Advanced Computer Architecture (0630561)

Lecture 8

Parallel Computer Models

Prof. Kasim M. Al-Aubidy

Computer Eng. Dept.

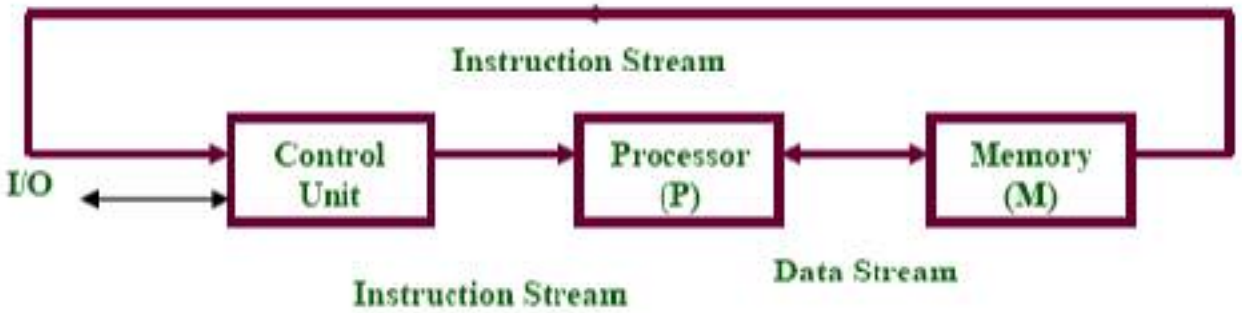
Classifications of Computer Architecture:

Flynn's classified architectures in terms of streams of data and instructions;
Stream of Instructions (SI): sequence of instructions executed by the computer.
Stream of Data (SD): sequence of data including input, temporary or partial results referenced by instructions.

Computer architectures are characterized by the multiplicity of hardware to serve instruction and data streams.

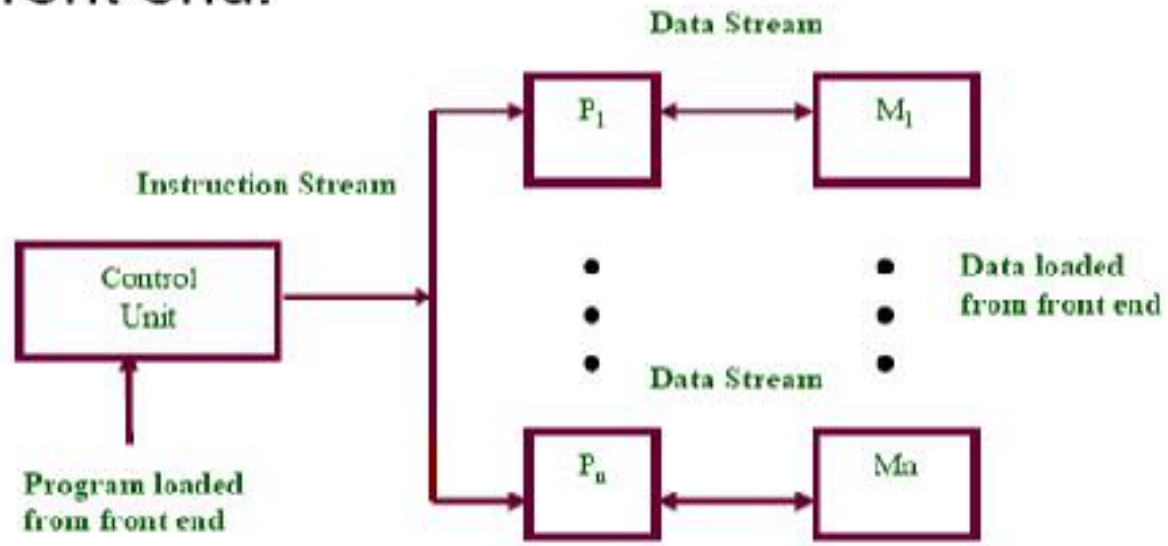
1. Single Instruction Single Data (SISD)
2. Single Instruction Multiple Data (SIMD)
3. Multiple Instruction Multiple Data (MIMD)
4. Multiple Instruction Single Data (MISD)

- Single Instruction Single Data (SISD)
- Conventional single-processor von Neumann computers are classified as SISD systems.



Single Instruction Multiple Data (SIMD)

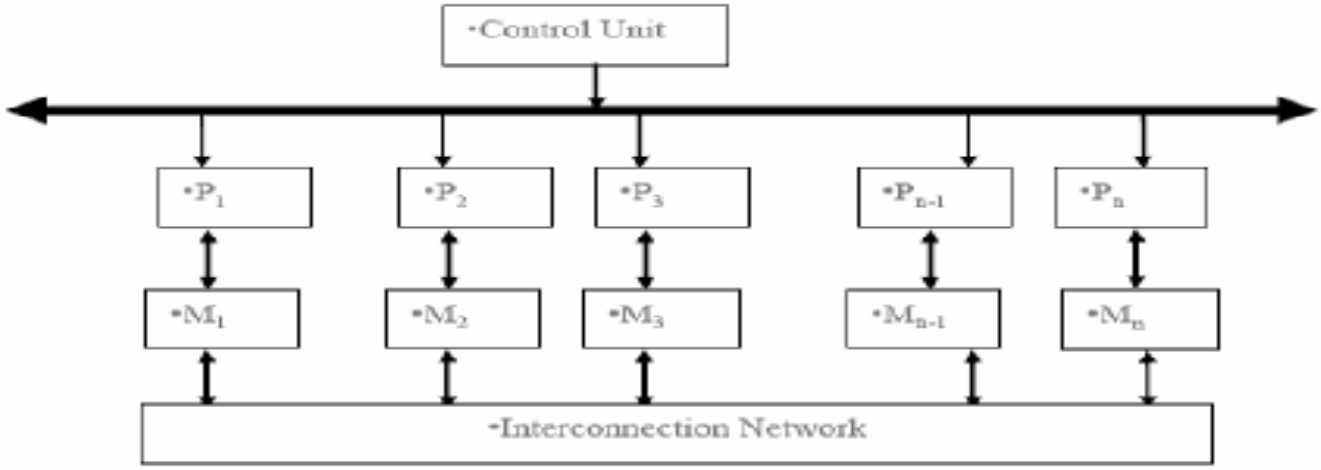
- Consists of 2 parts:
  - a front-end Von Neumann computer.
  - A processor array: connected to the memory bus of the front end.



