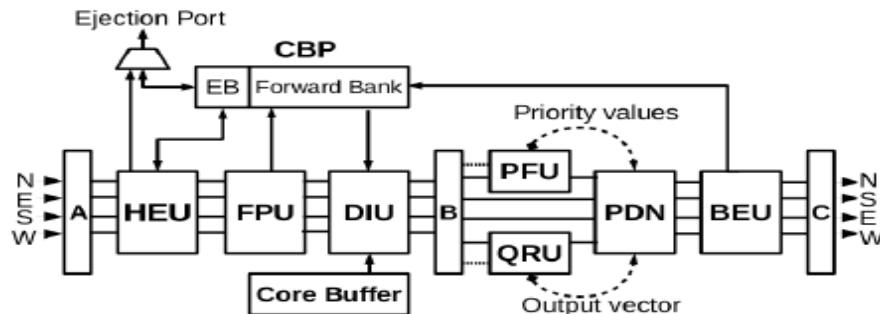
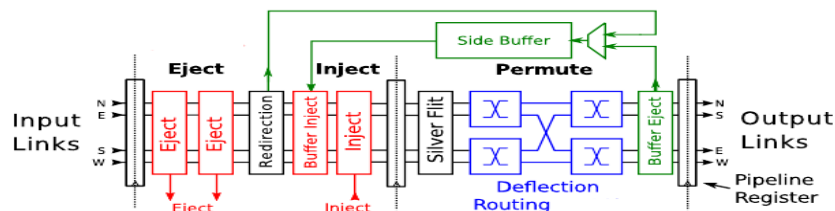


Bit Complement



Hardware Overhead Impact

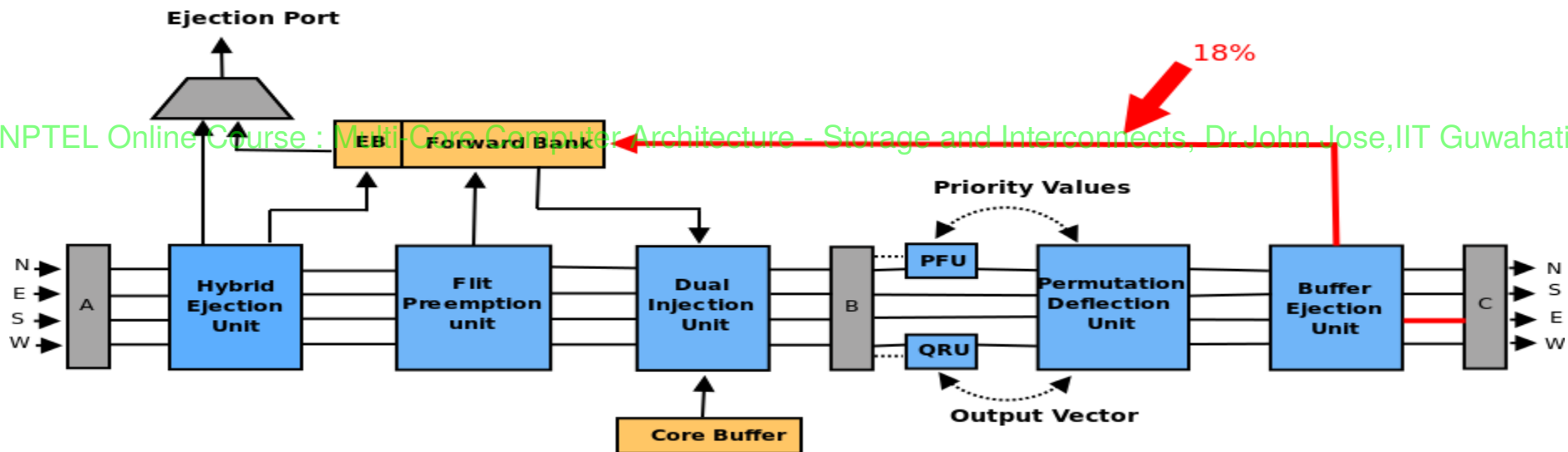
Analysis of effect on router critical path, power and area



- NPTel Online Course : Multi-Core Computer Architecture - Storage and Interconnects, Dr. John Jose, IIT Guwahati
- ❖ 17% latency reduction in stage -1
 - ❖ Overall latency of stage-2 is same for DeBAR and MinBD.
 - ❖ Since stage 2 dominates over stage 1 in latency, DEBAR has the same operating frequency as that of MinBD.
 - ❖ 18% reduction in channel wiring due to single ejection port

