Fundamentals of Computer Systems

IE 496 Lecture 2

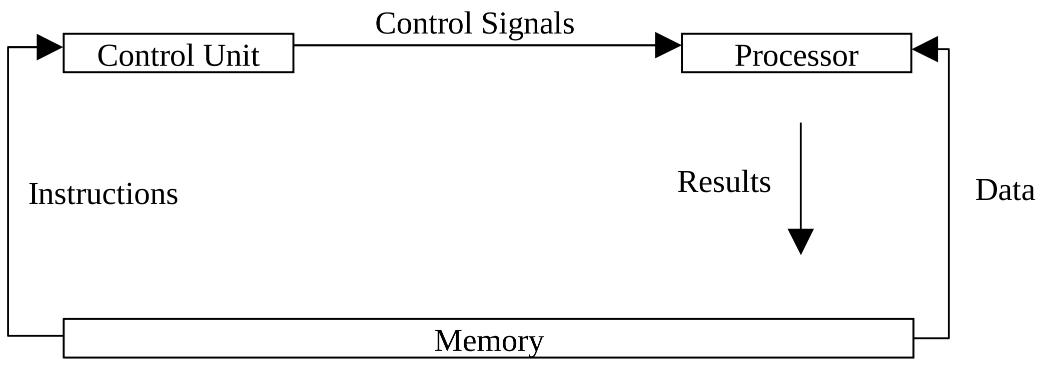
Reading for this lecture

- Miller and Boxer, Chapter 5
- Fountain, Chapter 4

Computer Architecture

Flynn's Taxonomy

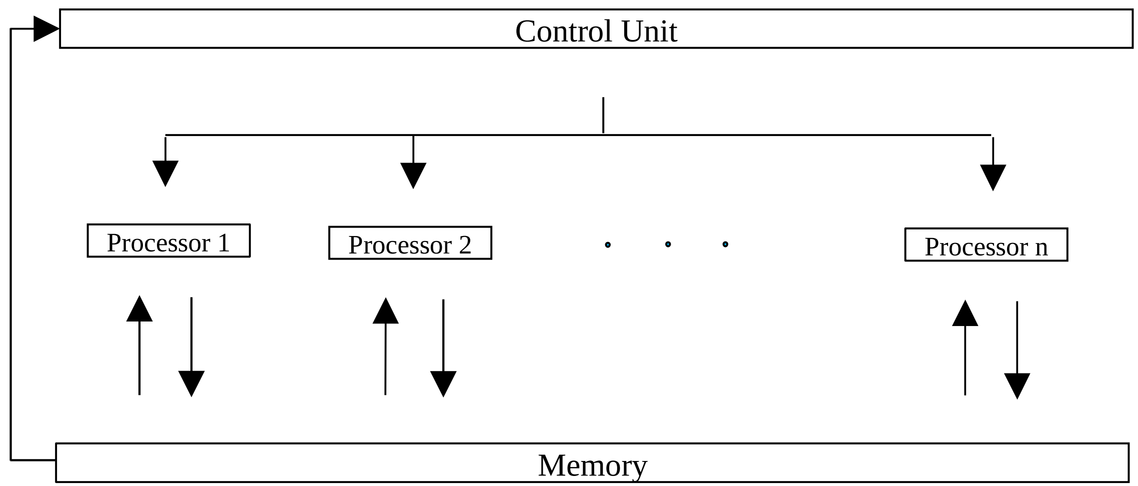
Single Instruction Stream, Single Data Stream
(Serial Computer)



Computer Architectures

Flynn's Taxonomy

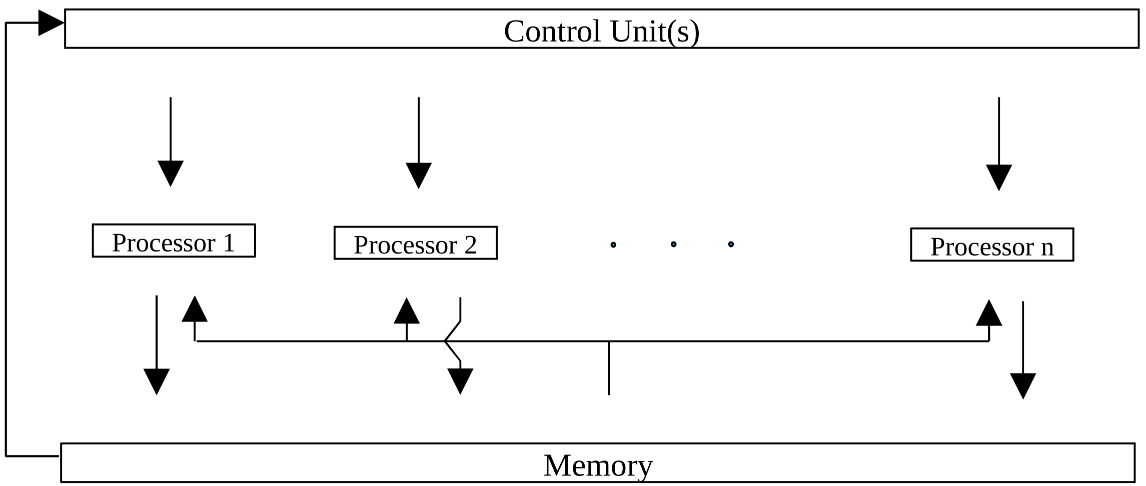
Single Instruction Stream, Multiple Data Stream (SIMD)



