ARCHITECTURAL CLASSIFICATION

Mariam A. Salih

- Basic types of architectural classification
  - ***FLYNN'S TAXONOMY OF COMPUTER ARCHITECTURE***
  - **FENG'S CLASSIFICATION**
  - **Handler Classification**
- Other types of architectural classification
  - **Classification based on coupling between processing elements**
  - **Classification based on mode of accessing memory**

ARCHITECTURAL CLASSIFICATION

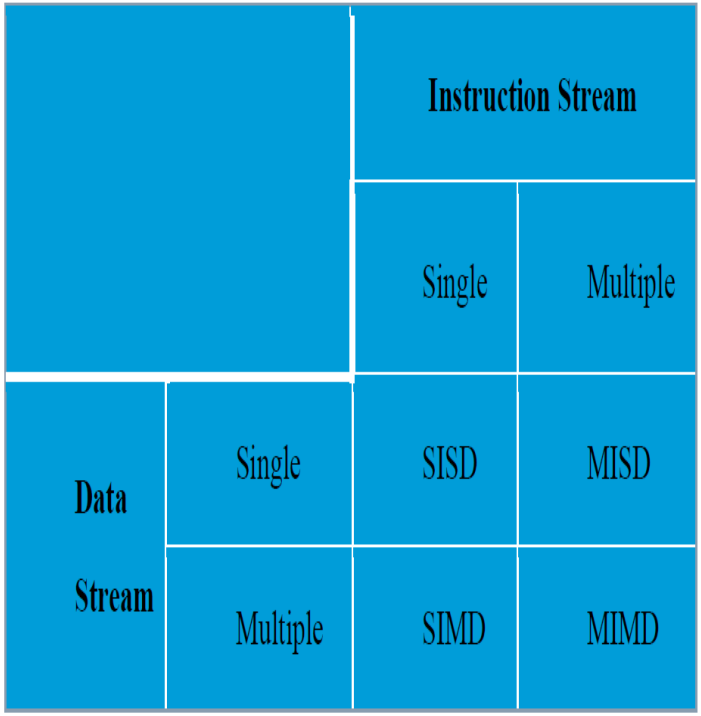
- **Flynn classification**: (1966) is based on multiplicity of instruction streams and the data streams in computer systems.
- **Feng's classification**: (1972) is based on serial versus parallel processing.
- **Handler's classification**: (1977) is determined by the degree of parallelism and pipelining in various subsystem levels.

FLYNN'S TAXONOMY OF COMPUTER ARCHITECTURE…

- The most popular taxonomy of computer architecture was defined by Flynn in 1966.
- Flynn's classification scheme is based on the notion of a stream of information. Two types of information flow into a processor: instructions and data.
- ***The instruction stream*** is defined as the sequence of instructions performed by the processing unit.
- The ***data stream*** is defined as the data traffic exchanged between the memory and the processing unit.

Types of ***FLYNN'S TAXONOMY***

- to Flynn's classification, either of the instruction or data streams can be single or multiple. Computer architecture can be classified into the following four distinct categories:
  - single-instruction single-data streams (**SISD**);
  - single-instruction multiple-data streams (**SIMD**);
  - multiple-instruction single-data streams (**MISD**); and
  - multiple-instruction multiple-data streams (**MIMD**).



SISD

- Conventional single-processor von Neumann computers are classified as SISD systems.

UNIVAC1

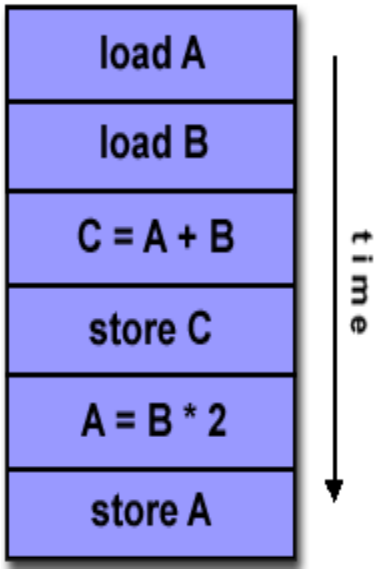
IBM 360

CRAY1

CDC 7600

PDP1

Dell Laptop



SIMD ARCHITECTURE

- The SIMD model of parallel computing consists of two parts: a front-end computer of the usual von Neumann style, and a processor array.
- The ***processor array*** is a set of identical synchronized processing elements capable of simultaneously performing the same operation on different data.