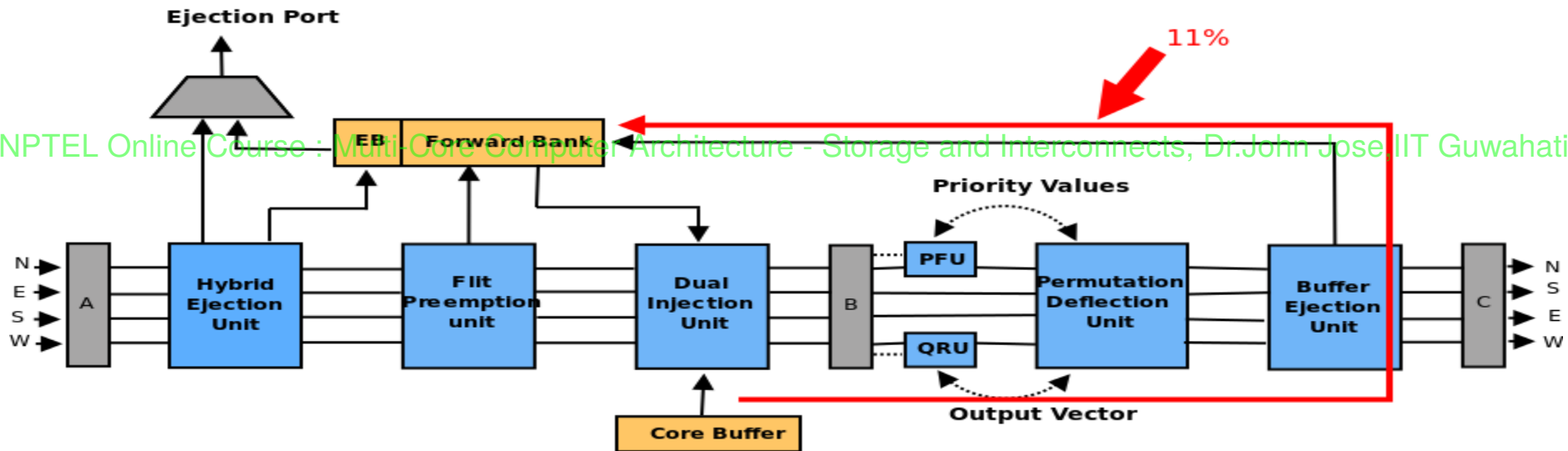
Limitations of DeBAR

❖ Problem of channel wastage



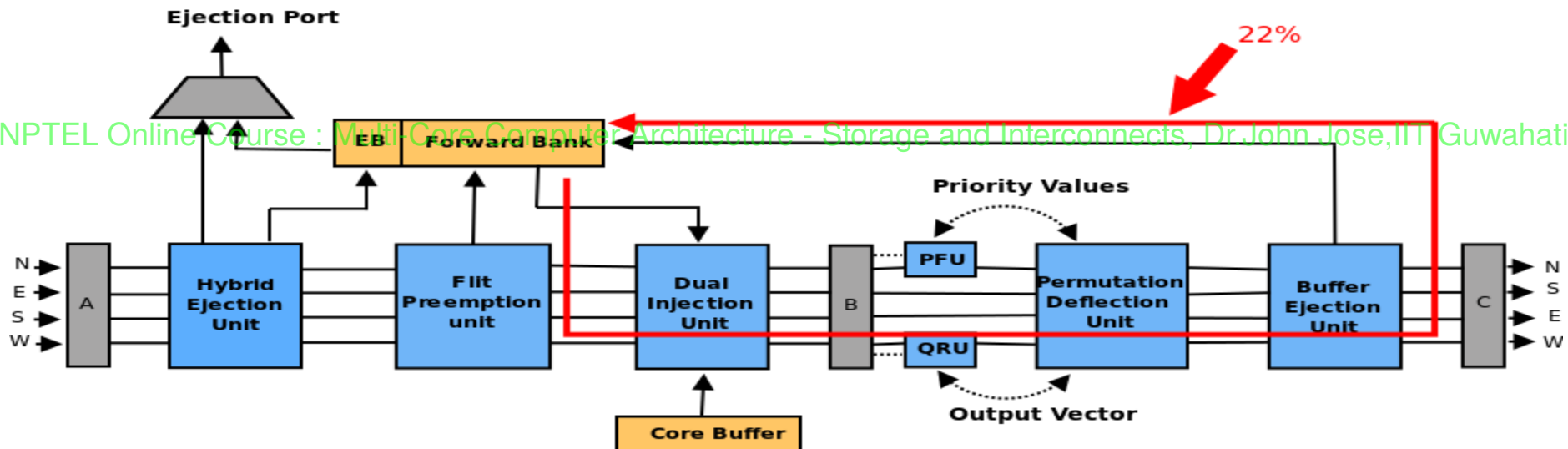
Limitations of DeBAR

❖ Problem of internal movements of flits (Core buffer to Side buffer)