Computer Architectures

Flynn's Taxonomy

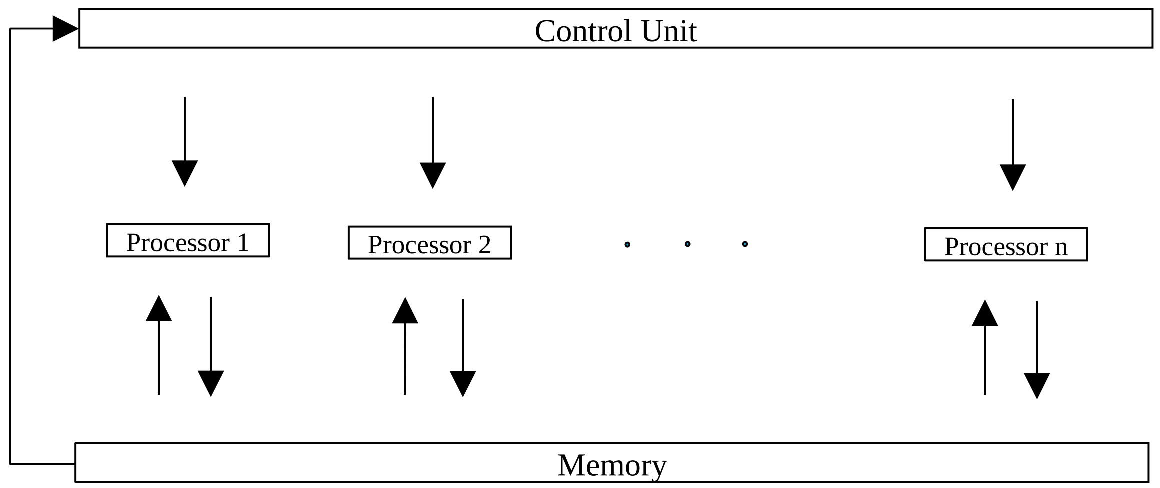
Multiple Instruction Stream, Single Data Stream (MISD)



Computer Architectures

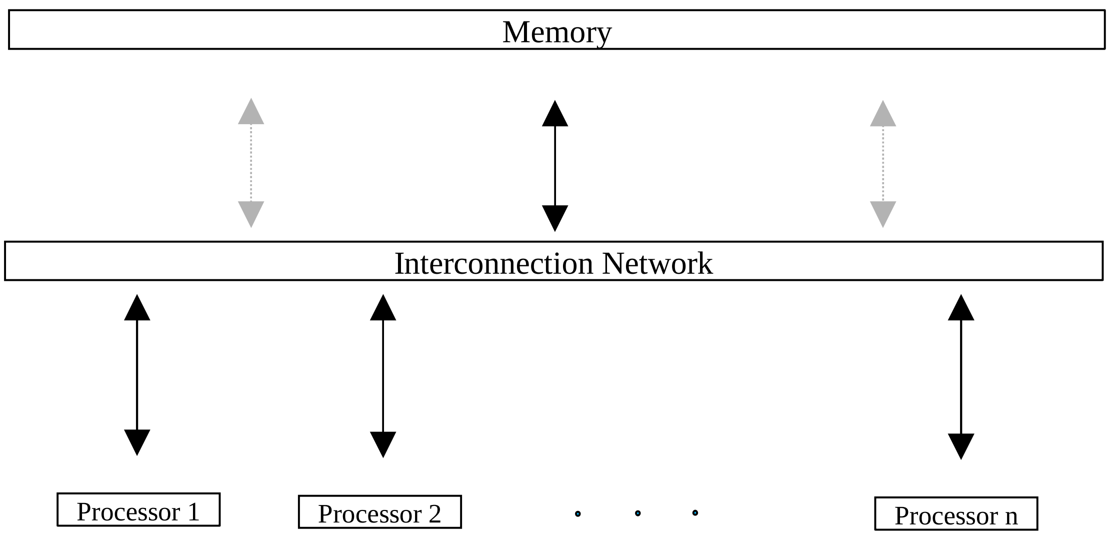
Flynn's Taxonomy

Multiple Instruction Stream, Multiple Data Stream (MIMD)