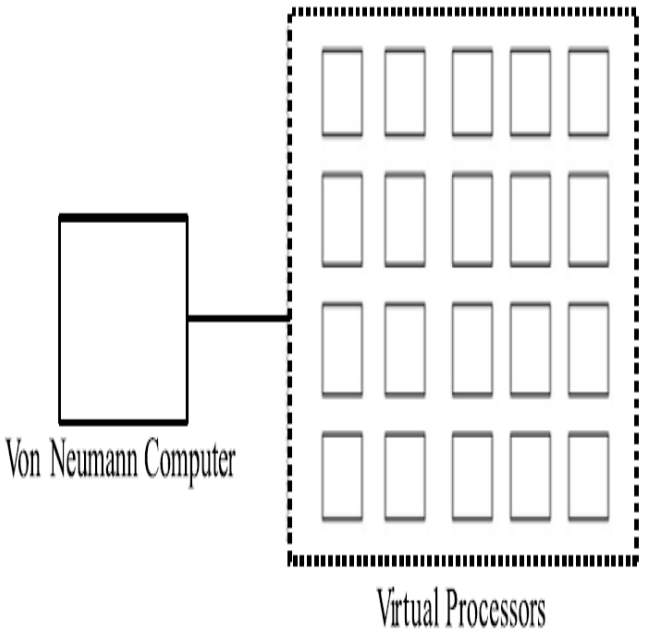


Figure 1.4 SIMD architecture model.

SIMD ARCHITECTURE

- Each processor in the array has a small amount of local memory where the distributed data resides while it is being processed in parallel.
- The processor array is connected to the memory bus of the front end so that the front end can randomly access the local processor memories as if it were another memory.

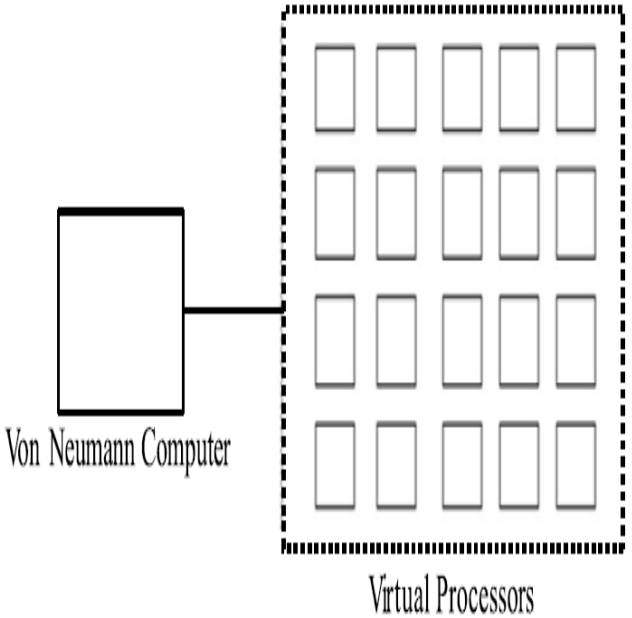


Figure 1.4 SIMD architecture model.

SIMD ARCHITECTURE

- The front end can issue special commands that cause parts of the memory to be operated on simultaneously or cause data to move around in the memory.
- The application program is executed by the front end in the usual serial way, but issues commands to the processor array to carry out SIMD operations in parallel.

