High Performance Computing Systems

- Networking and Topologies

Doug Shook

Topologies

- How do processors communicate with one another?

- Consider two schemes:
  - Ethernet
  - Fully connected

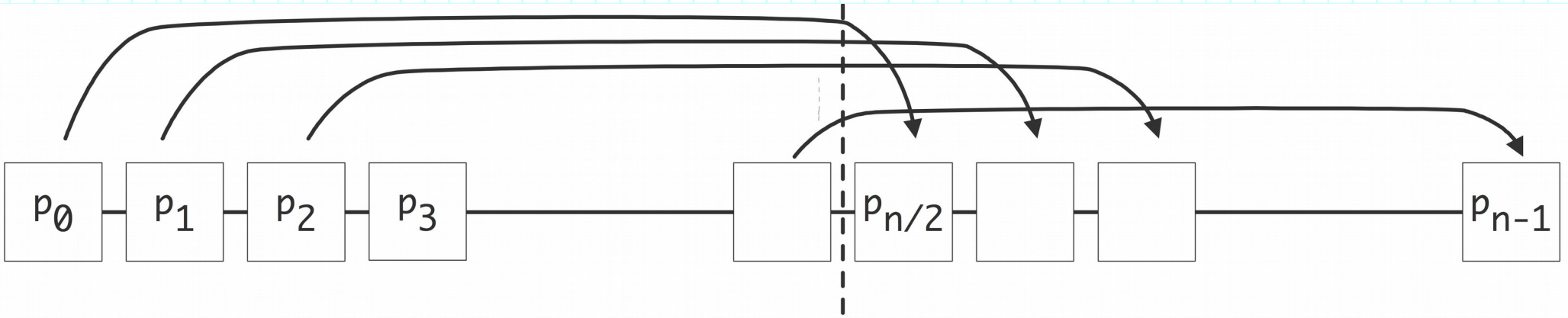
- What is the problem with these schemes?

Some Graph Theory

- We can represent the topology as a graph
  - What are the nodes?
  - What are the vertices?
- Important values to consider:
  - Degree
  - Diameter

Congestion

- What is the problem with the following?



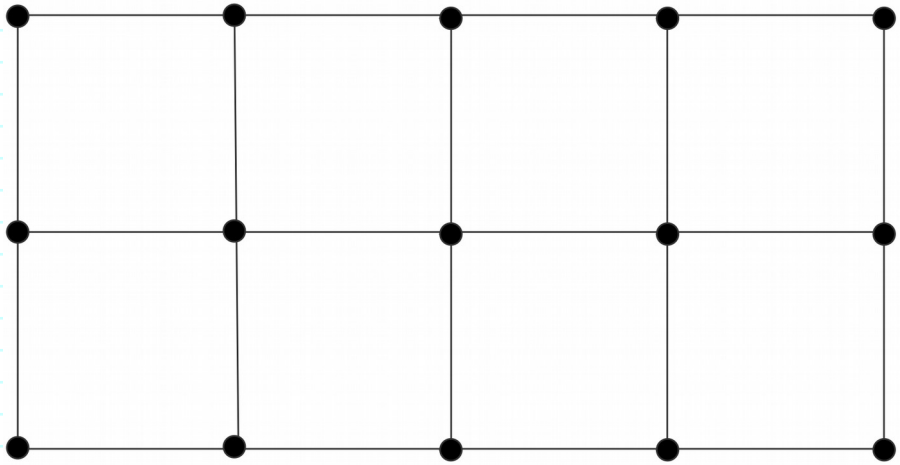
Congestion

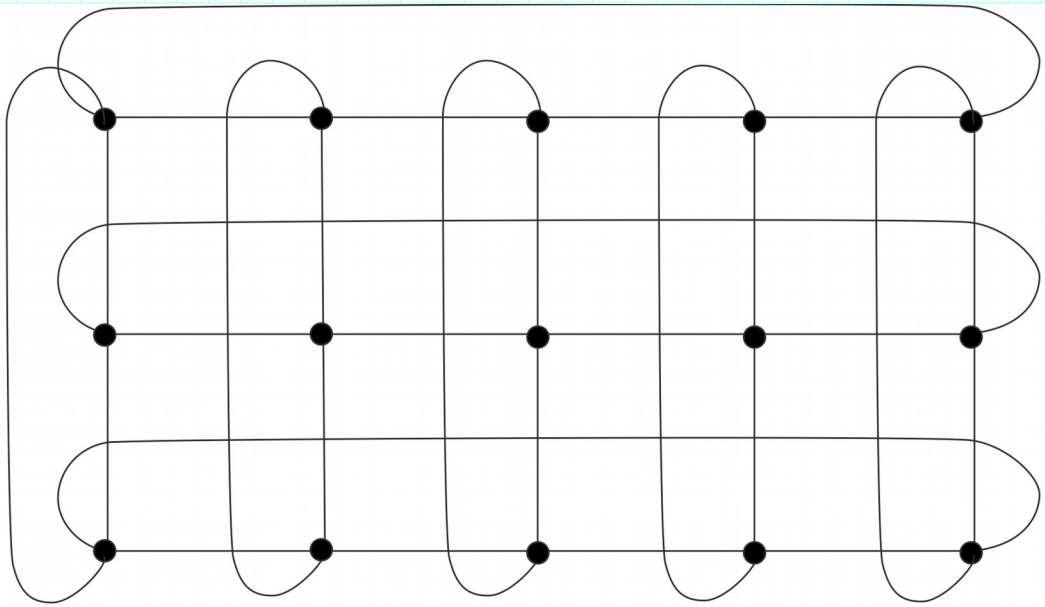
- Characterizing congestion:
  - Bisection width
    - Or more accurately, bisection bandwidth
  - Redundancy

Linear Array or Ring

- What is the bisection width of these topologies?