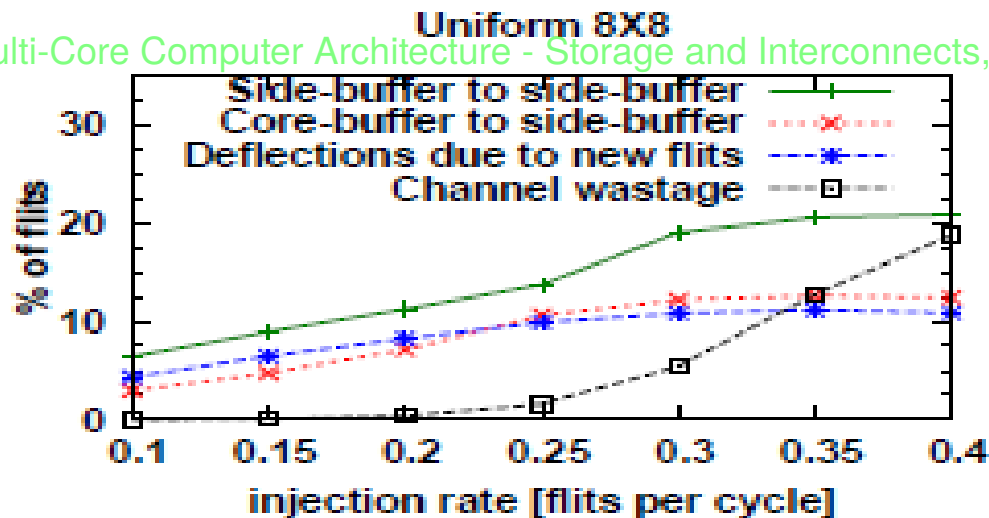
Limitations of DeBAR

❖ Problem of internal movement of flits (Side buffer to Side buffer)



Limitations of DeBAR

- ❖ Problem of channel wastage
- ❖ Problem of internal movements of flits
- ❖ Problem of older flit penalization

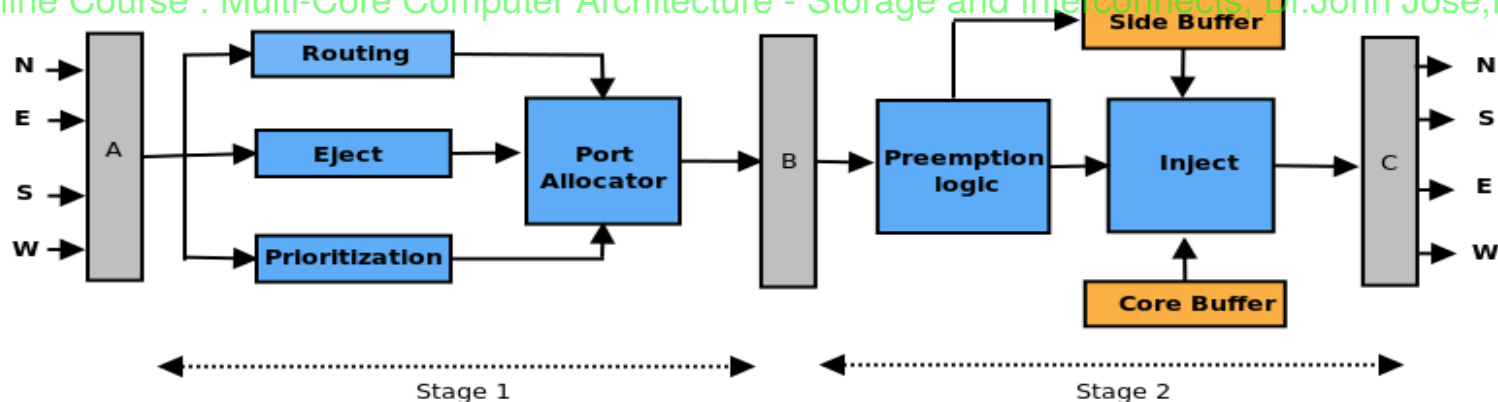


Proposed Router - SLIDER

SLIDER: Smart Late Injection DEflection Router

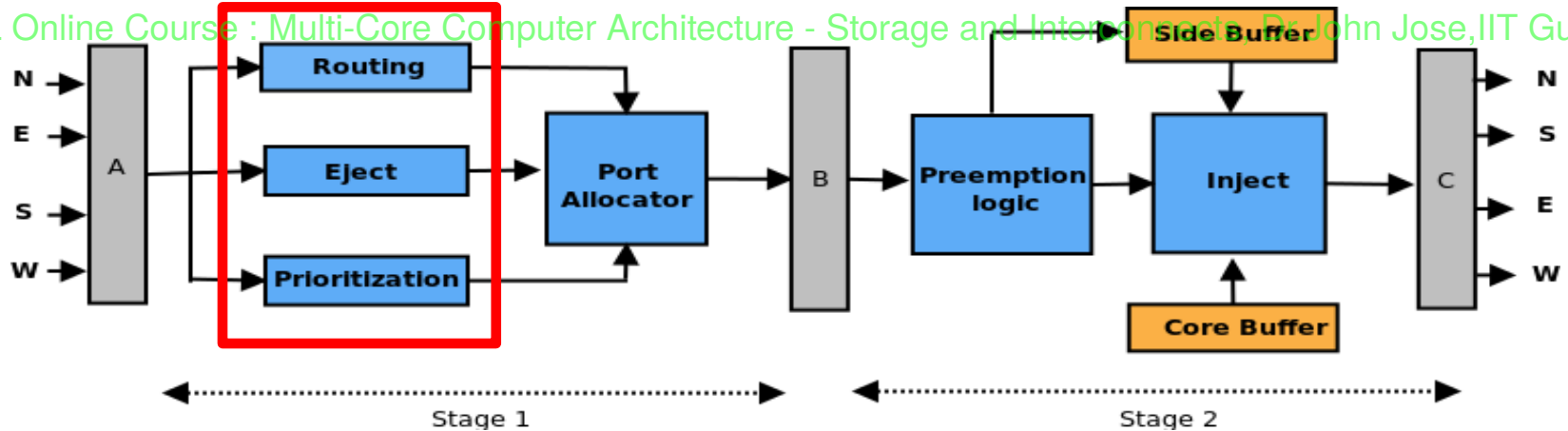
- ❖ Parallelizing independent operations
- ❖ Concept of selective preemption
- ❖ Concept of smart late injection