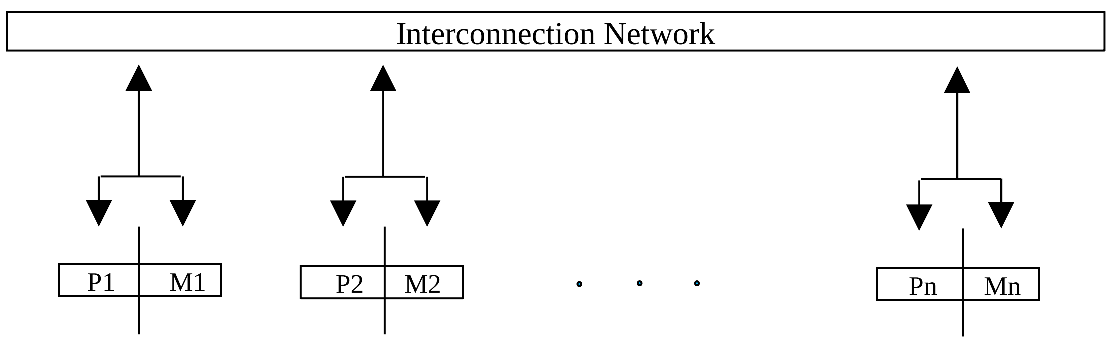
Memory Configurations

Shared Memory

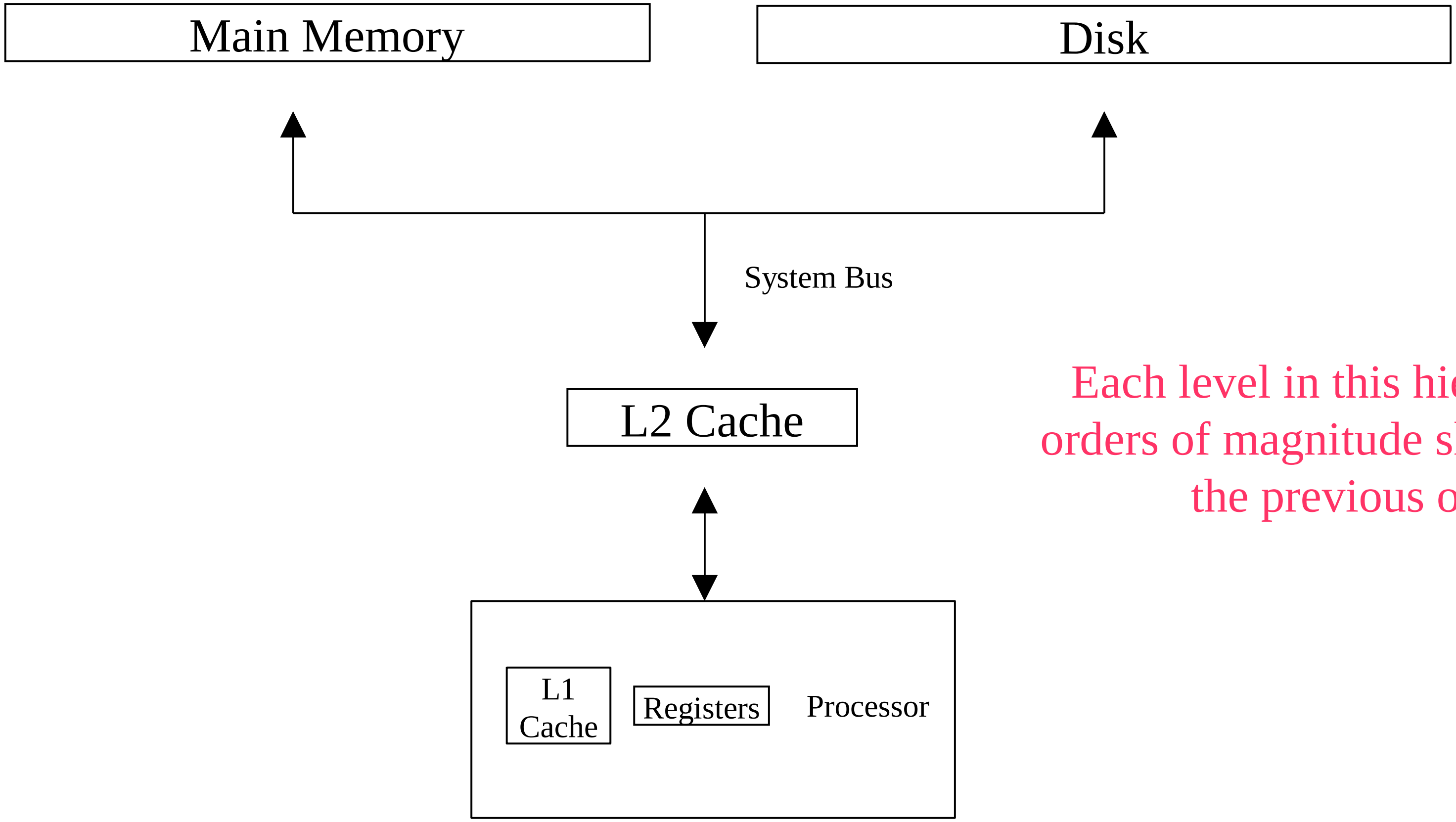


Memory Configurations

Distributed Memory



Memory Hierarchies



Each level in this hierarchy is orders of magnitude slower than the previous one.

Importance of the memory hierarchy

- The gap between the speeds of processors and memory has grown 50% per year.
- Cache memory tries to overcome this performance gap.
- However, it is relatively easy to defeat it.
- For this reason, it's important to be aware of it.
- We'll see an example later.

Interconnection Networks

Aside: Introduction to Graphs

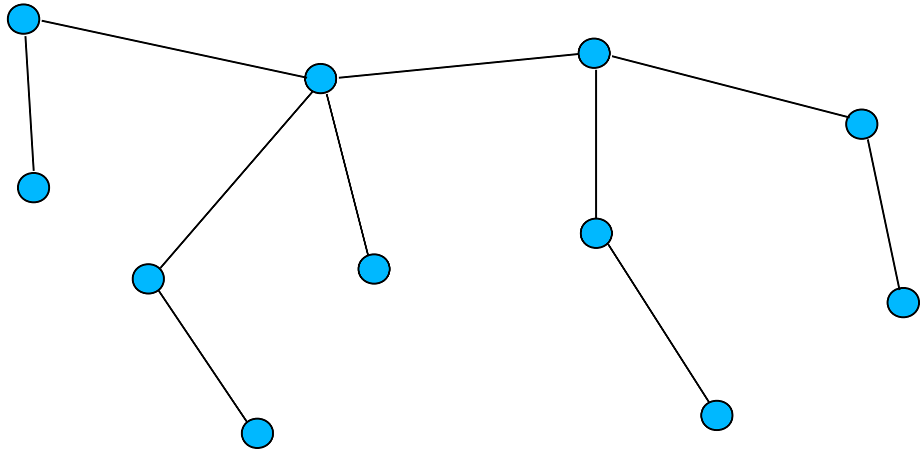
- A graph $G = (V, E)$ is defined by two sets, a finite, nonempty set V of *vertices* (or nodes) and a set $E \subseteq V \times V$ of *edges*.
- Example: A road network.
- The edges can be either ordered pairs or unordered pairs.
- If the edges are ordered pairs, then they are usually called *arcs* and the graph is called a *directed graph*.
- Otherwise, the graph is called *undirected.*
- See AHU, Section 2.3

(Undirected) Graph Terms

- Vertices *u* and *v* are *endpoints* of the edge (u, v).
- We say an edge e = (u, v) is *incident to* its endpoints.
- Two vertices *u* and *v* are *adjacent* if $(u, v) \in E$.
- The *degree* of a vertex is the number of edges incident to it (equivalently, the number of vertices adjacent to it).
- A *path* is a sequence of edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n)$
- The *length* of such a path is $n-1$.
- Often, we represent a path simply as a sequence of vertices.

Applications of Graph Theory

- Graph theory is a very rich subject area.
- Sample Applications
  - Shortest Path Problem
  - Minimum Spanning Tree
  - Traveling Salesman Problem



What is an interconnection network?

- A graph (directed or undirected)
  - The nodes are the processors
  - The edges represent direct connections
- Properties and Terms
  - Degree of the Network
  - Communication Diameter
  - Bisection Width
  - Processor Neighborhood
  - Connectivity Matrix
  - Adjacency Matrix

Measures of Goodness

- Communication diameter: The maximum shortest path between two processors.
- Bisection width: The minimum cut such that the two resulting sets of processors have the same order of magnitude.
- Connectivity Matrix
- Adjacenecy Matrix

