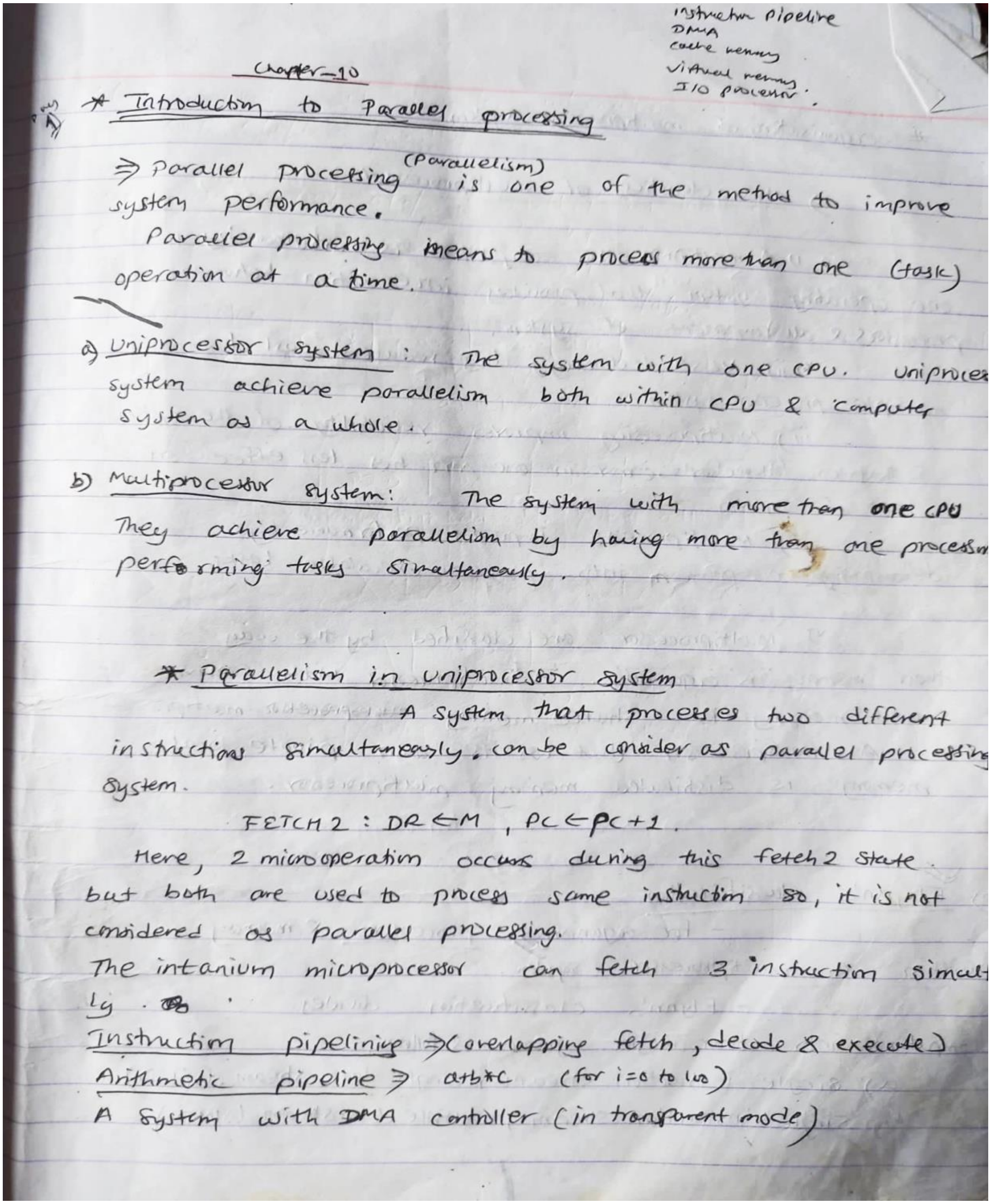
CHAPTER 10: INTRODUCTION to PARALLEL PROCESSING



instruction pipeline
DMA
cache memory
virtual memory
I/O processor.

Chapter-10

* Introduction to Parallel processing

⇒ Parallel processing (Parallelism) is one of the method to improve system performance.

