Parallel and Distributed Computing
CS3006 (BCS-6C/6D)
Lecture 05

Instructor: Dr. Syed Mohammad Irteza

Assistant Professor, Department of Computer Science, FAST

07 February, 2023

Previous Lecture

- Multi-processors (UMA, NUMA)
- Flynn's Taxonomy:
  - SISD, SIMD
  - MISD, MIMD
- Shared Memory Interconnection Networks
  - Central data bus
  - Crossbar
  - Multi-stage Omega network

Network Topologies: Multistage Omega Network

- Machines like IBM eServer p575 and SGI Altix 4000 use Ω-network
- Ω-network is much more interesting for a large number of processors
- Problem: the switches have to be fast enough, and also the width of the links is important. 16 bit parallel is better than serial links
- Multiprocessor vector-processors use crossbars instead (because at most only 32 processors)
- Synchronization is usually performed with special communication registers (CPU to CPU); if there is little synchronization shared memory is admitted

Advantages of shared memory machines

- User-friendly programming environment due to global address space
- Data sharing is fast and uniform due to proximity of memory to CPUs
- Memory coherence is managed by the operating system

Disadvantages of shared memory machines

- Performance degradation due to “memory contention” (several processors try to access the same memory location)
- Programmer’s responsibility for synchronization constructs (correct access to memory)
- Expensive to design shared memory computers with increasing numbers of processors
- Adding processors can geometrically increase traffic on the shared memory-CPU path and for cache coherence management

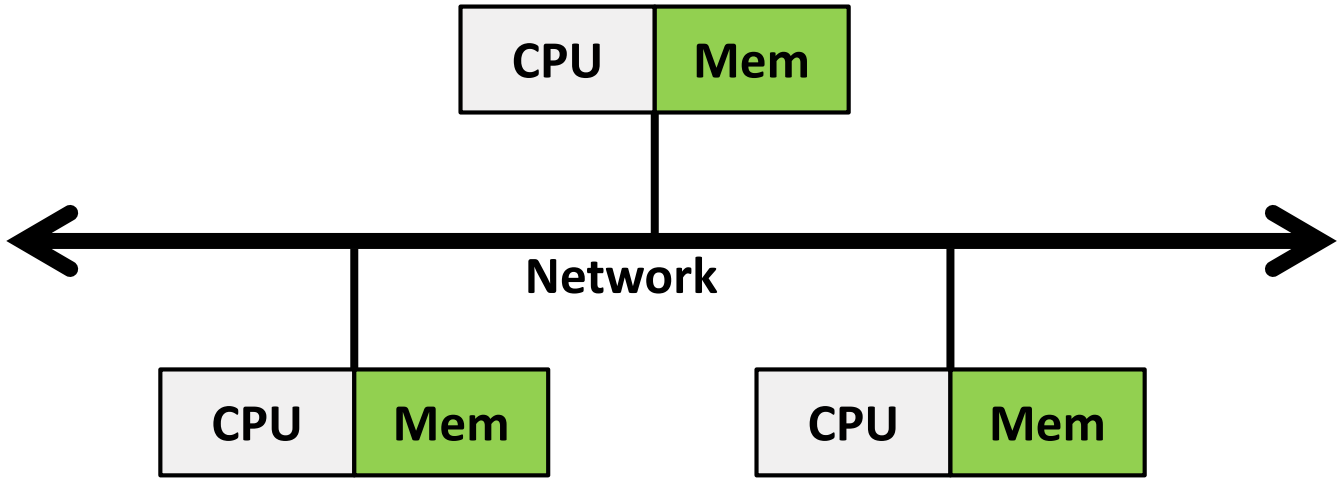
Distributed-memory SIMD (1)

- These machines are sometimes also known as processor-array machines
- They work in lock-step, i.e. all processors execute the same instruction at the same time (but on different data items); no synchronization is required
- A control processor issues the instructions that are to be executed by the processors in the processor array
- Processors sometimes are of a very simple bit-serial type (i.e., the processors operate on the data items bitwise, irrespective of their type), which can operate on operands of any length (when operands are little, this may result in speedups)

