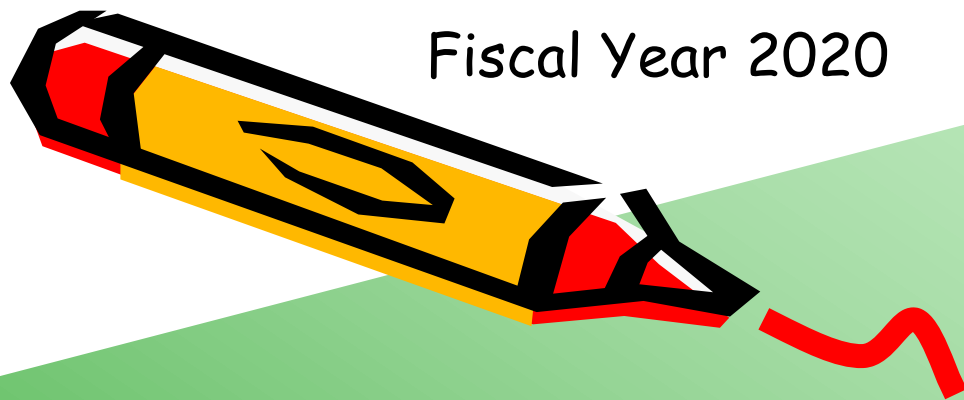


Fiscal Year 2020

Ver. 2021-01-27a



Course number: CSC.T433
School of Computing,
Graduate major in Computer Science

Advanced Computer Architecture

11. Thread Level Parallelism: Interconnection Network

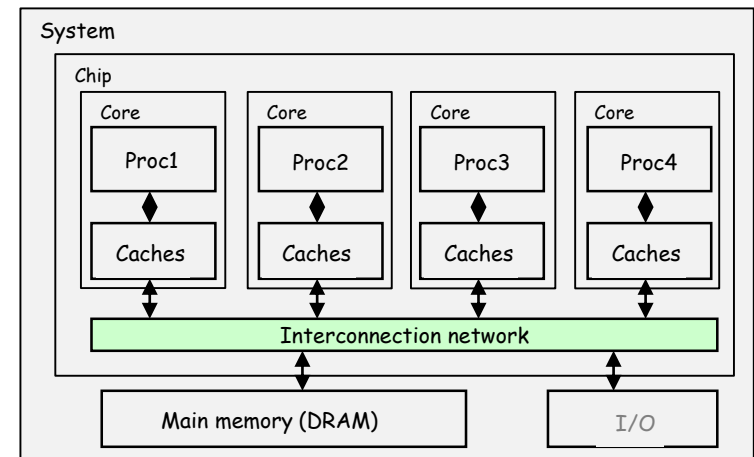


www.arch.cs.titech.ac.jp/lecture/ACA/
Room No.W936
Mon 14:20-16:00, Thr 14:20-16:00

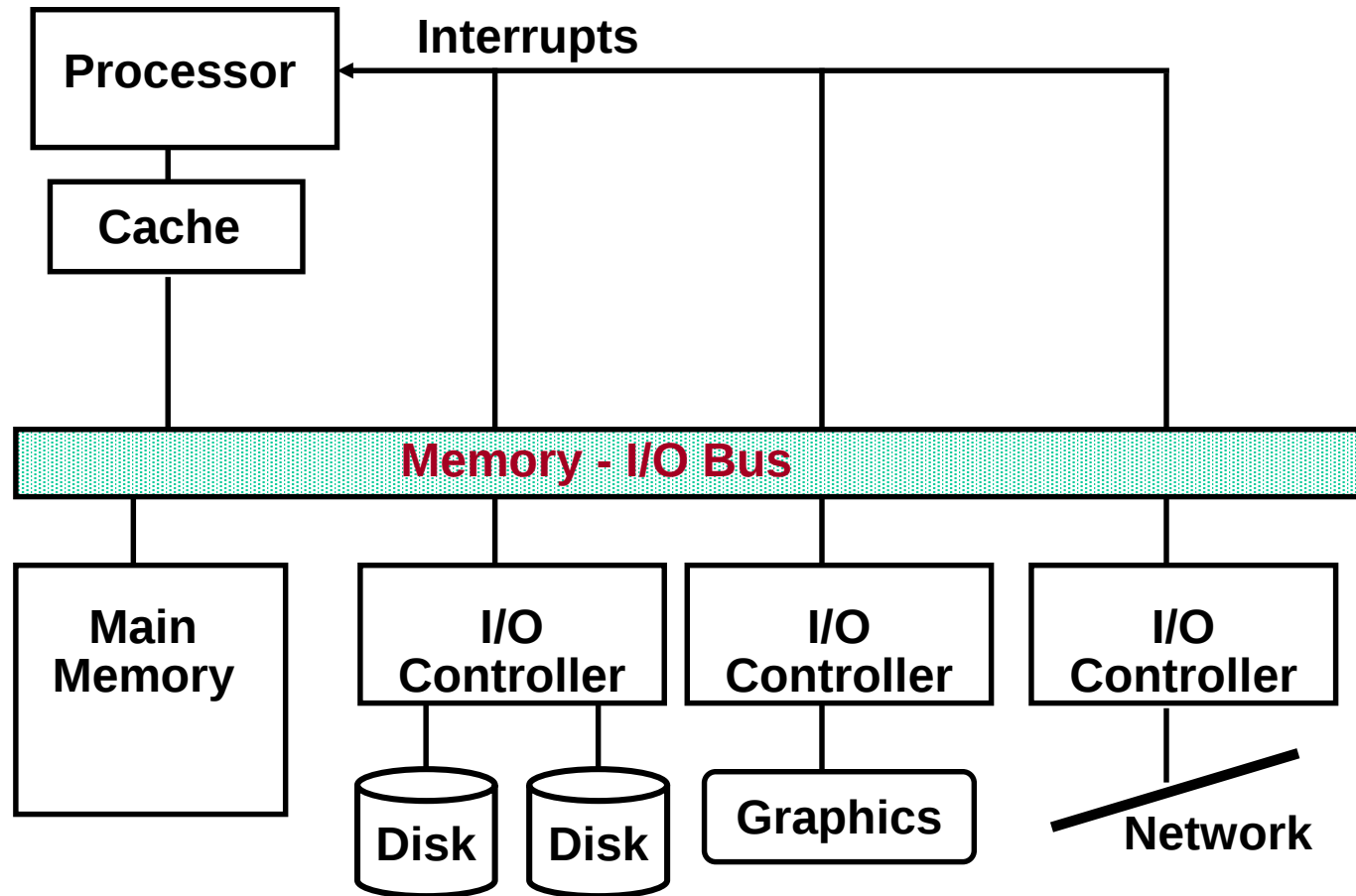
Kenji Kise, Department of Computer Science
kise_at_c.titech.ac.jp

Key components of many-core processors

- **Interconnection network**
 - connecting many modules on a chip achieving **high throughput** and **low latency**
- **Main memory and caches**
 - Caches are used to reduce latency and to lower network traffic
 - A parallel program has private data and shared data
 - New issues are cache coherence and memory consistency
- **Core**
 - High-performance superscalar processor providing a hardware mechanism to support thread synchronization



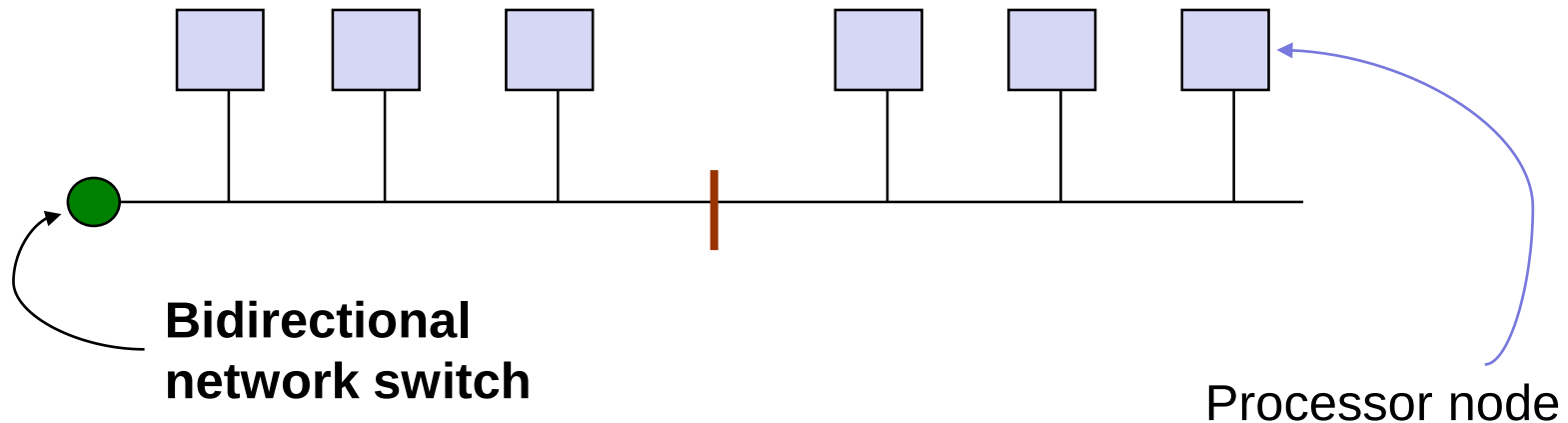
A Typical computer System





Bus Network

- N processors, 1 switch (●), 1 link (the bus)
- Only 1 simultaneous transfer at a time
 - NB (best case) = link (bus) bandwidth $\times$ 1
 - BB (worst case) = link (bus) bandwidth $\times$ 1
- All processors can **snoop** the bus