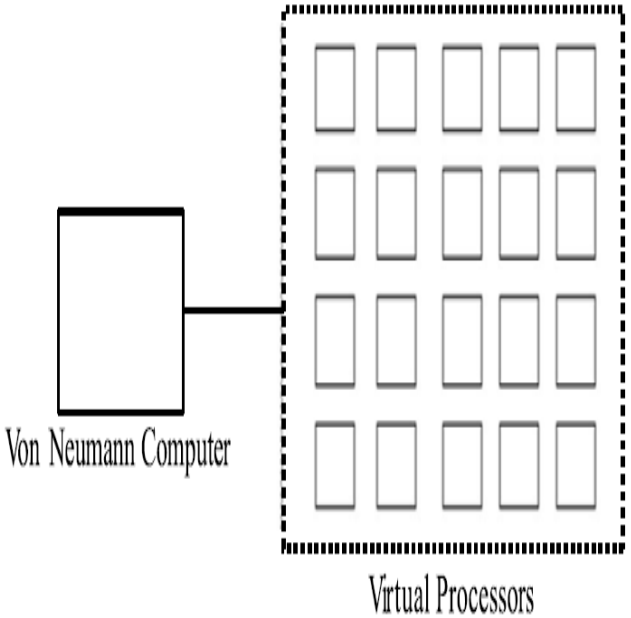
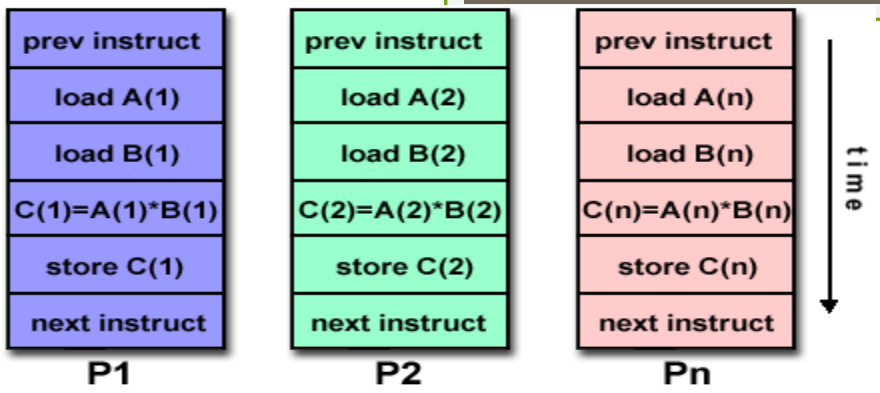
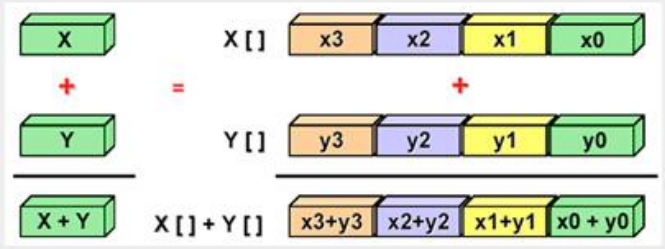


Figure 1.4 SIMD architecture model.