Distributed-memory SIMD (2)

- Graphical Processing Units (GPUs) are similar to processor array systems
- DM-SIMD are specialized on digital signal and image processing and on certain types of Monte Carlo simulations with virtually no exchange between processors
- Operations that cannot be executed by the processor array or by the control processor are off-loaded to the front-end system
- I/O may be through:
  - the front end system
  - the processor array (it is more efficient in providing data directly to the memory of the processor array)
  - both

Distributed-memory MIMD



- Processors have their own local memory
- No concept of global address space across all processors
- Distributed memory systems require a communication network to connect inter-processor memory

Examples

Shared-memory SIMD systems

- The Hitachi S3600 series

Distributed-memory SIMD systems

- The Alenia Quadrics
- The Cambridge Parallel Processing Gamma II
- The Digital Equipment Corp. MPP series
- The MasPar MP-1
- The MasPar MP-2

Shared-memory MIMD systems

- The Cray Research Inc. Cray J90-series, T90 series
- The Hitachi S3800 series
- The HP/Convex C4 series
- The Digital Equipment Corp. AlphaServer
- The NEC SX-4
- The Silicon Graphics Power Challenge
- The Tera MTA

Distributed-memory MIMD systems

- The Alex AVX 2
- The Avalon A12
- The C-DAC PARAM 9000/SS
- The Cray Research Inc. T3E
- The Fujitsu AP1000
- The Fujitsu VPP300 series
- The Hitachi SR2201 series
- The HP/Convex Exemplar SPP-1200
- The IBM 9076 SP2
- The Intel Paragon XP
- The Matsushita ADENART
- The Meiko Computing Surface 2
- The nCUBE 3
- The NEC Cenju-3
- The Parallel Computing Industries system
- The Parsys TA9000
- The Parsytec GC/Power Plus

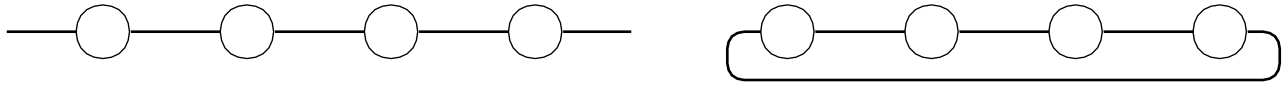
DM-MIMD Routing Mechanism

- Routing mechanism determines the path a message takes through network to reach from source to destination node.
- Data and task decomposition have to be dealt with explicitly!
- The topology and speed of the data paths are crucial and have to be balanced with costs.
- Routing can be classified as:
  - Minimal
  - Non-minimal
- It can also be classified as:
  - Deterministic routing
  - Adaptive routing

Network Topologies

Linear Arrays

- In a linear array, each node has two neighbors, one to its left and one to its right.
- If the nodes at either end are connected, we refer to it as a 1-D torus or a ring.

