UNIT 10
MULTIPROCESSORS

CO(2140707)

Prepared By: Prof. Chirag Bhalodia

Outline

- Characteristics of Multiprocessors
- Interconnection Structures
- Inter-processor Arbitration
- Inter-processor Communication and Synchronization
- Cache Coherence
- Shared Memory Multiprocessors

Characteristics of Microprocessor

- If a microprocessor system contains two or more components that can executes instructions independently, then the system is called **multiprocessor** system.

- This system has following **advantages**:
  - Improve performance ratio
  - To avoid expense of centralized system
  - Tasks are divided in module
  - Improves Reliability

- Multiprocessor can be **characterized** by two main categories:
  - It is a single computer that includes multiple processors
  - Communication between them may occur by sending messages from one processor to the other or by sharing common memory.

- Multiprocessor system implemented using to basic architecture:
  1. Tightly coupled multiprocessor (Closely)
  2. Loosely coupled multiprocessor

Tightly coupled multiprocessor

- In this system the processors shares clock generator, bus control logic, entire memory and I/O system.

- This systems communicate through a memory.

- One of the limitation of this system is the performance degradation due to memory contentions which occur when two or more processors attempt the same memory unit simultaneously.

- When high speed of real-time processing is desired, this system may be used.

- There are two models of this system
  - Without private cache
 - With private cache

Without Private cache
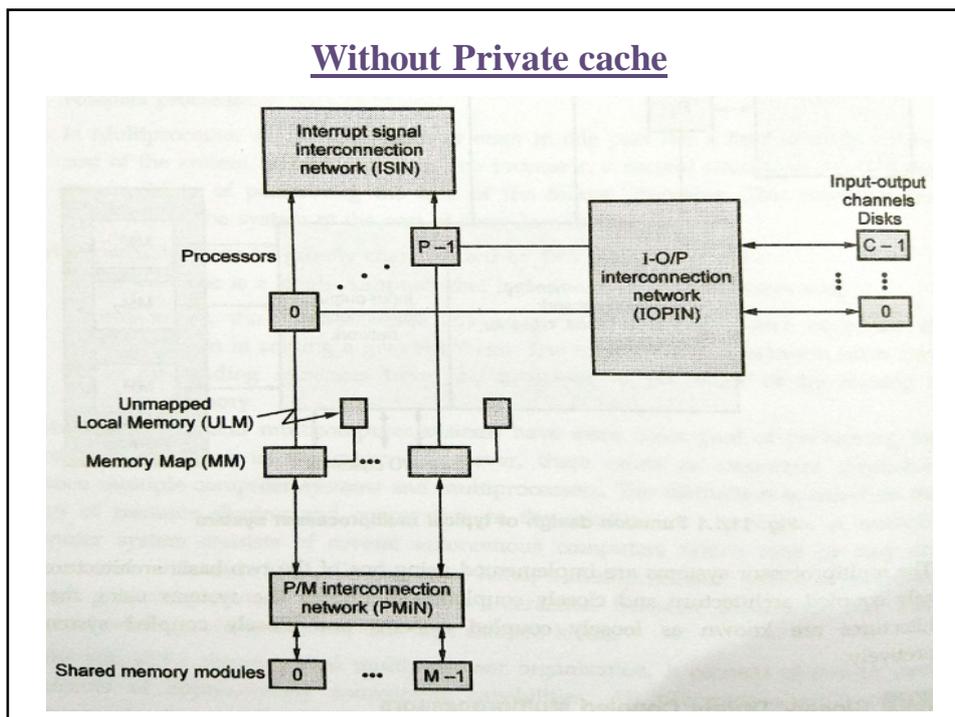
Without Private cache

- It consist of P processors, M memory modules and C input-output channels. These units are connected through a set of three interconnected networks.
  - The processor-memory interconnection network (PMIN)
  - The input-output processor interconnection network (IOPIN)
  - The interrupt-signal interconnection network (ISIN)

- The PMIN is a switch which is used to connect every processor to every memory module.

- The IOPIN is used to allow a processor to communicate with an I/O channel which is connected to I/O devices.

- The ISIN is used for two purposes : To direct an interrupt to any other inter-processor network and to initiate hardware alarm in case of processor failure.
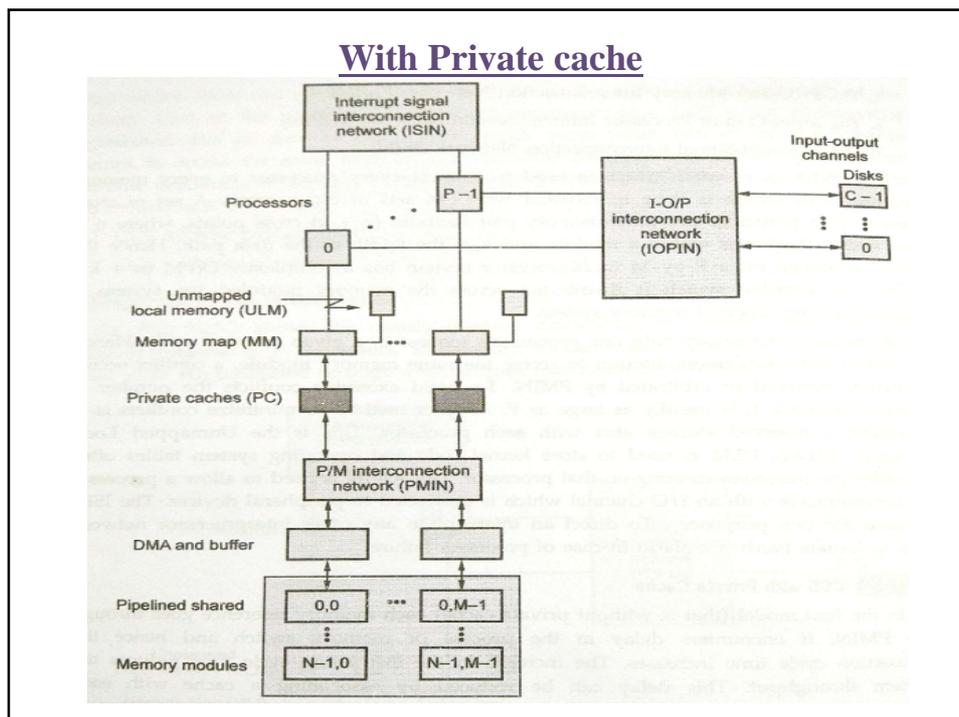
With Private cache