Parallel processing means to process more than one (task) operation at a time.

a) Uniprocessor system : The system with one CPU. Uniproce system achieve parallelism both within CPU & Computer system as a whole.

b) Multiprocessor system: The system with more than one CPU They achieve parallelism by having more than one processo performing tasks simultaneously.

* Parallelism in uniprocessor system

A system that processes two different instructions simultaneously, can be consider as parallel processing system.

FETCH 2 : DR ← M , PC ← PC + 1.

Here, 2 microoperation occurs during this fetch 2 state. but both are used to process same instruction so, it is not considered as parallel processing.

The intanium microprocessor can fetch 3 instruction simult ly.

Instruction pipelining ⇒ (overlapping fetch, decode & execute)

Arithmetic pipeline ⇒ a+b*c (for i=0 to 100)

A System with DMA controller (in transparent mode)

organisation of multiprocessor system

a) characteristics of multiprocessor

i) A multiprocessor system is controlled by one operating system that provides interaction between processors & all component of system.

ii) VLSI technology has reduced cost of multiple processor system

iii) Multiprocessing improves reliability of the system so that a failure or error in one part has less effect on rest of system

iv) Multiprocessor improves performance by decomposing a program into parallel executable tasks.

v) Multiprocessor are classified by the way their memory is organised. A multiprocessor with common shared memory is shared memory ~~microprocessor~~ multiprocessor. A multiprocessor that has its own private local memory is distributed memory multiprocessor.

b) Flynn's classification

- for organising processors & memory within a multiprocessor system.
- Flynn's classification divides computers into four major groups as follows:-

a) single instruction stream, single data stream. (SISD)

b) single instruction stream, multiple data stream (SIMD)

c) Multiple instruction stream, single data stream (MISD)
d) Multiple instruction stream, multipledata stream (MIMD)

SISD represents the organisation of single computer containing a control unit, a processor unit & memory unit. Instructions are executed sequentially (one by one) & system may or may not have internal parallel processing capabilities Parallel Processing may be achieved by means of multiple functional units or by pipelining processing

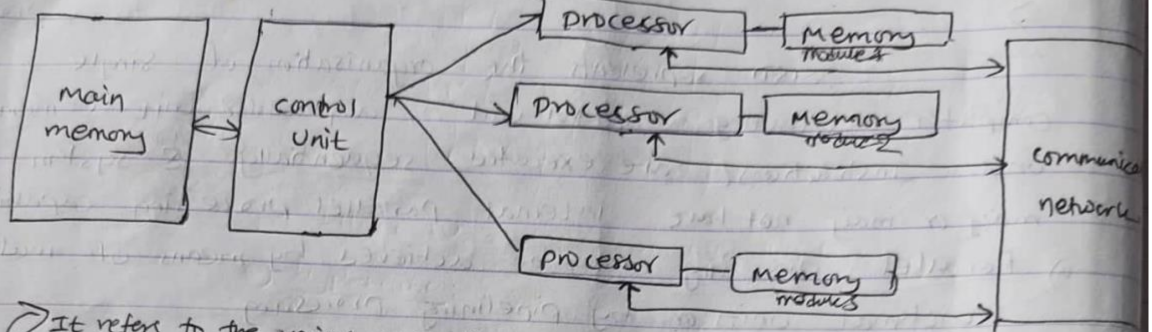
SIMD array processor

✓ SIMD represents an organisation that includes many processing units under the supervision of a common control unit. All the processors receive the same instruction from control unit but operate on different items of data. The shared memory unit must contain multiple modules so that it can communicate with all processors & simultaneously.

MISD structure is only of theoretical interest since no practical system ~~[illegible]~~ has been constructed using this.

MIMD organisation refers to a computer system capable of processing several programs at a same time. Most multiprocessor & multicomputer system are under this.

Generic SIMD organisation



It refers to the various ways of connection of processor.

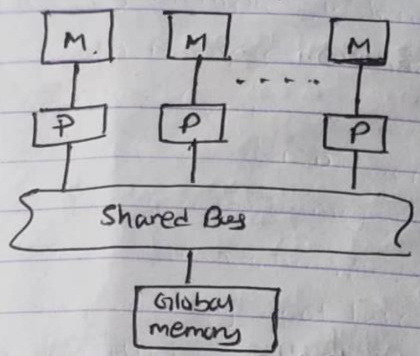
System topologies :-

Various arrangement of processor — MIMD system topologies

a) shared Bus topology
b) Ring Topology
c) Tree Topology
d) Mesh Topology
e) Hypercube Topology
f) completely connected Topology.