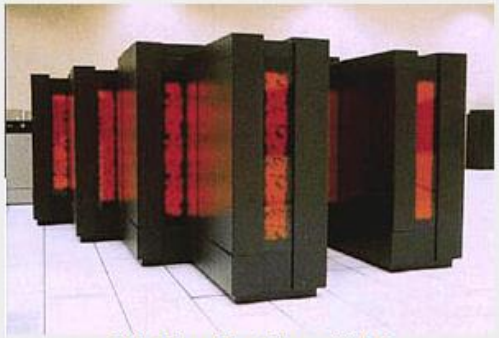
ILLIAC IV

MasPar

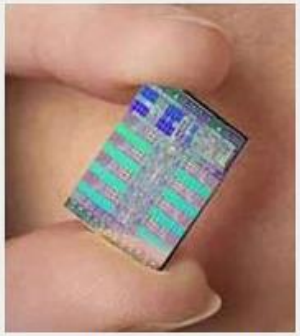


Cray X-MP

Cray Y-MP



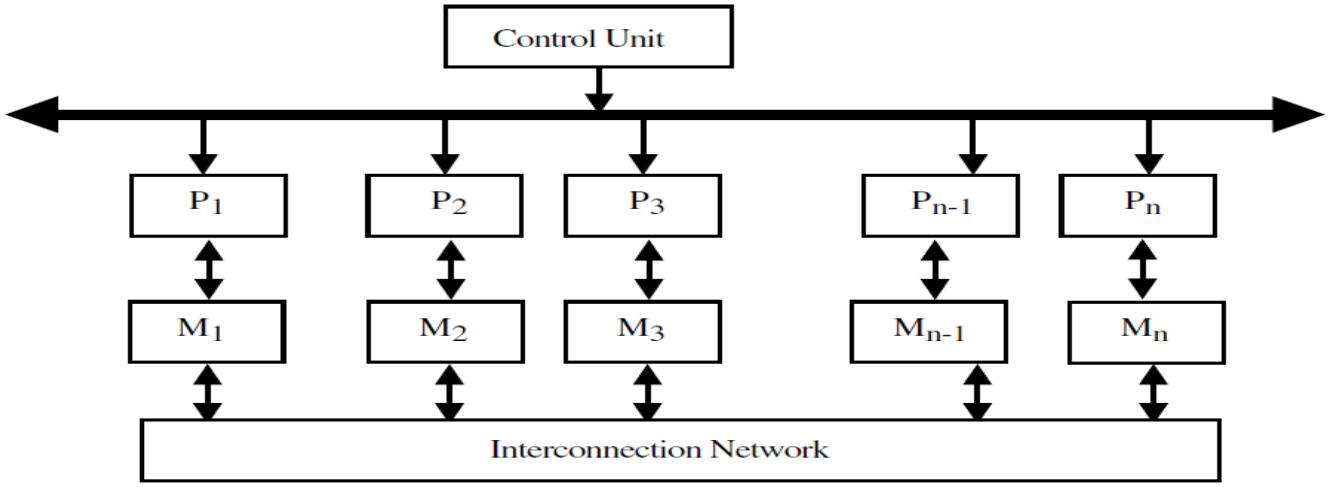
Thinking Machines CM-2



Cell Processor (GPU)

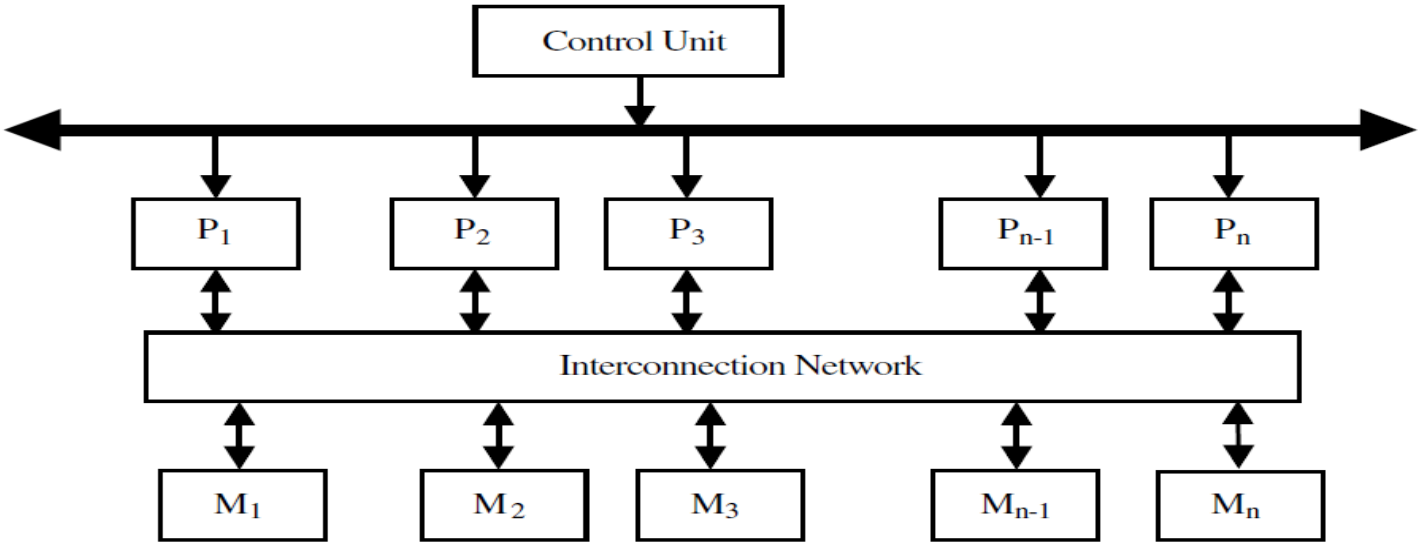
There are two main configurations that have been used in SIMD machines.