2D and 3D Arrays





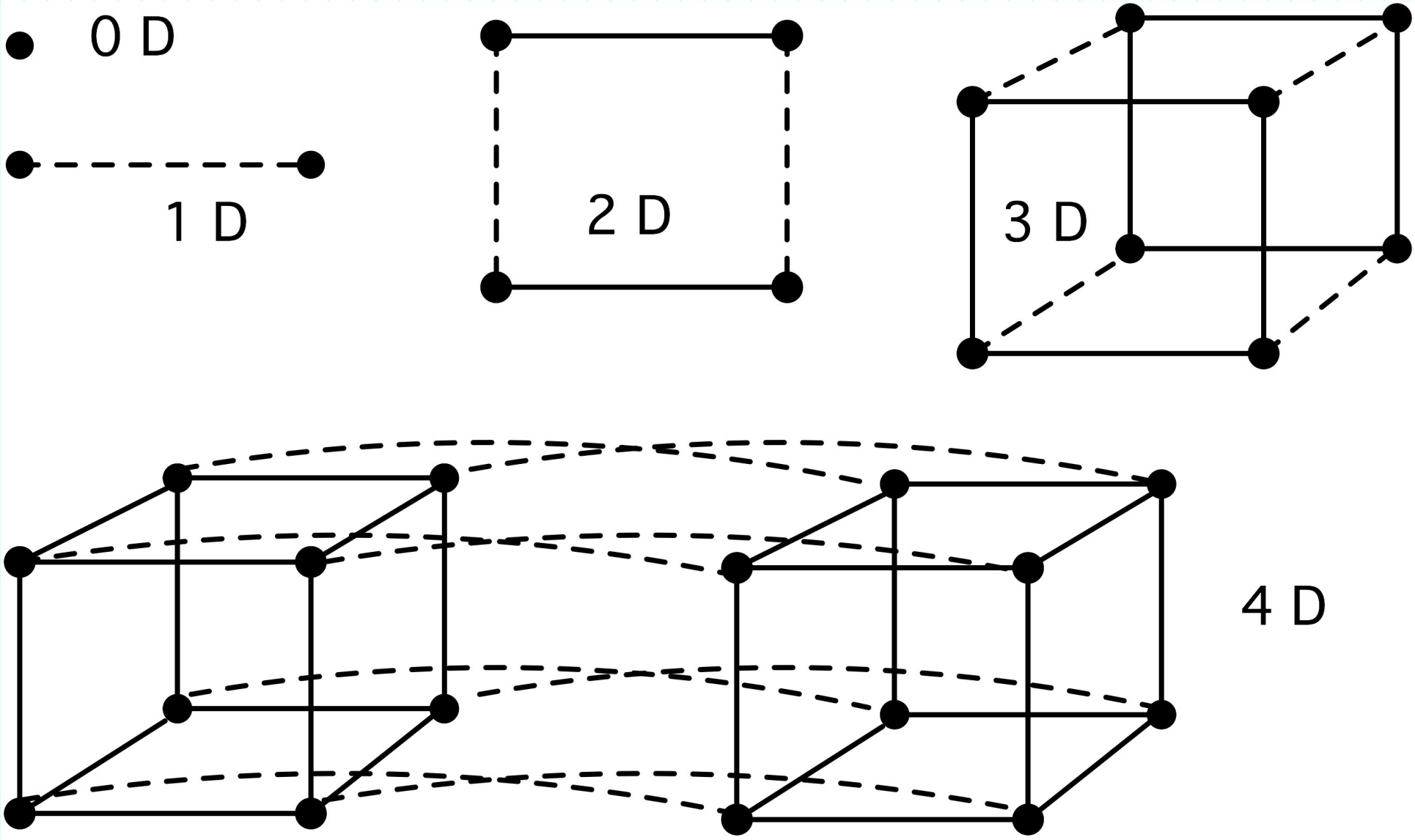
Exercises

- What is the bisection width and diameter of the two topologies on the previous slide?

- What is the bisection width and diameter of a 4 x 4 x 4 3D cube topology?

- What are some limitations of the 2D/3D approach?