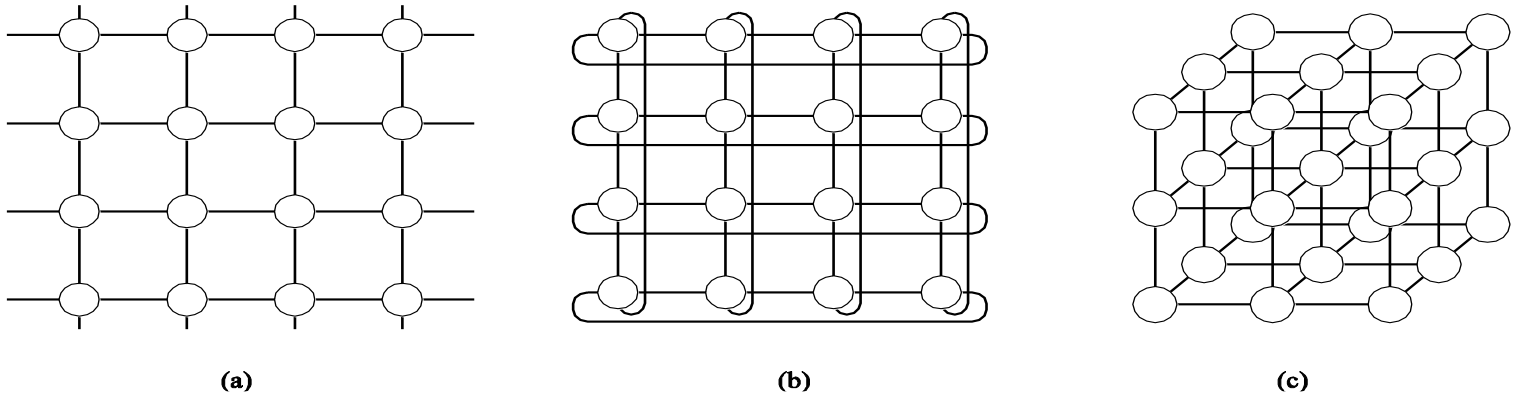


Linear arrays: (a) with no wraparound links; (b) with wraparound link.

Network Topologies

K-d Meshes

- A generalization has nodes with 4 neighbors, to the north, south, east, and west.
- A further generalization to *d* dimensions has nodes with *2d* neighbors (i.e., 6 neighbors in case of 3d cube).



Two and three dimensional meshes: (a) 2-D mesh with no wraparound; (b) 2-D mesh with wraparound link (2-D torus); and (c) a 3-D mesh with no wraparound.

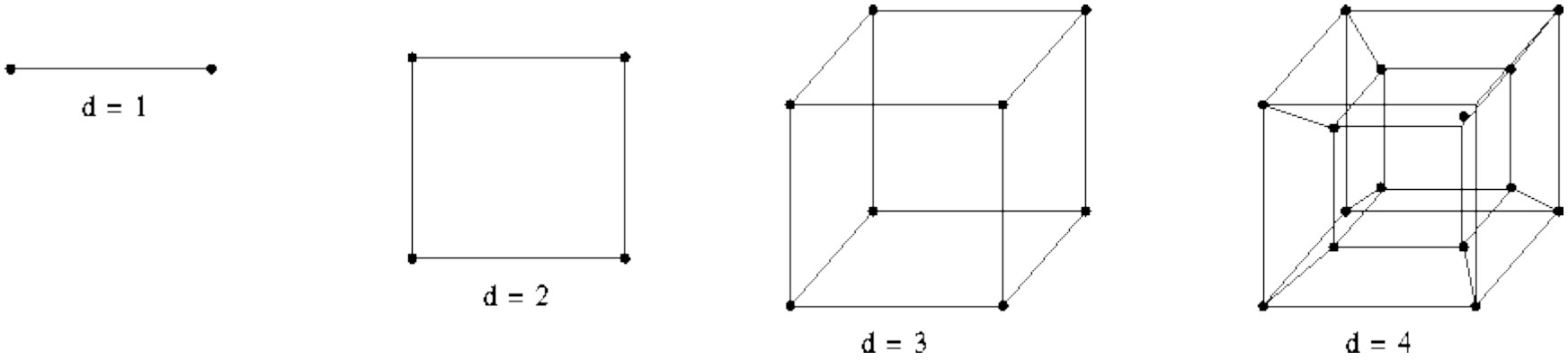
Network Topologies

Hypercube

- For a hypercube with 2^d nodes the number of steps to be taken between any two nodes is at most d (logarithmic grow)