With Private cache

- In previous method each memory reference goes through the PMIN, it encounters delay in the processor or memory switch and hence the instruction cycle time is increases.

- It reduces system throughput. This delay can be reduced by associating a cache with each processor.

- The another advantage of the cache is that the traffic through the crossbar switch can be reduced.

- More than one inconsistent copy of data may exist in the system as this multiprocessor system encounter cache coherence problem.

Loosely coupled multiprocessor

(a) A computer module

(b) Loose coupling of computer modules
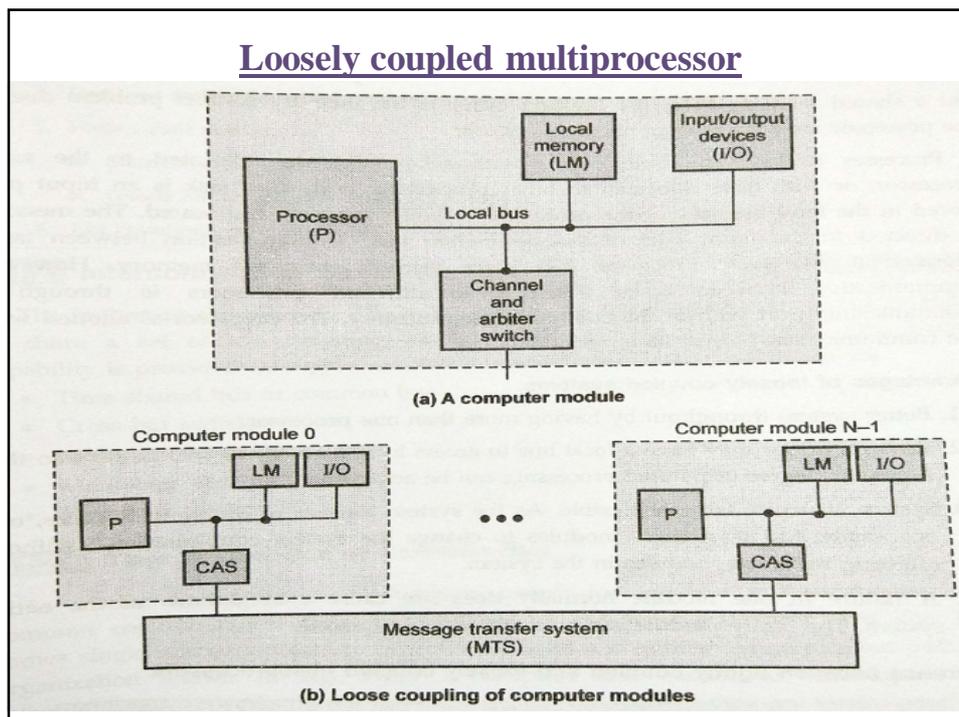
Loosely coupled multiprocessor

- In this system each processor has a set of input-output devices and a large local memory and input-output interfaces are together called **computer module**.

- Processes which executes on different computer modules communicate by exchanging messages through a **Message Transfer System (MTS).**

- This coupling in such a system is very loose. Hence, this system are also called as **distributed systems.**

- In diagram, module contains processor, local memory and I/O devices connected through **Channel and Arbiter Switch (CAS)**

- CAS consists of a high speed communication memory which is used for buffering block transfer of message.

Loosely coupled multiprocessor

- The MTS are divided in to two categories:
  - Simple time shared bus
  - Shared memory system

- The time shared bus is a common communication path connecting all functional units.

- On the other hand shared memory system consists of a set of memory modules and a processor-memory interconnection network.

- The configuration of MTS is one of the most important factors that determines the performance of the multiprocessor system.

Loosely coupled multiprocessor

- Communication between tasks allocated to the same processor take place through the local memory.

- However, communication between task allocated to different processors is through a communication port with in the communication memory.

- Advantages:
  - Better system throughput
  - Parallel Processing
  - More flexible
  - More reliable

## Tightly coupled vs Loosely coupled multiprocessor

Tightly coupled multiprocessors	Loosely coupled multiprocessors
A large main memory is shared among all processors.	Each processor has local main memory.
Degree of interaction among tasks is high.	Degree of interaction among tasks is low.
Because of frequent sharing of codes between two processors, bus conflicts are high.	Bus conflict problem due to sharing does not exists in loosely coupled multiprocessors.
Used in parallel processing systems.	Used in distributed computing systems.