Hypercubes

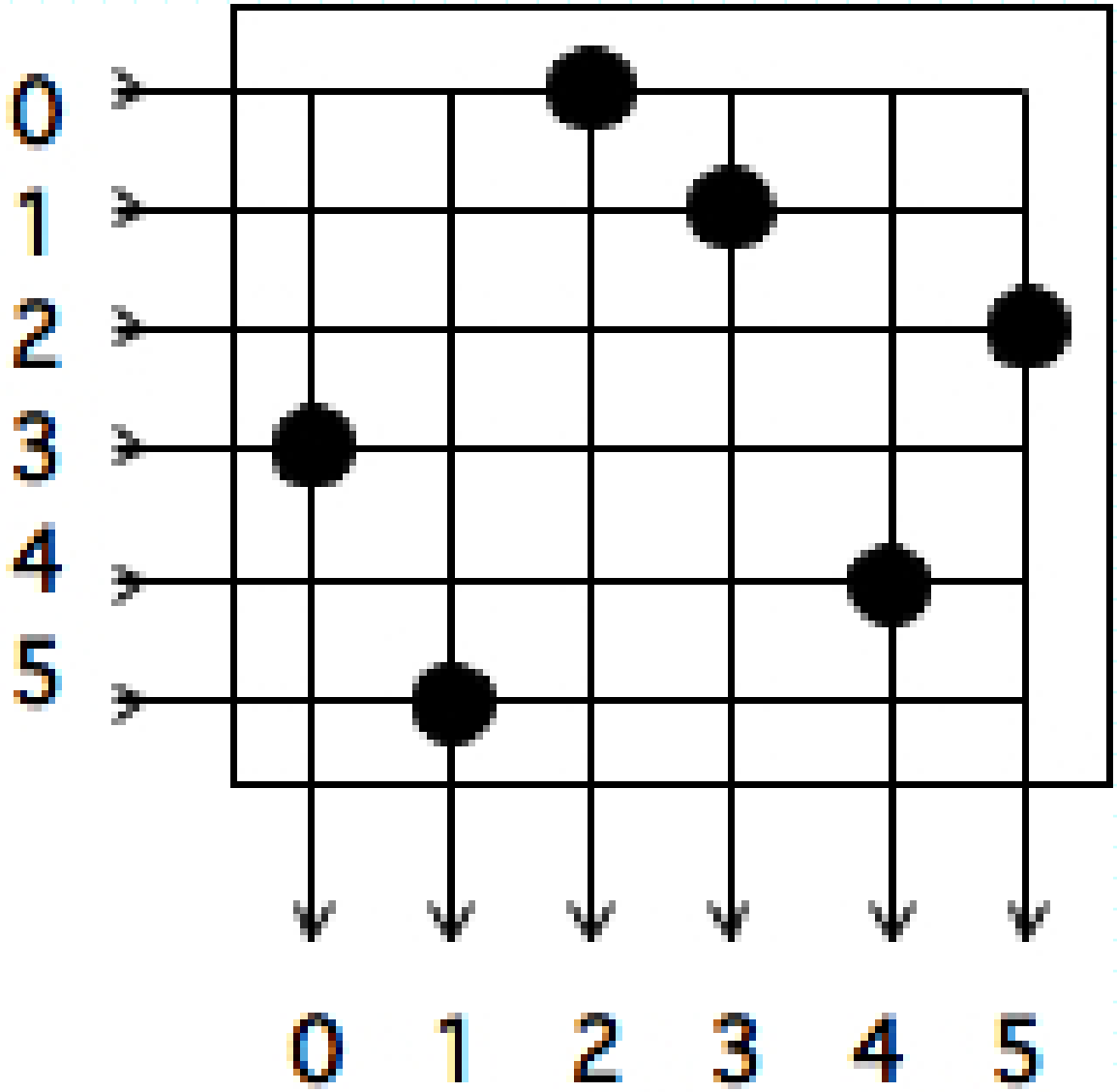


Hypercube

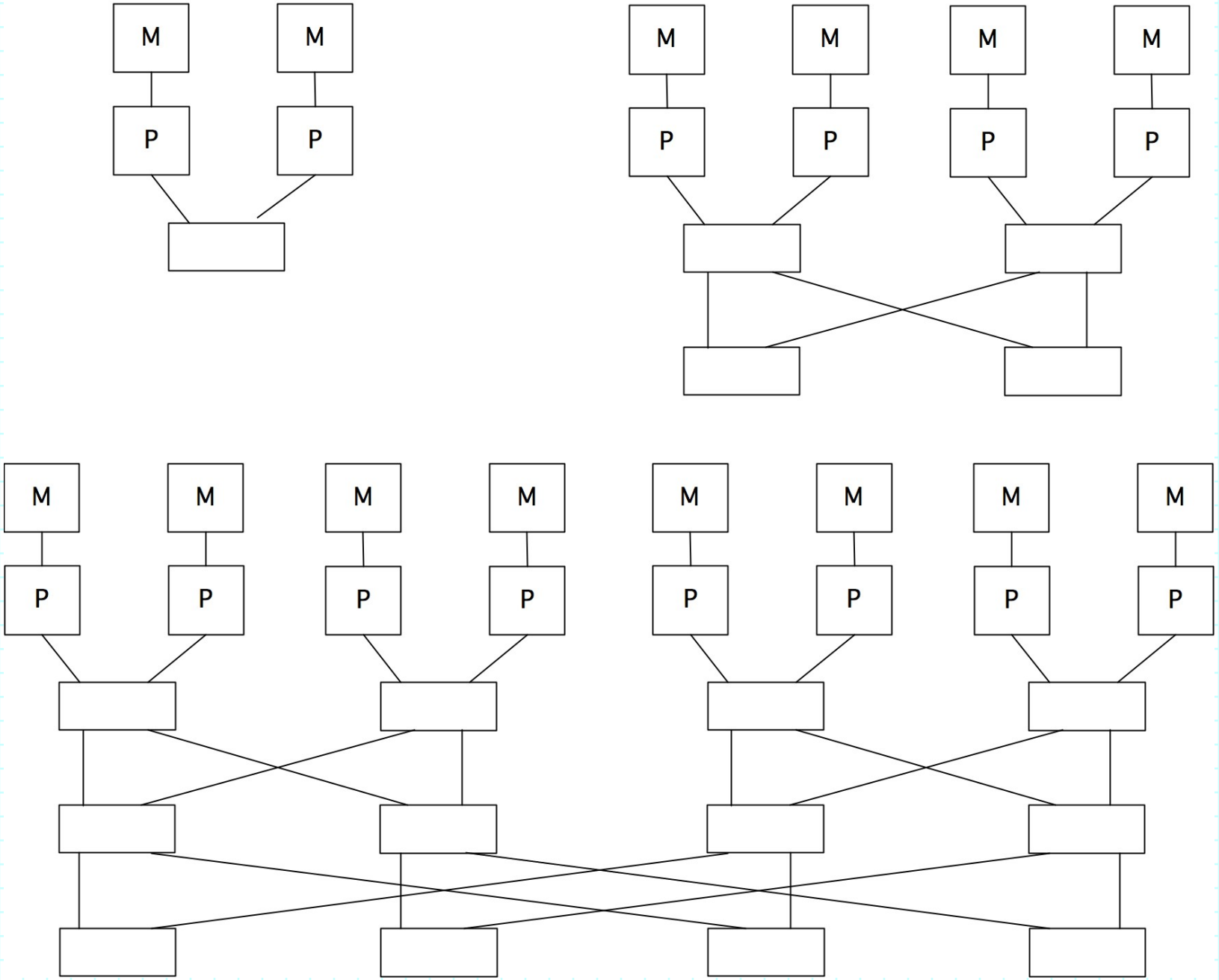
- What is the diameter of the topology on the previous slide? The bisection width?

- What if we want to do grid-like computations on a hypercube. Is this possible?