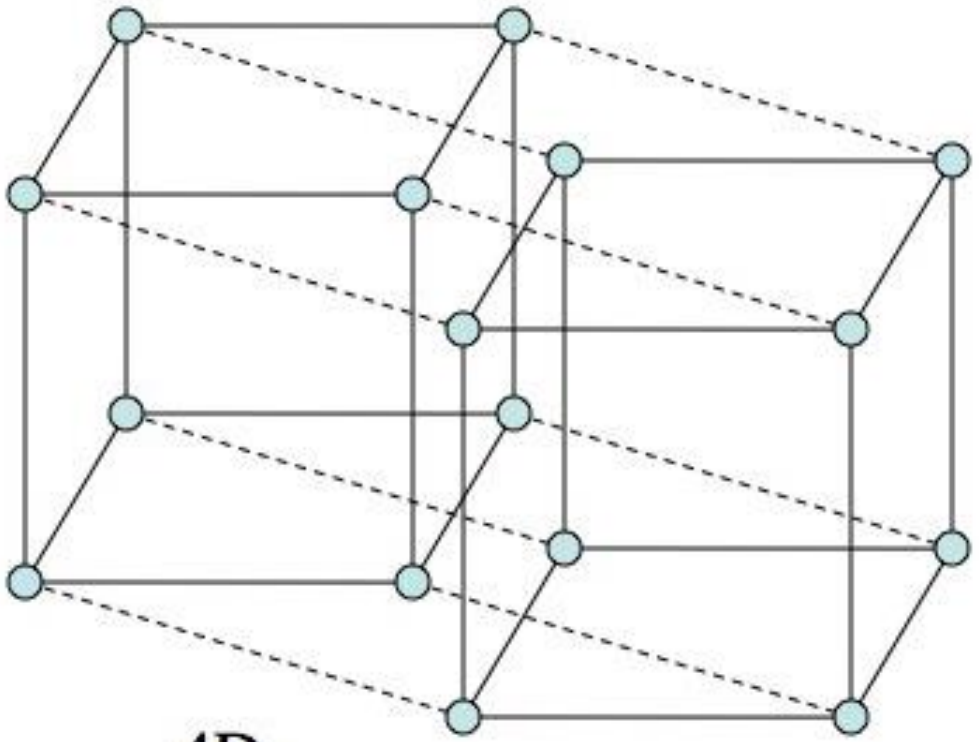
  d = log p (dimensions = log(nodes))

- The distance between any two nodes is at most *log p*.
- Each node has *log p* neighbors.



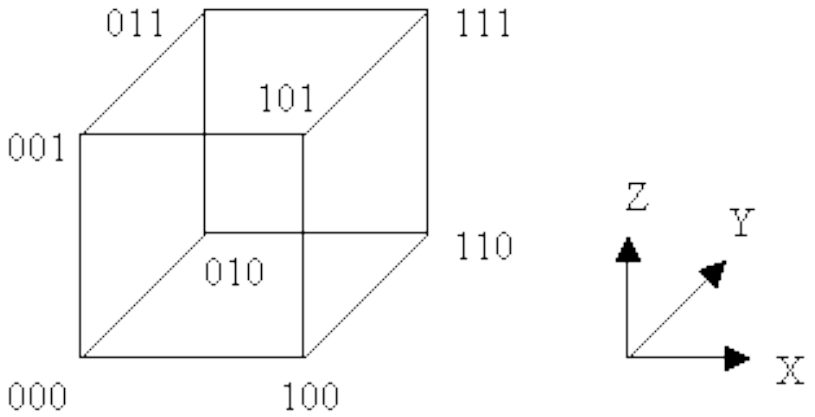
Network Topologies: hypercube

- Rule of thumb is: “d-dimensional hypercube can be constructed by connecting corresponding nodes of two (d-1)-dimensional hypercubes”



Network Topologies: hypercube

- The processors are numbered with 3-bit binary numbers which represent the X-Y-Z coordinates
- The distance between two nodes is given by the number of bit positions at which the two nodes differ.