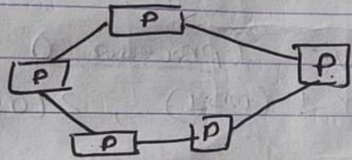
a) Shared Bus topology :-

- simplest topology
- A processor communicate with each other through this bus.
- This bus can handle only one data transmission at a time.
- Easy to expand.
- if more processors are added then demand of bus will be high & result in delay.



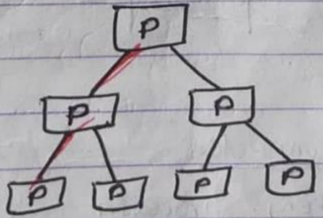
b) Ring Topology :- It has dedicated connections between processors.

A data need to be travel through several processors to reach from source to final destination.



c) Tree Topology :- It also uses direct connection between processors.

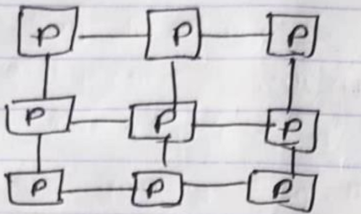
There is only one unique path between any pair of processors.



d) Mesh Topology :

In this topology every processor is connected to its below & above processor as well as left & right processor.

- given mesh topology is 3x3 mesh



e.g:- IlliacIV multiprocessor uses this topology

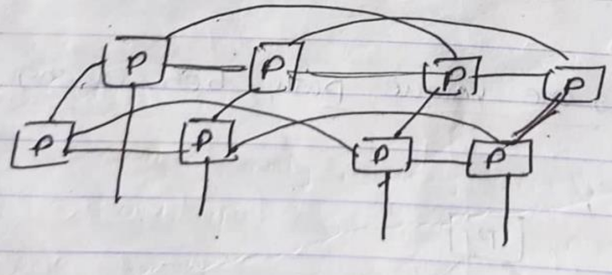
e) Hypercube:

- a multidimensional mesh.
- ~~every~~ Each processor connects to all processor whose binary values differ only by one bit.

For e.g:-

In fig:- processor 0 (0000) connects to processor 1 (0001), 2(0010), 4(0100) and 8(1000)

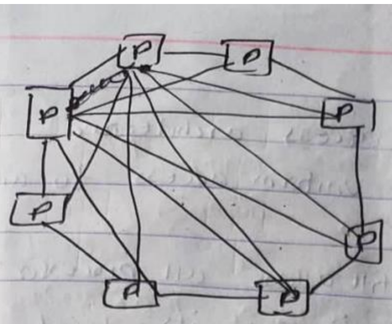
e.g:- nCUBE system use hypercube topology



f) completely connected

Each & Every processor is connected to all processor.

- Increases complexity but offers max^m communication.



* **MIMD system Architecture :**

It refers to its connections with respect to system memory.

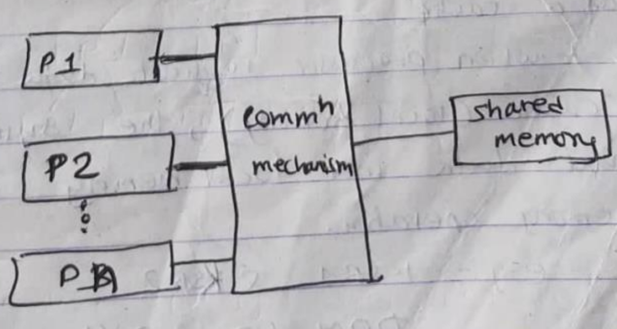
Symmetric multiprocessor (SMP)

A computer system that has two or more processors with comparable capabilities. All processors must be capable of performing the same functions. An Integrated OS controls entire computer system. All processor have access to same I/O devices & memory modules.