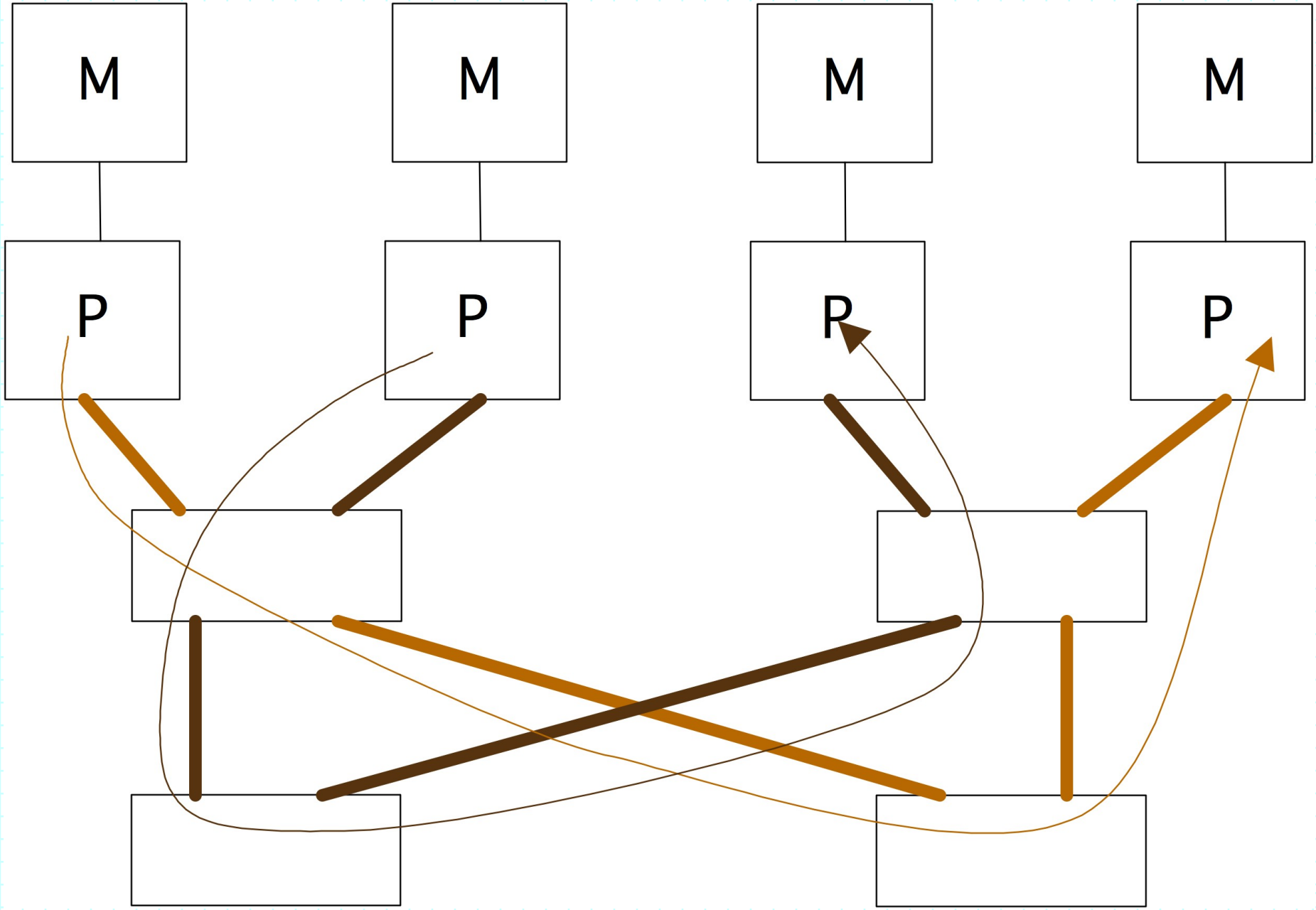
Switched Networks



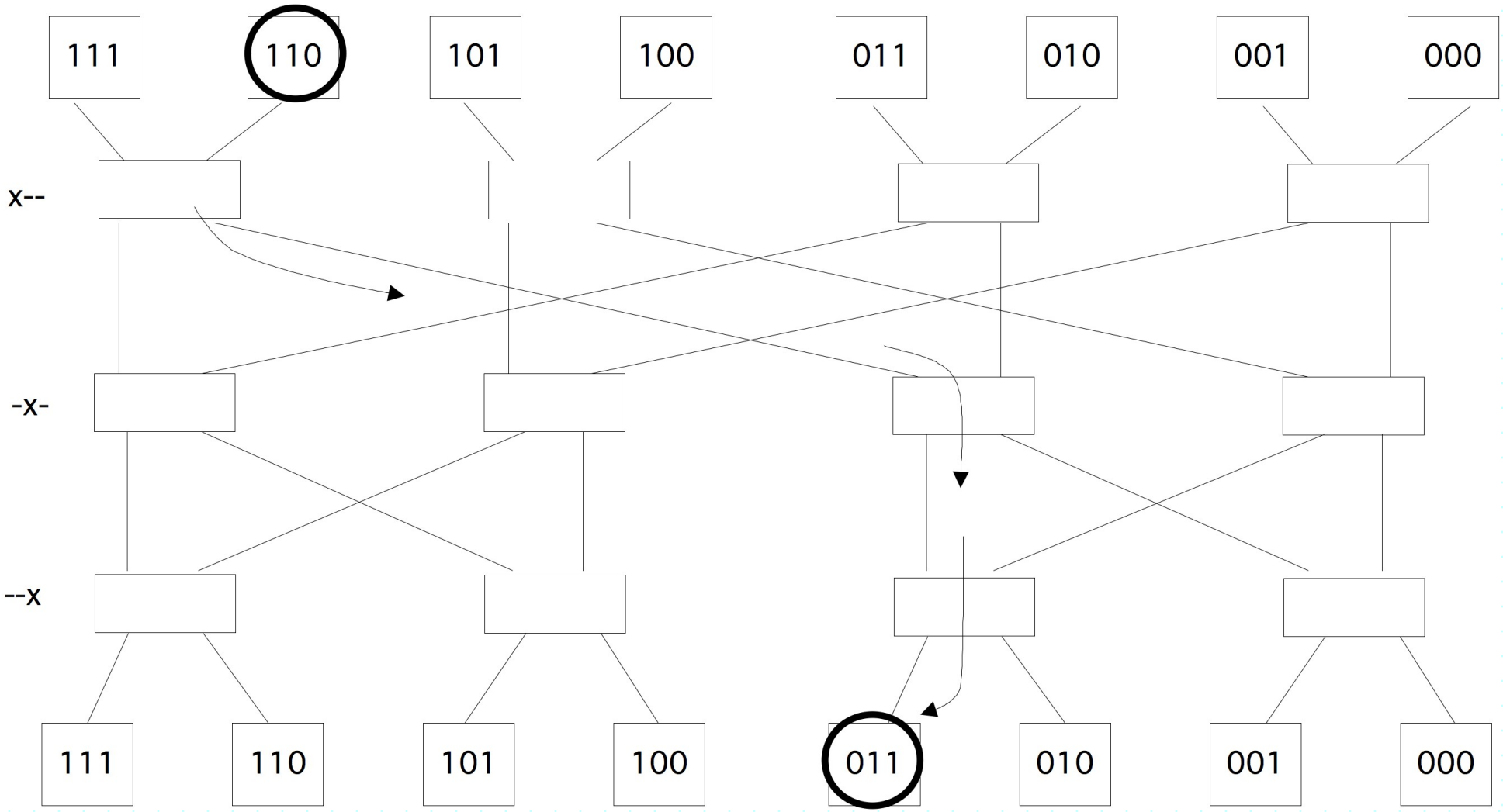
Butterfly Exchange



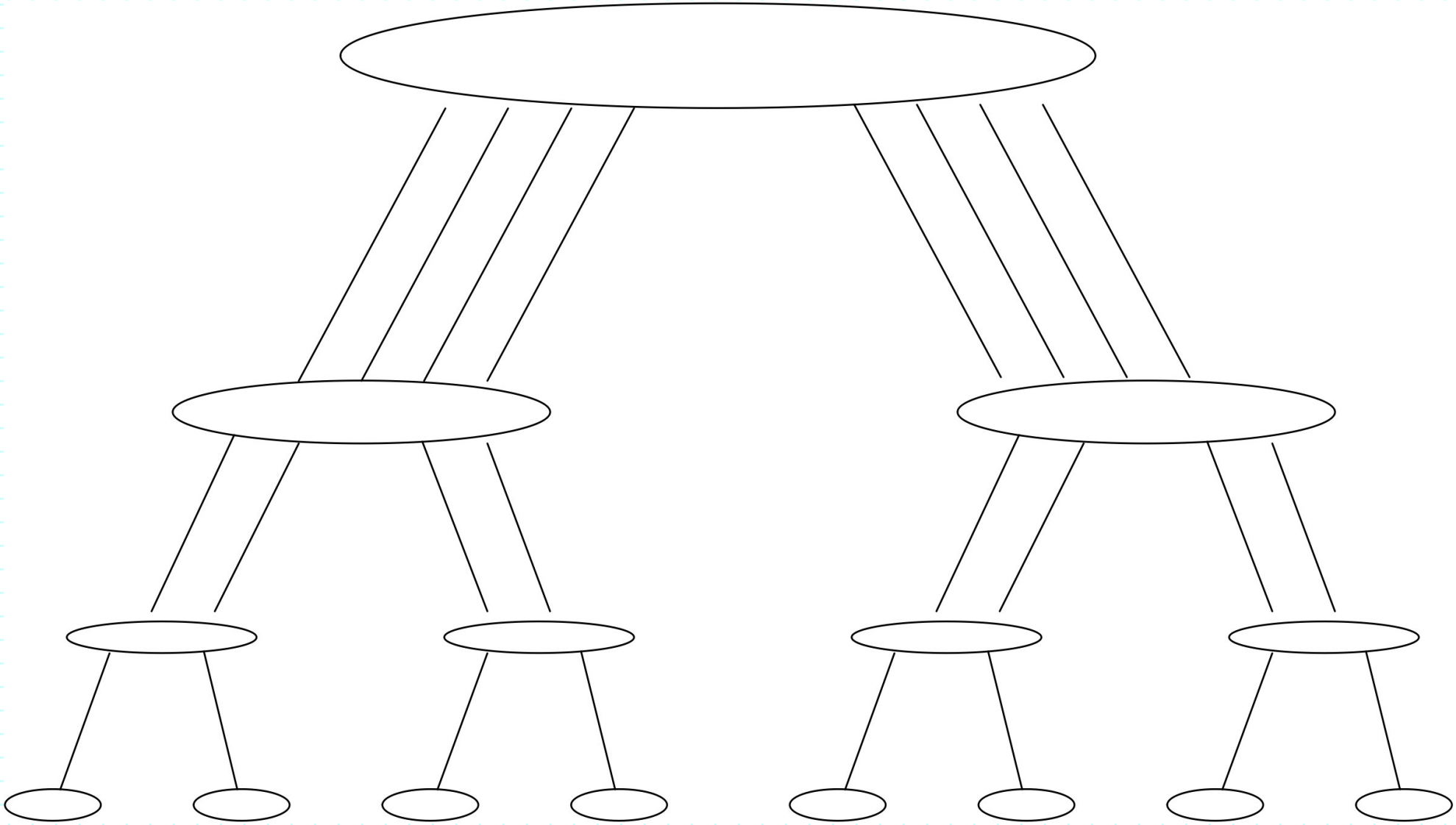
Butterfly Exchange