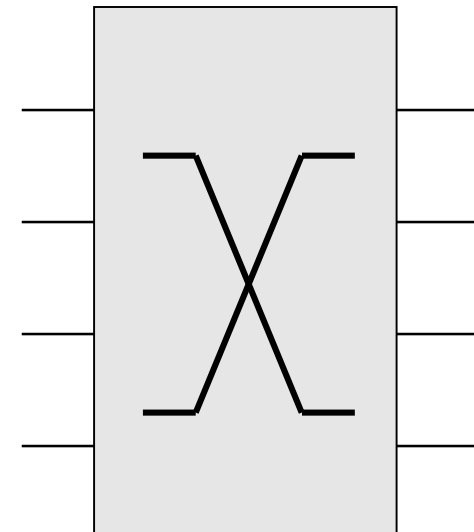
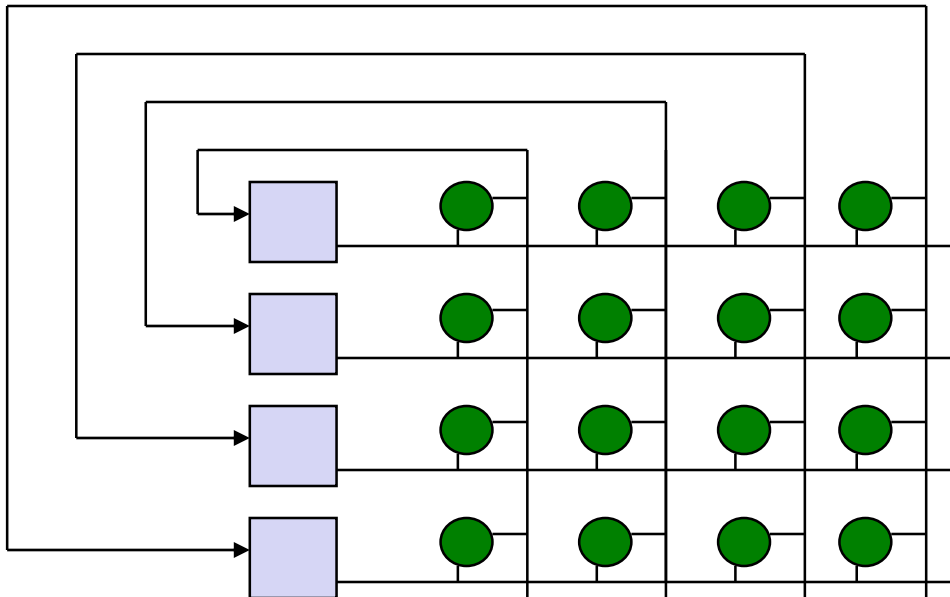
Performance metrics of interconnection network

- Network cost
 - number of links on a switch to connect to the network (plus one link to connect to the processor)
 - width in bits per link, length of link
 - number of switches
- Network bandwidth (NB)
 - represents the best case
 - bandwidth of each link $\times$ number of links
- Bisection bandwidth (BB)
 - represents the worst case
 - divide the machine in two parts, each with half the nodes and sum the bandwidth of the links that cross the dividing line



Crossbar (Xbar) Network

- N processors, N^2 switches (unidirectional), 2 links/switch, N^2 links
- N simultaneous transfers
 - NB = link bandwidth $\times N$ (best case)
 - BB = link bandwidth $\times N$ (worst case)

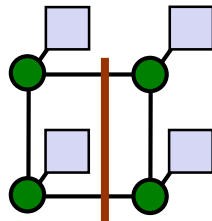


A symbol of Xbar

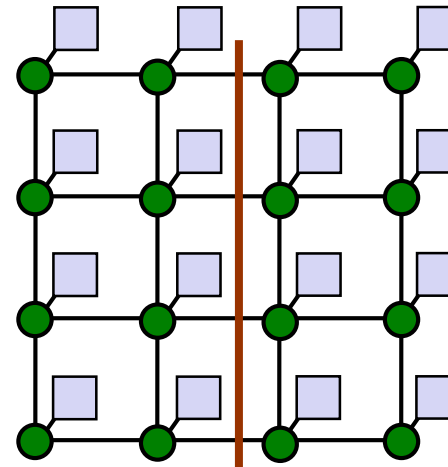


Mesh Network

- N processors, N switches, 4 links/switch, $N \times (N^{1/2} - 1)$ links
- N simultaneous transfers
 - $NB = \text{link bandwidth} \times 2N$ (best case)
 - $BB = \text{link bandwidth} \times 2N^{1/2}$ (worst case)



$N = 4$

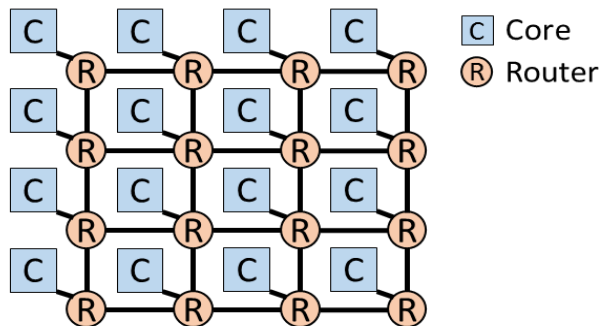


$N = 16$

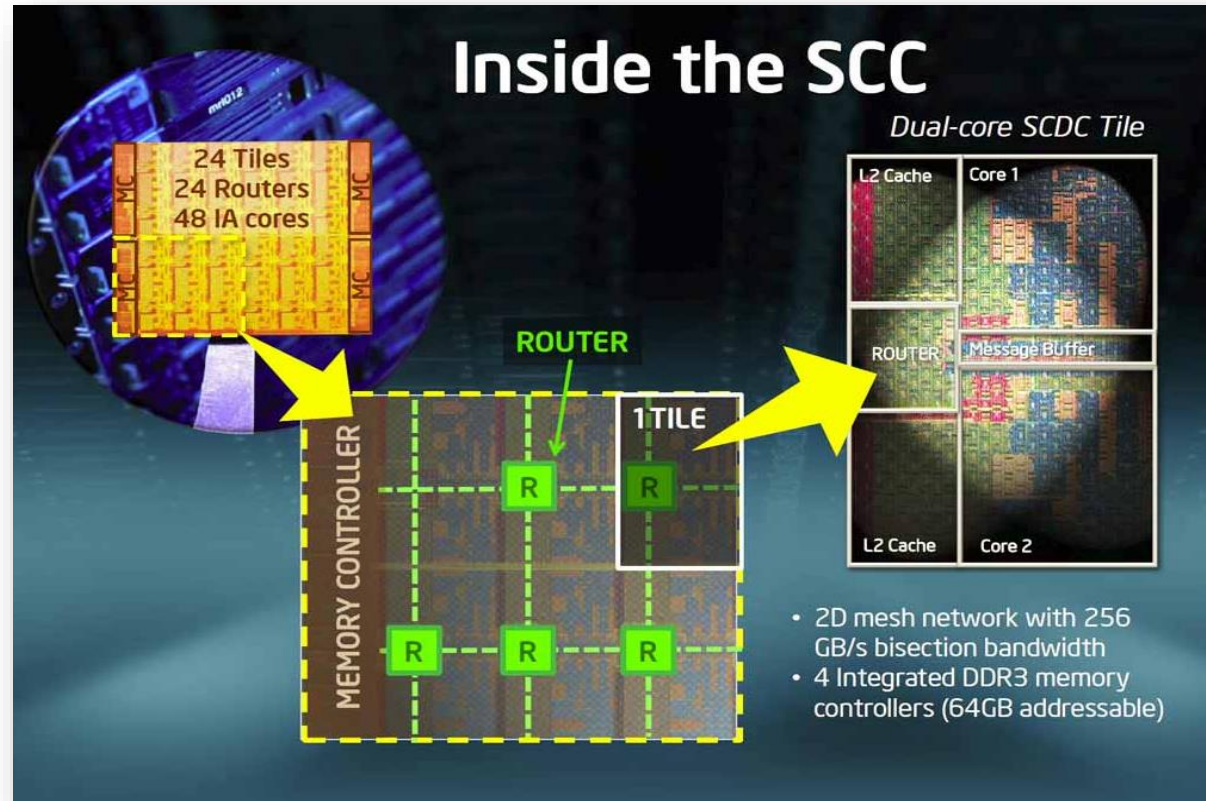


Intel Single-Chip Cloud Computer (2009)

- To research multi-core processors and parallel processing.



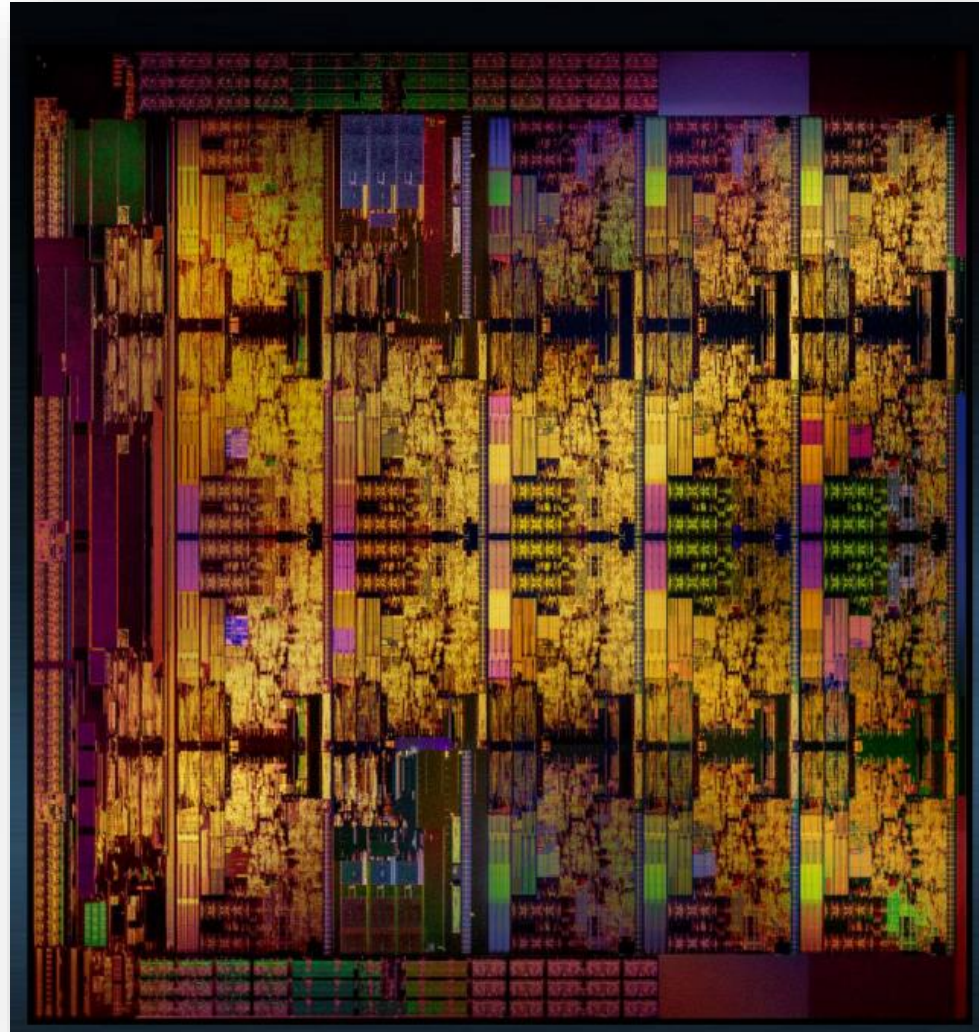
**A many-core architecture
with 2D Mesh NoC**