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



SLIDER - Features

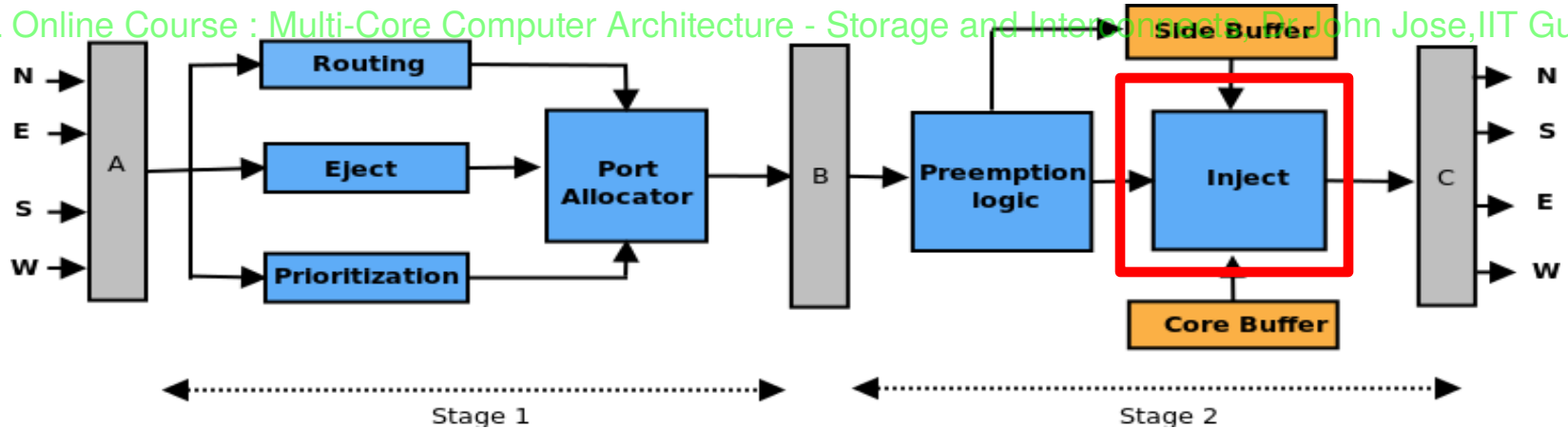
❖ Parallelizing independent operations



SLIDER - Features

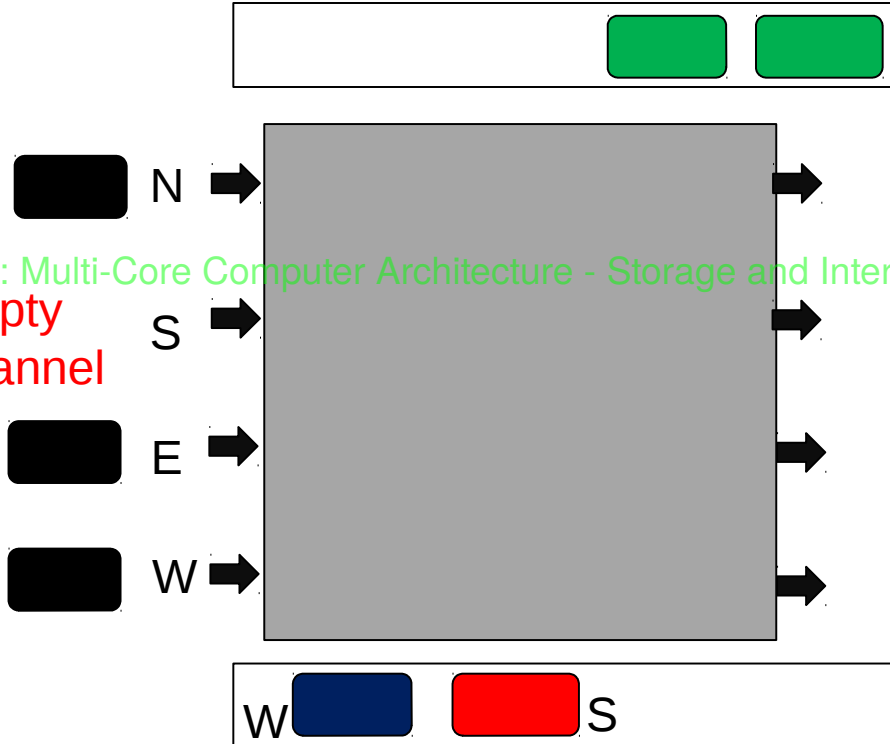
❖ Concept of smart late injection