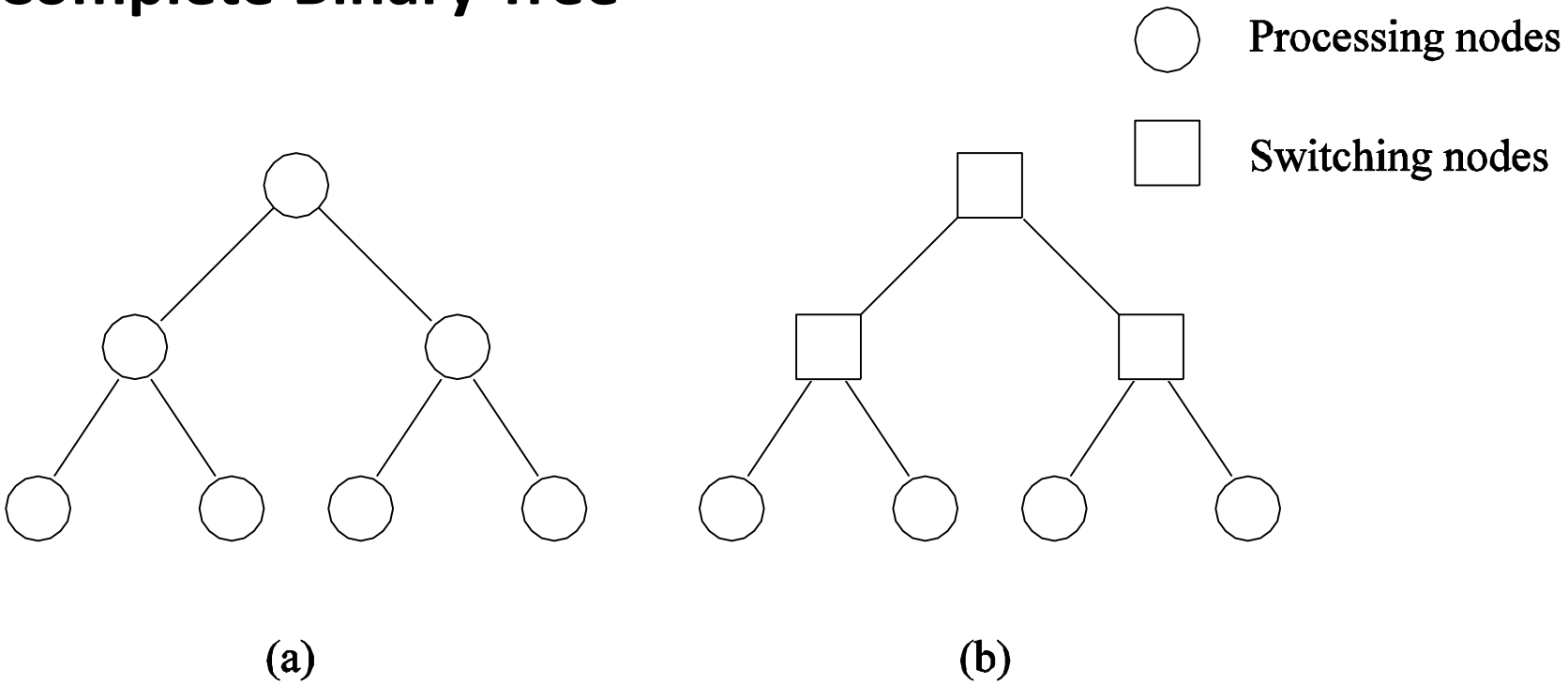


Network Topologies: Tree based Networks

- A tree network is one in which there is one path between any pair of nodes
- Linear arrays and star-connected networks are special cases of tree-based networks
- In static tree network, each node represent a processing element
- In dynamic tree network, leaf nodes represent processing element while internal nodes are switching elements.
- The source node sends the message up the tree until it reaches the node at the root of the smallest subtree containing both the source and destination nodes.
- Trees can be laid out in 2D with no wire crossings. This is an attractive property of trees.
- The distance between any two nodes is no more than $(2 * \log_2 p)$

Network Topologies: Tree based Networks

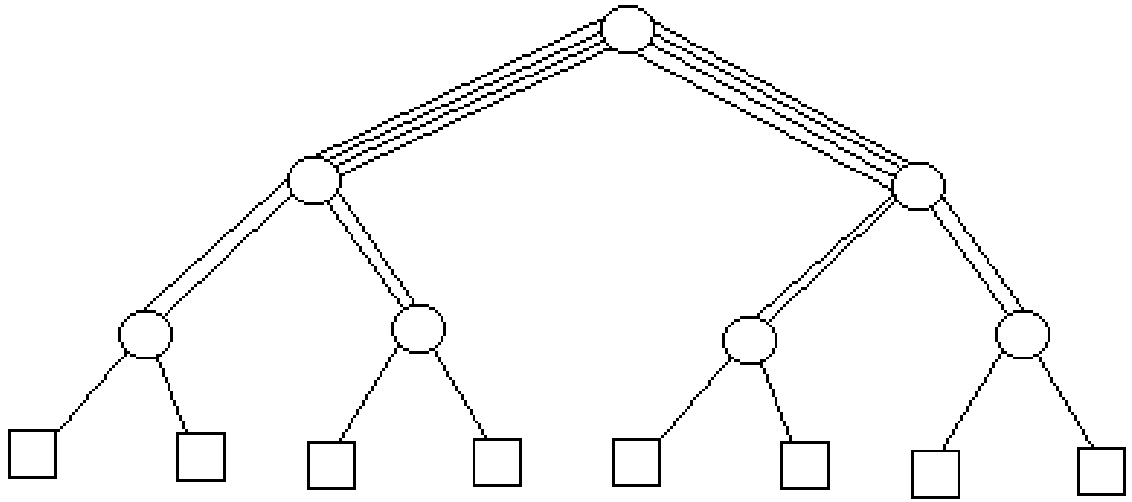
Complete Binary Tree



Complete binary tree networks: (a) a static tree network; and (b) a dynamic tree network.

Routing - Fat Tree

- Another topology is the “fat tree”
- In a tree, a node can speak to another node passing through the root so we have higher traffic at the root node.
- Fat tree amends this by providing more bandwidth (with multiple connections) in the higher levels of the tree
- N-ary fat tree is when the levels towards the root have N times the number of connections in the level below it



Here, leaf nodes are processing nodes, and all intermediate nodes are switches

Evaluating Static Interconnections

The parameters to evaluate a static interconnection:-

- **Cost:** Usually depends on number of links for communication. E.g., cost for linear array is p-1.
  - Lower values are favorable

- **Diameter:** The shortest distance between the farthest two nodes in the network. The diameter of a linear array is p − 1.
  - Lower values are favorable

- **Bisection Width:** The minimum number of wires (i.e., links) you must cut to divide the network into two equal parts. The bisection width of a linear array is 1.
  - What does this tell us about the performance of a topology?

Evaluating Static Interconnections

- **Arc-connectivity:** The minimum number of arcs or links that must be removed from the network, to break the network into two disconnected networks
  - Higher values are desirable
  - It is the minimum number of the links that must be cut to separate the single node from the network
  - Higher values means, that incase of link failure there are multiple other routes to the node.
  - Arc-connectivity of linear array is 1 and 2 for ring.

Evaluating Static Interconnections

Network	Diameter	Bisection Width	Arc Connectivity	Cost (No. of links)
Completely-connected	1	$p^2/4$	$p-1$	$p(p-1)/2$
Star	2	1	1	$p-1$
Complete binary tree	$2\log((p+1)/2)$	1	1	$p-1$
Linear array	$p-1$	1	1	$p-1$
2-D mesh, no wraparound	$2(\sqrt{p}-1)$	$\sqrt{p}$	2	$2(p-\sqrt{p})$
2-D wraparound mesh	$2\lfloor\sqrt{p}/2\rfloor$	$2\sqrt{p}$	4	$2p$
Hypercube	$\log p$	$p/2$	$\log p$	$(p\log p)/2$

Sources

- Slides of Dr. Rana Asif Rahman & Dr. Haroon Mahmood, FAST