K. S. Institute of Technology

Department of Computer Science and Engineering

Advanced Computer Architecture – 18CS733

Faculty Name: Dr. Vijayalaxmi Mekali

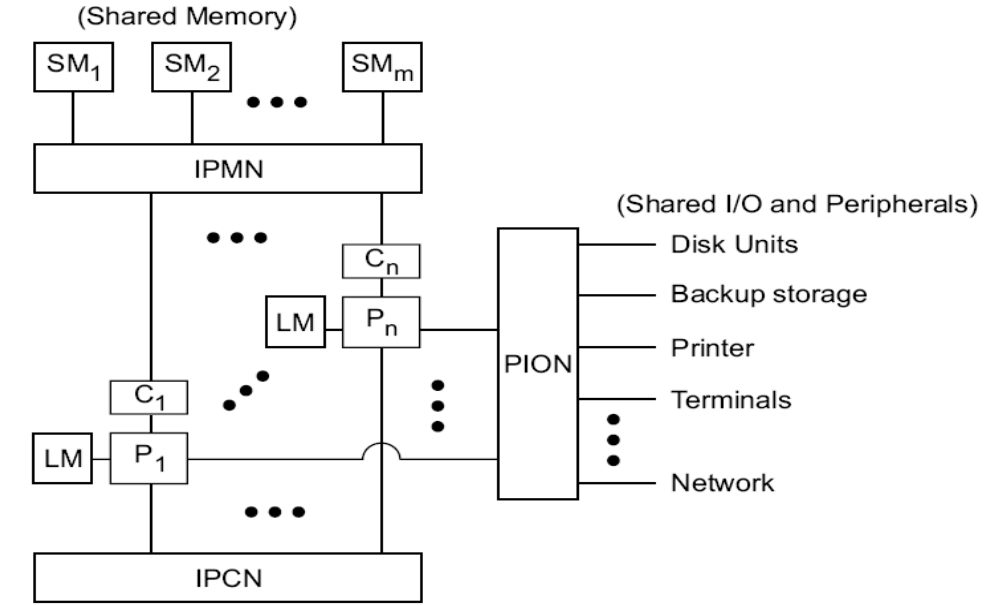
Associate Professor, Dept. of CSE

KSIT, Bangalore

Multiprocessor System Interconnects

- Parallel processing demands the use of efficient system interconnects for fast communication among multiple processors and shared memory, I/O and peripheral devices.
- Hierarchical buses, crossbar switches and multistage networks are often used for this purpose.
- A generalized multiprocessor system is depicted in Fig. 7.1. This architecture combines features from the UMA, NUMA and COMA models

Multiprocessor System Interconnects



Legends:
IPMN (Inter-Processor-Memory Network)
PION (Processor- I/O Network)
IPCN (Inter-Processor Communication Network)
P (Processor)
C (Cache)
SM (Shared Memory)
LM (Local Memory)

Fig. 7.1 Interconnection structures in a generalized multiprocessor system with local memory, private caches, shared memory, and shared peripherals

Multiprocessor System Interconnects

- Each processor Pi is attached to its own local memory and private cache.
- These multiple processors connected to share memory through interprocessor memory network (IPMN).
- Processors share the access of I/O and peripheral devices through Processor-I/O Network (PION). Both IPMN and PION are necessary in a shared-resource multiprocessor.
- An optional Interprocessor Communication Network (IPCN) can permit processor communication without using shared memory.

Module-IV

Parallel and Scalable Architectures: Multiprocessors and Multicomputers

Multiprocessor System Interconnects

a) Network Characteristics

The networks are designed with many choices like timing, switching and control strategy like in case of dynamic network the multiprocessors interconnections are under program control.

Dynamic networks are used in multiprocessors in which the interconnections are under program control.

Timinng, switching, and control are three major operational characteristics of an interconnection networks.

- **Timing**
  - **Synchronous – controlled by a global clock which synchronizes all network activity.**
  - **Asynchronous – use handshaking or interlock mechanisms for communication and especially suitable for coordinating devices with different speed.**
- **Switching Method**
  - **Circuit switching – Once a device is granted a path in the network, it occupies the path for the entire duration of the data transfer. A pair of communicating devices control the path for the entire duration of data transfer**
  - **Packet switching – large data transfers broken into smaller pieces, each of which can compete for use of the path in the network.**
- **Network Control**
  - **Centralized – global controller receives and acts on requests**
  - **Distributed – requests handled by local devices independently**