Bottlenecks

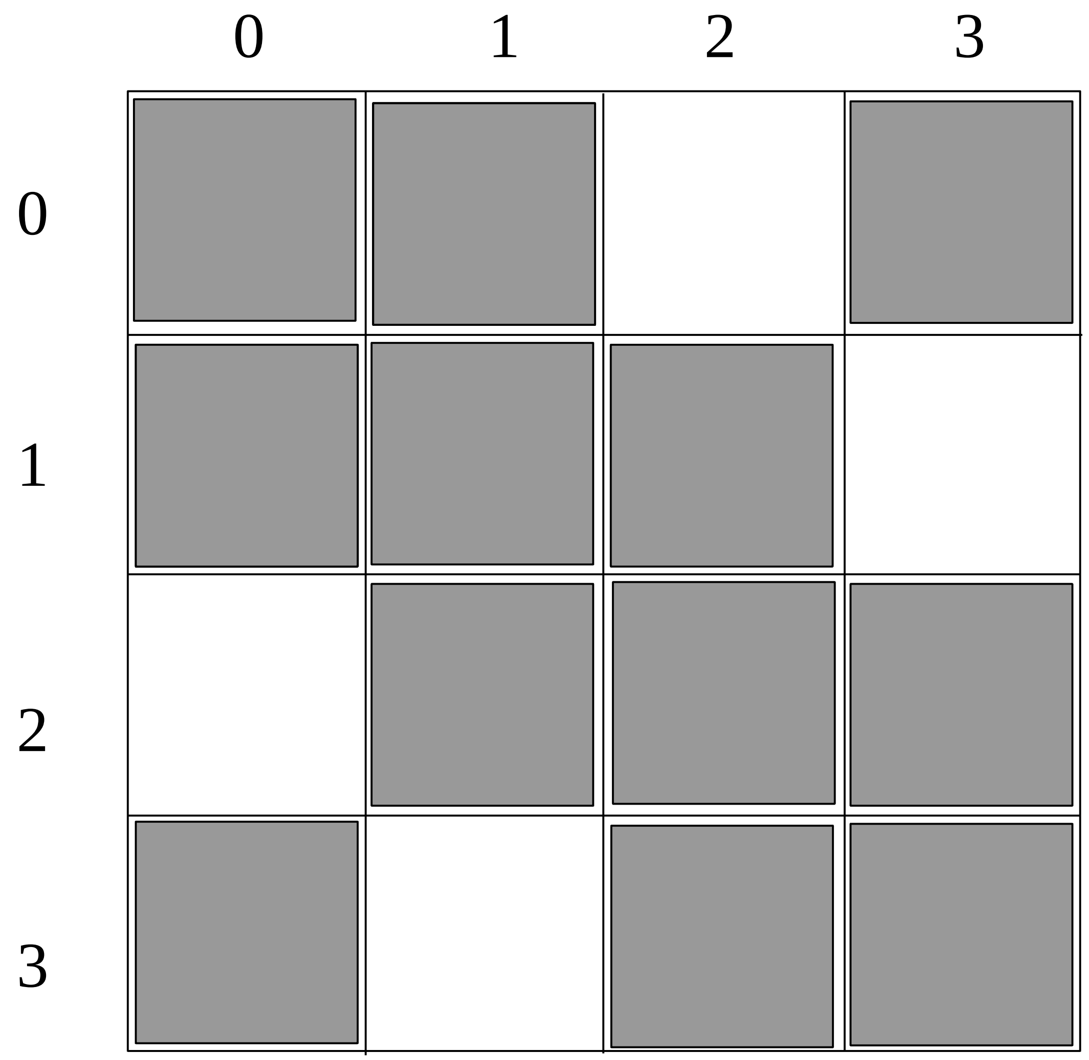
- The communication diameter indicates how long it may take to send information from one processor to another.
- Thus, it may be the bottleneck in any algorithm in which the data are initially distributed equally.
- The bisection width is the bottleneck when processors must exchange large amounts of information.
- The bisection width is a lower bound for sorting.

Benchmark Problems

- Broadcast:
  - Send value from one processor to all others
  - Limited by diameter
- Sorting:
  - Sort a list of values
  - Limited by bissection bandwidth
- Semigroup
  - Combine values using a binary associative operator
  - Requires bandwidth and diameter to be balanced

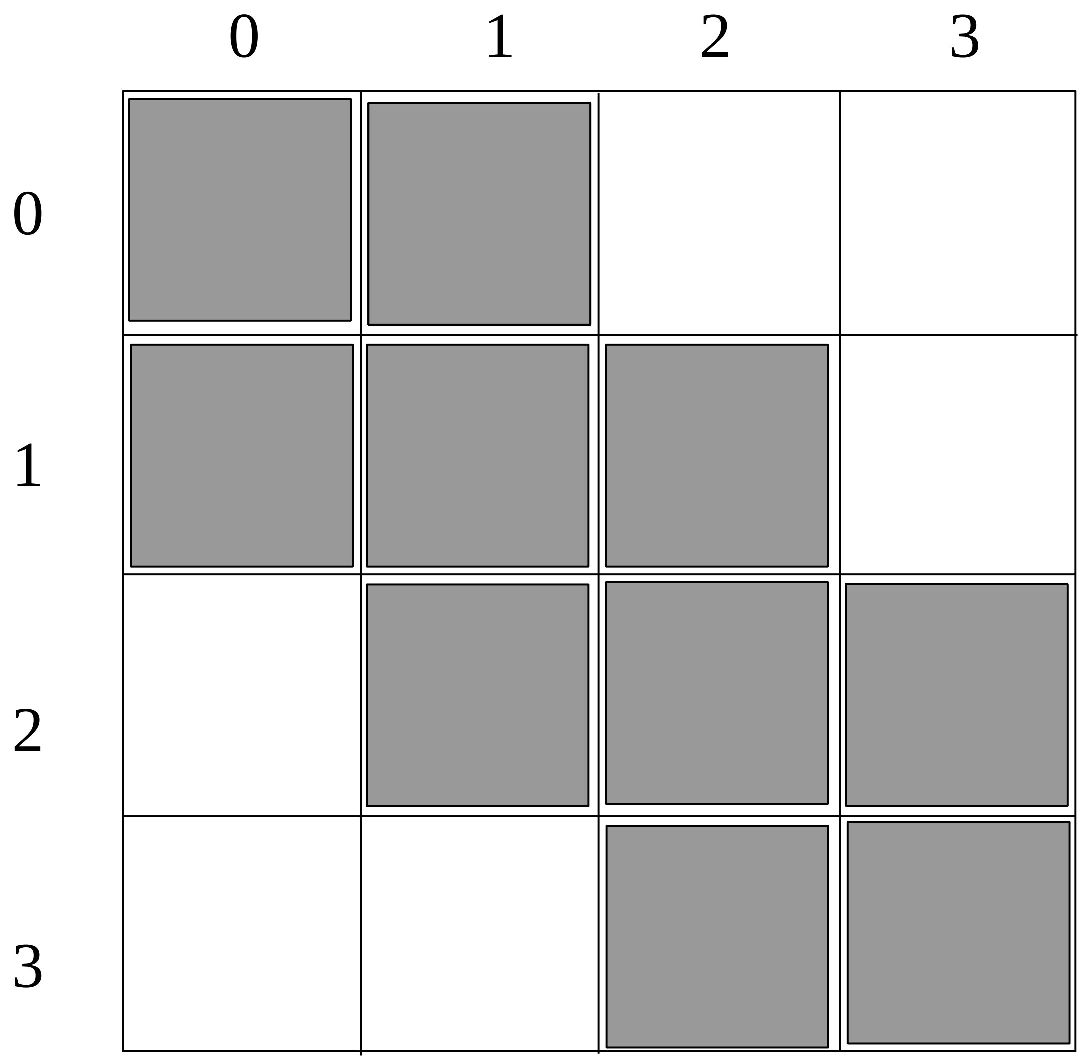
Connectivity Matrices

Example 1



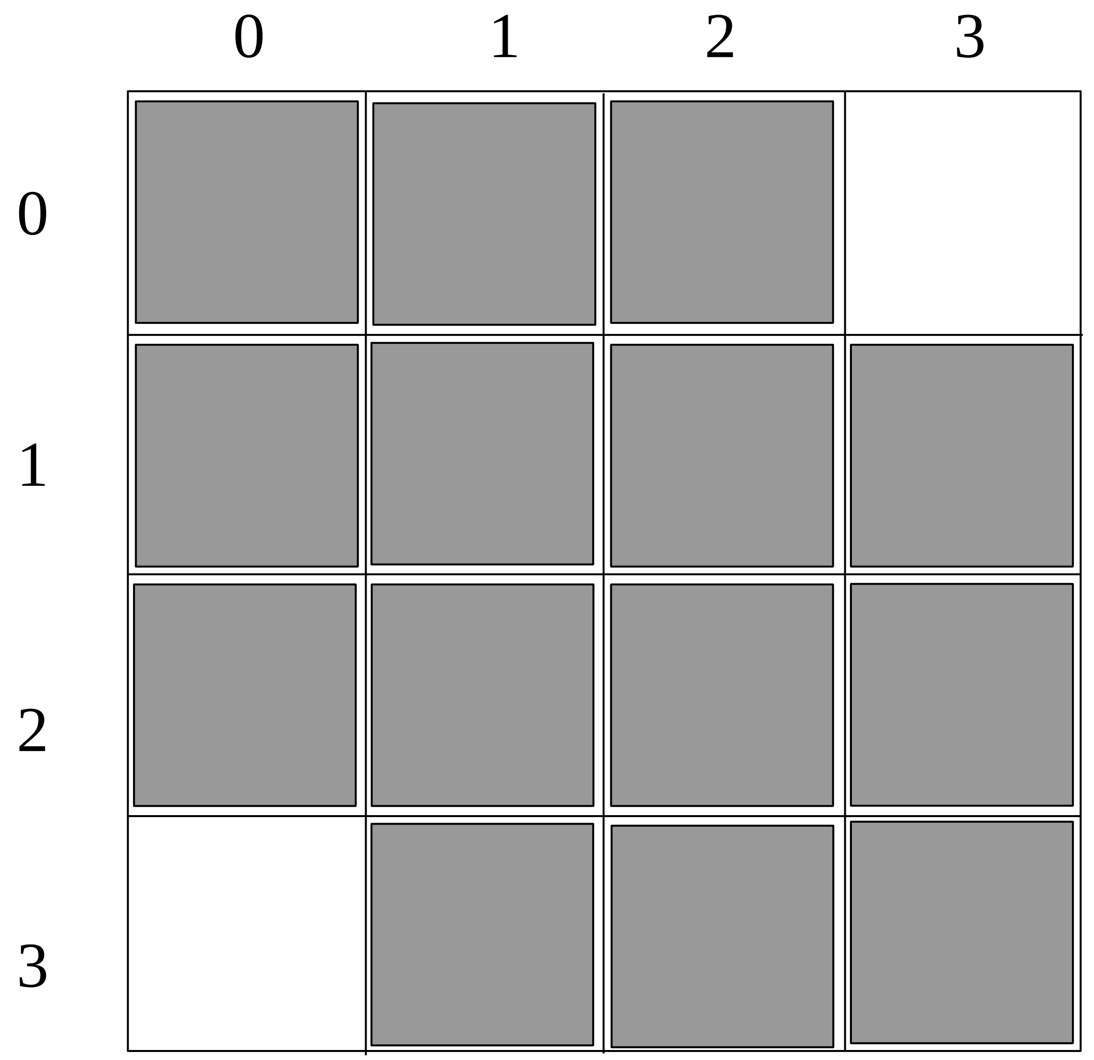
Connectivity Matrices

Example 2



2-step Connectivity Matrices

Example 2



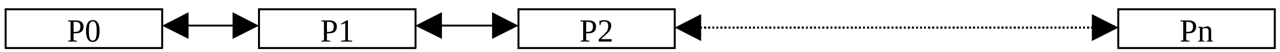
N-step Connectivity Matrices

- Indicates the processor pairs that can reach each other in N steps
- Computed using Boolean matrix multiplication
- The corresponding adjacency matrix indicates how many disjoint paths connect each pair.

	0	1	2	3
0	1	1	1	
1	1	1	1	1
2	1	1	1	1
3		1	1	1

	0	1	2	3
0	1	1	2	1
1	1	1	1	2
2	2	1	1	1
3	1	2	1	1

Linear Array

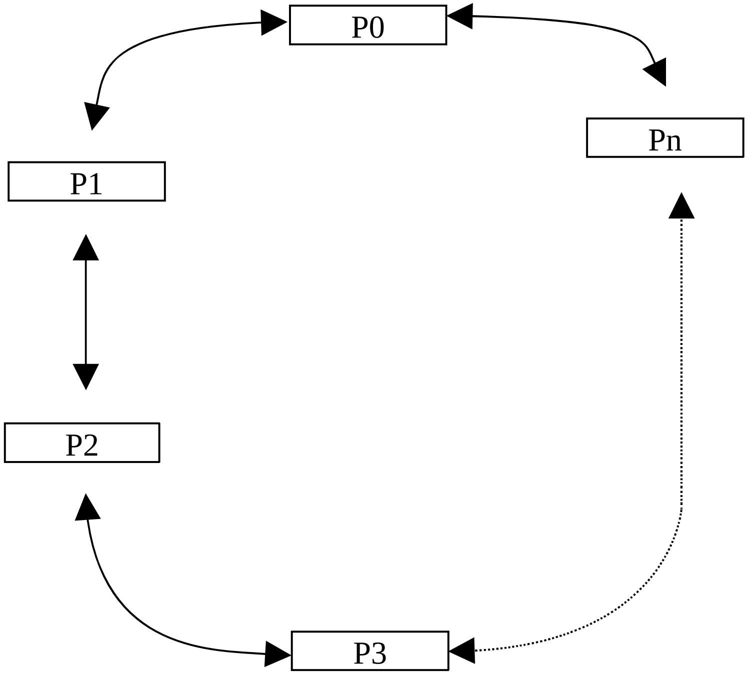


Diameter

Bisection Width

Degree

Ring



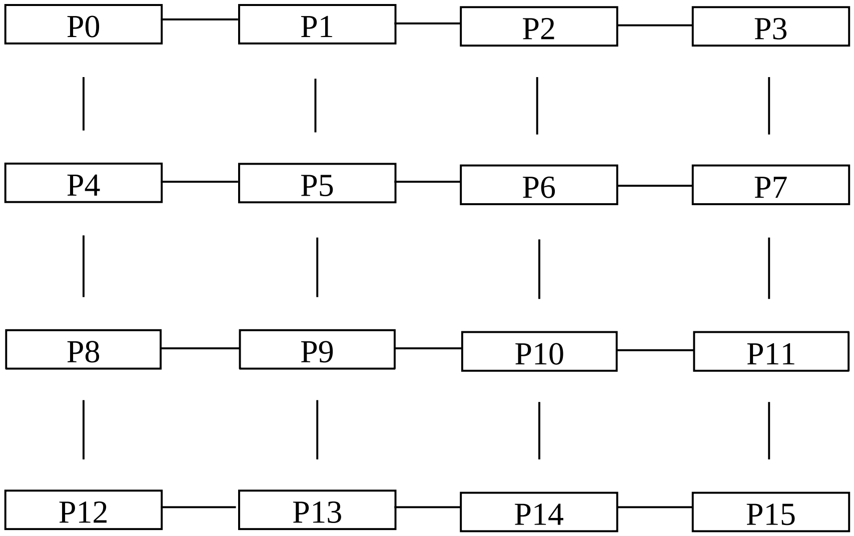
?

Diameter

Bisection Width

Degree

Mesh



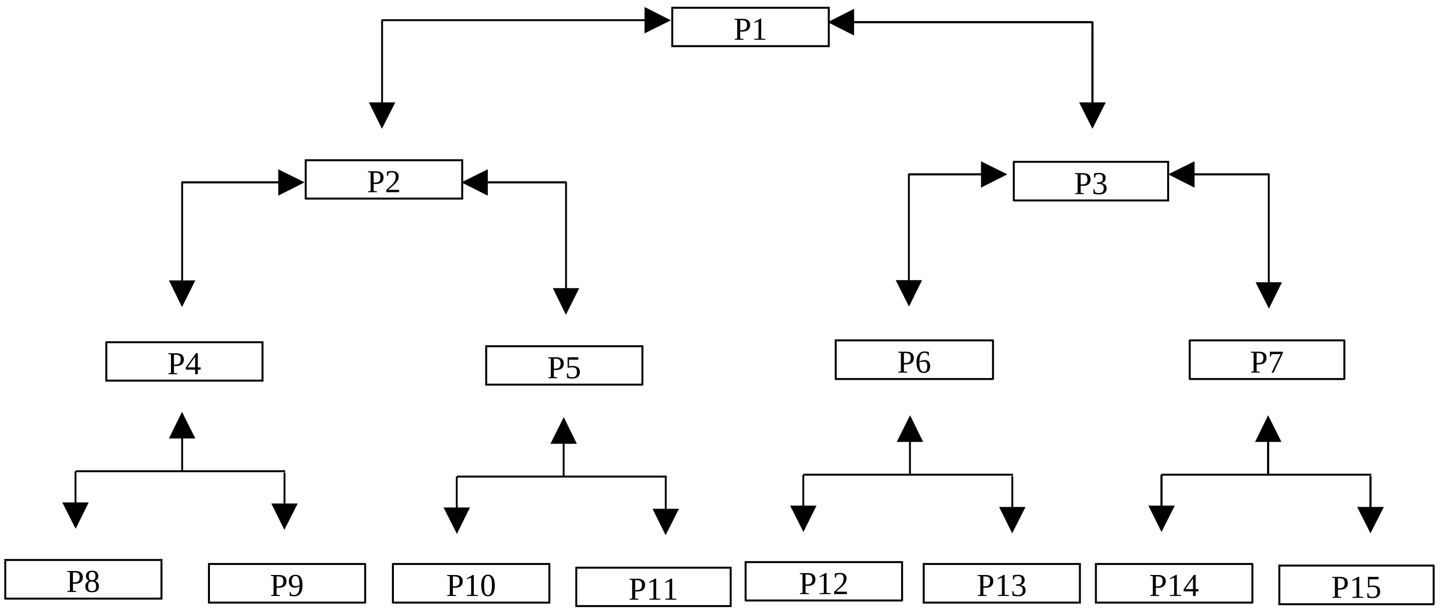
?

Diameter

Bisection Width

Degree

Tree



?

Diameter

Bisection Width

Degree

Other Schemes

- Pyramid: A 4-ary tree where each level is connected as a mesh
- Hypercube: Two processors are connected if and only if their ID #'s differ in exactly one bit.
  - Low communications diameter
  - High bisection width
  - Doesn't have constant degree
- Perfect Shuffle: Processor *i* is connected *one-way* to processor *2i mod(N-1)*.
- Others: Star, De Bruijn, Delta, Omega, Butterfly