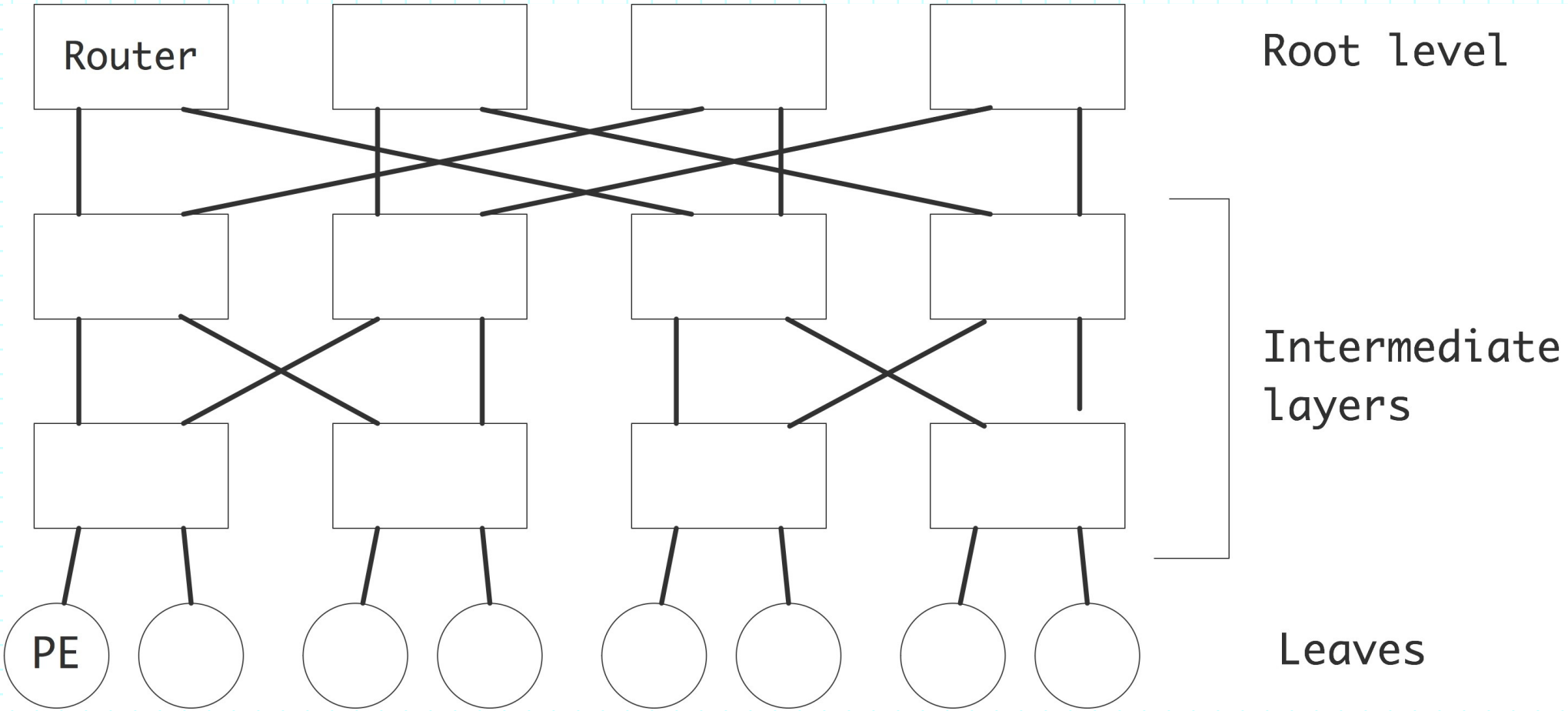
Butterfly Exchange



Fat Trees



Fat Trees



Bandwidth and Latency

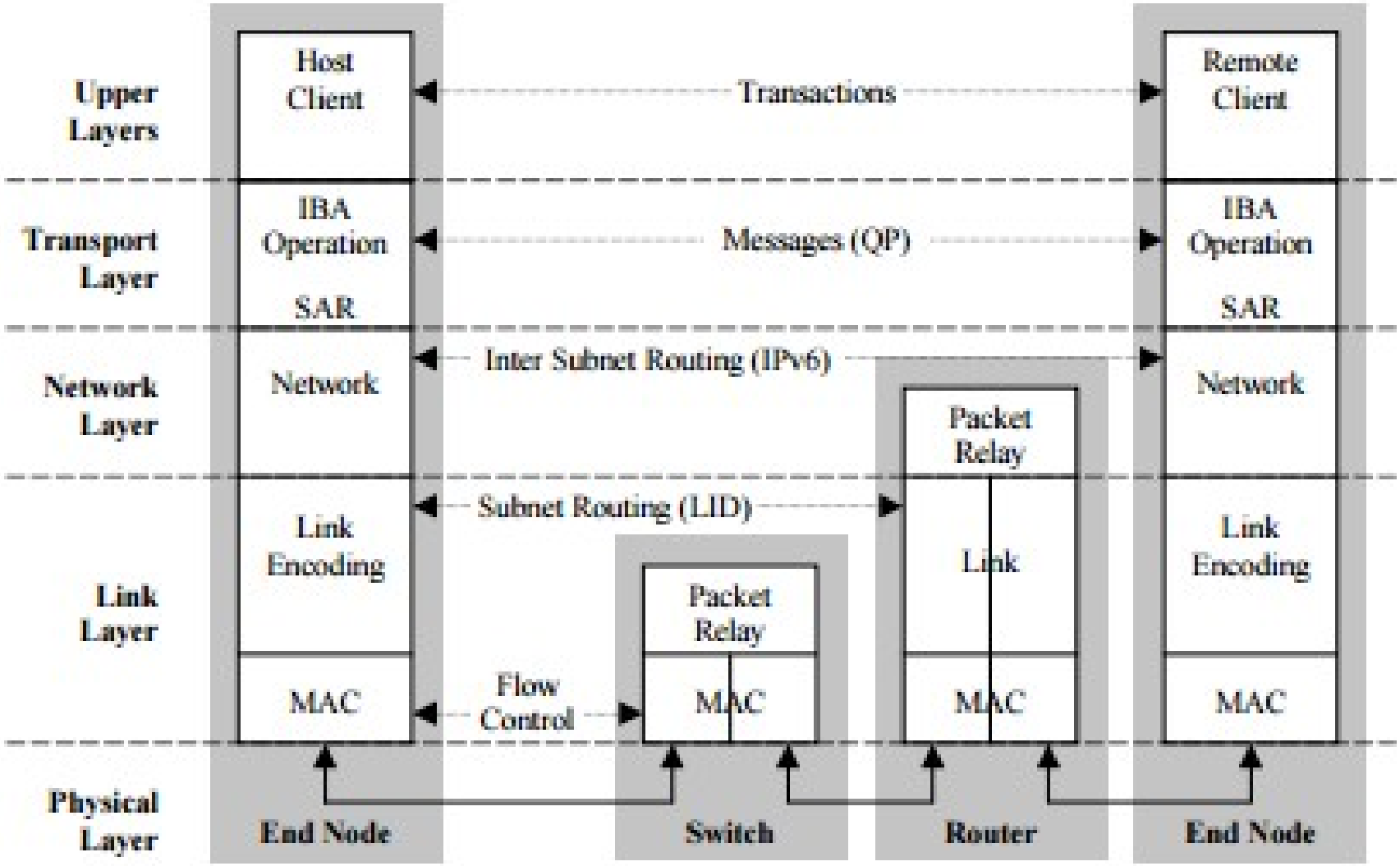
- All connections have some sort of latency
  - Why?