Intel Single-Chip Cloud Computer (48 Core)

Intel Skylake-X, Core i9-7980XE, 2017

- 18 core



Bus vs. Networks on Chip (NoC) of mesh topology



intersection

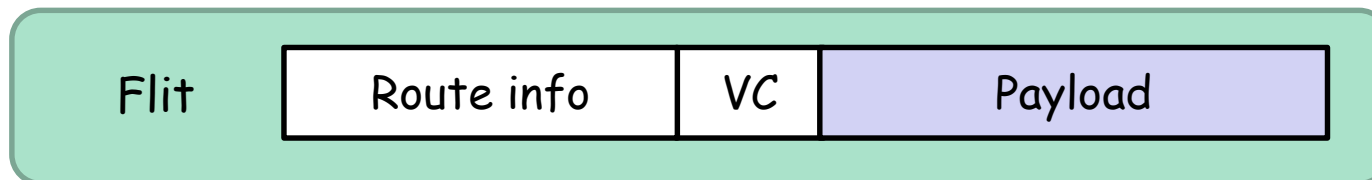


Packet organization (Flit encoding)

- A **flit** (flow control unit or flow control digit) is a link-level atomic piece that forms a network packet.
 - A packet has one head flit and some body flits.
- For simplicity, assume that a packet has only one flit.
 - Later we see a packet which has some flits.
- Each flit has typical three fields:
 - Payload (data)
 - Route information
 - Virtual channel identifier (VC)

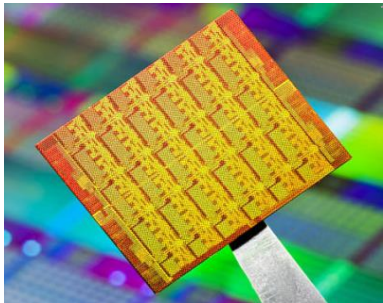


Packet (tag + data)

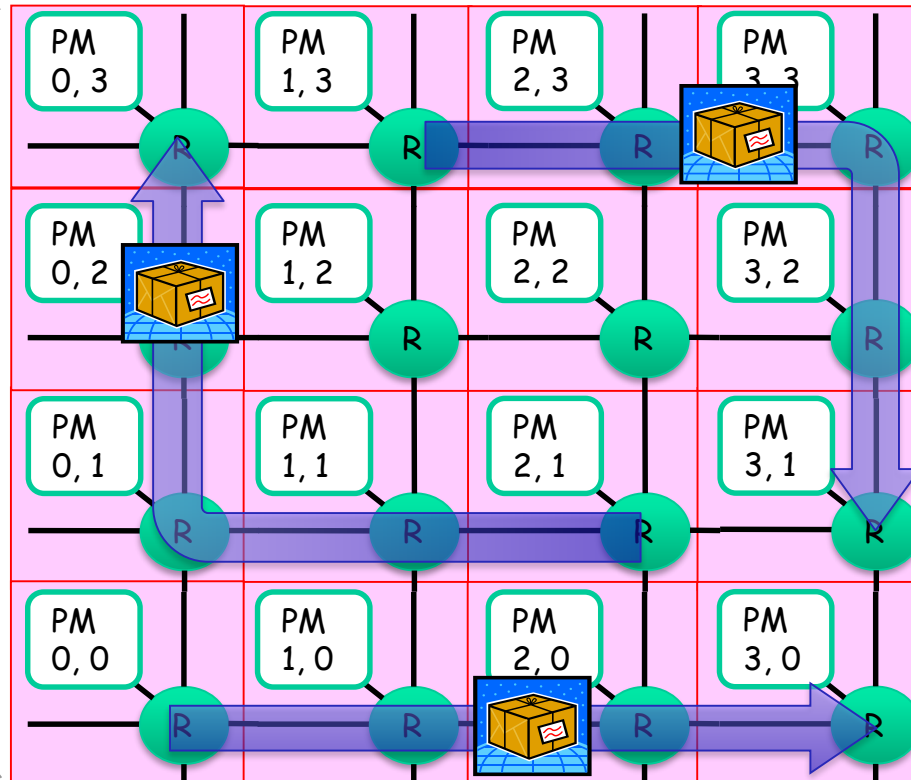


Typical NoC architecture of mesh topology

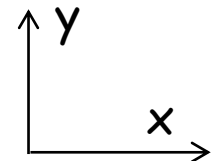
- NoC requirements: low latency, high throughput, low cost
- Packet based data transmission via NoC routers and XY-dimension order routing



PM: Processing Module or Core,
R: Router

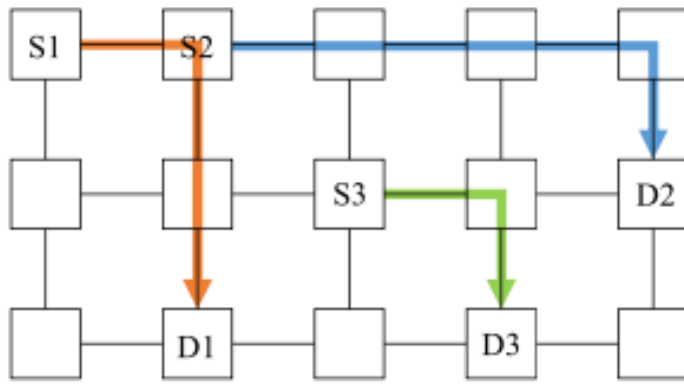


Packet
(tag + data)

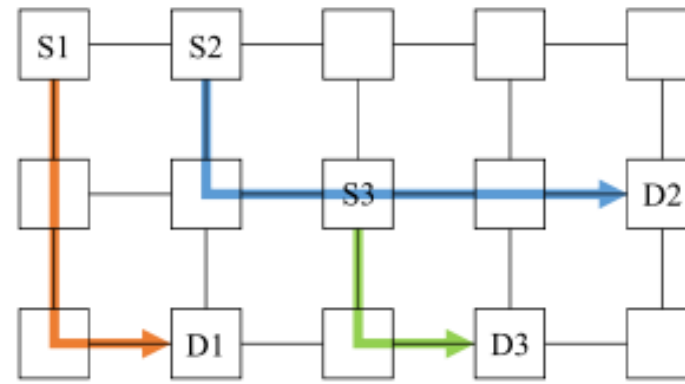


Routing

- XY dimension order routing (DOR), YX DOR



(a) XY routing

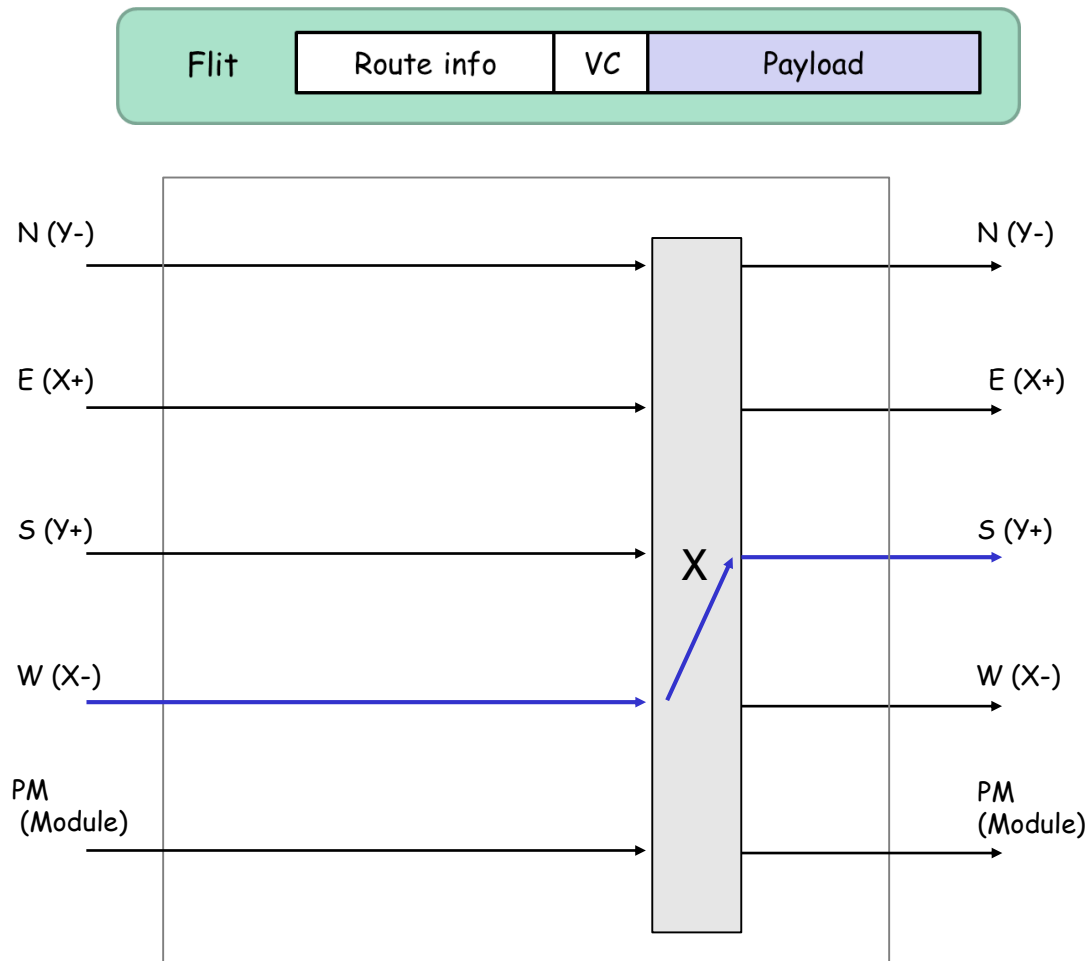


(b) YX routing

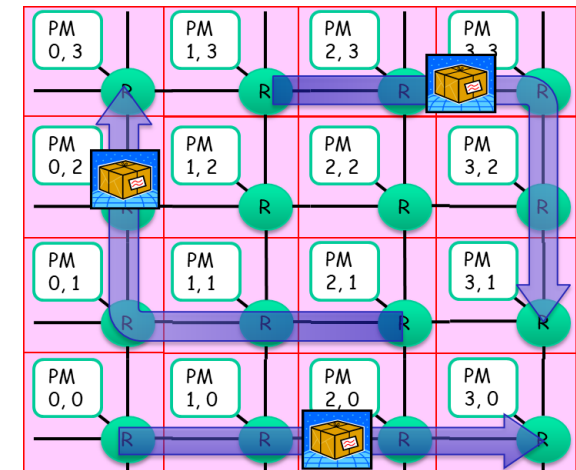
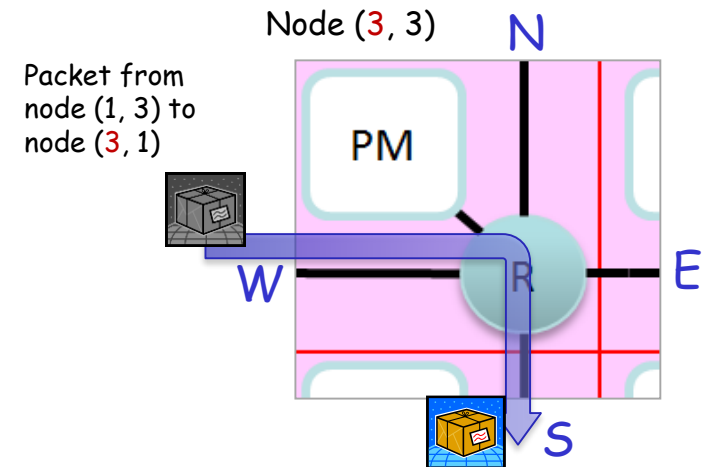


Simple NoC router architecture

- Routing computation for **XY-dimension order**

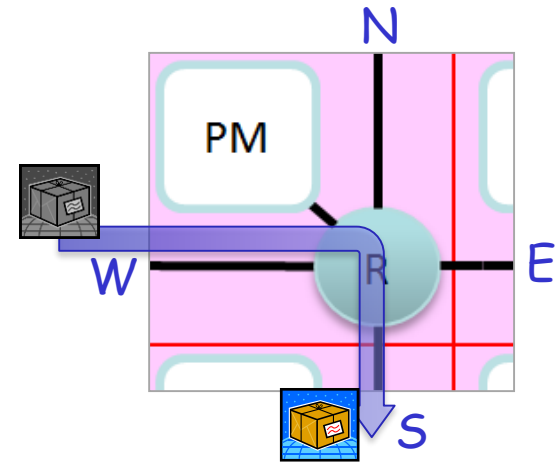
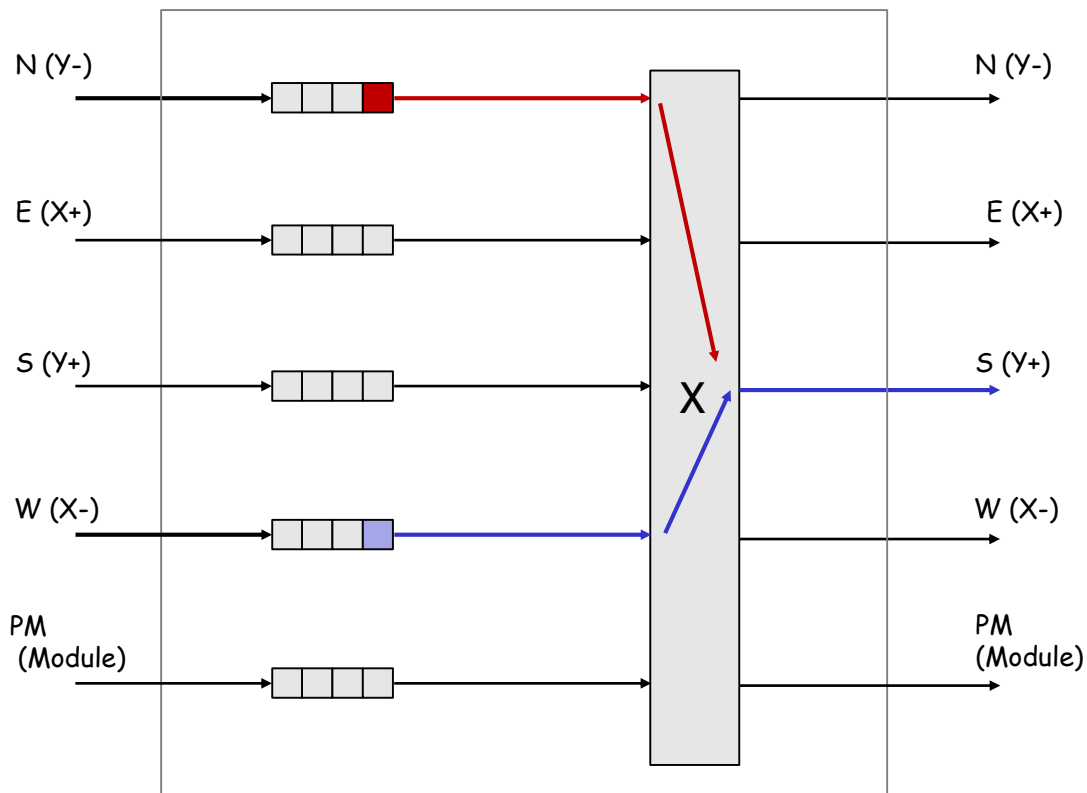


NoC router



Simple NoC router architecture

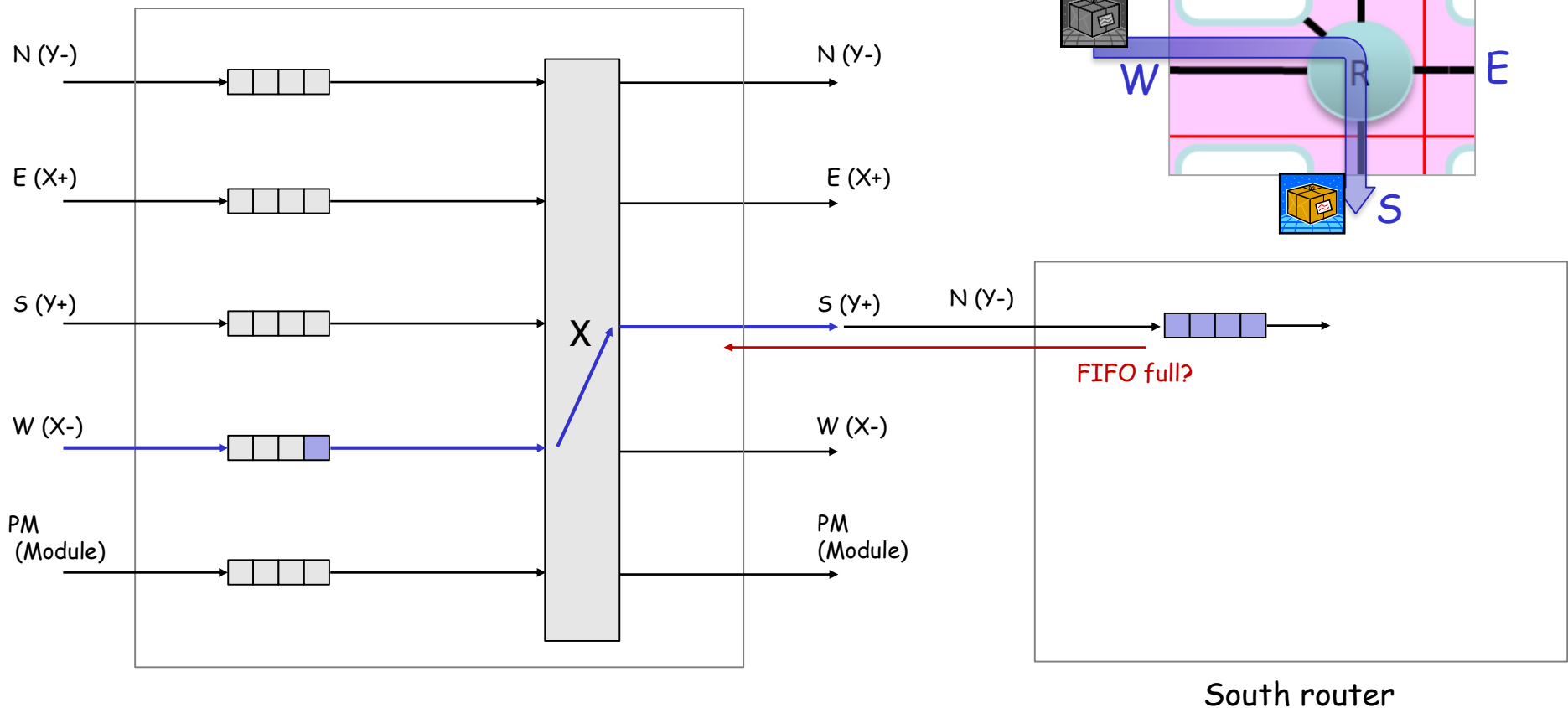
- **Buffering and arbitration**
 - time stamp based, round robin, etc.





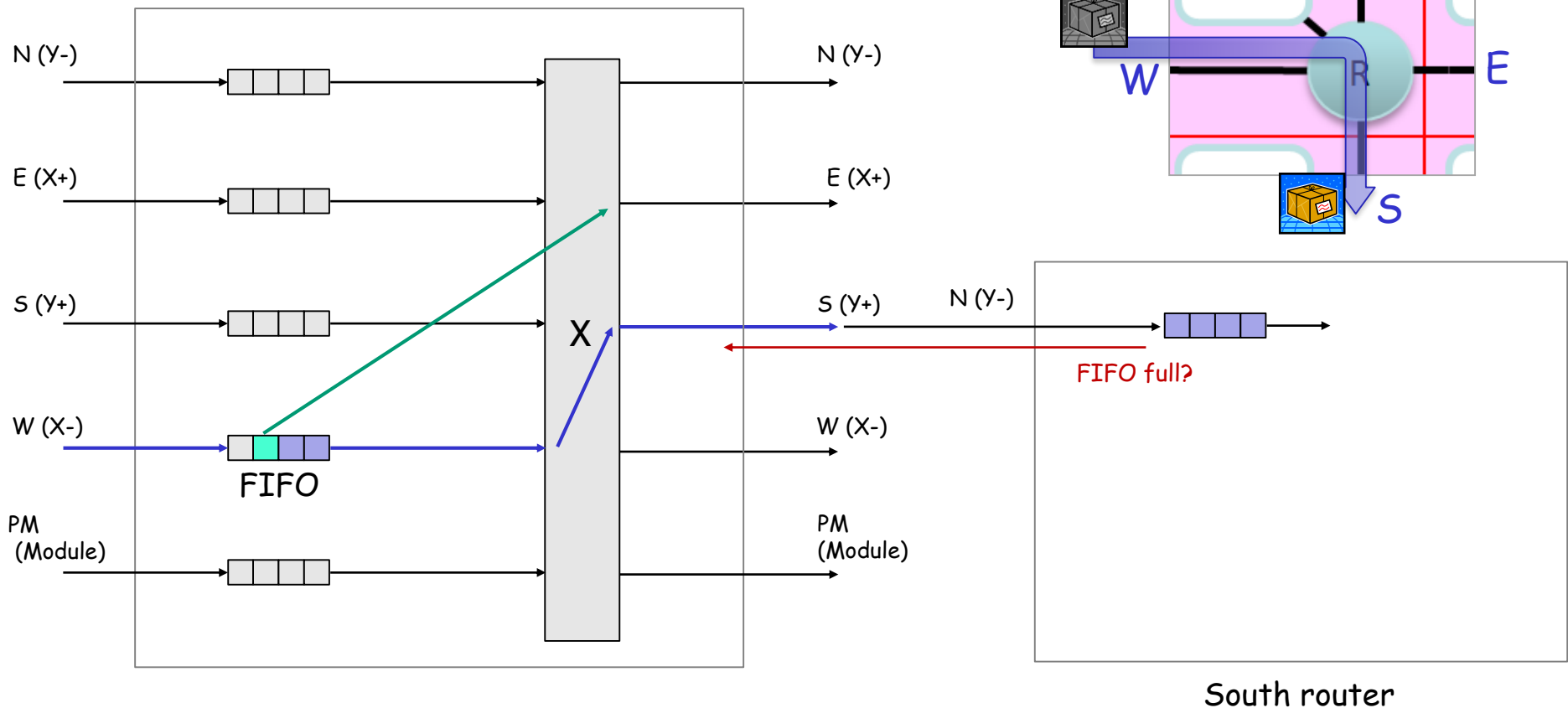
Simple NoC router architecture

- Flow control



Simple NoC router architecture

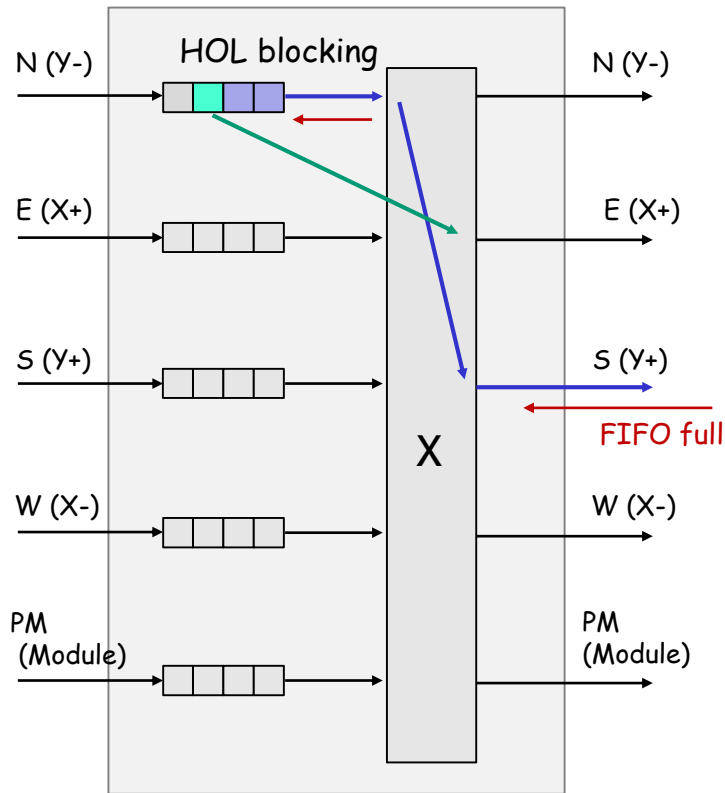
- Problem: Head-of-line (HOL) blocking



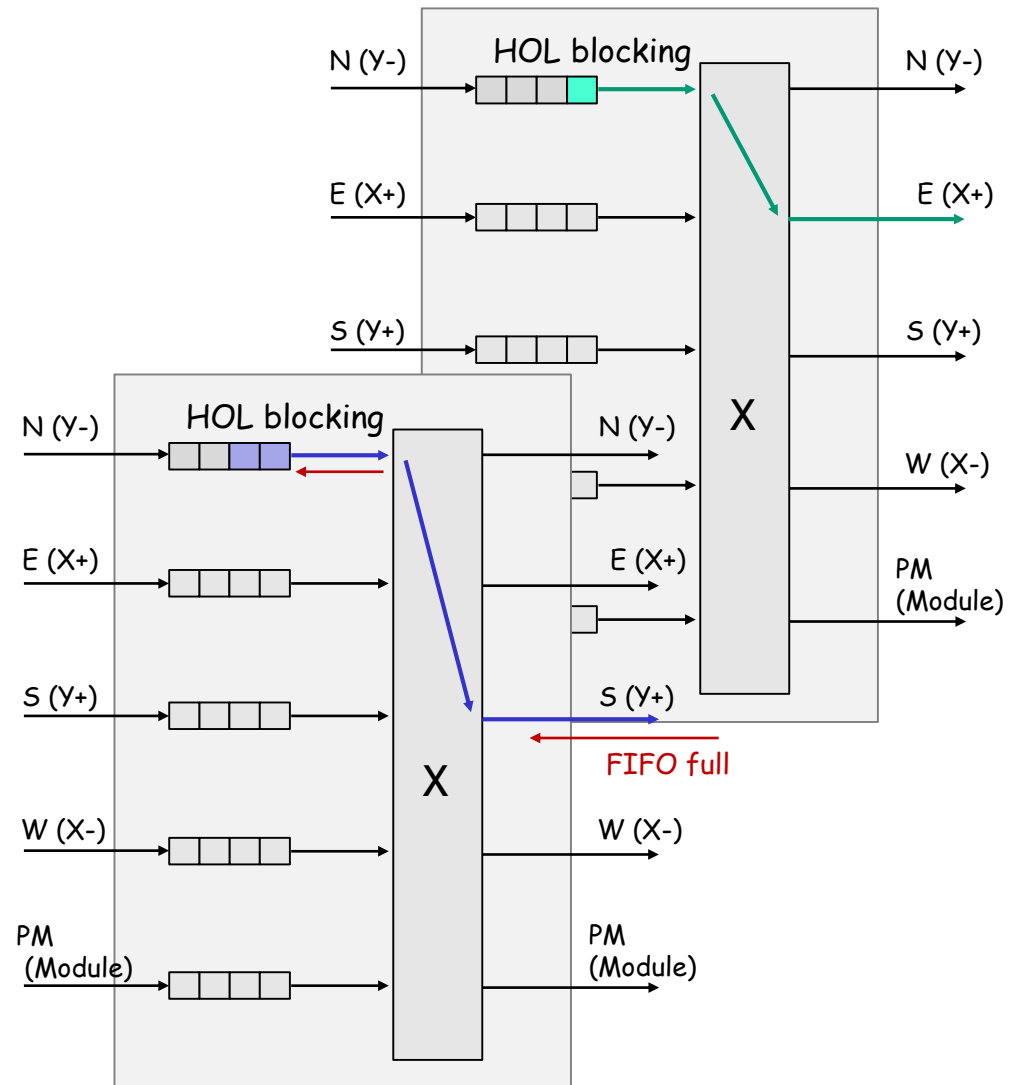
South router



Two (physical) networks to mitigate HOL ?

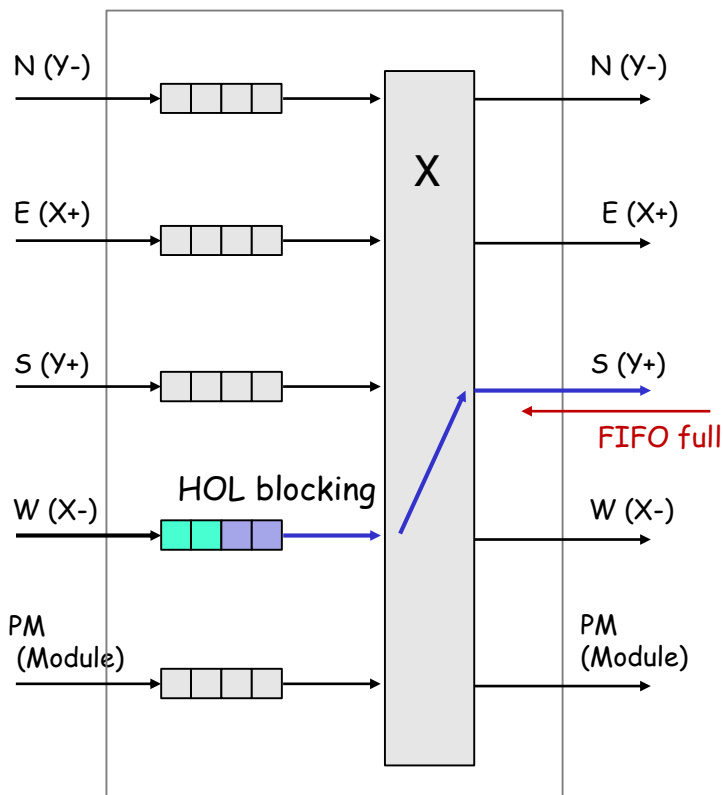
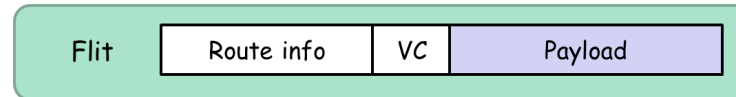


Simple NoC router

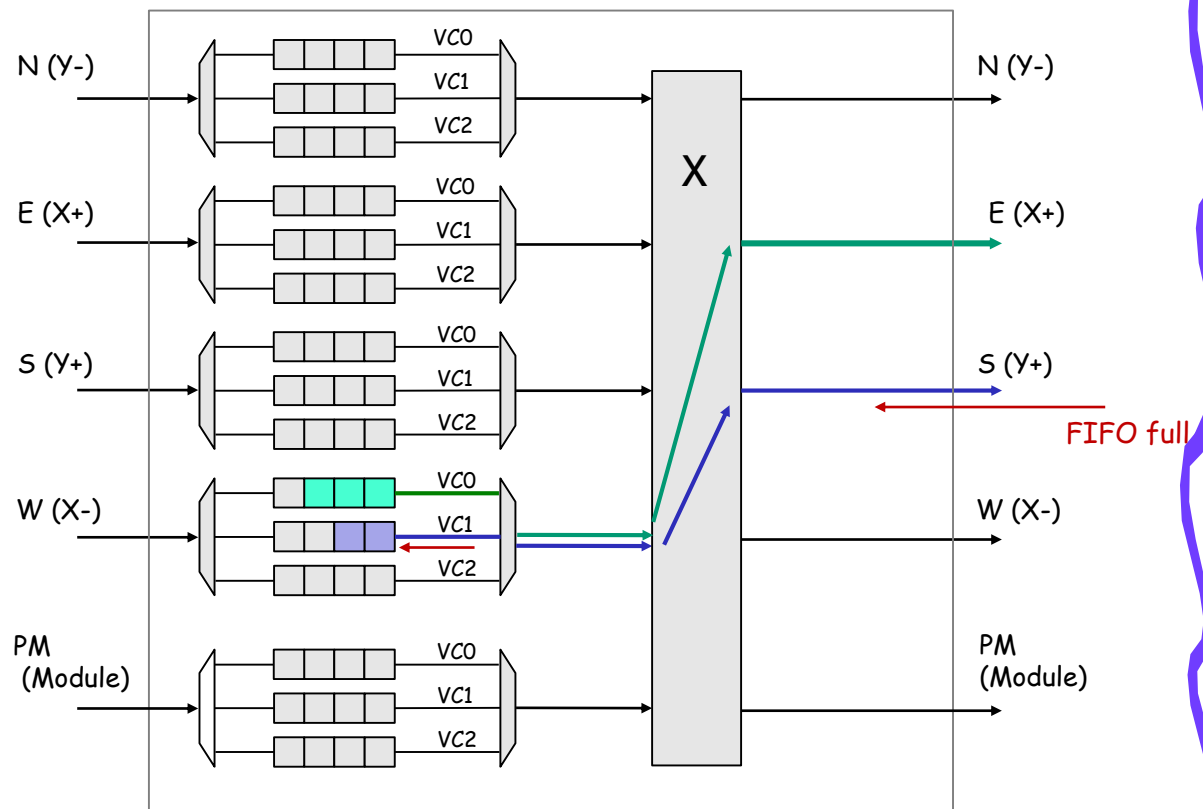


Datapath of Virtual Channel (VC) NoC router

- To mitigate **head-of-line (HOL) blocking**, virtual channels are used



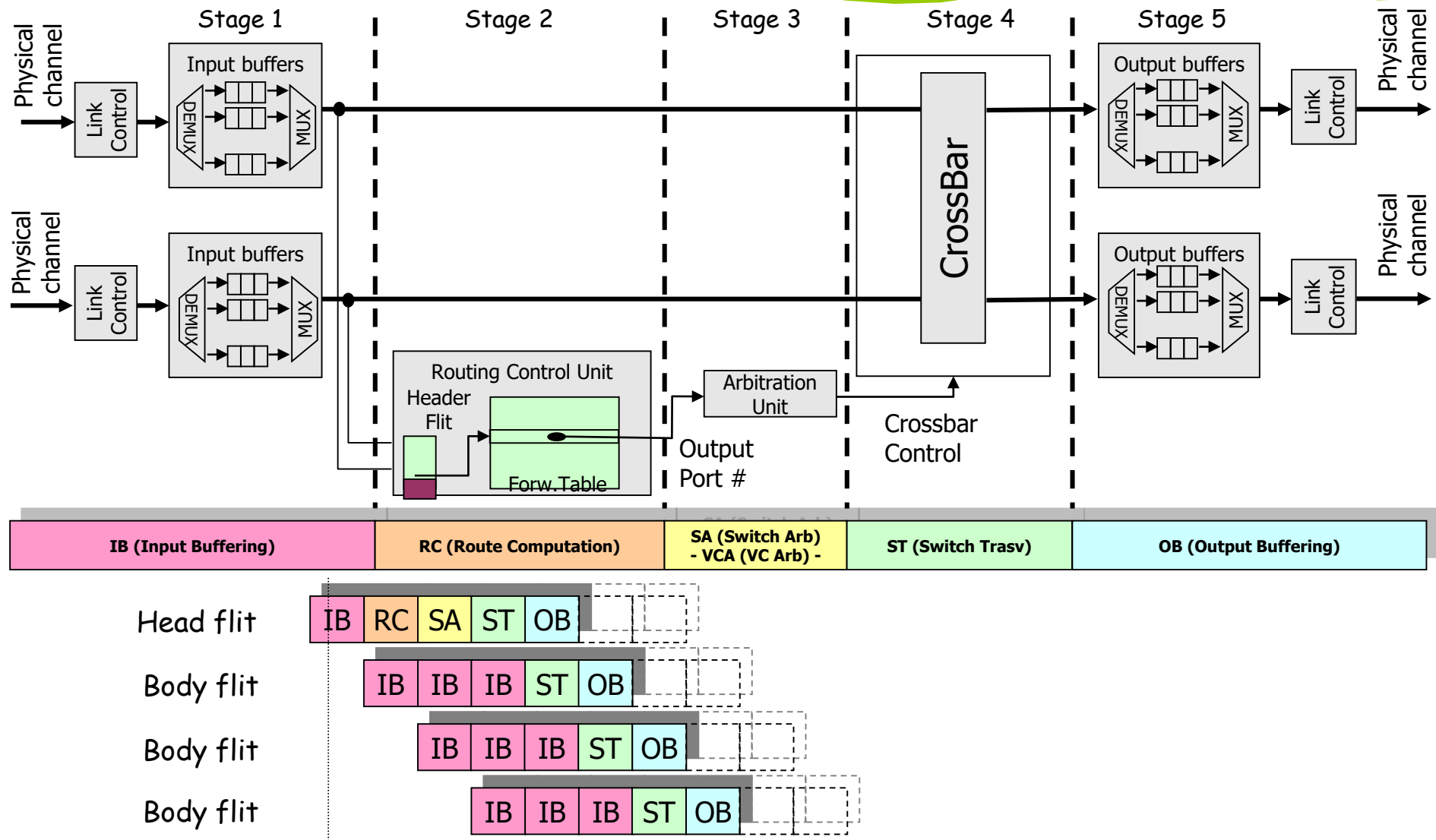
Simple NoC router



VC NoC router



Pipelining the NoC router microarchitecture



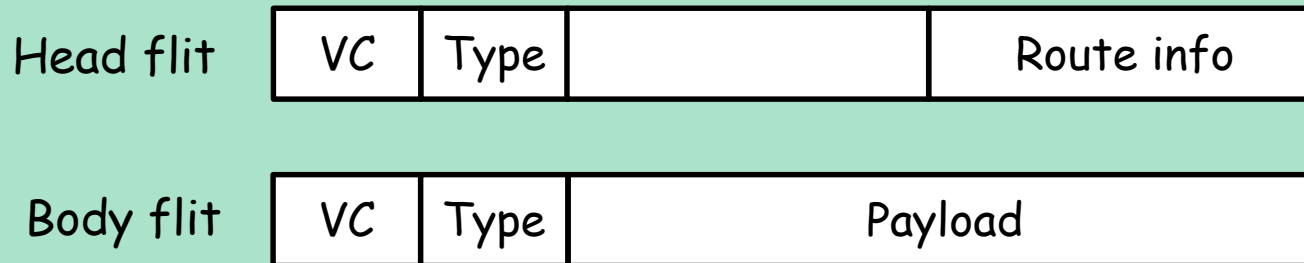
"A Delay Model and Speculative Architecture for Pipelined Routers," L. S. Peh and W. J. Dally, Proc. of the 7th Int'l Symposium on High Performance Computer Architecture, January, 2001.

Packet organization (Flit encoding)

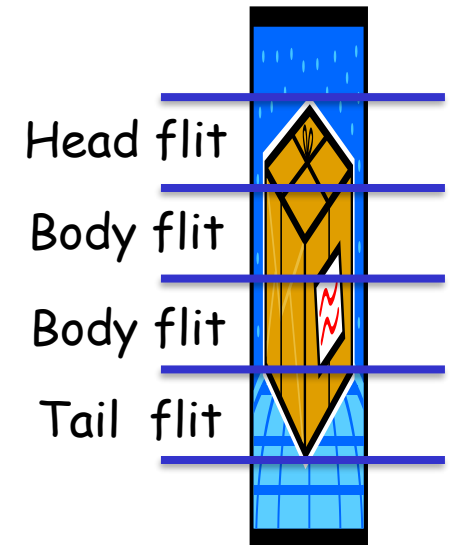
- A **flit** (flow control unit or flow control digit) is a link-level atomic piece that forms a network packet.
 - A packet has one head flit and some body flits.
- Each flit has typical three fields:
 - payload(data) or route information(tag)
 - flit type : head, body, tail, etc.
 - virtual channel identifier



Packet (tag + data)



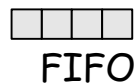
Head and body flit formats



Bus vs. Networks on Chip (NoC) of mesh topology



Distributed system



Packet
(tag + data)



intersection

Bus vs. Networks on Chip (NoC) of mesh topology



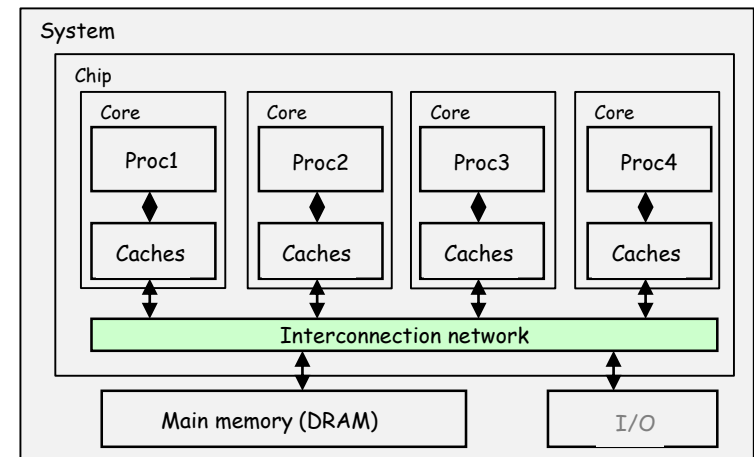
To mitigate
head-of-line (HOL) blocking

Virtual Channel



Key components of many-core processors

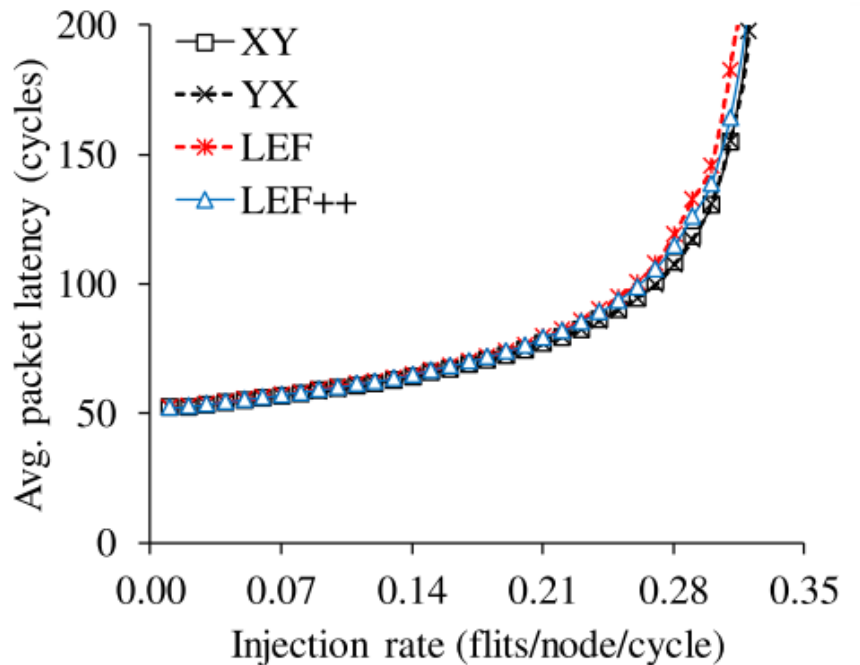
- **Interconnection network**
 - connecting many modules on a chip achieving **high throughput** and **low latency**
- **Main memory and caches**
 - Caches are used to reduce latency and to lower network traffic
 - A parallel program has private data and shared data
 - New issues are cache coherence and memory consistency
- **Core**
 - High-performance superscalar processor providing a hardware mechanism to support thread synchronization





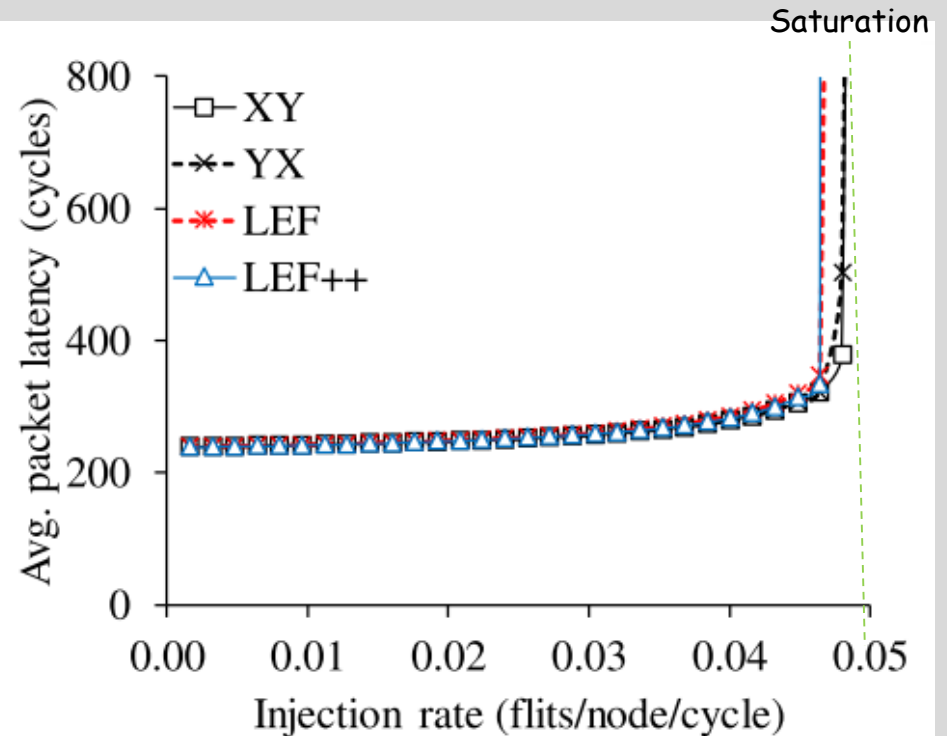
Average packet latency of mesh NoCs

- 5 stage router pipeline
- Uniform traffic (destination nodes are selected randomly)



(a) Average packet latency under uniform traffic

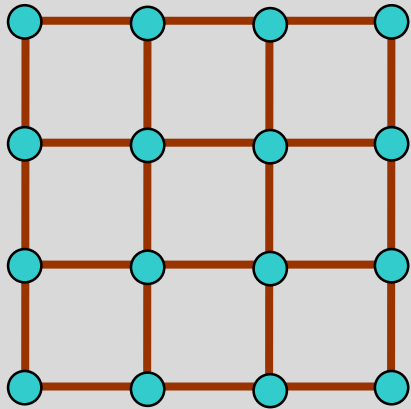
8x8 NoC



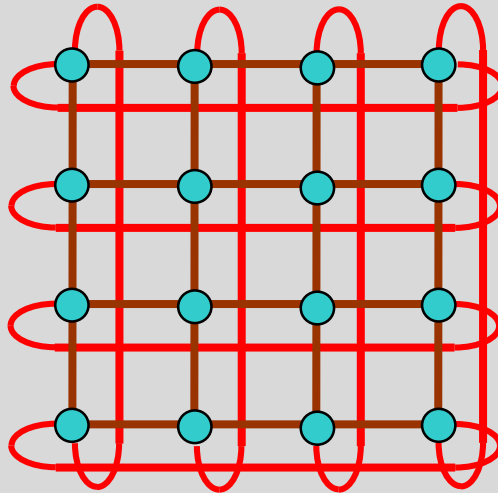
(a) Average packet latency under uniform traffic

64x64 NoC (4096 nodes)

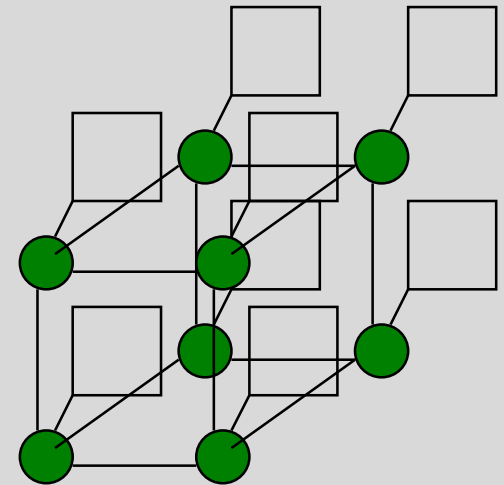
2D and 3D Mesh / Torus Network



2D Mesh



Torus

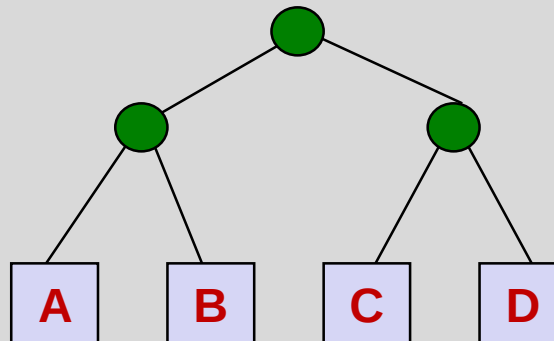


3D Mesh



Fat Tree (1)

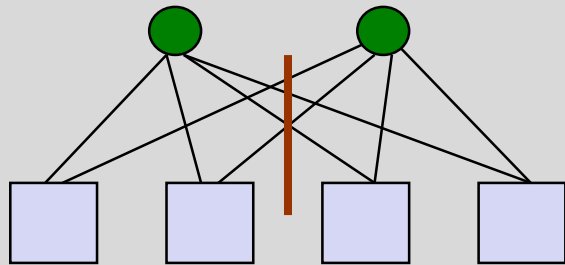
- Trees are good structures. People in CS use them all the time. **Suppose we wanted to make a tree network.**
- Any time A wants to send to C, it ties up the upper links, so that B can't send to D.
 - The bisection bandwidth on a tree is horrible - 1 link, at all times
- The solution is to '**thicken**' the upper links.
 - More links as the tree gets thicker increases the bisection bandwidth



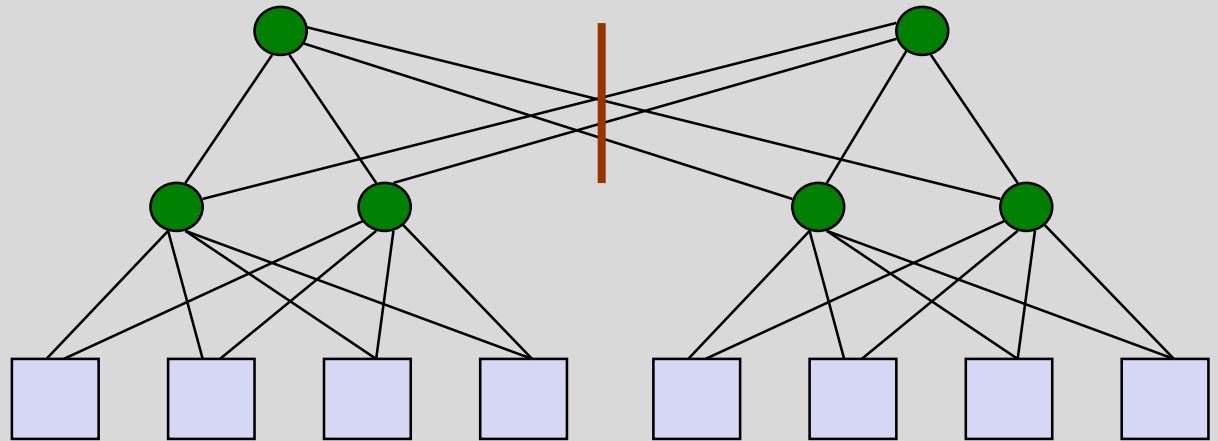
$N = 4$

Fat Tree

- N processors, $\log(N-1) \times \log N$ switches, 2 up + 4 down = 6 links/switch, $N \times \log N$ links
- N simultaneous transfers
 - $NB = \text{link bandwidth} \times N \log N$
 - $BB = \text{link bandwidth} \times 4$



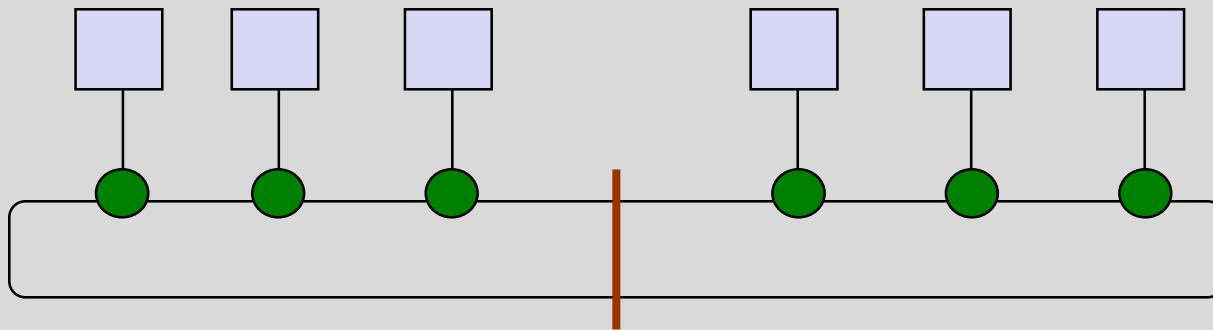
$N = 4$



$N = 8$

Ring Network

- N processors, N switches, 2 links/switch, N links
- N simultaneous transfers
 - NB (best case) = link bandwidth $\times$ N
 - BB (worst case) = link bandwidth $\times$ 2
- If a link is as fast as a bus, the ring is only twice as fast as a bus in the worst case, but is N times faster in the best case



Cell Broadband Engine (2005)

- Cell Broadband Engine (2005)
 - 8 core (SPE) + 1 core (PPE)
 - each SPE has 256KB memory
 - PS3, IBM Roadrunner (12k cores)



PlayStation3
from PlayStation.com (Japan)

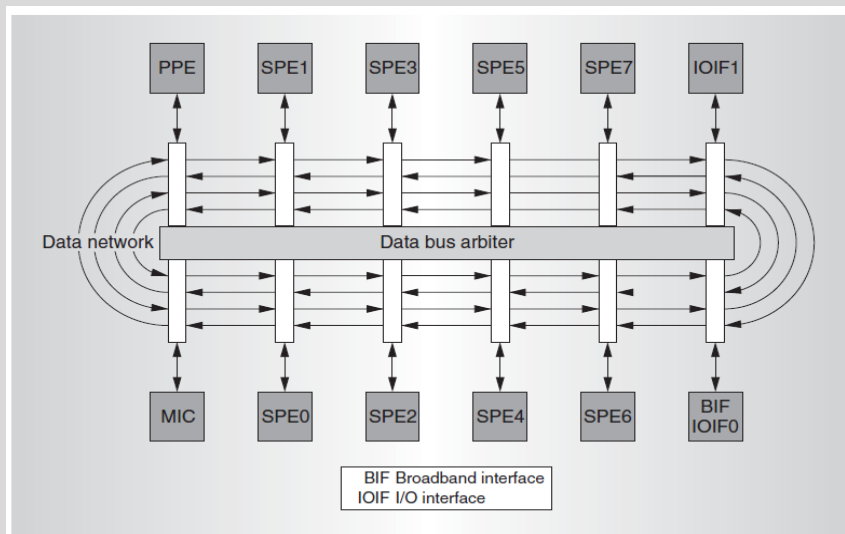


Figure 2. Element interconnect bus (EIB).

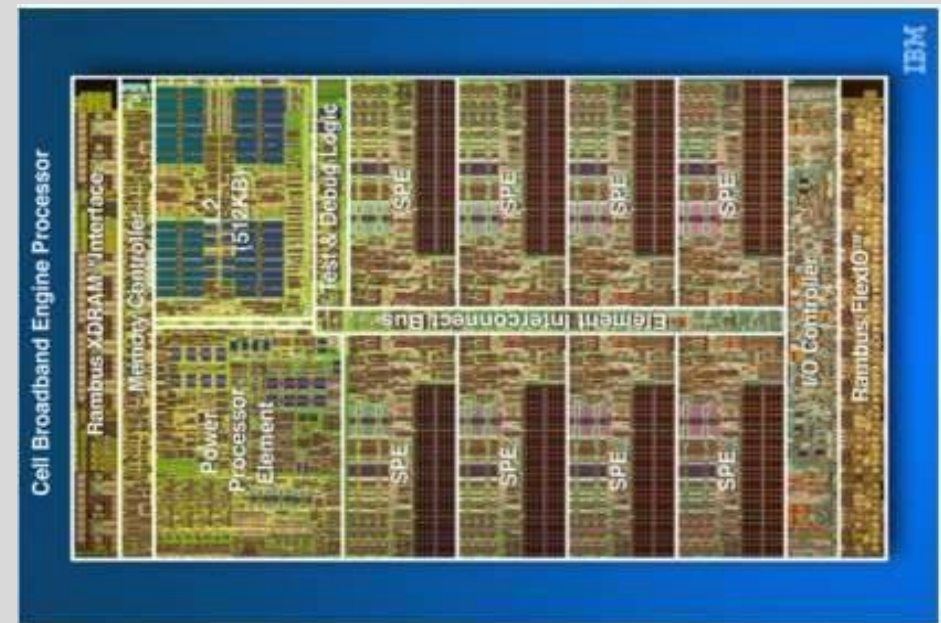


Diagram created by IBM to promote the CBEP, ©2005 from WIKIPEDIA

Intel Xeon Phi (2012)



Intel® Xeon Phi™ Coprocessor Block Diagram

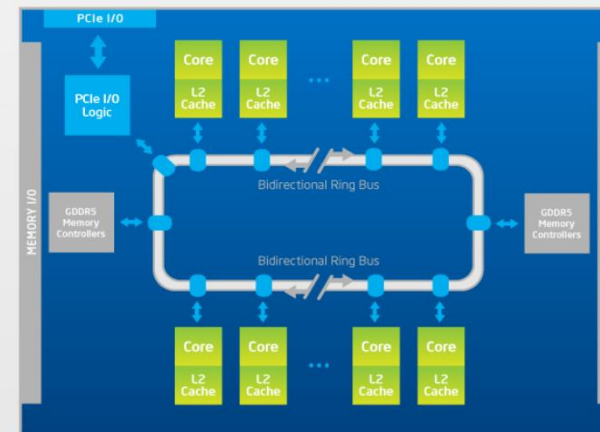
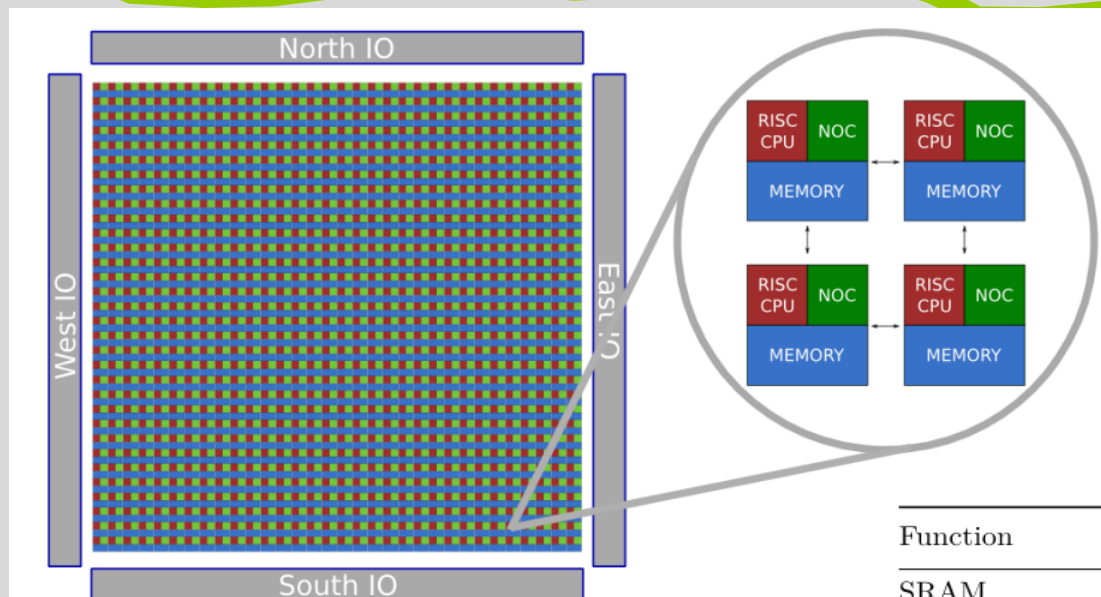


Table 2. Intel® Xeon Phi™ Product Family Specifications

PRODUCT NUMBER	FORM FACTOR &, THERMAL SOLUTION ⁴	BOARD TDP (WATTS)	NUMBER OF CORES	FREQUENCY (GHz)	PEAK DOUBLE PRECISION PERFORMANCE (GFLOP)	PEAK MEMORY BANDWIDTH (GB/s)	MEMORY CAPACITY (GB)	INTEL® TURBO BOOST TECHNOLOGY
3120P	PCIe, Passive	300	57	1.1	1003	240	6	N/A
3120A	PCIe, Active	300	57	1.1	1003	240	6	N/A
5110P	PCIe, Passive	225	60	1.053	1011	320	8	N/A
5120D	Dense form factor, None	245	60	1.053	1011	352	8	N/A
7110P	PCIe, Passive	300	61	1.238	1208	352	16	Peak turbo frequency: 1.33 GHz
7120X	PCIe, None	300	61	1.238	1208	352	16	

Epiphany-V: A 1024 core 64-bit RISC system-on-chip



Summary of Epiphany-V features:

- 1024 64-bit RISC processors
- 64-bit memory architecture
- 64/32-bit IEEE floating point support
- 64MB of distributed on-chip memory
- 1024 programmable I/O signals
- Three 136-bit wide 2D mesh NOCs
- 2052 Independent Power Domains
- Support for up to 1 billion shared memory processors
- Binary compatibility with Epiphany III/IV chips
- Custom ISA extensions for deep learning, communication, and cryptography

Function	Value (mm ²)	Share of Total Die Area
SRAM	62.4	53.3%
Register File	15.1	12.9%
FPU	11.8	10.1%
NOC	12.1	10.3%
IO Logic	6.5	5.6%
“Other” Core Stuff	5.1	4.4%
IO Pads	3.9	3.3%
Always on Logic	0.66	0.6%

Table 5: Epiphany-V Area Breakdown