GTU Questions

1. Discuss the differences between tightly-coupled multiprocessor and loosely-coupled multiprocessor.

2. Explain tightly coupled system

3. Explain loosely coupled system

Interconnection Structures

- The important characteristics of processor used in multiprocessor system is its ability to share a set of main memory modules and possibly I/O devices. These sharing capability is provided through a set of two interconnection networks. These are
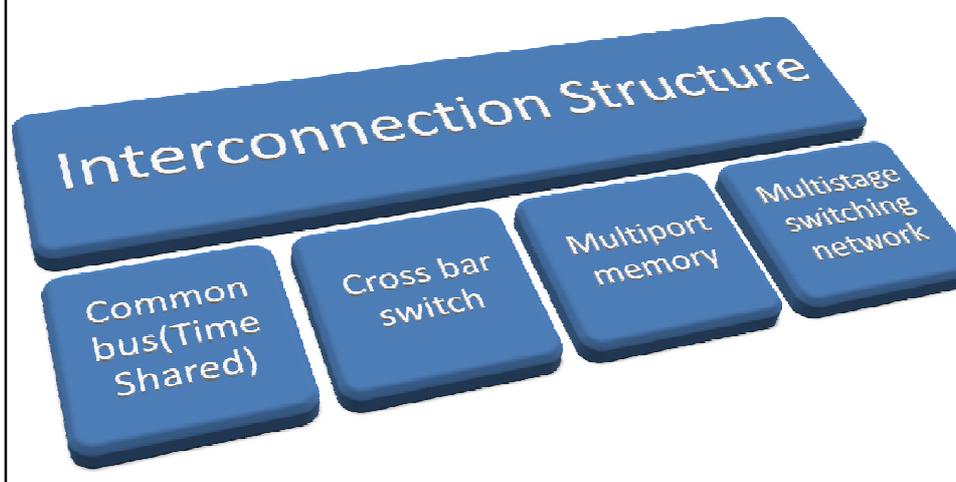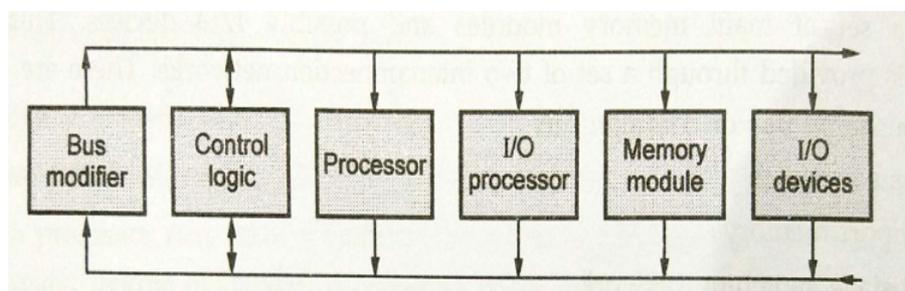
Common Bus (Time Shared Bus)
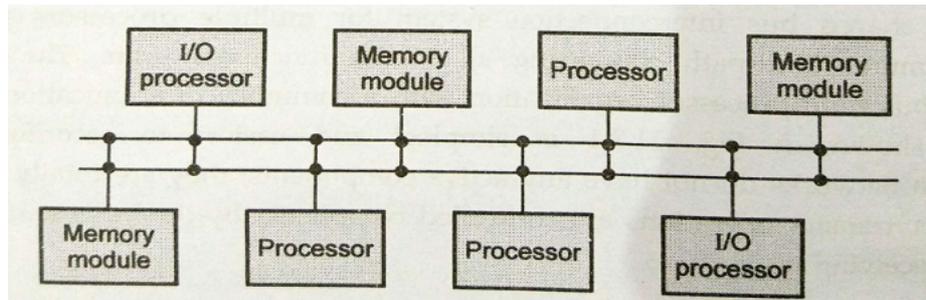
• Single bus multiprocessor organization
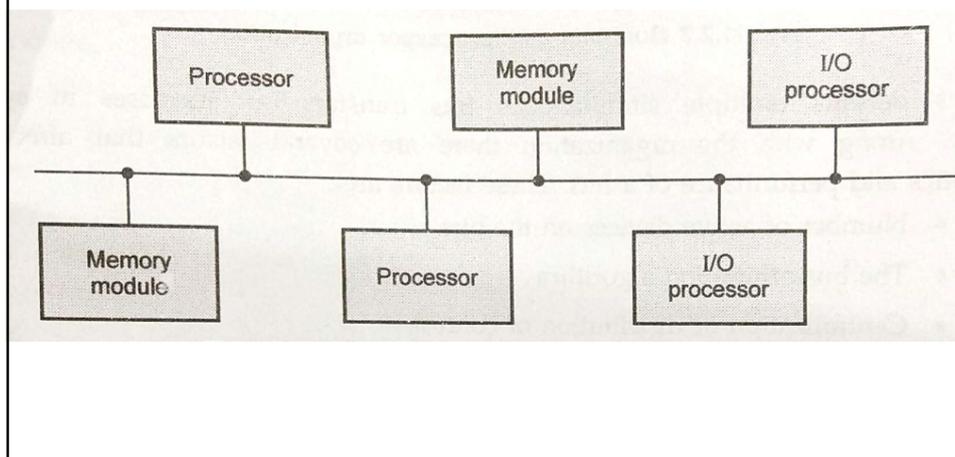
Common Bus (Time Shared Bus)