In the first scheme, each processor has its own local memory. Processors can communicate with each other through the interconnection network.



There are two main configurations that have been used in SIMD machines.

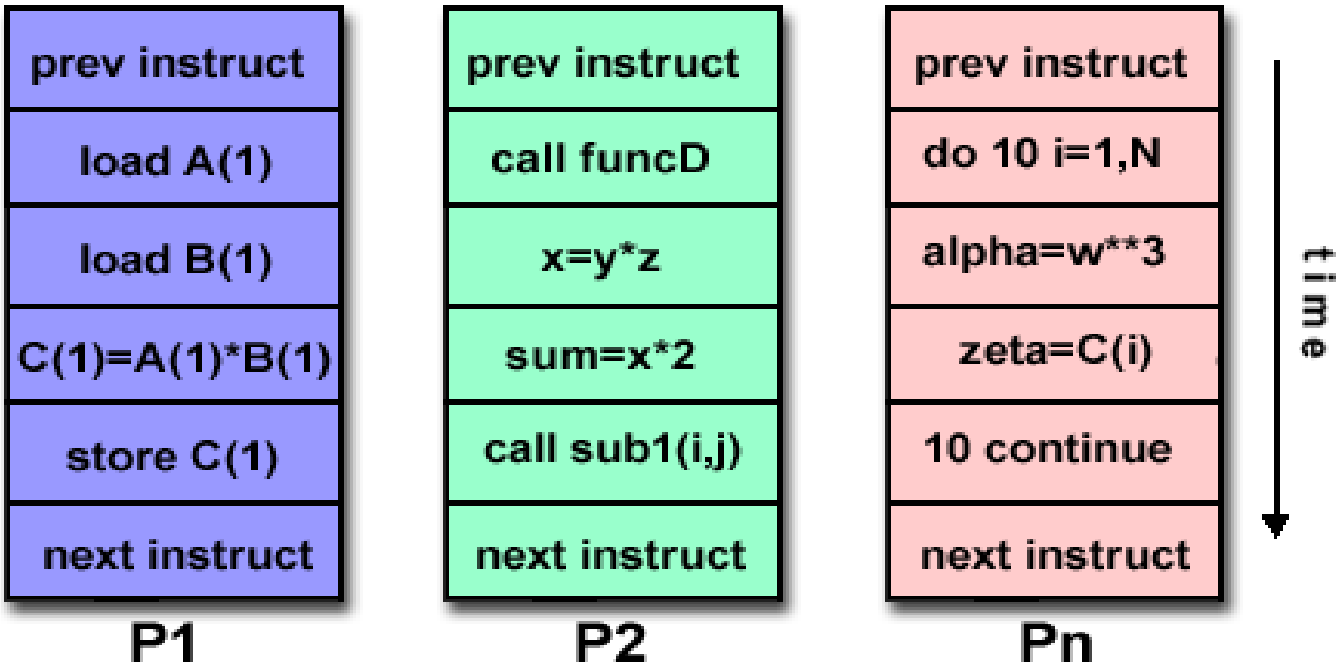
second SIMD scheme, processors and memory modules communicate with each other via the interconnection network.



MIMD ARCHITECTURE

- Multiple-instruction multiple-data streams (MIMD) parallel architectures are made of multiple processors and multiple memory modules connected together via some interconnection network. They fall into two broad categories: **shared memory** or **message passing**.
- ***Processors exchange information through their central shared memory in shared memory systems, and exchange information through their interconnection network in message passing systems.***