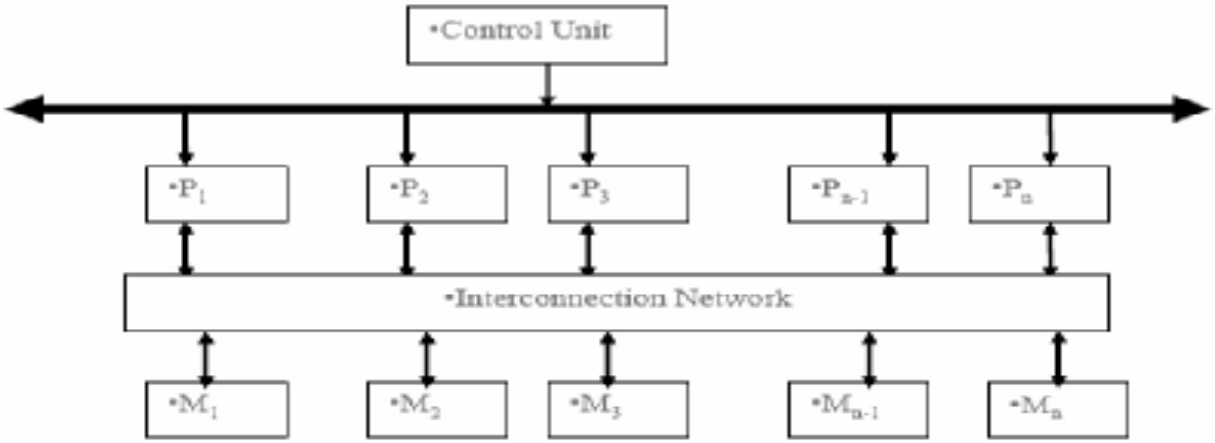
- Vector computers are equipped with scalar and vector hardware or appear as SIMD machines.