- Characterizing communication:
  - $T(n)=\alpha+\beta n$
  - $T(n)=\alpha+\beta n+\gamma n$
  - $T(n,p)=\alpha+\beta n+\delta p$

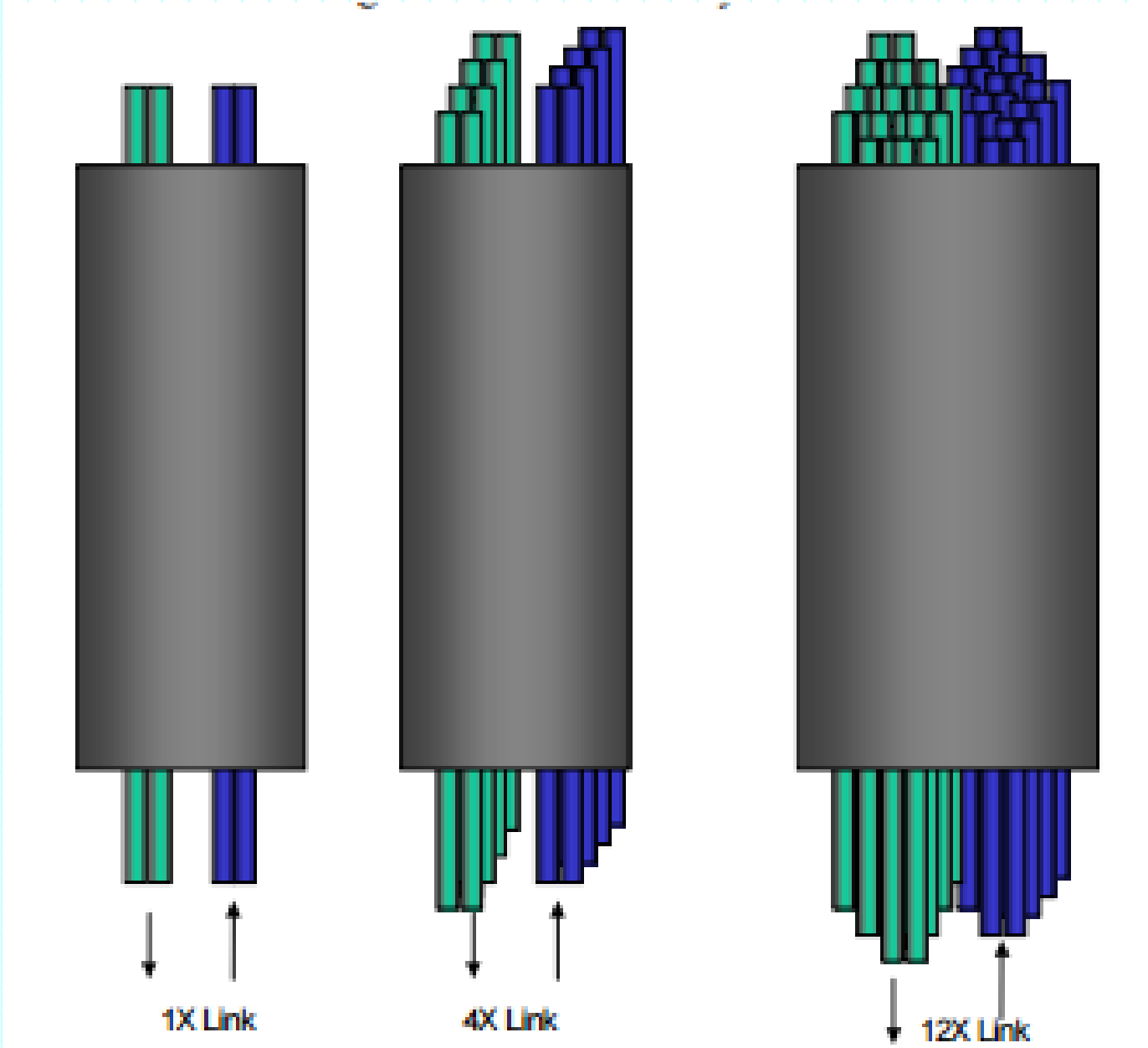
Infiniband

- Most popular interconnect on the TOP500
  - Not by much though…
- Originated in 1999
- Features:
  - Switch based
  - Point to point
- Mellanox Technologies Inc. "Introduction to Infiniband" 2003