Types of SMP

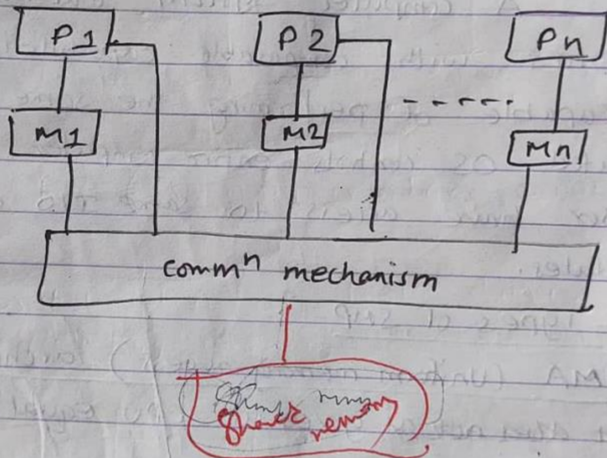
i) **UMA (Uniform memory access) architecture.**

- It ~~does not~~ gives all CPU equal access to all locations in shared memory.
- They interact with shared memory through some commn mechanism.



NUMA

- non uniform memory access architecture.
- It does not allow uniform access to all shared memory locations.
- This architecture still allows all processor to memory module closest to it but its local memory more quickly than other. Hence, memory access time are non-uniform.
- NUMA computers be not like SMP.
- NUMA has better performance than UMA
- e.g:- CRAY T3E



- COMA

- cache only memory access architechere.
- Each processor's local memory is treated as a cache.
- when processor request data that is not in cache (local memory), the system loads that data into local memory as part of memory operation.
- e.g:- KSR1 & KSR2
  DDM (Data Diffusion machine)

day3

* Commⁿ in multiprocessor system

→ It is a key factor in determining system overall performance.

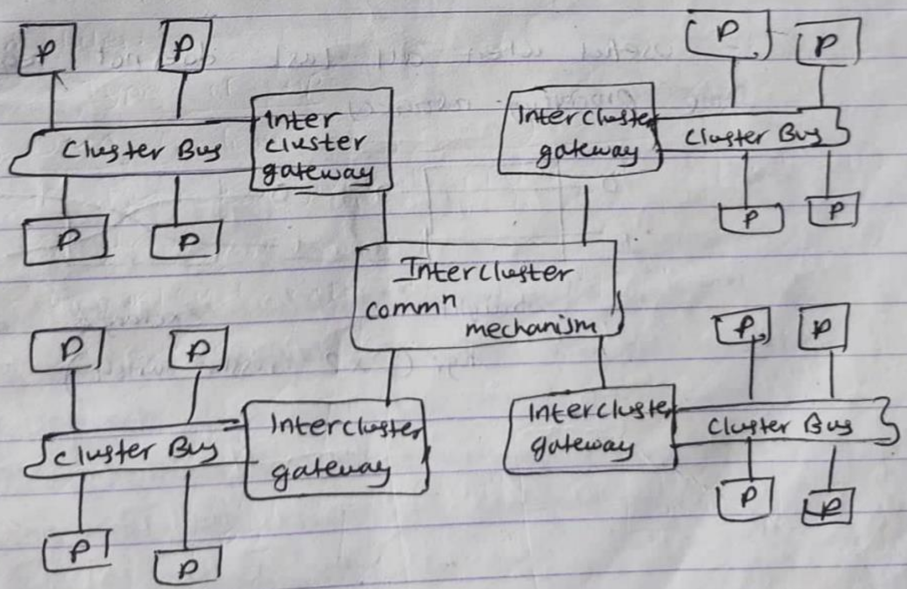
Two ways of communication:-

i) Fixed connections

ii) Reconfigurable connections

i) Fixed connection:-

- The connection that never change.
- Inflexible for some system but sufficient for many systems.
- Less costly than reconfigurable connections.
- e.g:- system with shared bus.
- clustering is one of the fixed connection topology



(Fig - A 16 processor multiprocessor that uses clustering)

ii) Reconfigurable connections (changing)

In this, there is a ability of reconfiguring connections between processors & memory, I/o devices & other processor can allow it to meet the needs of individual tasks & maximize system performance.

- cross bar switch mechanism is used.

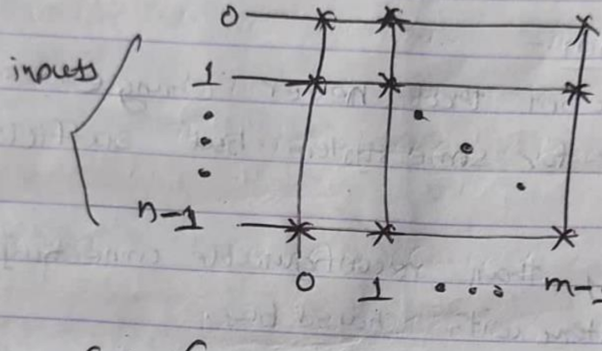


fig:- (A n×m cross bar switch)

- useful when all task does not require same processing resources

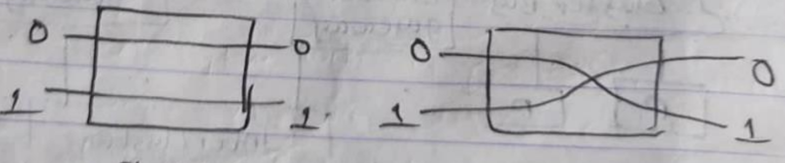


fig:- (2×2 crossbar switch)

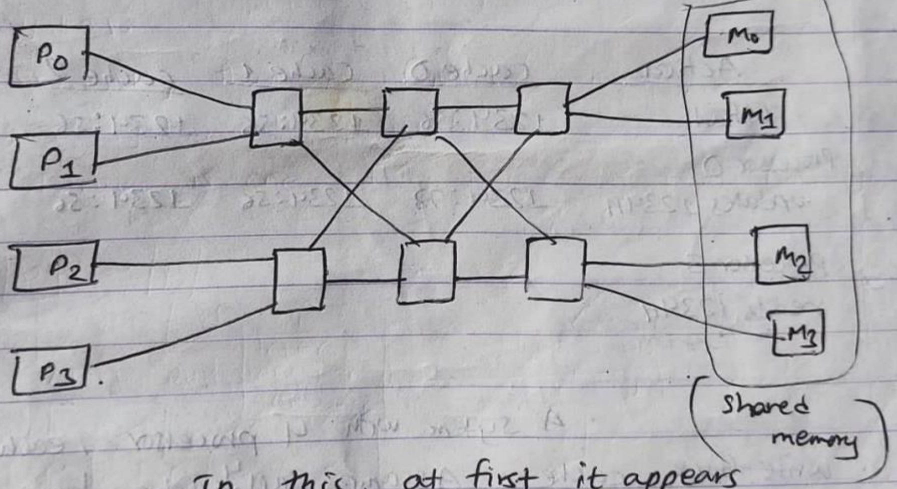
* Memory organisation in multiprocesor system :

1) Shared memory :-

- Both UMA & NUMA architecture uses shared memory. Processors can ~~share~~ access shared programs & data.

- Processors can also use Shared memory. to communicate each other through message passing.

- In add^n to message passing, the OS uses shared memory to store information about its current state. which ~~can~~ be access by processor.



In this, at first it appears like all processors try to access a single Shared memory module & that only one can be successful at given time. In practise, shared memory is partitioned into several modules (M1, M2, M3) all of which can accessed simultaneously.

ii) Cache coherence :

It is a problem in multiprocessor system. Multiprocessor have individual cache for each processor. This can lead to problem when two or more caches hold the value of same memory location simultaneously.

As one processor stores a value to that location in its cache, the other cache will have an invalid value in its location. The extra writes to main memory is needed which decrease system performance. This is cache coherence.

Action	cache 0	Cache 1	cache 2	cache 3
Initial	1234:56	1234:56	1234:56	1234:56
Processor 0 updates 1234H	1234:78	1234:56	1234:56	1234:56
processor 3 reads 1234H location				Reads 56H instead of 78H.

. A sytem with 4 processor, each of which has write-back cache.. Assume all 4 cache have loades the content of shared memory location 1234H which is 56H Then one of the processor, processor 0 writes value 78H to this location in its cache. But caches 1, 2, & 3 do not have correct value. if one of the other processor reads then it will read the old incorrect value 56H insteat of correct value 78H.