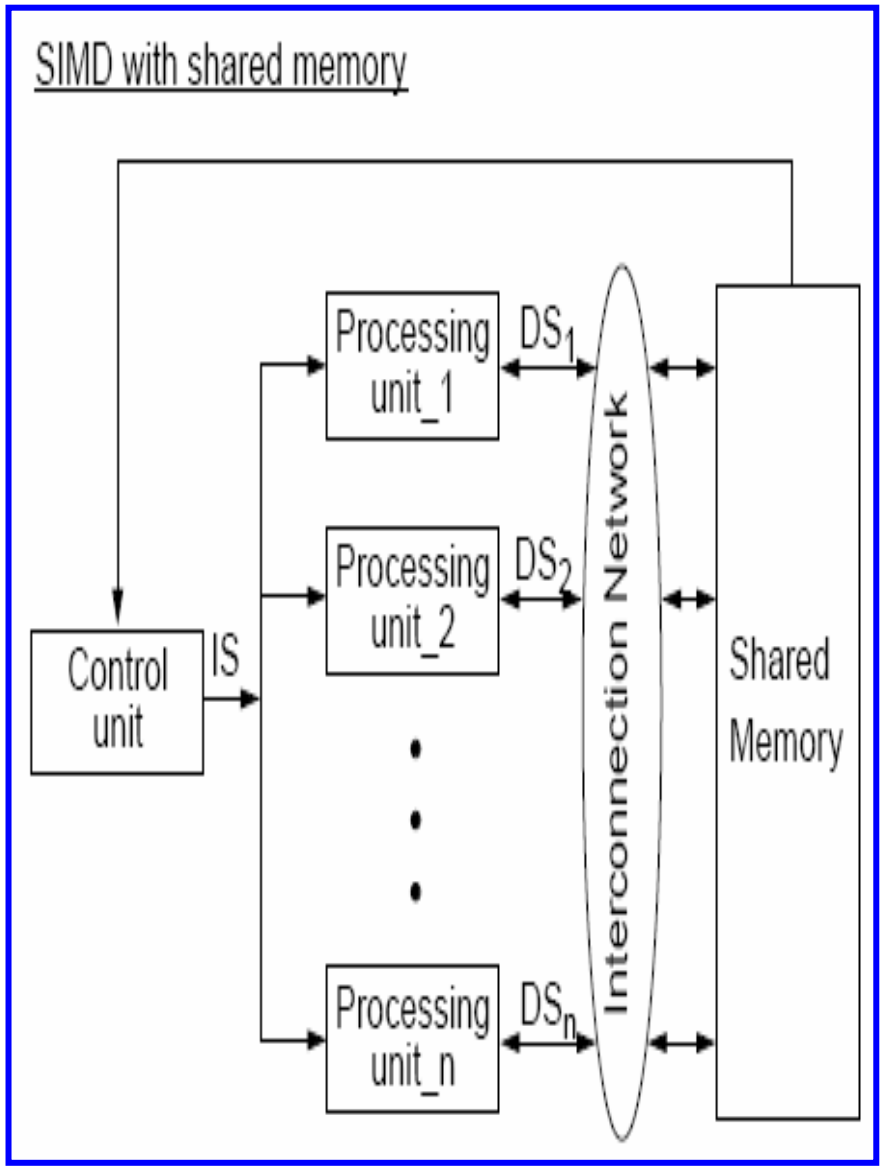
SIMD Architecture

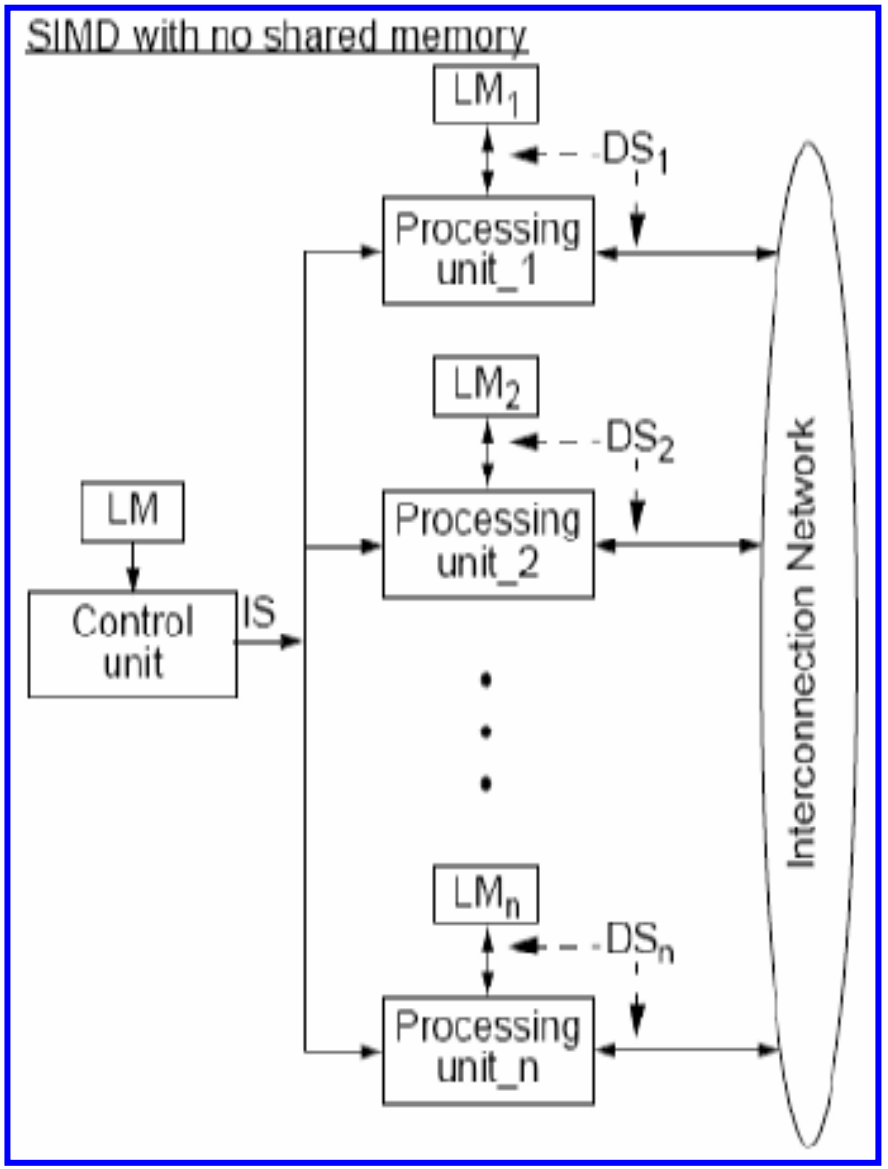
Scheme1 Each processor has its own local memory.



Scheme2 Processors and memory modules communicate with each other via interconnection network.

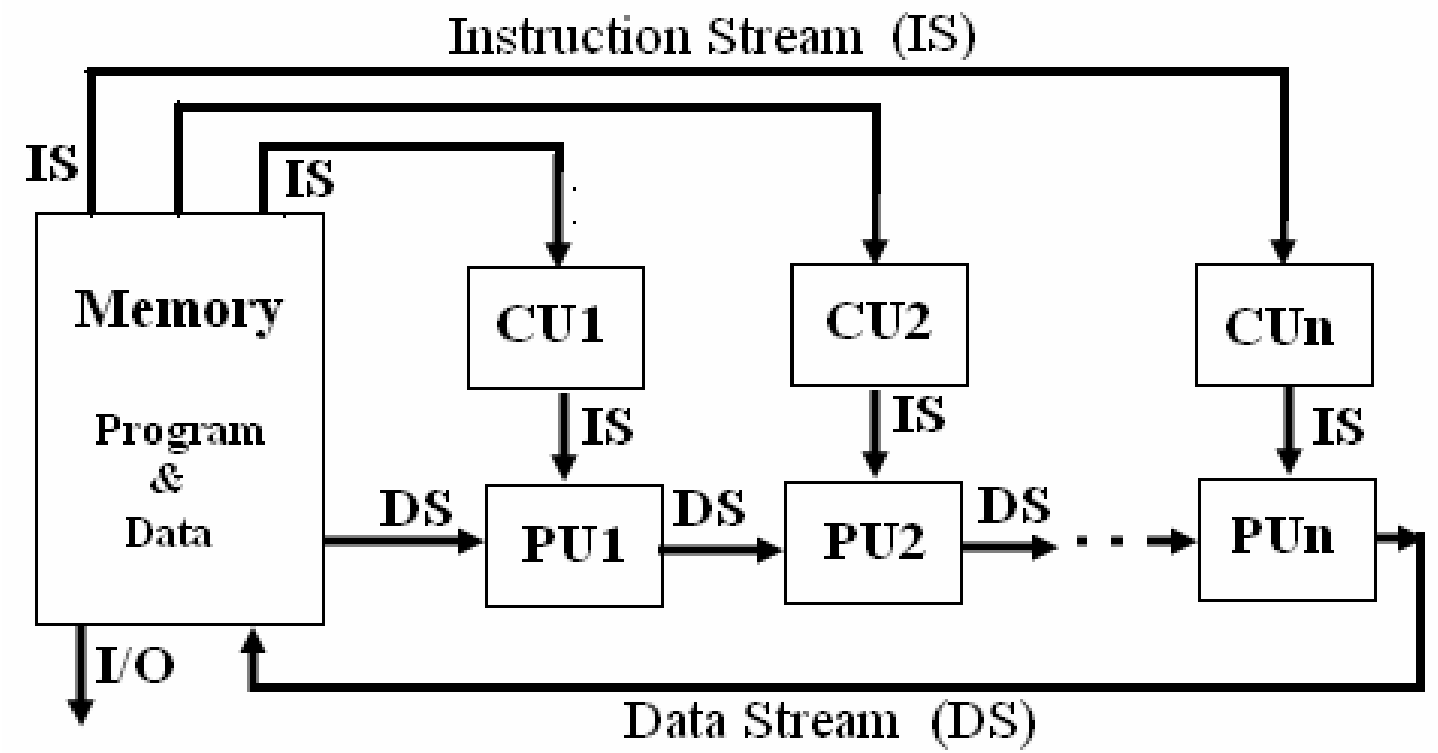



SIMD with shared memory
Control unit
IS
Processing unit_1
Processing unit_2
Processing unit_n
DS1
DS2
DSn
Interconnection Network
Shared Memory


SIMD with no shared memory
LM1
LM2
LMn
LM
Control unit
IS
Processing unit_1
Processing unit_2
Processing unit_n
DS1
DS2
DSn
Interconnection Network

MISD Architecture:

- Many functional units perform different operations on the same data.
- The same data stream flows through a linear array of processors executing different instruction streams.
- This architecture is also known as **SYSTOLIC ARRAYS** for pipelined execution of specific algorithms.