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



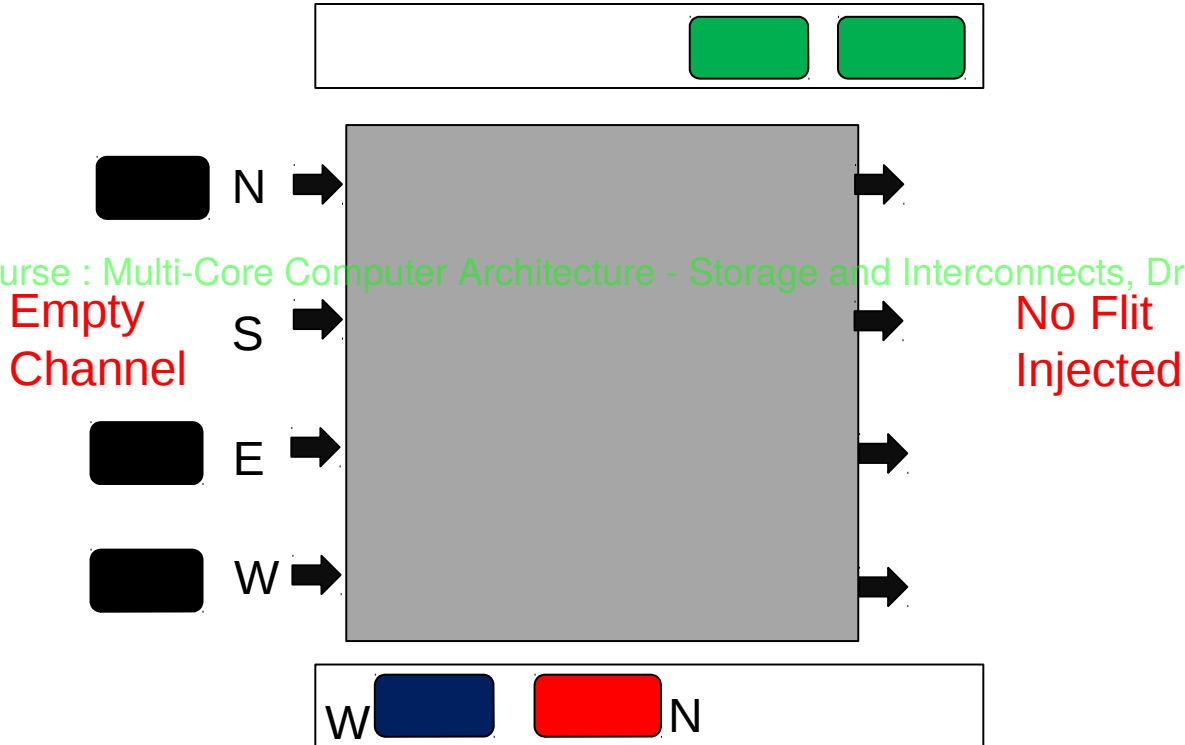
Inject Unit

❖ Restricted Injection



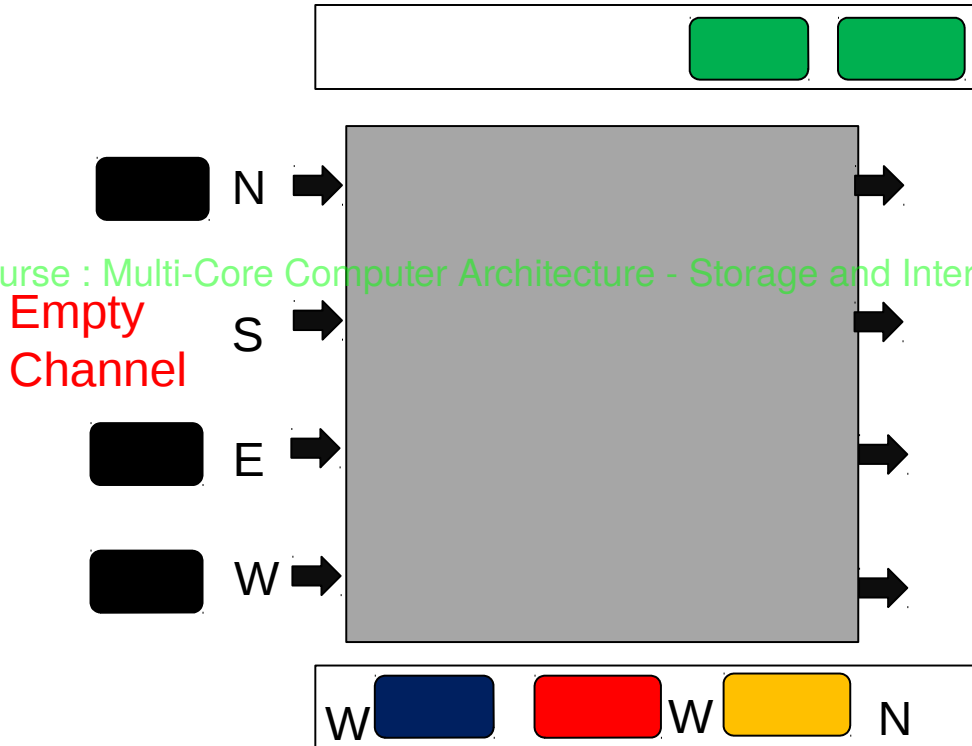
Inject Unit

❖ Restricted Injection