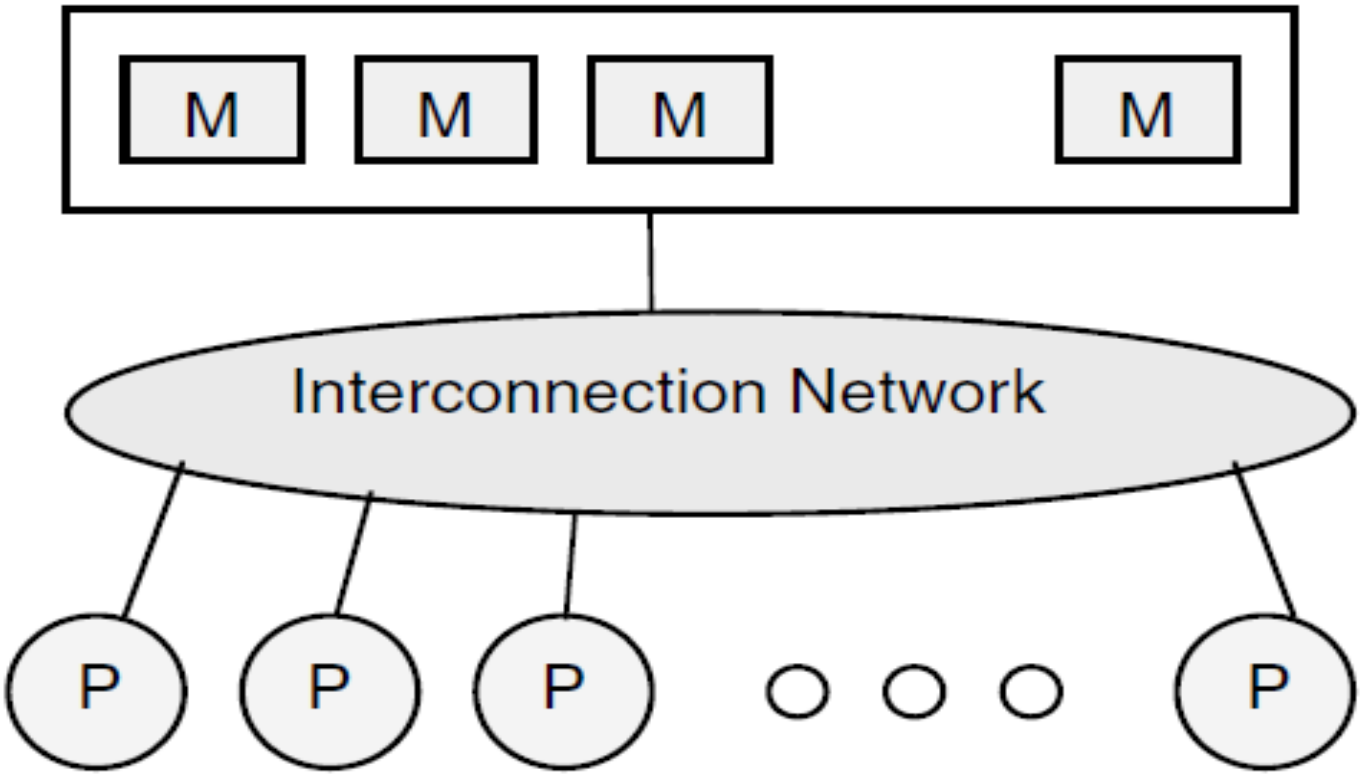
MIMD


prev instruct
load A(1)
load B(1)
C(1)=A(1)*B(1)
store C(1)
next instruct
P1
prev instruct
call funcD
x=y*z
sum=x*2
call sub1(i,j)
next instruct
P2
prev instruct
do 10 i=1,N
alpha=w**3
zeta=C(i)
10 continue
next instruct
Pn
time

MIMD "shared memory system"

- A shared memory system typically accomplishes inter-processor coordination through a global memory shared by all processors.

