

# NEBULA @ BITS TEAM BOB

Meghadri Ghosh  
Pramit Pal  
Pranav Chandra N. V.

BITS Pilani, Pilani Campus

# AGENDA

1. Project Overview
2. The Packet
  - Structure
  - Assembly & Disassembly
3. Router Microarchitecture
  - Router Overview
  - Pipeline Stages
  - VC Allocation FSM
  - Routing Algorithms
  - Arbitration Switch Allocation
  - AXI & CHI Algorithm Usage
4. Mesh Architecture
5. Q & A

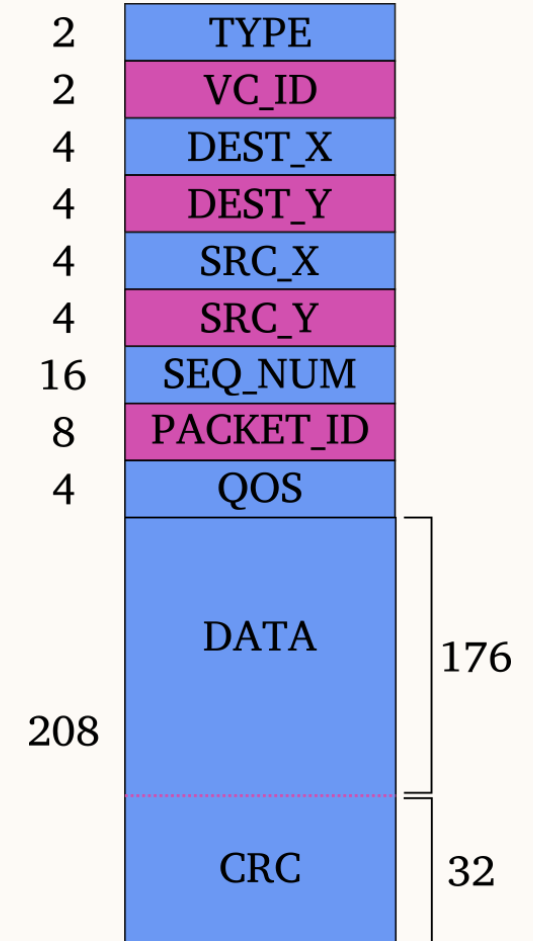
# HIGH LEVEL OVERVIEW

- **Output:** A scalable NOC design for multi-GPU system
- **Topology:** 2D mesh, 2x2 to 8x8 grid (up to 64 GPUs)
- **Protocols:** AXI4 (non-coherent), CHI (coherent)
- **Languages Used:** SystemVerilog (RTL), Python (Analysis)
- **Key Features:**
  - Five-Stage Router Pipeline
  - Adaptive and Deterministic Routing
  - Credit Based Flow and Arbitration

# **THE PACKET**

# PACKET STRUCTURE

- Packets are decomposed into FLITs
  - 48-bit Header
  - 208-bit body
- May be single- or multi-flit based on size
- Routers are configured to handle QoS for prioritization



# ASSEMBLY AND DISASSEMBLY

- Converts coordinates and metadata into predefined packets.
- Payload segmentation splits large payloads across multiple flits to optimize transmission.
- Ensures proper packet conversion for both AXI4 and CHI protocols.