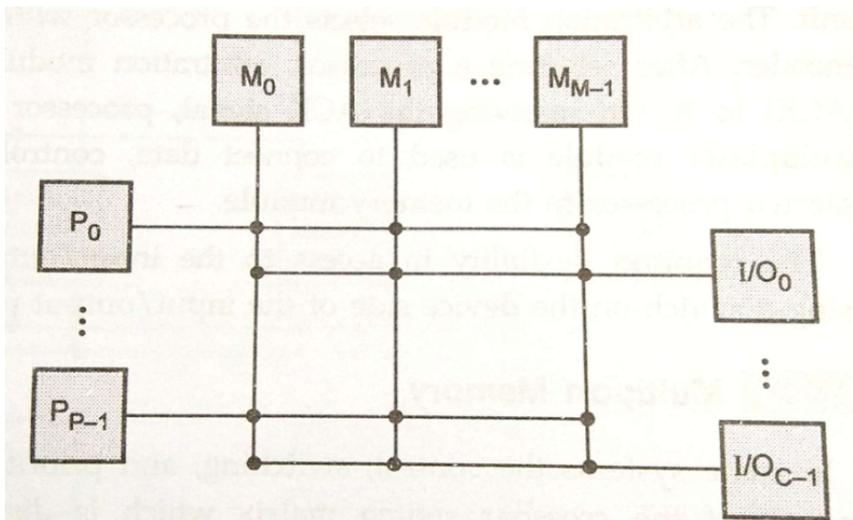• Multiprocessor with unidirectional buses

Common Bus (Time Shared Bus)

- Multi-bus multiprocessor organization

Crossbar Switch

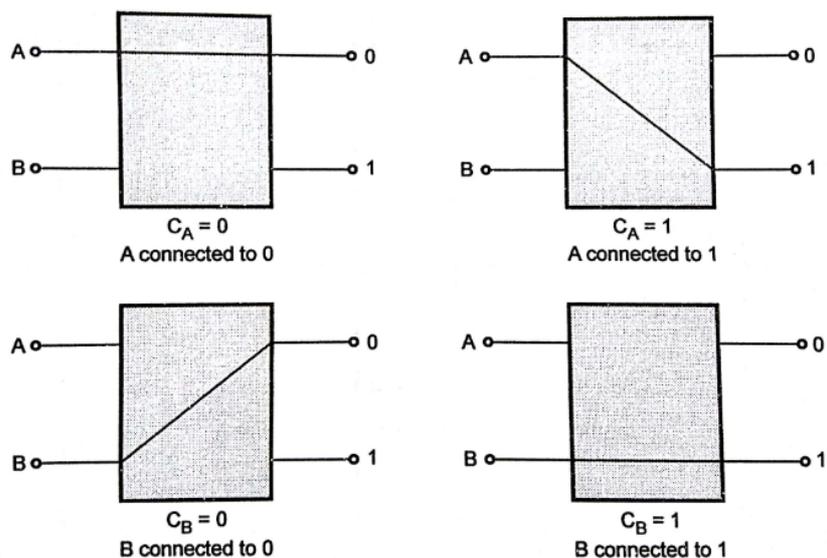# Crossbar Switch

- If the number of buses in a common bus system is increased, a point is reached at which there is a separate path available for each memory module.

- Diagram shows the cross bar switch system organization for multiprocessors which provides separate path for each memory module.

- The interconnection network shown in diagram is called non-blocking crossbar.

Crossbar Switch

Crossbar Switch

- In multistage switching networks a 2x2 crossbar switch is used as basic components. It has two input label A and B, and two output label 0 and 1.

- The control inputs C_A and C_B associated with the switch establish the connection between the input and output terminal shown in diagram.

- If control input is zero it connects the input to the 0 output and if it is one it connects the input to the 1 output.

- If both input A and B require the same output terminal, then only one of them will be connected and the other will be blocked or rejected.

Multiport Memory

- Multiport memory without fixed priority
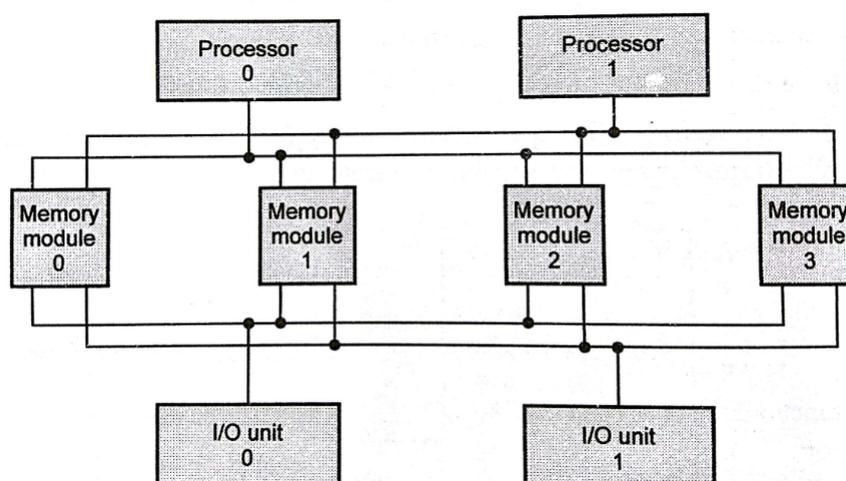
Multiport Memory