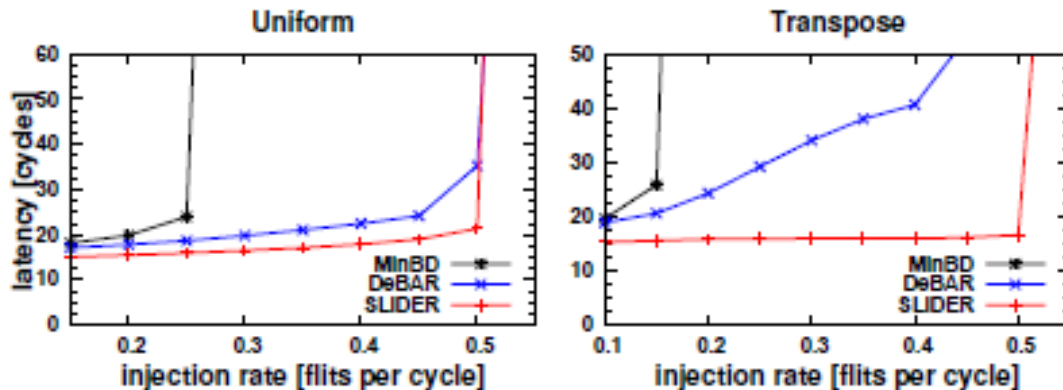
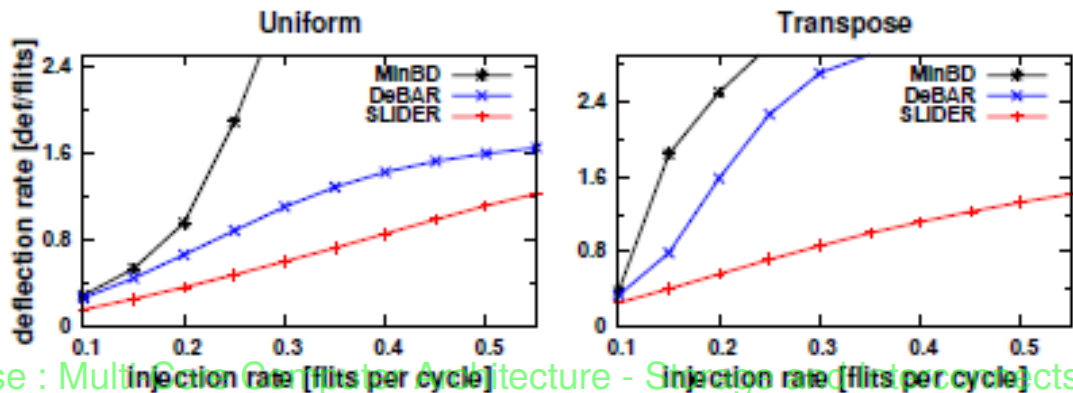


Inject Unit

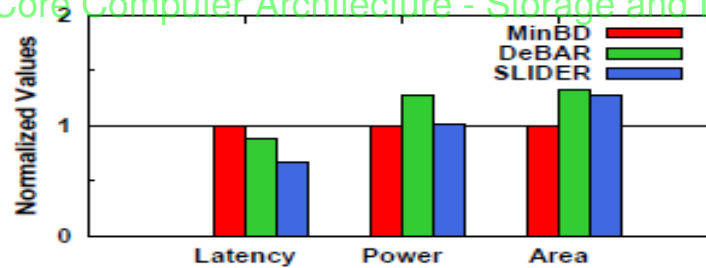
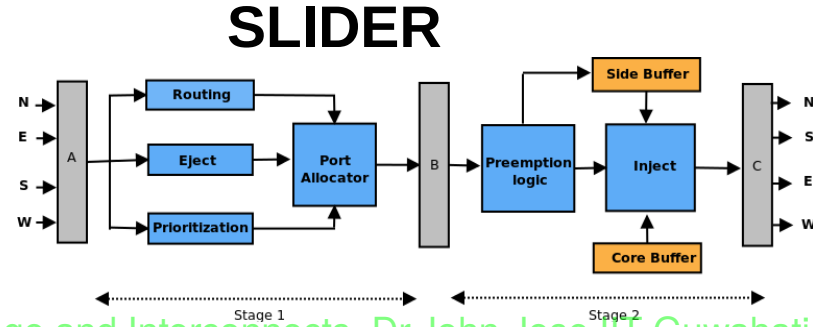
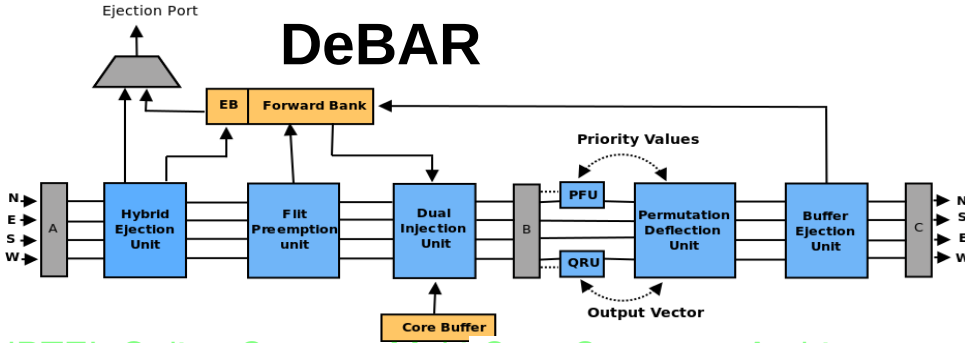
❖ Non-Restricted Injection