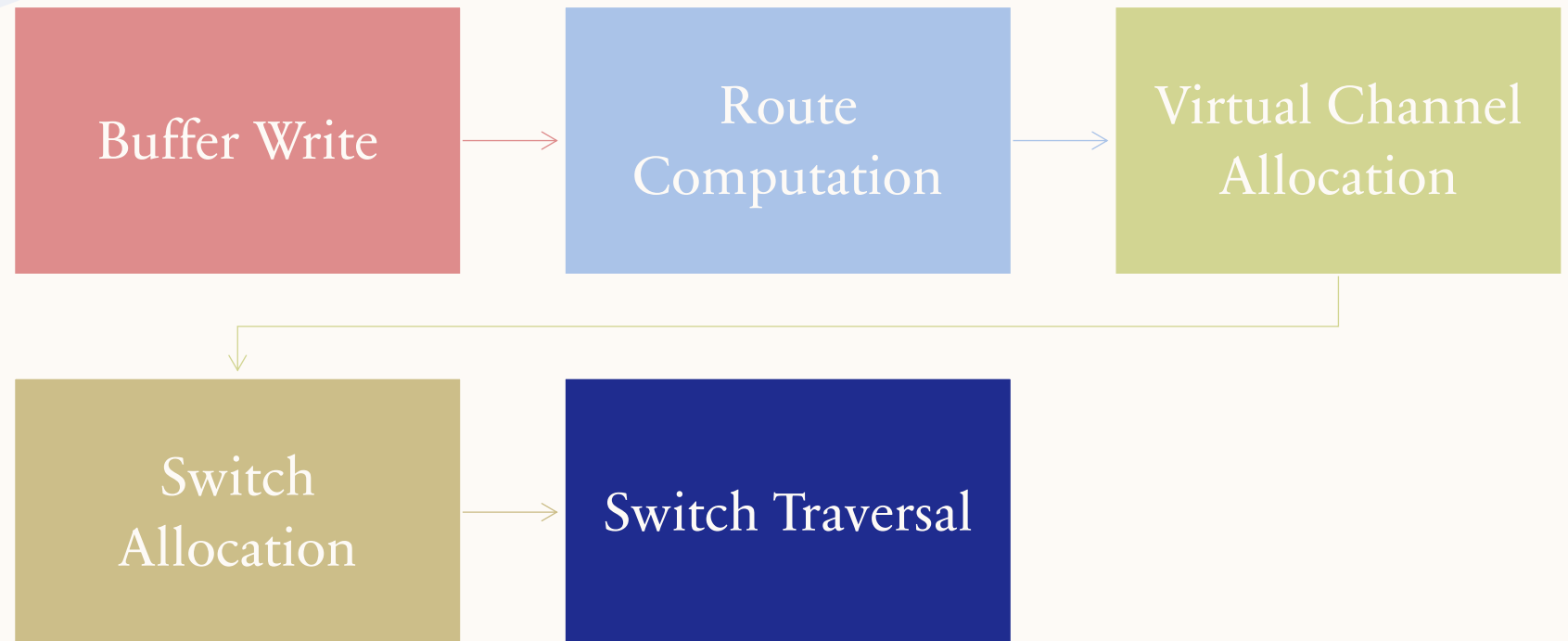
# **ROUTER MICROARCHITECTURE**

# ROUTER OVERVIEW

- Five-stage pipeline to process packets:
  1. Buffer Write
  2. Route Computation
  3. Virtual Channel Allocation
  4. Switch Allocation
  5. Switch Traversal
- 5 ports per router, 4 VCs per port
- Four-stage FSM for VC control



# ROUTER PIPELINE



# ROUTING ALGORITHMS

- Basic XY Routing for most packet transmission
- Basic heuristic correction to prevent clustering in mesh



Figure 1: Only XY Routing



Figure 2: Adaptive Routing

# ARBITRATION SWITCH ALLOCATION

- Our system uses a Round Robin Arbiter
- Tree Arbiters are faster, but more hardware heavy

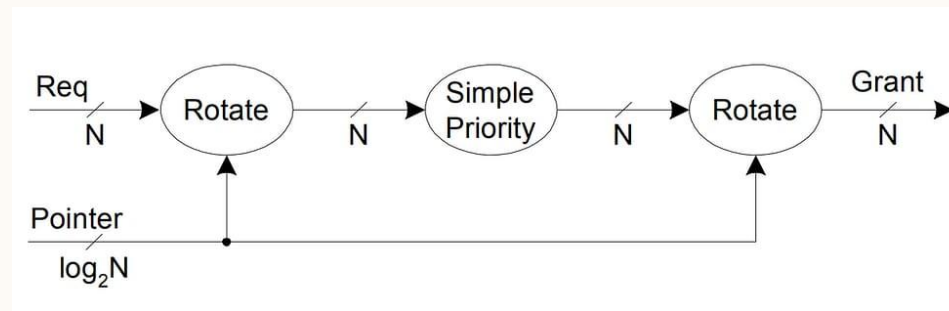


Figure 1: Round Robin Arbiter

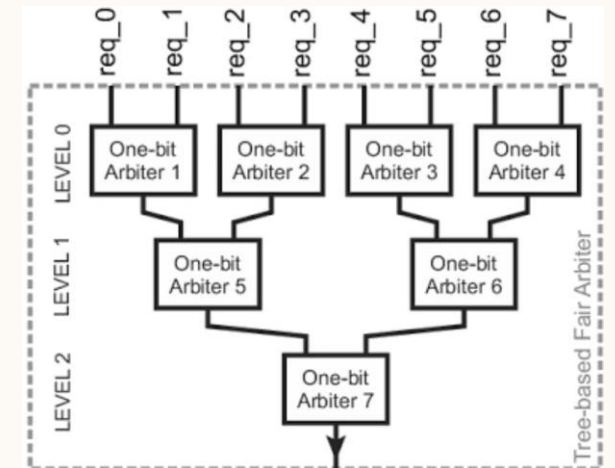


Figure 2: Tree Arbiter

# AXI4 AND CHI PROTOCOL

- AXI-4 is used for high frequency; low latency communication of information packets between nodes
- CHI is used to maintain memory coherence between nodes

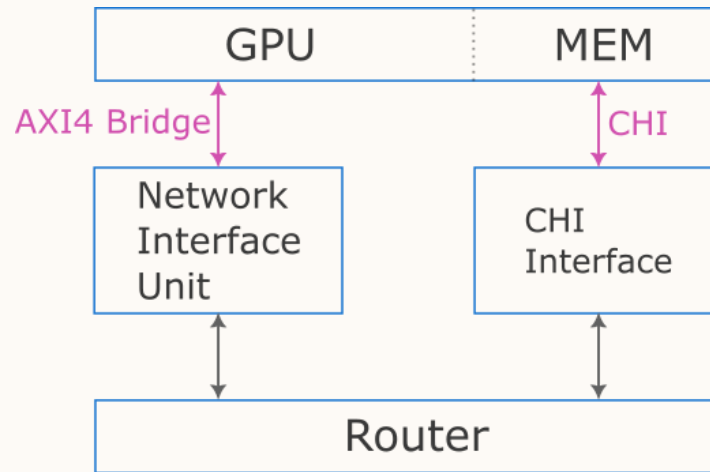


Figure 1: Single Node Architecture

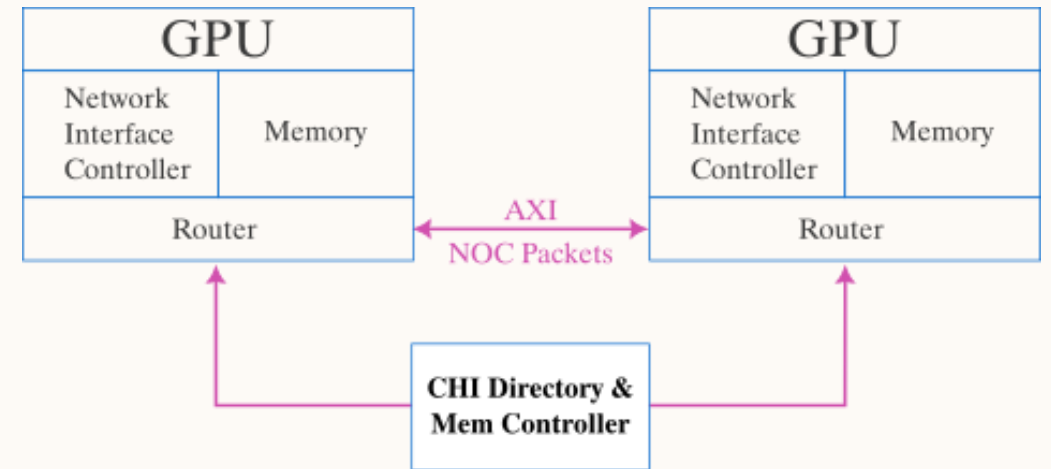


Figure 2: Mesh Connection Architecture

# Q&A SESSION

Meghadri Ghosh

Pramit Pal

Pranav Chandra N. V.

Team Bob (Group 2)

Bits Pilani, Pilani Campus