Multiprocessor System Interconnects

b) Hierarchical Bus Systems

- A bus system consists of a hierarchy of buses connecting various system and subsystem components in a computer.
- Each bus is formed with a number of signal, control, and power lines. Different buses are used to perform different interconnection functions.
- In general, the hierarchy of bus systems are packaged at different levels as depicted in Fig. 7.2, including local buses on boards, backplane buses, and I/O buses.
-

Multiprocessor System Interconnect

b) Hierarchical Bus Systems

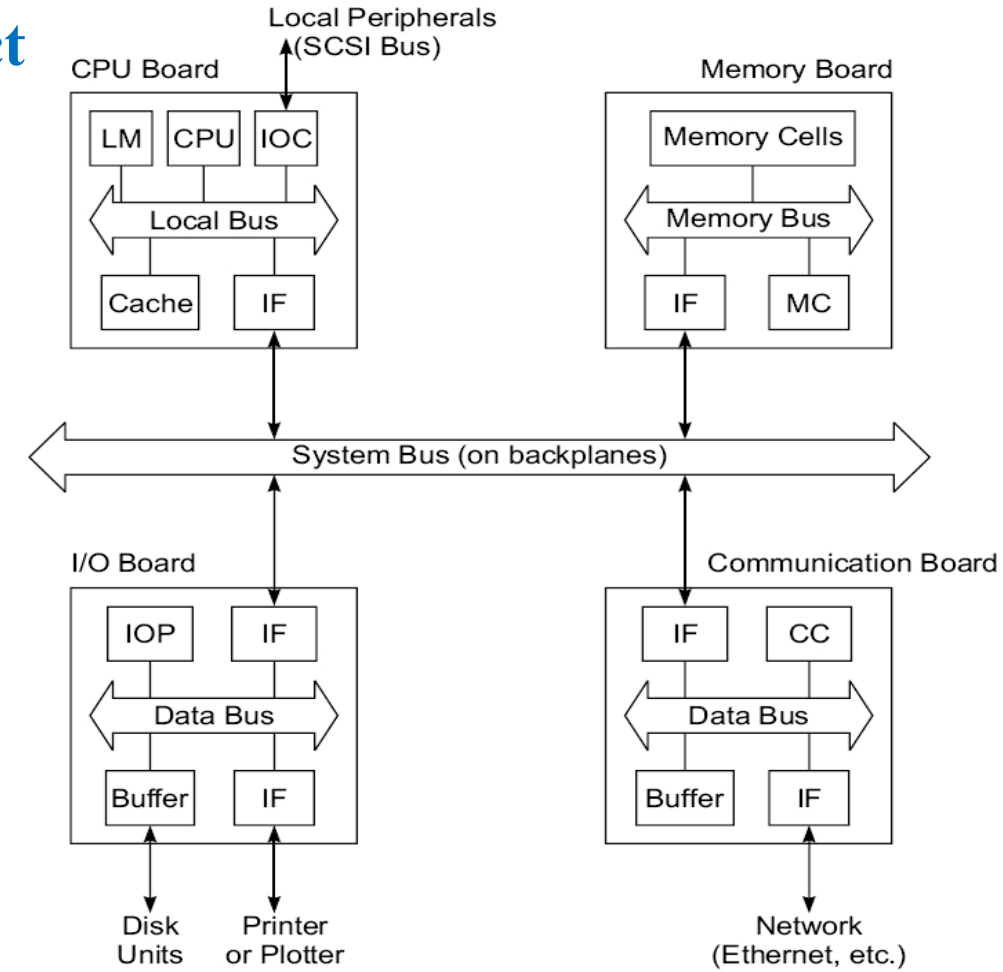


Fig. 7.2 Bus systems at board level, backplane level, and I/O level

Multiprocessor System Interconnects

b) Hierarchical Bus Systems

- **Local Bus:** Buses implemented within processor chip or on printed-circuit boards are called local buses.
- On a processor board one often finds a local bus which provides a common communication path among major components (chips) mounted on the board.
- A memory board uses a memory bus to connect the memory with the interface logic.
- An I/O or network interface chip or board uses a data bus. Each of these board buses consists of signal and utility lines.
- **Backplane Bus**
  - A backplane is a printed circuit on which many connectors are used to plug in functional boards.
  - A system bus, consisting of shared signal paths and utility lines, is built on the backplane. This system bus provides a common communication path among all plug-in boards.
- **I/O Bus**
  - Input/Output devices are connected to a computer system through an I/O bus such as the SCSI(Small Computer Systems Interface) bus.
  - This bus is made of coaxial cables with taps connecting disks, printer and other devices to a processor through an I/O controller.
  - Special interface logic is used to connect various board types to the backplane bus.

Multiprocessor System Interconnects

b) Hierarchical Bus Systems

- Complete specifications for a bus system includes electrical and mechanical properties, various application profiles and interface requirements.
- Digital bus interconnects can be adopted in commercial systems ranging from workstations to minicomputers, mainframes and multiprocessors.
- Hierarchical bus system can be used to build medium sized multiprocessors with less than 100 processors.

Multiprocessor System Interconnects

b) Hierarchical Bus Systems

➢ Hierarchical Buses and Caches

- This is a multilevel tree structure in which the leaf nodes are processors and their private caches (denoted Pj and C1j in Fig. 7.3). These are divided into several clusters, each of which is connected through a cluster bus.
- An intercluster bus is used to provide communications among the clusters.
- Second level caches (denoted as C2i) are used between each cluster bus and the intercluster bus.
- Each second level cache must have a capacity that is at least an order of magnitude larger than the sum of the capacities of all first-level caches connected beneath it.

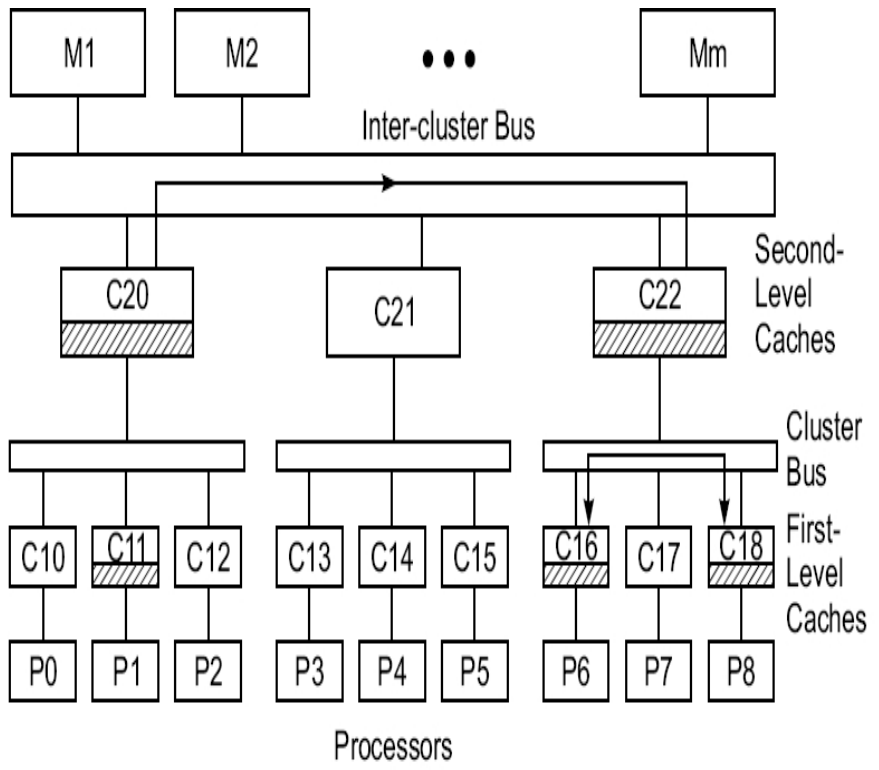


Fig. 7.3 A hierarchical cache/bus architecture for designing a scalable multiprocessor (Courtesy of Wilson; reprinted from *Proc. of Annual Int. Symp. on Computer Architecture, 1987*)

Multiprocessor System Interconnects

b) Hierarchical Bus Systems

➢ Hierarchical Buses and Caches

- Each single cluster operates as a single-bus system. Snoopy bus coherence protocols can be used to establish consistency among first level caches belonging to the same cluster.
- Second level caches are used to extend consistency from each local cluster to the upper level.
- The upper level caches form another level of shared memory between each cluster and the main memory modules connected to the intercluster bus.
- Most memory requests should be satisfied at the lower level caches.
- Intercluster cache coherence is controlled among the second-level caches and the resulting effects are passed to the lower level.

Multiprocessor System Interconnects

c) Crossbar Switch and Multiport Memory

- **Switched networks provides the dynamic interconnection between the inputs and outputs.**
- **Crossbar networks are mostly used in small or medium size systems**

➢ Network Stages

- **Single stage networks are sometimes called recirculating networks because data items may have to pass through the single stage many times before reaching their destination.**
- **Single stage networks are cheaper to build, but multiple passes may be needed to establish certain connections.**
- **The crossbar switch and the multiported memory organization are both single-stage networks.**
- **This is because even if two processors attempted to access the same memory module (or I/O device) at the same time, only one of the requests is serviced at a time.**

➢ Multistage Networks

- **Multistage networks consist of more than one stage of switch boxes and should be able to connect any input to any output.**

Dynamic Interconnection Network –

Dynamic interconnection networks are connections between processing nodes and memory nodes that are usually connected through switching element nodes.

It is scalable because the connections can be reconfigured before or even during the execution of a parallel program.

Here instead of fixed connections, the switches or arbiters are used.

The dynamic networks are normally used in shared memory (SM) multiprocessors.

These networks use configurable paths and do not have a processor associated with each node.

Multiprocessor System Interconnects

c) Crossbar Switch and Multiport Memory

➢ Multistage Networks

➢ Blocking versus Non-blocking Networks

❖ Blocking Networks

- A multistage network is called blocking if the simultaneous connections of some multiple input/output pairs may result in conflicts in the use of switches or communication links.
- Example of blocking networks includes Omega, Baseline, Banyan and Delta networks.
- Most of multistage networks are blocking in nature.
- In a blocking networks multiple passes through the network may be needed to achieve certain input-output connections.

❖ Non-Blocking Networks

- A nonblocking multistage network can perform all possible connections between inputs and outputs by rearranging its connections.
- In this networks connections are always be established between any input-output pair.
- Example: Benes network,

Multiprocessor System Interconnects

c) Crossbar Switch and Multiport Memory cond..

➢ Crossbar Networks

- Crossbar Networks is single stage, nonblocking, permutation network
- In Crossbar networks, every input port is connected to a free output through a crosspoint switch without blocking.
- A crossbar network is a single stage network built with unary switches at the crosspoints. These switches can be set open or closed providing point to point connections path between the source and destination.
- Once the data is read from the memory its value is returned to the requesting processor along the same crosspoint switch.
- This type of crossbar networks requires the use of n X m crosspoints switches.
- A square crossbar (n=m) can implement any of the n! permutation without blocking.
- In an n-processor, m-unary switch which can be open or closed, providing a point-to-point connection path between the processor and a memory module.
- All processor can send the memory request independently and asynchronously. This poses the problem of multiple requests denied for the same memory module at the same time. In such a cases, only one of the requests is served at a time.

Multiprocessor System Interconnects

c) Crossbar Switch and Multiport Memory cond..

- **Crossbar Networks**
  - **Crosspoint Switch Design**
    - Out of n crosspoint switches in each column of an n *m crossbar mesh, only one can be connected at a time.
    - Each crosspoint switch requires the use of a large number of connecting lines accommodating address, data path and control signals. That is crosspint has a complexity matching that of a bus of the same width.
    - For n*n crossbar network, this implies that n^2 sets of crosspoints switches and a large number of lines are needed.
    - On each row of the crossbar mesh multiple crosspoints switches can be connected simultaneously. Simultaneous data transfers can take place in a crossbar between n pairs of processors and memories.
    - Crosspoint switches must be designed with extra hardware to handle the potential contention for each memory module. A crossbar switch avoids competition for bandwidth by using O(N) switches to connect N inputs to Noutputs.
    - Although highly non-scalable, crossbar switches are a popular mechanism for connecting a small number of workstations, typically 20 or fewer

Multiprocessor System Interconnects

c) Crossbar Switch and Multiport Memory cond..

➢ Crosspoint Switch Design cond..

- Each processor provides a request line, a read/write line, a set of address lines, and a set of data lines to a crosspoint switch for a single column.
- The crosspoint switch eventually responds with an acknowledgement when the access has been completed.

❖ Disadvantages or limitation of Crossbar

- Crossbar networks are cost effective only for small multiprocessors with a few processors accessing a few memory modules.
- A single stage crossbar network is not expandable once it is built.
- Redundancy at parity check lines can be built into each crosspoint switch to enhance the fault tolerance and reliability of the crossbar network.
- Crossbar switches are expensive and not suitable for systems with many processors or memory modules