- Multiport memory with private memories

Fig. 11.2.7 Multiport ... ...
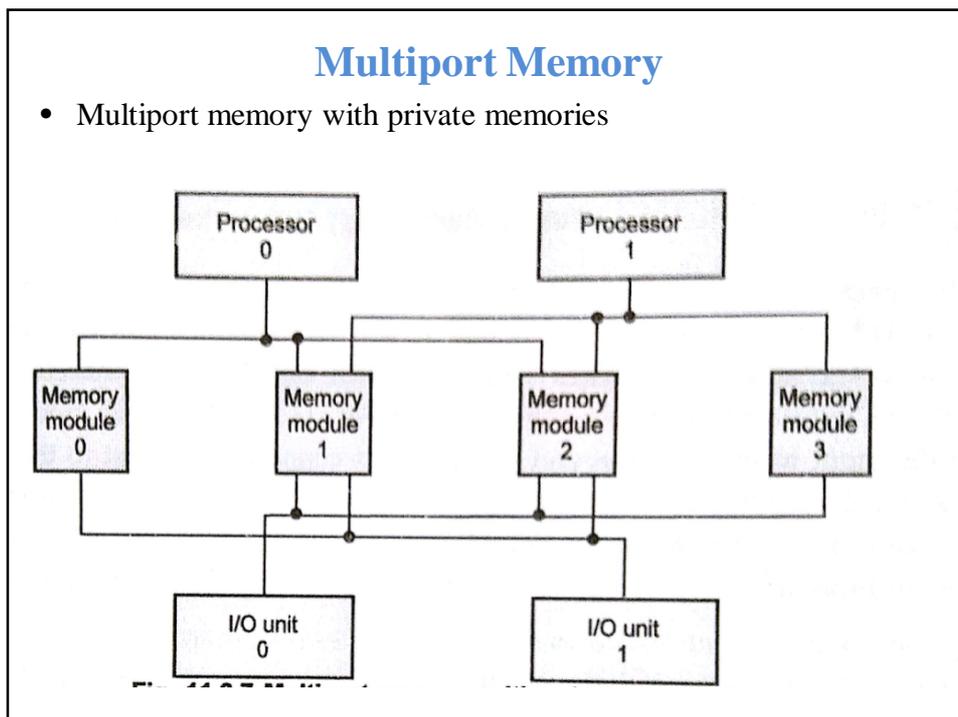
Multistage Switching Network

- 1 to 8 way switch using basic 2x2 switch (Binary tree)
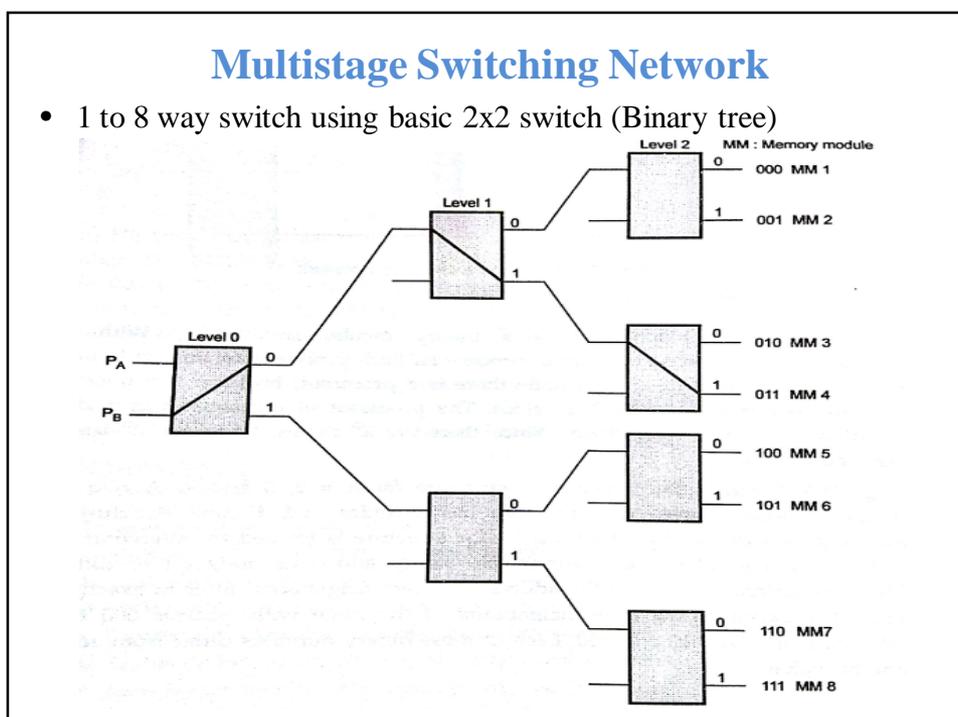
GTU Questions

1. Explain any two interconnection structures that make it possible to form a multiprocessor system, with diagrams.

2. Write a note on crossbar switch interconnection structure with block diagram.

Inter-processor Arbitration

- The processor, main memory and I/O devices can be interconnected by means of a common bus.