Performance Comparison

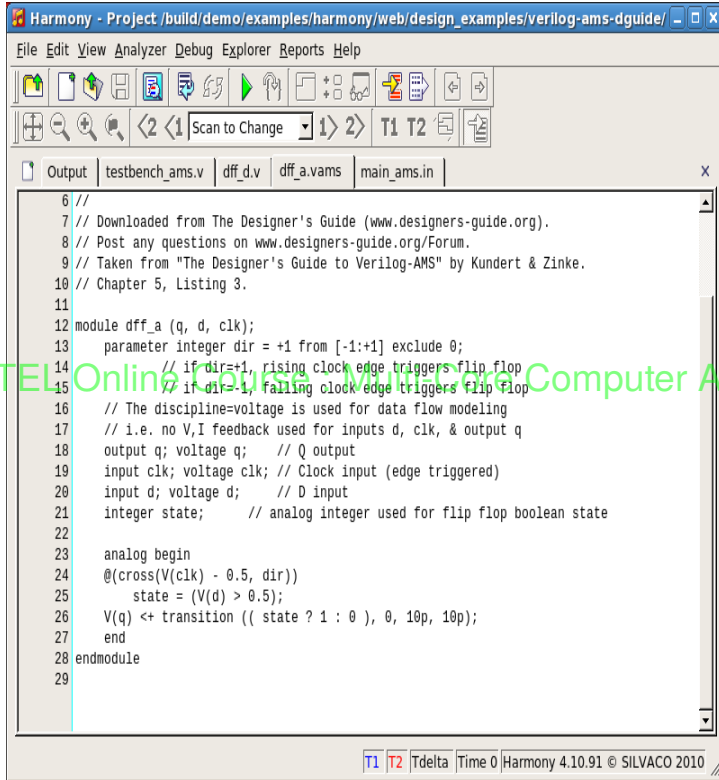


Experimental Analysis



- ❖ Router pipeline latency reduction 25% w.r.t. DeBAR.
- ❖ Power reduction 19% and area reduction 3.5% w.r.t. DeBAR.

Hardware Design Verification



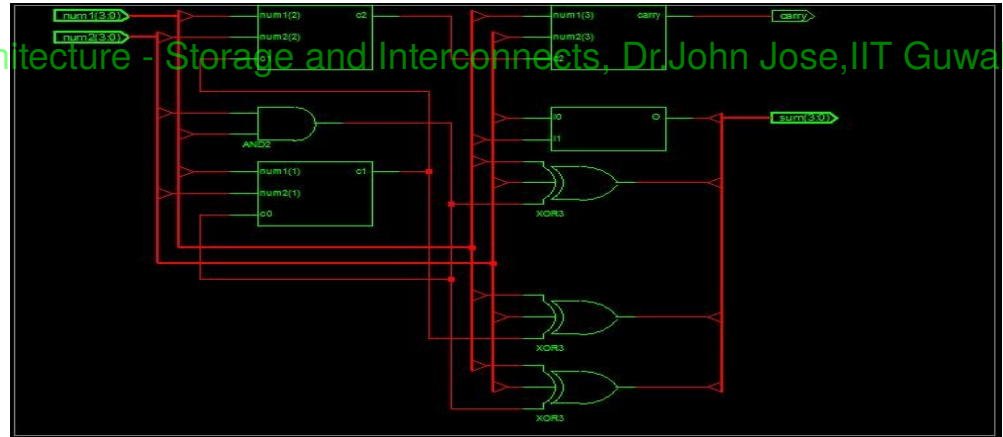
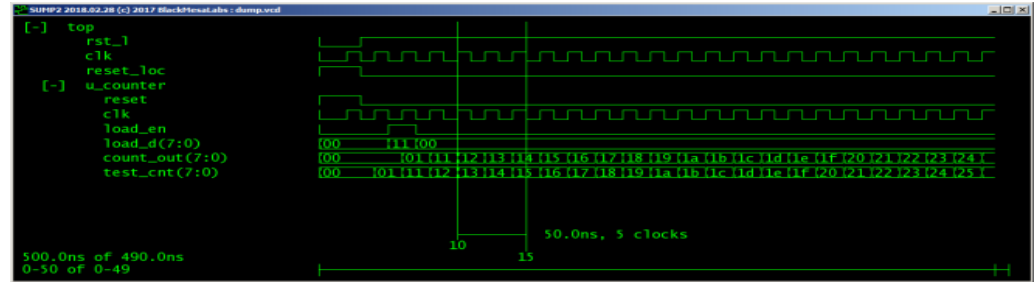
Harmony - Project /build/demo/examples/harmony/web/design_examples/verilog-ams-dguide/