Infiniband - Layers



Infiniband – Physical Layer



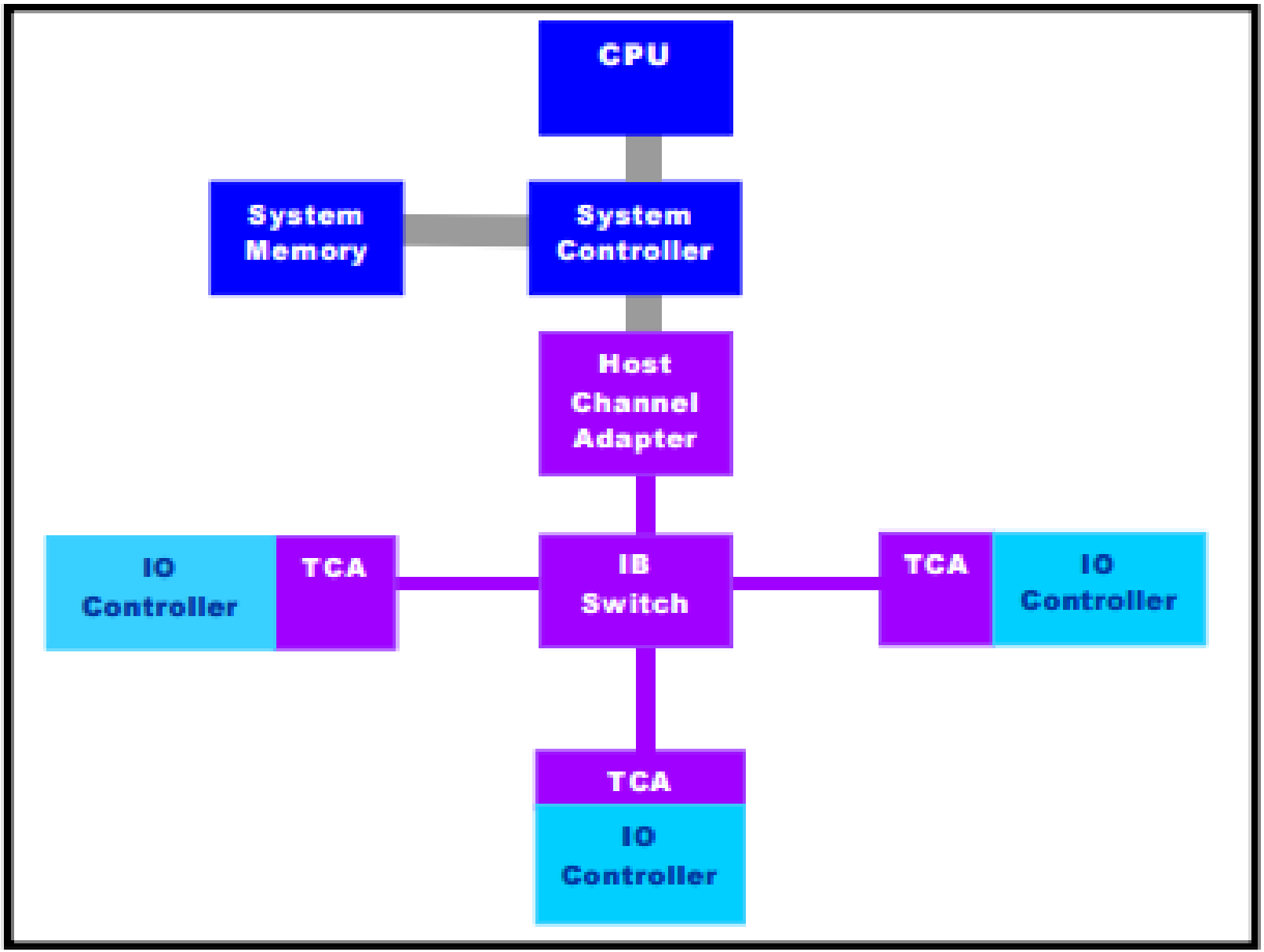
Infiniband – Link Layer

- Features:
  - Packets
  - Switching
  - QoS

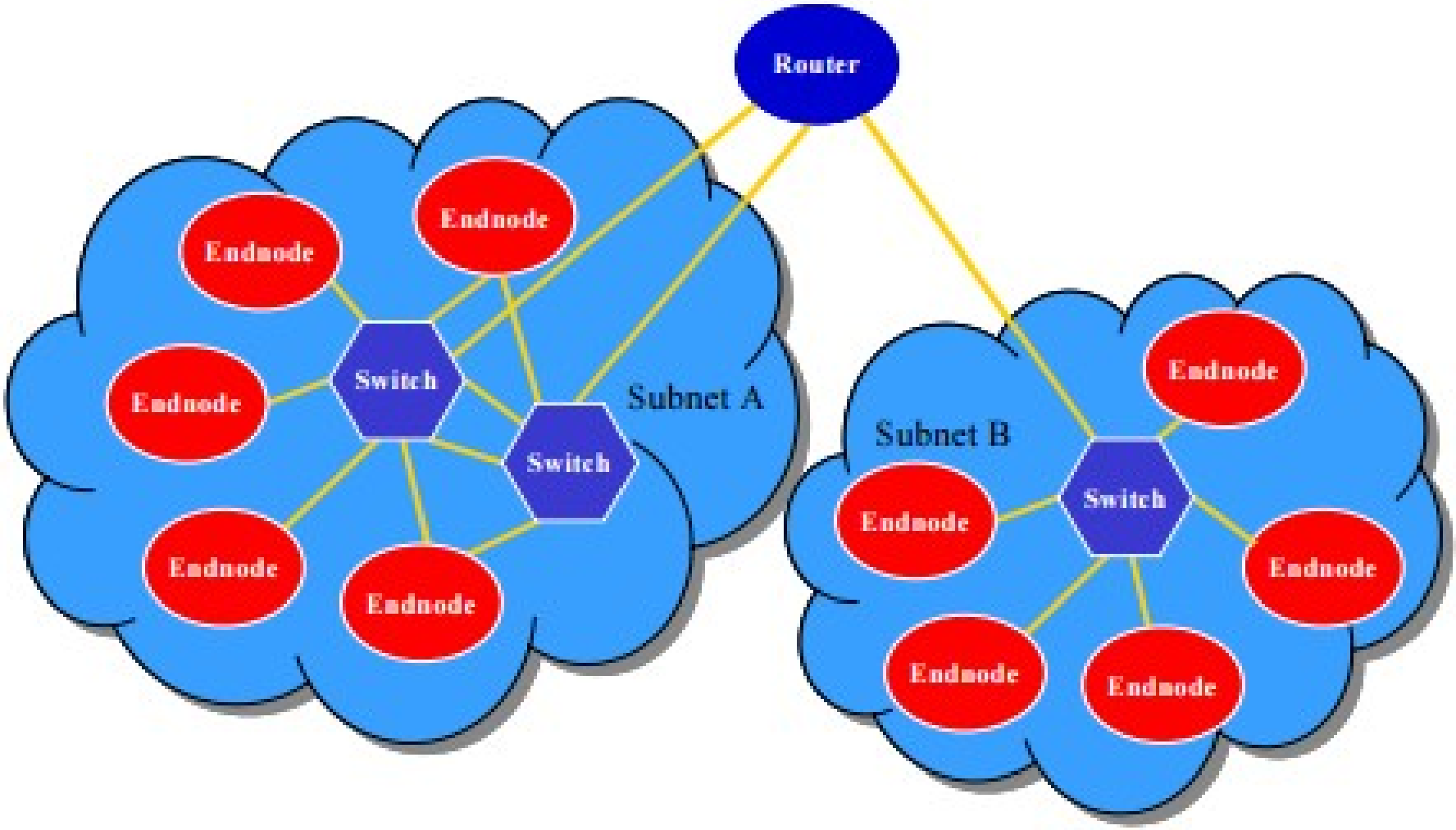
Infiniband - Layers

- Network Layer
- Transport Layer

Infiniband - Elements



Infiniband - Switches



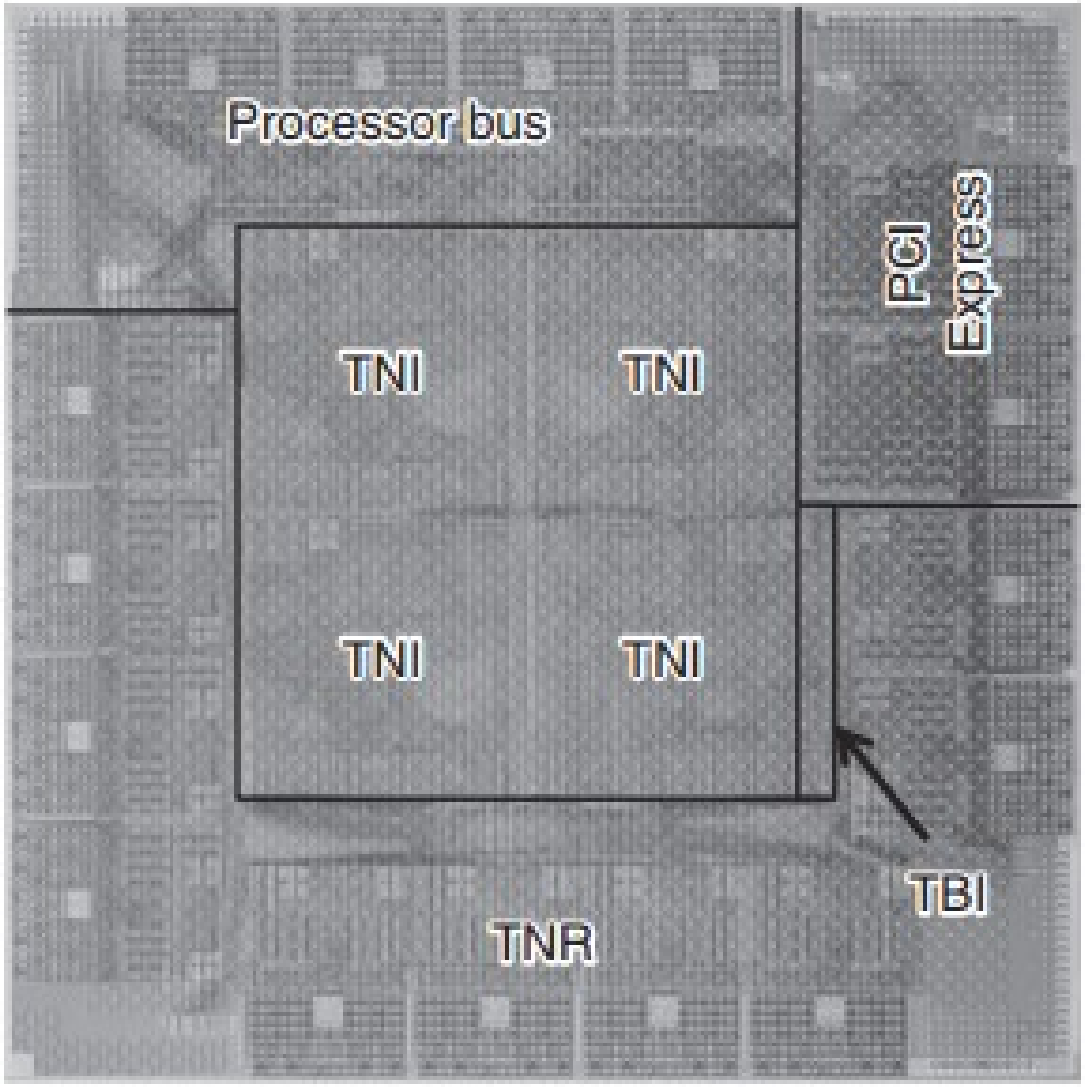
Tofu

- A custom interconnect designed specifically for the K Computer
  - #5
  - More than 80000 cores

- Ajima, et. al, "Tofu: Interconnect for the K Computer" 2012

Tofu – Interconnect Controller

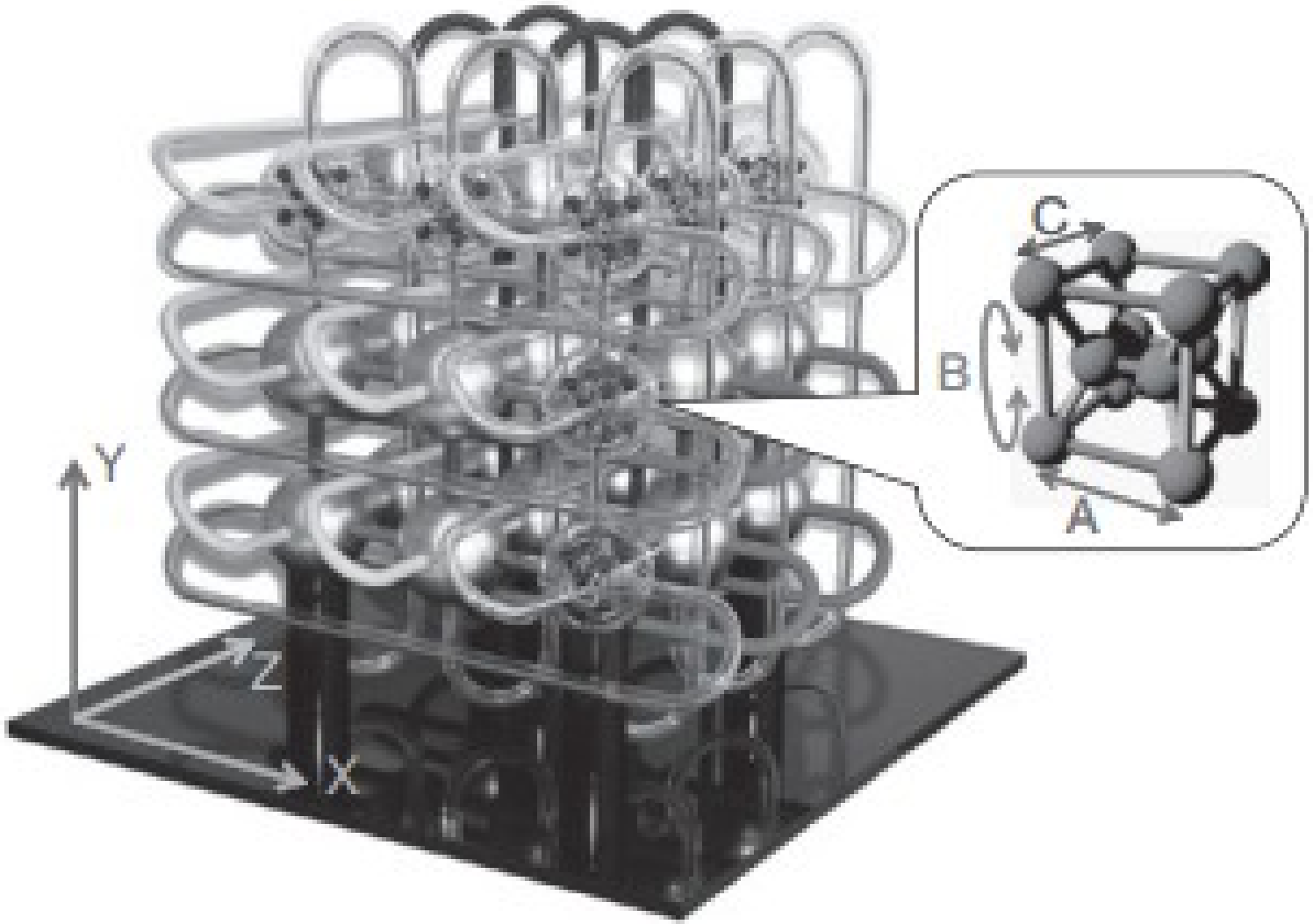
- Specialized chip for implementing the interconnect



Tofu – Interconnect Controller

Item	Specifications
No. of simultaneous communication	4 transmission connections + 4 reception connections
Operating frequency	312.5 MHz
Switching capacity	100 GB/s
Link rate	5 GB/s × Bidirectional
No. of ports	10
Process technology	65-nm CMOS
Die size	18.2 mm × 18.1 mm
No. of logic gates	48 million gates
No. of SRAM cells	12 million bits
Differential I/O signals	
Tofu link	6.25 Gb/s, 80 lanes
Processor bus	6.25 Gb/s, 32 lanes
PCI Express	5 Gb/s, 16 lanes

Tofu – 6D Network



Tofu – High Performance Communications

- RDMA
- Low Latency
- Four way simultaneous
- Virtual Channel
- High Reliability

Tofu Barrier Interface