- A bus is set of lines (wires) defined to transfer all bits of a word from a specified source to a specified destination. Thus bus provides a communication path for the transfer of data.
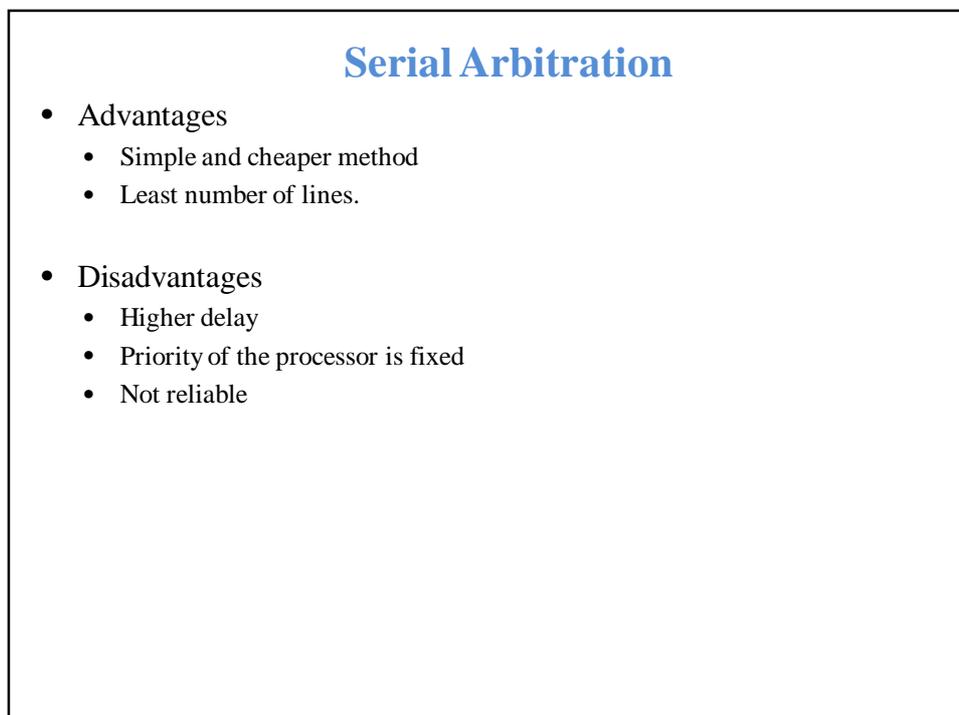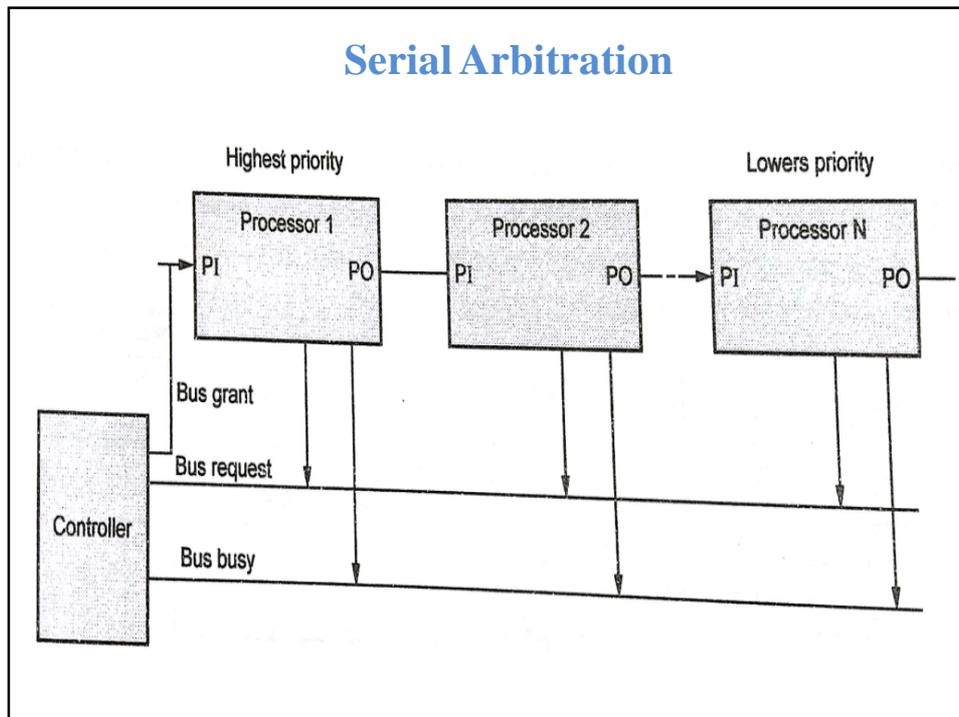
- The bus includes data lines, address lines and control lines. Such a bus known as **system bus**.

- Types of arbitration :
    - Bus arbitration
    - Serial arbitration
    - Parallel arbitration
    - Dynamic arbitration

Bus Arbitration

- The device that is allowed to initiate data transfer on the bus at any given time is called **bus master.**

- In a computer system there may be more than one bus master such as processor, DMA controller etc.

- Bus arbitration is the process by which the next device to become the bus master is selected and bus master-ship is transferred to it.

- The selection of bus master is usually done on the priority basis.

Serial Arbitration

Highest priority

Lowers priority

Processor 1
PI   PO

Processor 2
PI   PO

Processor N
PI   PO

Bus grant

Bus request

Controller

Bus busy

Serial Arbitration

- Advantages
    - Simple and cheaper method
    - Least number of lines.