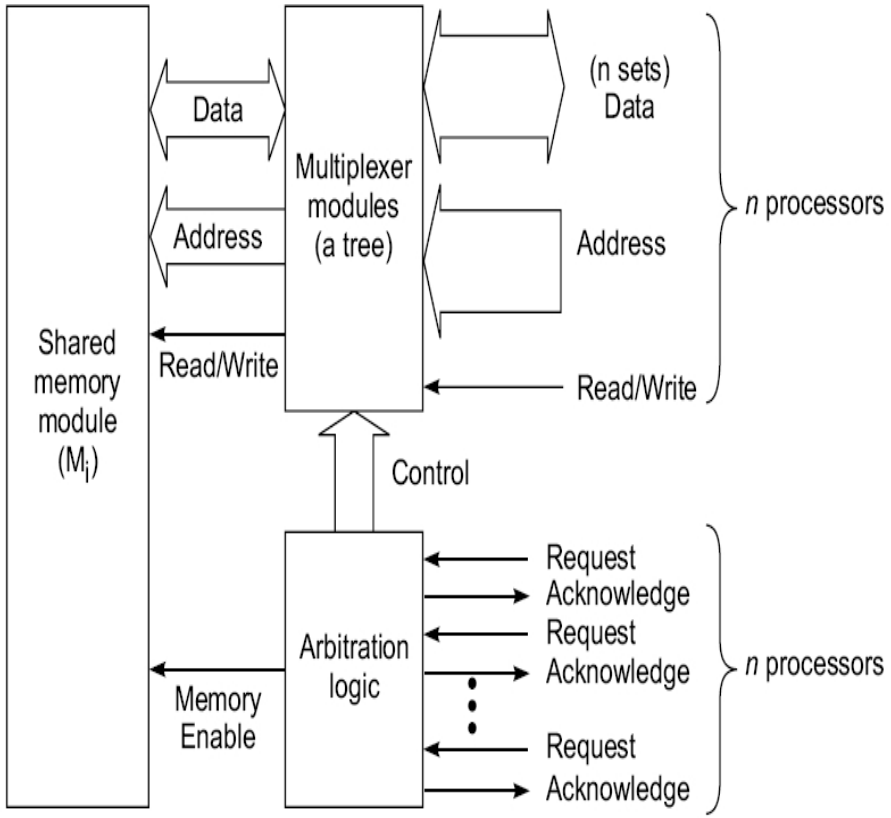


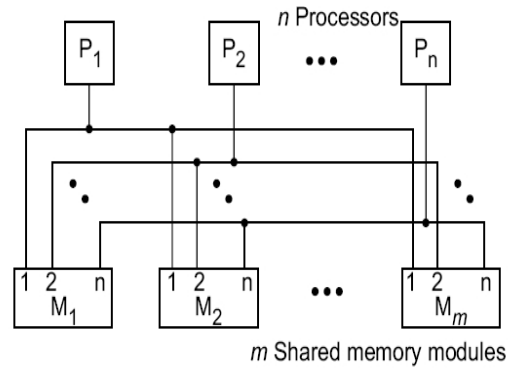
Fig. 7.6 Schematic design of a row of crosspoint switches in a crossbar network

Multiprocessor System Interconnects

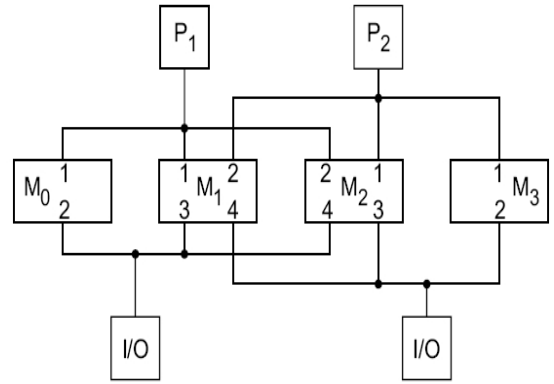
c) Crossbar Switch and Multiport Memory cond.

➢ Multiport Memory

- Since crossbar switches are expensive and not suitable for systen with many processors or memory modules, multiport memory modules may be used instead.
- A multiport memory module has multiple connection points for processors (or I/O devices), and the memory controller in the module handles the arbitration and switching that might otherwise have been accomplished by a crosspoint switch.
- A two function switch can assume only two possible state namely state or exchange states. However a four function switch box can be any of four possible states. A multistage network is capable of connecting any input terminal to any output terminal. Multi-stag networks are basically constructed by so called shuffle- exchange switching element, which is basically a 2 x 2 crossbar. Multiple layers of these elements are connected and form the network.



(a) *n*-port memory modules used



(b) Memory ports prioritized or privileged in each module by numbers

Fig. 7.7 Multiport memory organizations for multiprocessor systems (Courtesy of P. H. Enslow, *ACM Computing Surveys*, March 1977)