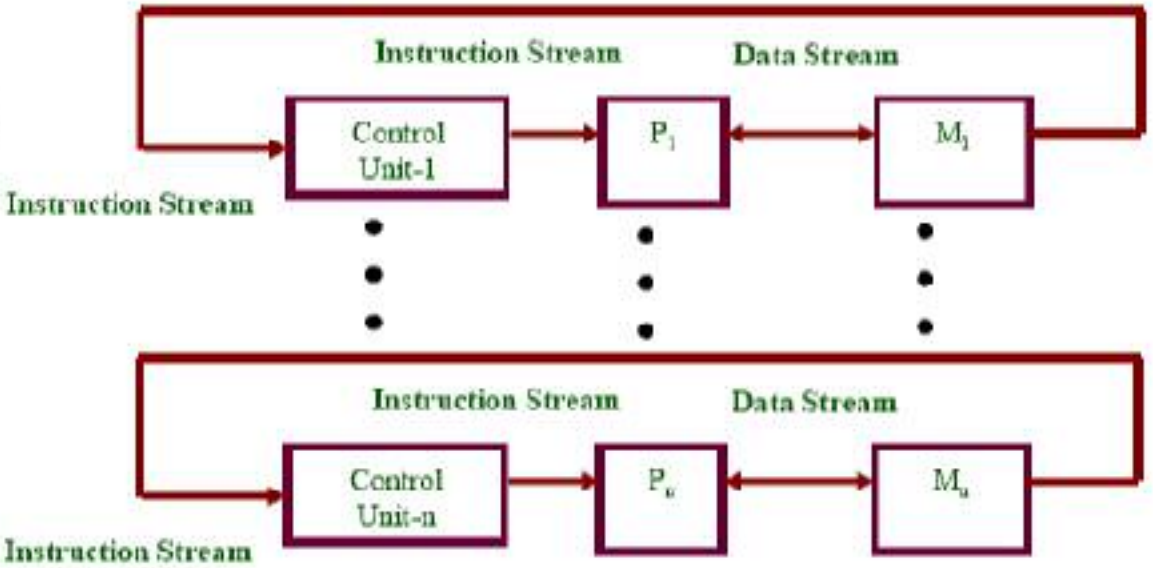


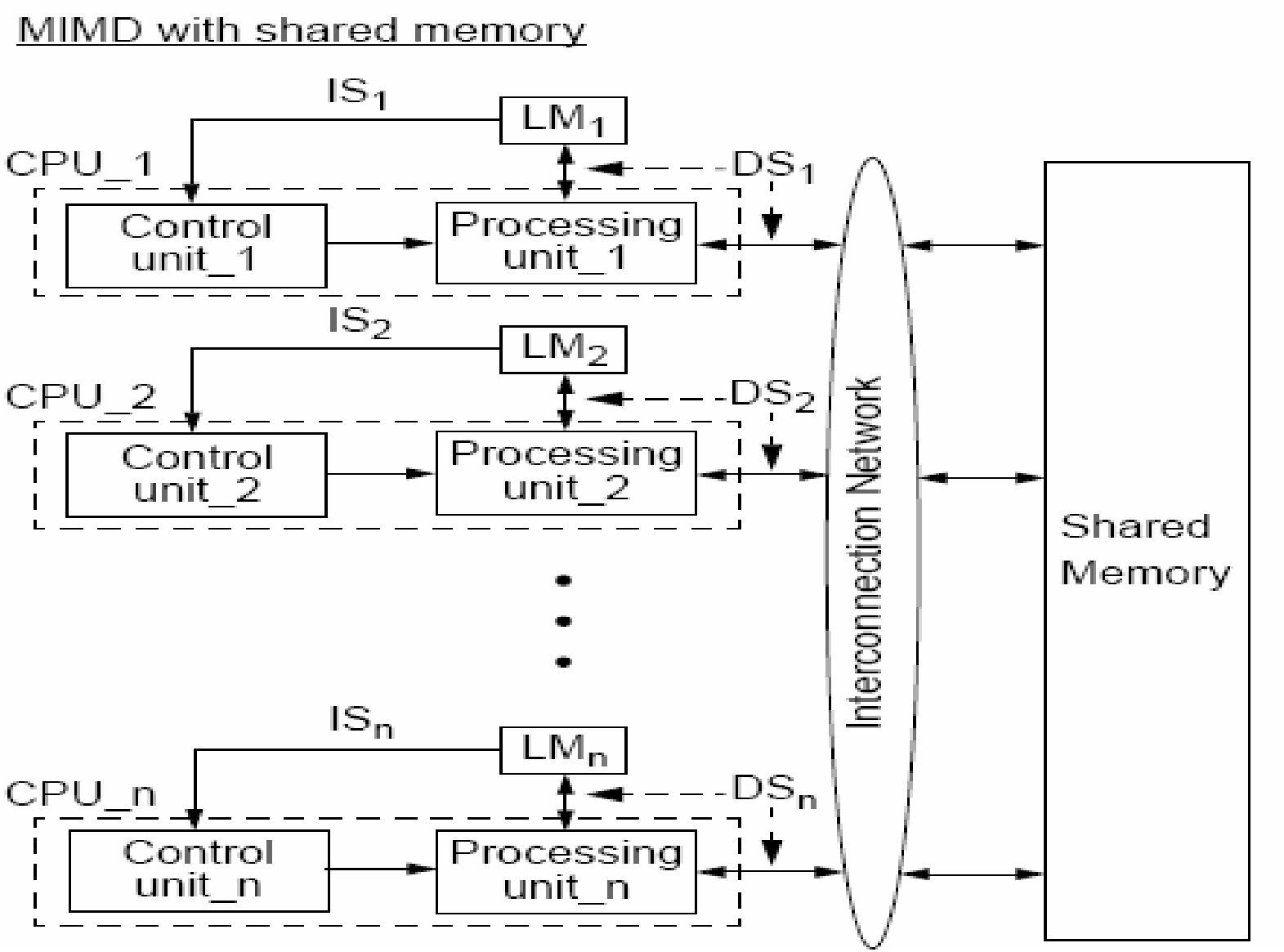
MIMD Architecture

– Made of multiple processors and multiple memory modules connected together via some interconnection network.

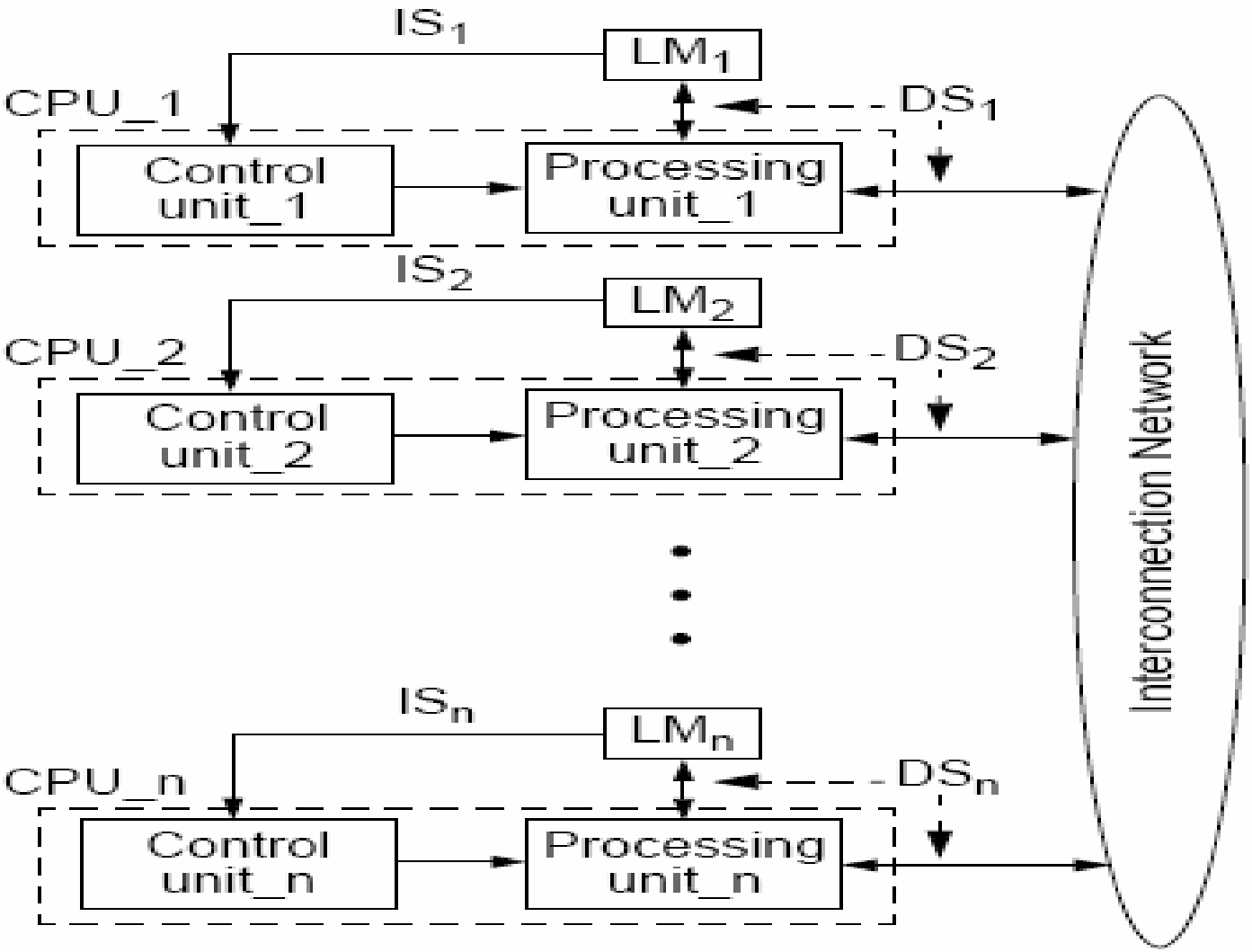
– 2 broad categories:

- Shared memory
- Message passing



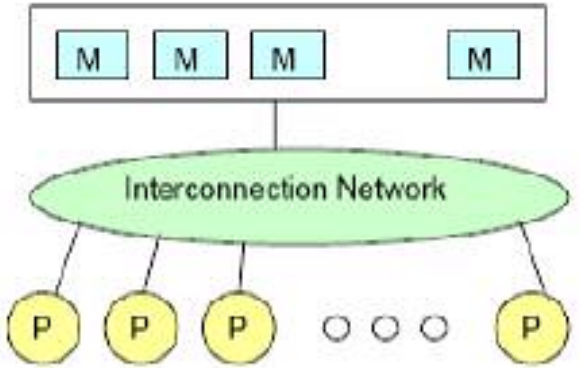

MIMD with shared memory
IS1
LM1
CPU_1
DS1
Control unit_1
Processing unit_1
IS2
LM2
CPU_2
DS2
Control unit_2
Processing unit_2
ISn
LMn
CPU_n
DSn
Control unit_n
Processing unit_n
Interconnection Network
Shared Memory

MIMD with no shared memory



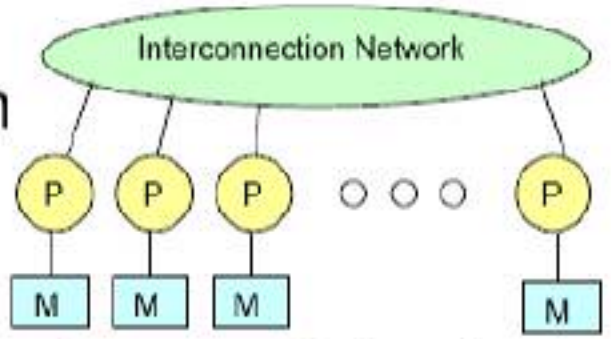
MIMD Architecture

- Shared Memory Organization
  - Inter-processor coordination is accomplished by reading and writing in a global memory shared by processors.



  - Typically consists of servers that communicate through a bus and cache memory controller.

- Message Passing Organization
  - Each processor has access to its own local memory.
  - Communications are performed via send and receive operations.



  - Message passing multiprocessors employ a variety of static networks in local communications.

Shared Memory Organization

A shared memory model is one in which processors communicate by reading and writing locations in a shared memory that is equally accessible by all processors. Each processor may have registers, buffers, caches, and local memory banks as additional memory resources. A number of basic issues in the design of shared memory systems have to be taken into consideration. These include access control, synchronization, protection, and security.

Message Passing Organization

Message passing systems are a class of multiprocessors in which each processor has access to its own local memory. Unlike shared memory systems, communications in message passing systems are performed via send and receive operations. A *node* in such a system consists of a processor and its local memory. Nodes are typically able to store messages in buffers (temporary memory locations where messages wait until they can be sent or received), and perform send/receive operations at the same time as processing.

INTERCONNECTION NETWORKS

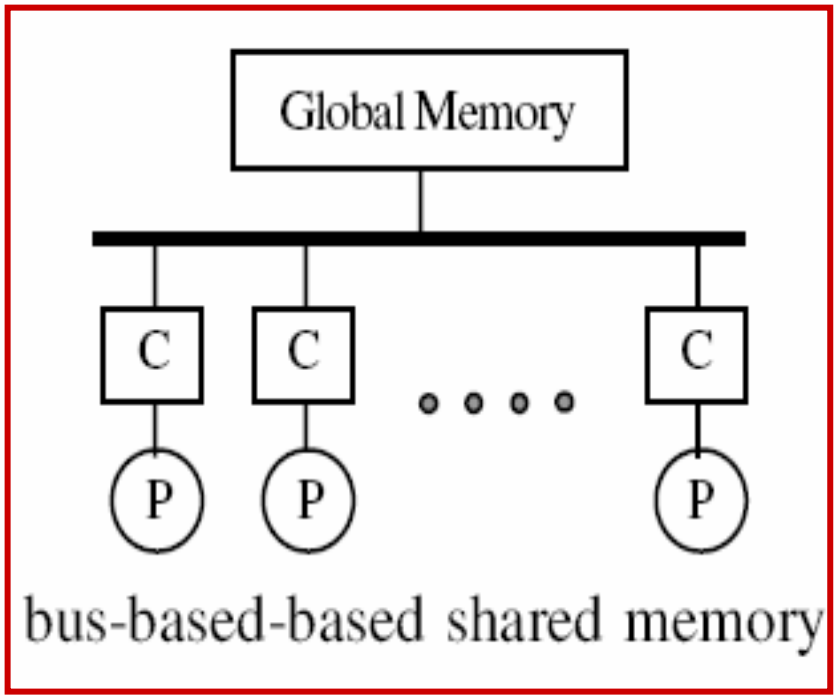
Multiprocessors interconnection networks (INs) can be classified based on a number of criteria. These include

(1) mode of operation (synchronous versus asynchronous),

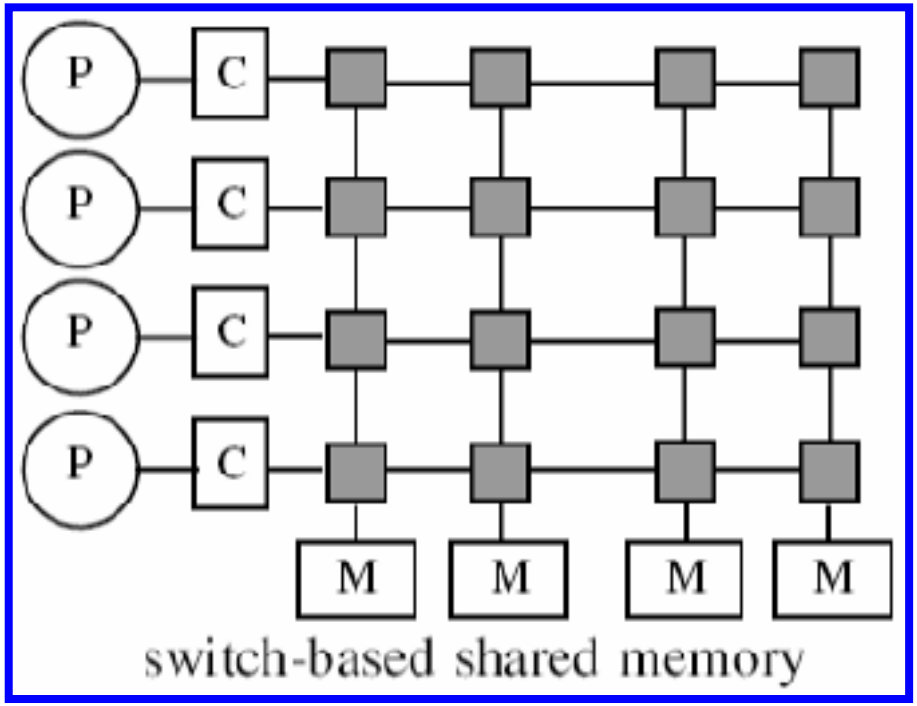
(2) control strategy (centralized versus decentralized),

(3) switching techniques (circuit versus packet), and

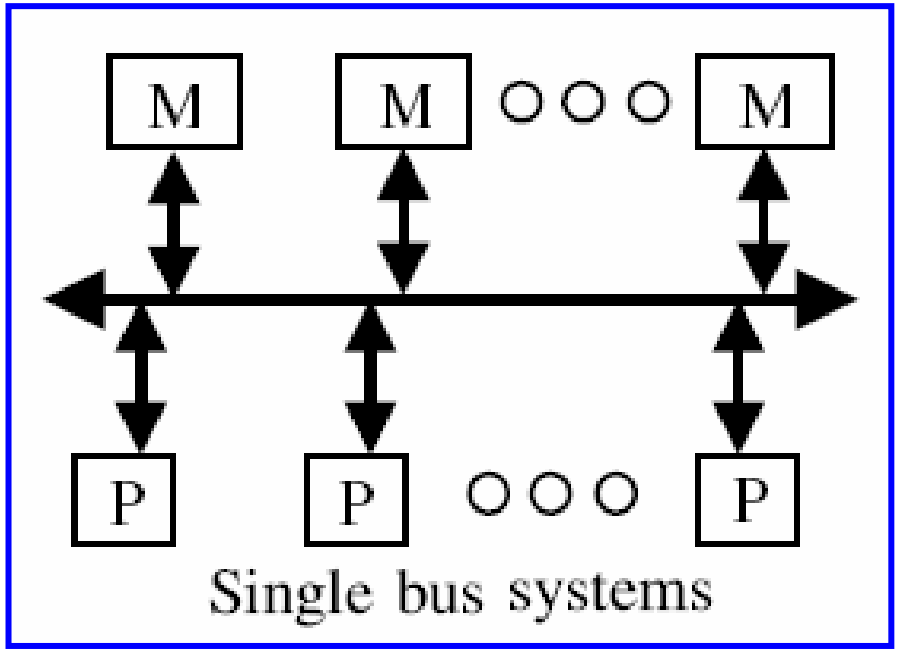
(4) topology (static versus dynamic).

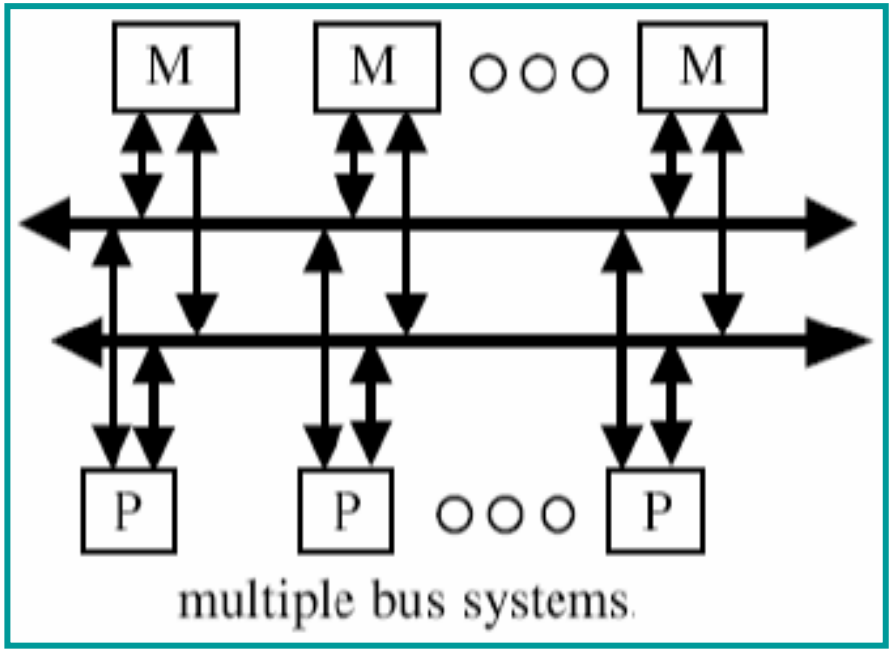


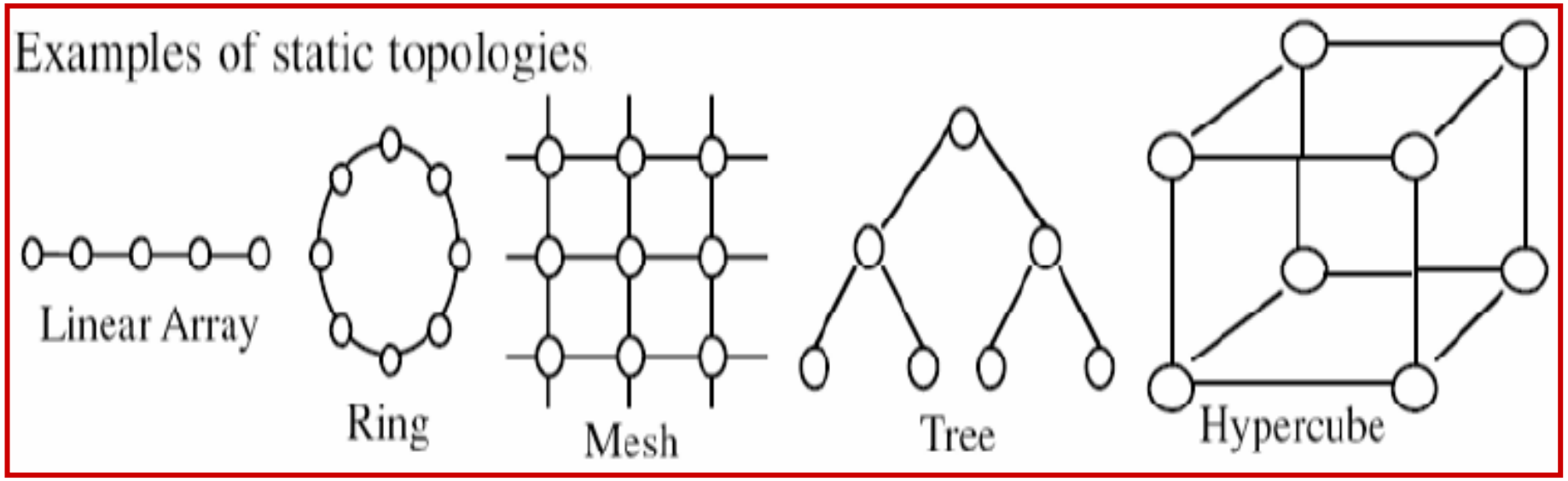
bus-based-based shared memory



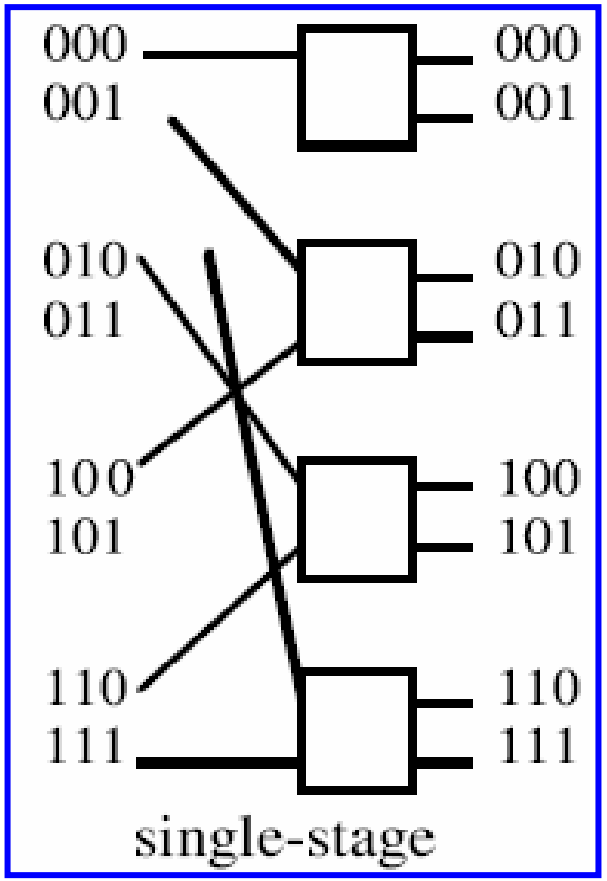
switch-based shared memory

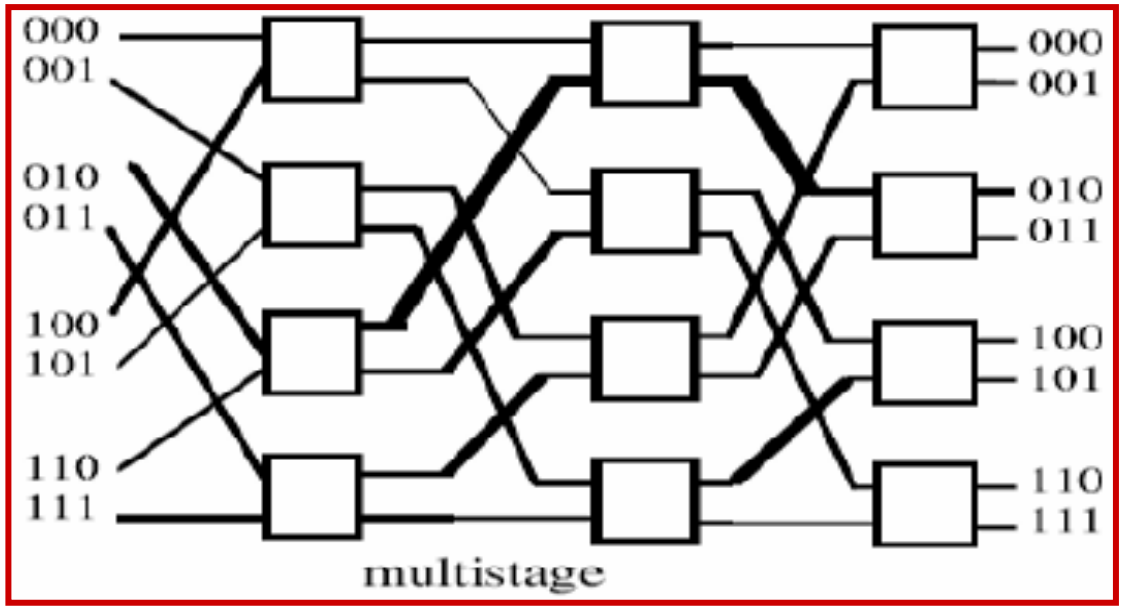

M
M
M
P
P
P
Single bus systems

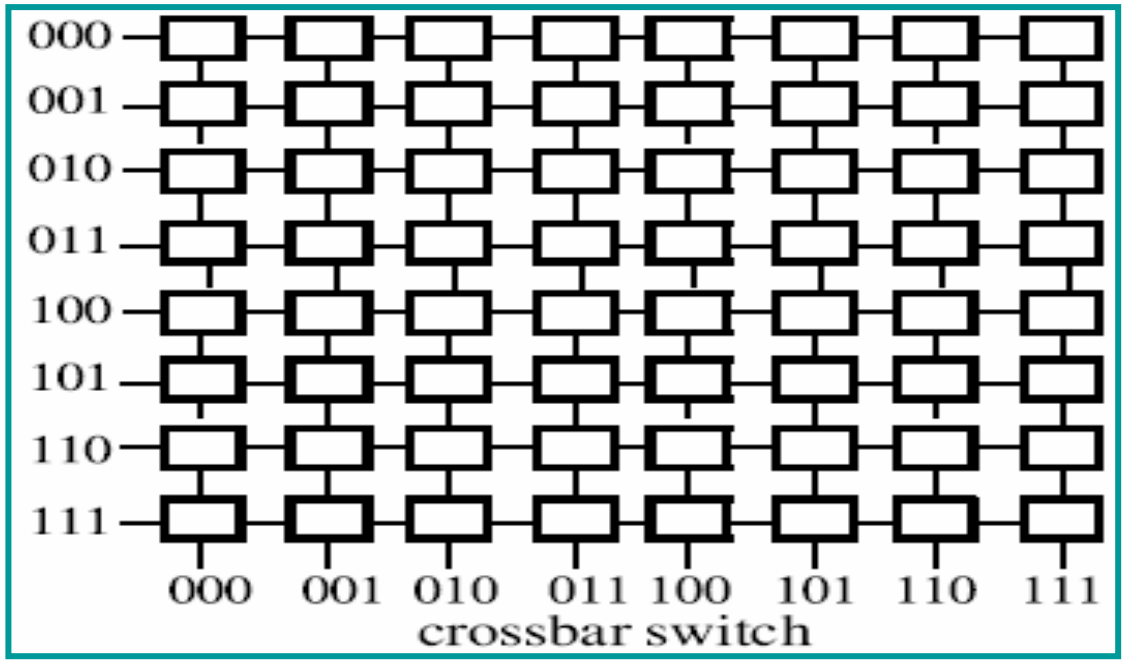

M
M
M
P
P
P
multiple bus systems


Examples of static topologies
Linear Array
Ring
Mesh
Tree
Hypercube

Dynamic INs:


000
001
010
011
100
101
110
111
000
001
010
011
100
101
110
111
single-stage


000
001
010
011
100
101
110
111
000
001
010
011
100
101
110
111
multistage


000
001
010
011
100
101
110
111
000 001 010 011 100 101 110 111
crossbar switch

Performance Comparison of Some Dynamic INs

Network	Delay	Cost (Complexity)
Bus	$O(N)$	$O(1)$
Multiple-bus	$O(mN)$	$O(m)$
MINs	$O(\log N)$	$O(N \log N)$

In this table, m represents the number of multiple buses used, N represents the number of processors (memory modules) or input/output of the network.

Performance Characteristics of Static INs

Network	Degree	Diameter	Cost (#links)
Linear array	2	$N - 1$	$N - 1$
Binary tree	3	$2(\lceil \log_2 N \rceil - 1)$	$N - 1$
n-cube	$\log_2 N$	$\log_2 N$	$nN/2$
2D-mesh	4	$2(n - 1)$	$2(N - n)$

In this table, the degree of a network is defined as the maximum number of links (channels) connected to any node in the network. The diameter of a network is defined as the maximum path, p, of the shortest paths between any two nodes. Degree of a node, d, is defined as the number of channels incident on the node.

Summary:

- Of the four machine models, most parallel computers built in the past assumed the MIMD model for general purpose computations. For this reason, MIMD is the most popular model, SIMD next, and MISD the least popular model being applied in commercial machines.
- The SIMD and MISD models are more suitable for special-purpose computations.