- Disadvantages
    - Higher delay
    - Priority of the processor is fixed
    - Not reliable

Parallel Arbitration

Processor 1 — Arbitration Logic — Req ACK Busy
Processor 2 — Arbitration Logic — Req ACK Busy
Processor 3 — Arbitration Logic — Req ACK Busy
Processor 4 — Arbitration Logic — Req ACK Busy
Busy

4×2 Priority Encoder

2×4 Decoder

2 – bits bus to transfer
2 – bit code representing higher priority processor

Parallel Arbitration

- Advantages
  - Separate pair of bus request and bus grant signals, so it is faster.

- Disadvantages
  - Require more bus request and grant signal

Dynamic Arbitration

Parallel Arbitration

- Advantages
  - The priority can be changed by altering the sequence stored in controller.
  - More reliable.

Inter-processor Communication and Synchronization

Inter-Process Communication

- In a multiprocessing environment processors implies parallelism by concurrent processing.

- The concurrent processing requires sharing of resources between the processors and inter-processor communication.

- Basically there are two ways by which inter process communication is achieved
  - Using Shared Variables
  - Using Message Passing

IPC using shared variable
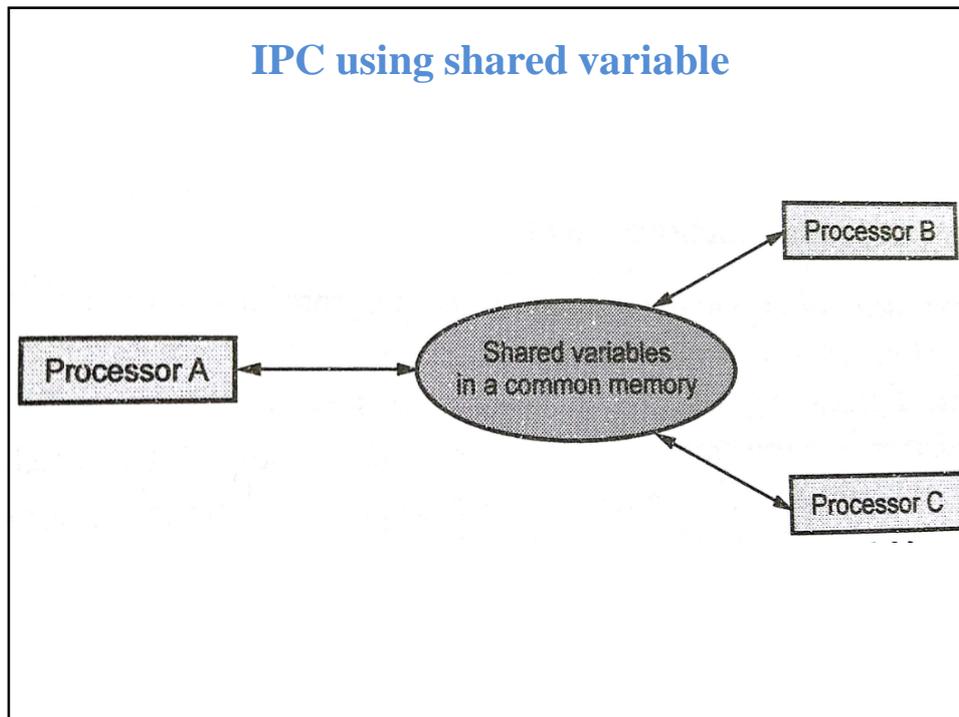
IPC using shared variable

- In this system shared variable are stored in common memory which is accessible to all processors in the system.

- While sharing common resources or shared variables conflict problem may arise.

- It is necessary to prevent conflict use of shared resources by several processors. This task is done by operating system.

1. Master-slave operating system
2. Separate operating system
3. Distributed operating system

IPC using shared variable

1. **Master-slave operating system**: In this processor that executes the operating system function is called **master.** The other processors are called **slave.** When slave needs operating system service it request it by interrupting the master.

2. **Separate operating system**: In each processor has entire copy of operating system can be execute operating system functions. This organization is more suitable for loosely couple system.

3. **Distributed operating system**: In this operating system, routines are distributed among the available processors. Such type operating system is also known as **floating operating system.**

IPC using message passing

Processor A ←→ Message (send / receive) (Communication channel) ←→ Processor B
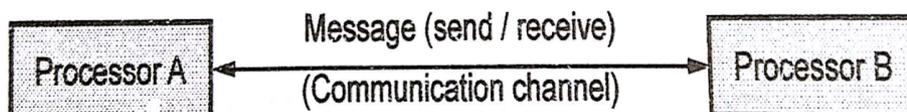
IPC using message passing

- In multiprocessor system with no shared memory we use message passing mechanism to perform inter-process communication.

- When processor wants to communicate with another processor it uses a special procedure which initiates communication.

- It identifies the destination processor and once source and destination processors are identified a communication channel is established.

- A message is then sent through the communication channel

Inter-Processor Synchronization

- At the higher level of parallelism, the program is partitioned into process that are executed in different processors.