File Edit View Analyzer Debug Explorer Reports Help

Output testbench_ams.v dff_d.v dff_a.vams main_ams.in

```
6 //
7 // Downloaded from The Designer's Guide (www.designers-guide.org).
8 // Post any questions on www.designers-guide.org/Forum.
9 // Taken from "The Designer's Guide to Verilog-AMS" by Kundert & Zinke.
10 // Chapter 5, Listing 3.
11
12 module dff_a (q, d, clk);
13     parameter integer dir = +1 from [-1:+1] exclude 0;
14     // if dir=+1, rising clock edge triggers flip flop
15     // if dir=-1, falling clock edge triggers flip flop
16     // The discipline=voltage is used for data flow modeling
17     // i.e. no V,I feedback used for inputs d, clk, & output q
18     output q; voltage q; // Q output
19     input clk; voltage clk; // Clock input (edge triggered)
20     input d; voltage d; // D input
21     integer state; // analog integer used for flip flop boolean state
22
23     analog begin
24         @(cross(V(clk) - 0.5, dir))
25             state = (V(d) > 0.5);
26         V(q) <+ transition (( state ? 1 : 0 ), 0, 10p, 10p);
27     end
28 endmodule
29
```

T1 T2 Delta Time 0 Harmony 4.10.91 © SILVACO 2010



Design Implementation in FPGA

Project: LDPC_OP
Solution: solution1
Product family: kintex7
Target device: xc7k325tffg900-2

Performance Estimates

Timing (ns)

Summary

Clock	Target	Estimated	Uncertainty
ap_clk	10.00	8.58	1.25

Latency (clock cycles)

Summary

Latency		Interval		Type
min	max	min	max	
214	242	215	243	none

Detail

Instance

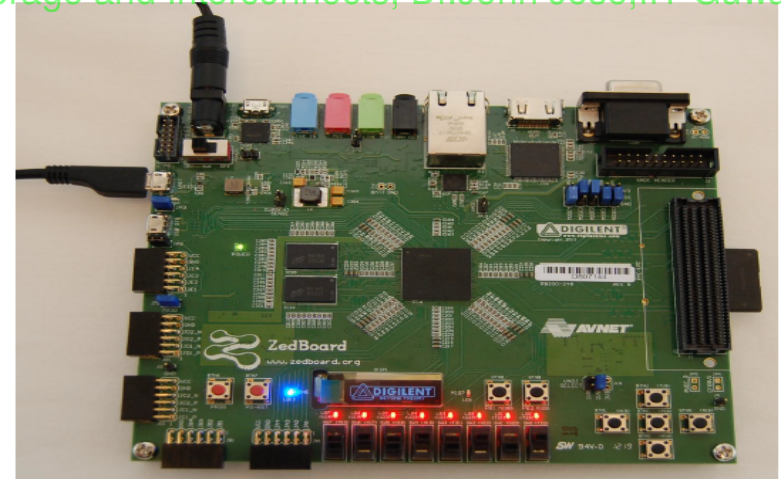
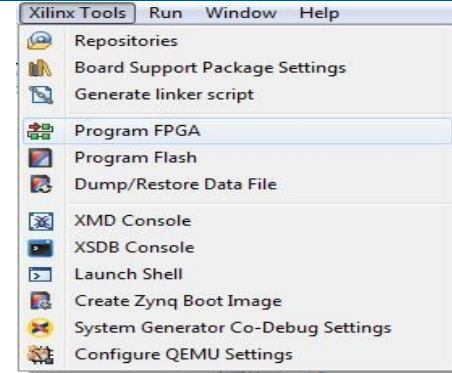
Loop

Loop Name	Latency		Iteration	Initiation Interval		Trip Count	Pipelined
	min	max		achieved	target		
- memset_countArray	7	7	2	-	-	4	no
- memset_fArray	7	7	2	-	-	4	no
- Loop 3	40	40	10	-	-	4	no
+ Loop 3.1	8	8	1	-	-	8	no
- Loop 4	72	72	18	-	-	4	no
+ Loop 4.1	16	16	2	-	-	8	no
- Loop 5	16	44	4 ~ 11	-	-	4	no
+ Loop 5.1	5	5	3	1	1	4	yes
- Loop 6	16	16	2	-	-	8	no
- Loop 7	24	24	3	3	3	8	yes

Utilization Estimates

Summary

Name	BRAM_18K	DSP48E	FF	LUT
DSP	-	-	-	-
Expression	-	-	0	857
FIFO	-	-	-	-
Instance	6	-	276	282
Memory	0	-	74	7
Multiplexer	-	-	-	569
Register	-	-	685	-



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

**THANK
YOU!**

FOR

**YOUR
PRESENCE**

johnjose@iitg.ac.in

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