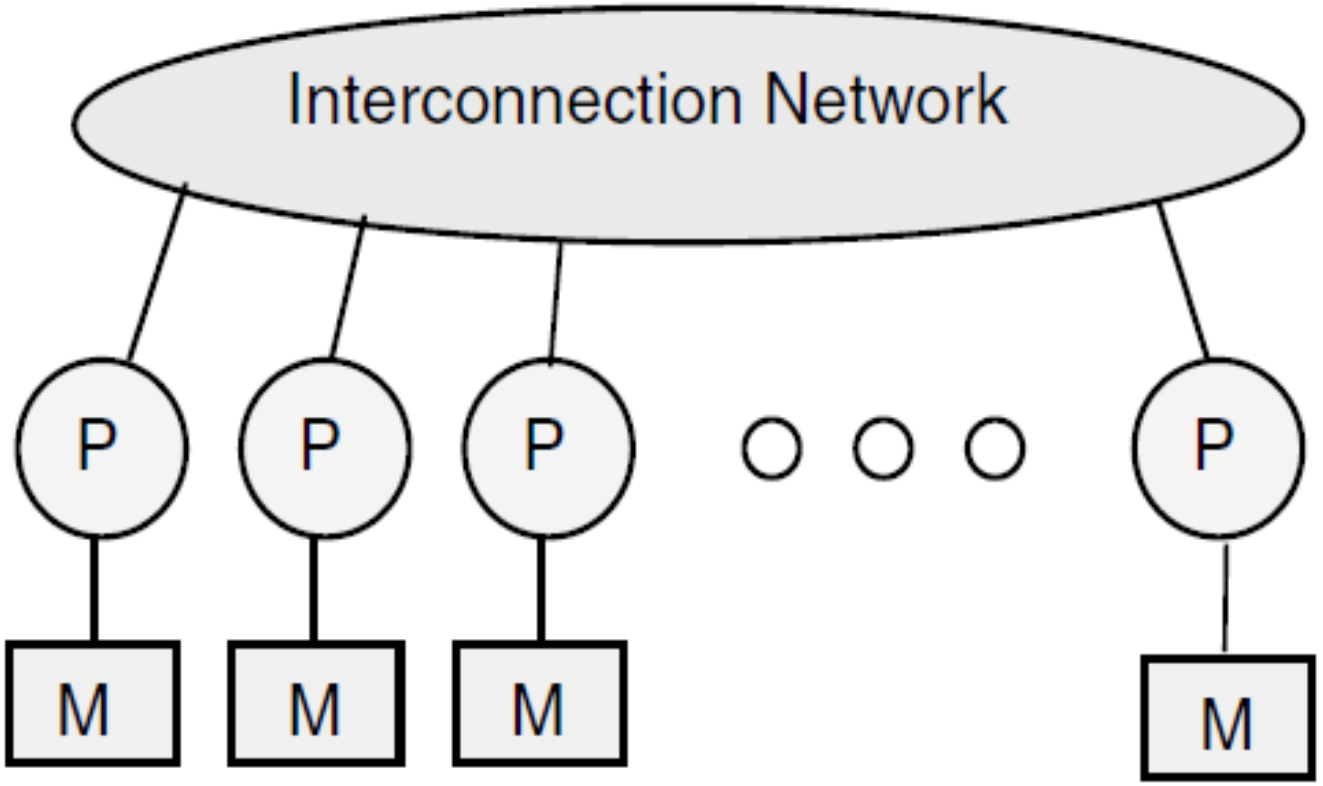
- Because access to shared memory is balanced, these systems are also called SMP (symmetric multiprocessor) systems.



Shared Memory MIMD Architecture

MIMD "message passing system"

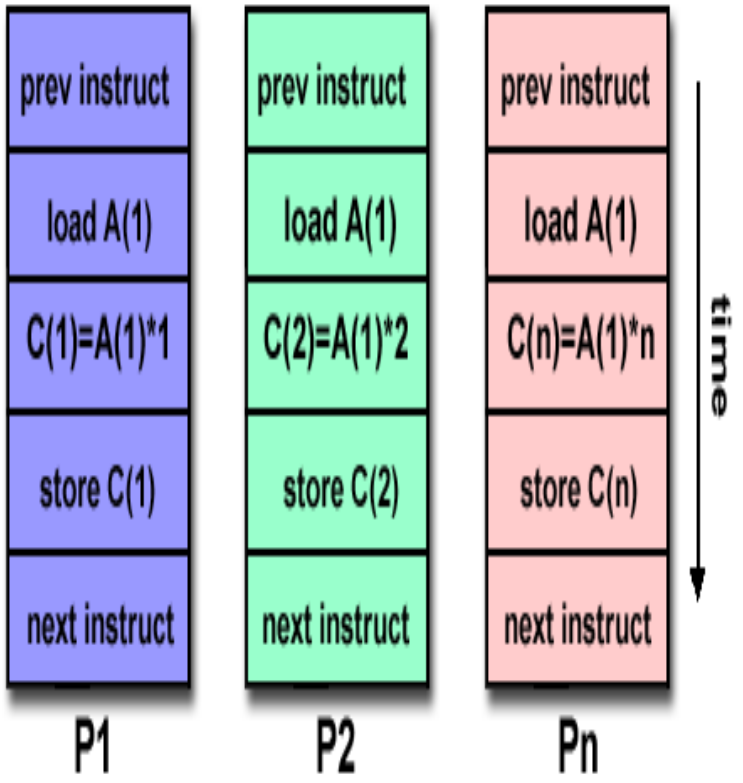
- A message passing system (also referred to as distributed memory) typically combines the local memory and processor at each node of the interconnection network.
- There is no global memory, so it is necessary to move data from one local memory to another by means of message passing.
- This is typically done by a Send/Receive pair of commands, which must be written into the application software by a programmer.



Message Passing MIMD Architecture

MISD ARCHITECTURE

- In the MISD category, the same stream of data flows through a linear array of processors executing different instruction streams.
- In practice, there is no viable MISD machine; however, some authors have considered pipelined machines (and perhaps systolic-array computers) as examples for MISD.



FENG'S CLASSIFICATION

- Tse-yun Feng suggested the use of degree of parallelism to classify various computer architectures.
- The maximum number of binary digits that can be processed within a unit time by a computer system is called the maximum parallelism degree P.
- A bit slice is a string of bits one from each of the words at the same vertical position.
- under above classification
  - **Word Serial and Bit Serial (WSBS)**
  - **Word Parallel and Bit Serial (WPBS)**
  - **Word Serial and Bit Parallel(WSBP)**
  - **Word Parallel and Bit Parallel (WPBP)**

- **WSBS** has been called bit parallel processing because one bit is processed at a time.
- **WPBS** has been called bit slice processing because m-bit slice is processes at a time.
- **WSBP** is found in most existing computers and has been called as Word Slice processing because one word of n bit processed at a time.
- **WPBP** is known as fully parallel processing in which an array on n x m bits is processes at one time.

Mode	Computer Model	Degree of parallelism
WSPS N = 1 M = 1	The 'MINIMA'	(1,1)
WPBS N=1 M>1	STARAN MPP DAP	(1,256) (1,16384) (1,4096)
WSBP n>1 m=1 (Word Slice Processing)	IBM 370/168 UP CDC 6600 Burrough 7700 VAX 11/780	(64,1) (60,1) (48,1) (16/32,1)
WPBP n>1 m>1 (fully parallel Processing)	Illiav IV	(64,64)

Handler Classification

- Wolfgang Handler has proposed a classification scheme for identifying the parallelism degree and pipelining degree built into the hardware structure of a computer system. He considers at three subsystem levels:
  - **Processor Control Unit (PCU)**
  - **Arithmetic Logic Unit (ALU)**
  - **Bit Level Circuit (BLC)**

- Each PCU corresponds to one processor or one CPU. The ALU is equivalent to Processor Element (PE). The BLC corresponds to combinational logic circuitry needed to perform 1 bit operations in the ALU.

Classification based on coupling between processing elements

Coupling refers to the way in which PEs cooperate with one another.

- **Loosely coupled**: the degree of coupling between the PEs is less.

Example: parallel computer consisting of workstations connected together by local area network such as Ethernet is loosely coupled. In this case each one of the workstations works independently.

If they want to cooperate they will exchange message. Thus logically they are autonomous and physically they do not share any memory and communication via I/O channels.

- **Tightly coupled**: a tightly coupled parallel computer, on the other hand shares a common main memory. Thus communication among PEs is very fast and cooperation may be even at the level of instructions carried out by each PE as they share a common memory.

Classification based on mode of accessing memory

- ***Uniform memory access parallel computers*** (UMC): in a shared memory computer system all processors share a common global address space. For these systems the time to access a work in memory is constant for all processors. Such a parallel computer is said to have a ***Uniform Memory Access*** (UMA).

- **Non uniform memory access parallel computers**: in a distributed shared memory computer system, each processor may its own local memory and may or may not share a common memory. For these systems, the time taken to access a word in local memory smaller than the time taken to access a word stored in memory of other computer or common shared memory. thus this systems said to have ***Non Uniform Memory Access*** (NUMA)

Simple Quiz..


THANK
YOU