- This technique is called **concurrent processing.**

- During concurrent processing when two or more processors need the same resource at a time, the contention problem arise. Such problem can be solved by **synchronization.**

- To achieve synchronization a set of hardware primitives are used to automatically read and modified a memory location without any interruption between the two operations.
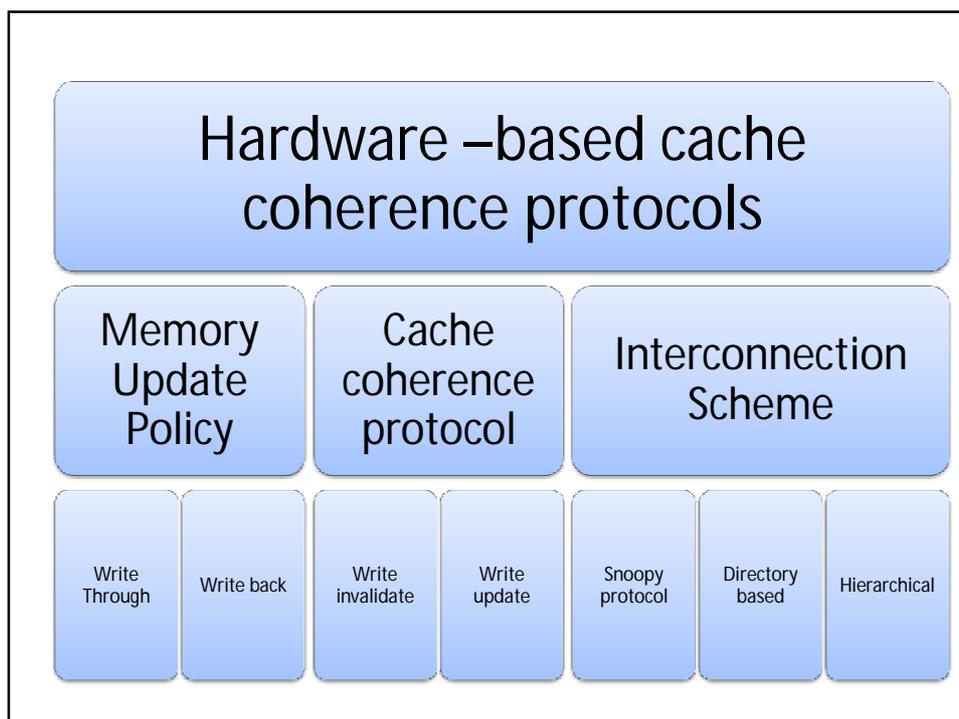
Inter-Processor Synchronization

- Such mechanisms are necessary to protect data from being changed simultaneously by two or more processors. This mechanism is known as **mutual exclusion.**

- A program sequence which accesses the shared resources, one begun, must complete execution before another processor accesses the same shared address.

- Thus program sequence which accesses the shared memory is known as **critical section** of the program.

- Inter-process communication done by mutual exclusion with semaphore and mutual exclusion using load and store conditional instruction.

GTU Questions

1. Explain inter process communication and synchronization in the context of multiprocessor.
2. What is mutual exclusion?
3. What is critical section?
4. Why program sequence is required in inter processor synchronization?

Cache Coherence

- In a multiprocessor system, two copies of same data, one in cache and another in main memory may become different. This data consistency is called as **cache coherence problem.**

- The protocol used to maintain coherence for multiple processors are called **cache coherence protocols.**

26

Hardware –based cache coherence protocols

Memory Update Policy	Cache coherence protocol	Interconnection Scheme
Write Through · Write back	Write invalidate · Write update	Snoopy protocol · Directory based · Hierarchical

---

Memory Update Policy

- Write through
- Write back

(Already studied in previous chapter (Memory Organization))

Cache coherence protocol

Write Update :

- In write update protocol whenever a processor updates a cache data, it immediately updates all other cached copies.

- Whether the shared memory copy is updated depends on the memory update policy.

- Network traffic is higher in compare to write invalidate protocol.

- Less time to required to read the data.

Cache coherence protocol

Write Invalidate :

- In write invalidate protocol, the updated cache block is not sent immediately to other caches.

- Invalidate command is sent to all other cache copies and to the original version in the shared memory so that they become invalidate.

- If later another processor wants to read data, it is provided by the updating processor.

- In this protocol network traffic is reduced in compare to write update protocol.

Interconnection Scheme

Snoopy cache coherence protocol :

- It is mostly used in shared bus multiprocessor.

- This protocol typically used in single bus shared memory

- In this protocol, every cache that has a copy of data from a block of physical memory also has a copy of the sharing status of the block.

Interconnection Scheme

Directory based cache coherence protocol :

- It is mostly used in multistage network.

- In large connection network cannot support broadcasting efficiently and therefore a mechanism is needed is needed that con directly forward command to those cache that contain copy of the updated data.

- For this purpose the sharing status of a block of physical memory is kept in location, called the directory.

Interconnection Scheme

Hierarchical cache coherence protocol :

- This protocol tries to avoid the application of the directory protocol but still provide high scalability.

- It purposes multiple-bus network with the application of hierarchical cache coherence protocol.

Shared memory Multiprocessors

- MIMD multiprocessor divided into two classes, depending on the number of processors involved.

  – **Centralized Share memory architecture**

    Multiprocessor with small number of processors and processors in the multiprocessor can share single centralized memory.

  – **Distributed shared memory architecture**

    Multiprocessors with large number of processors and physically distributed memory among the processors.

Centralized Shred Memory Architecture

Distributed-Shred Memory Architecture

Distributed-Shred Memory Architecture

Existing virtually memory: