Distributed Shared Memory Architecture

DSM Architecture

- Distributed shared memory (DSM) is a form of memory architecture where the (physically separate) memories can be addressed as one (logically shared) address space.

- The term "shared" does not mean that there is a single centralized memory.

- The address space is shared virtually as there is no physical memory shared across the nodes .

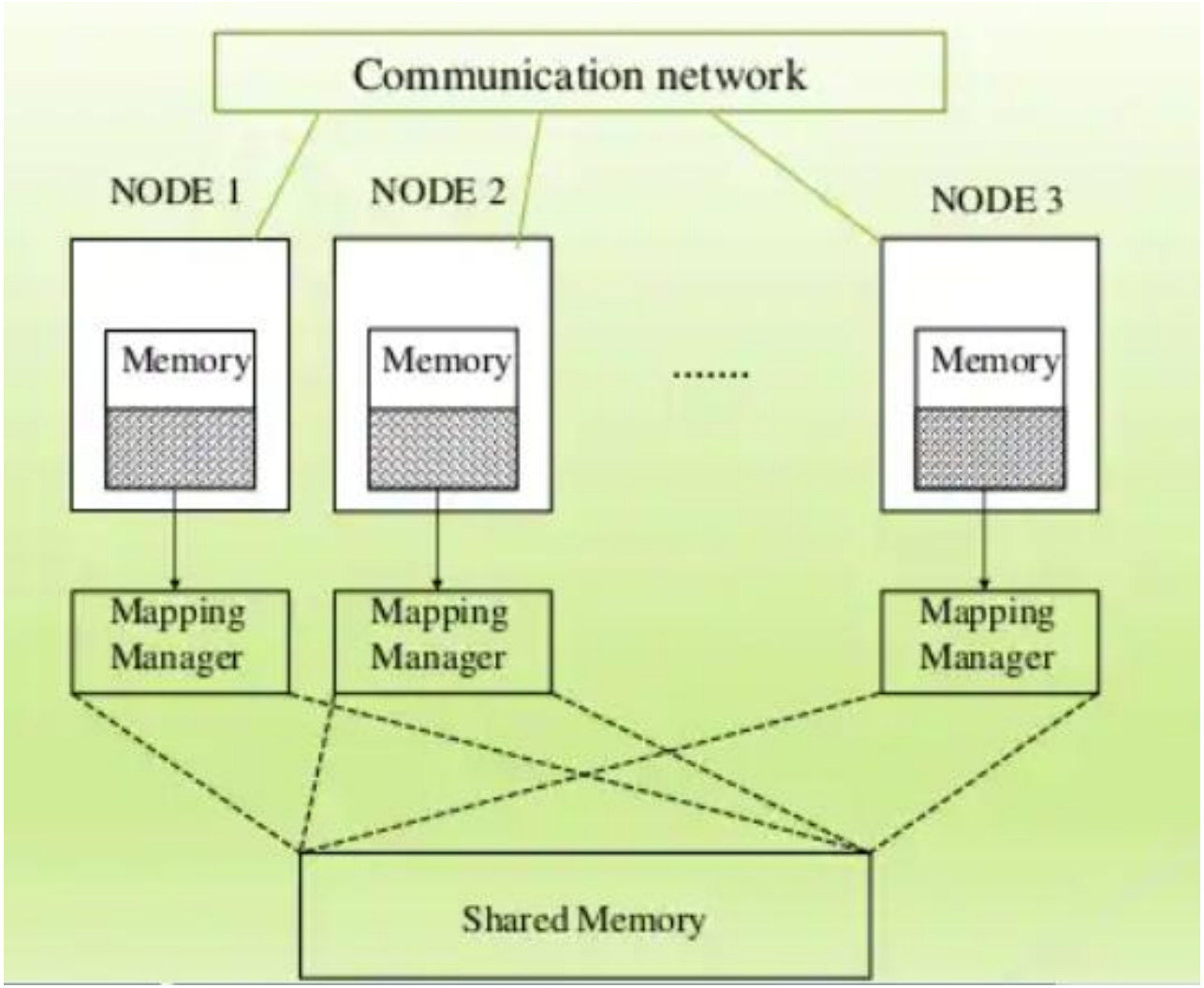
DSM Architecture: Working Principle

- Data moves between
  - main memory and secondary memory (within a node)
  - between main memories of different nodes

- When a process accesses data in the shared address space, the mapping manager maps shared memory address to physical memory (local or remote).

- Each data object is owned by a node :
  - Initial owner is the node that created object.
  - Ownership can change as object moves from node to node

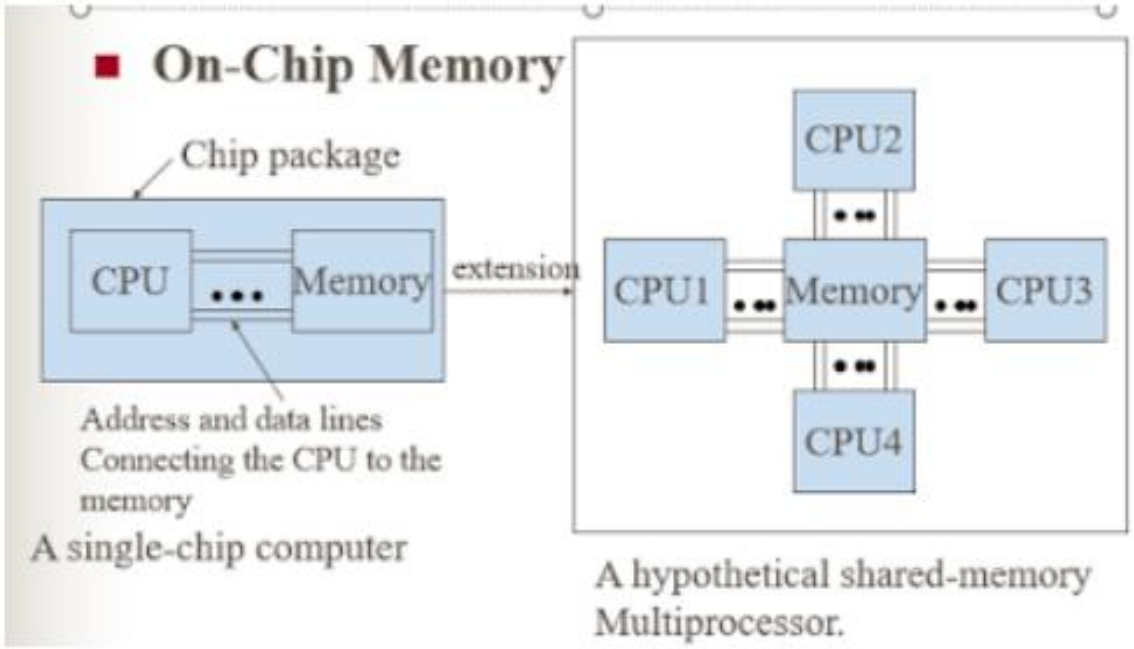
DSM Architecture: Working Principle



DSM Architecture

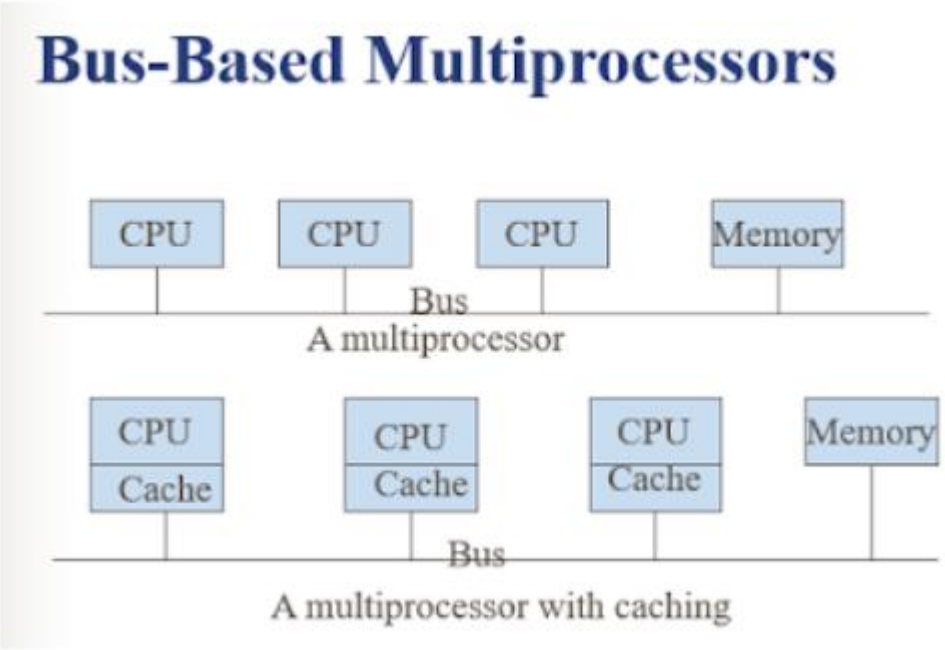
- High speed communication network is used for connecting nodes.
- A simple message passing system allows on different node to exchange message with each other.
- Memory mapping manager routine in each node maps local memory onto the shared virtual memory.
- The mapping manager is layer of software implemented either in the operating kernel or as runtime library routine.
- For mapping operation, the shared memory space is partitioned into blocks.
- The main memory of individual nodes is used to cache the shared memory space.

DSM Architecture



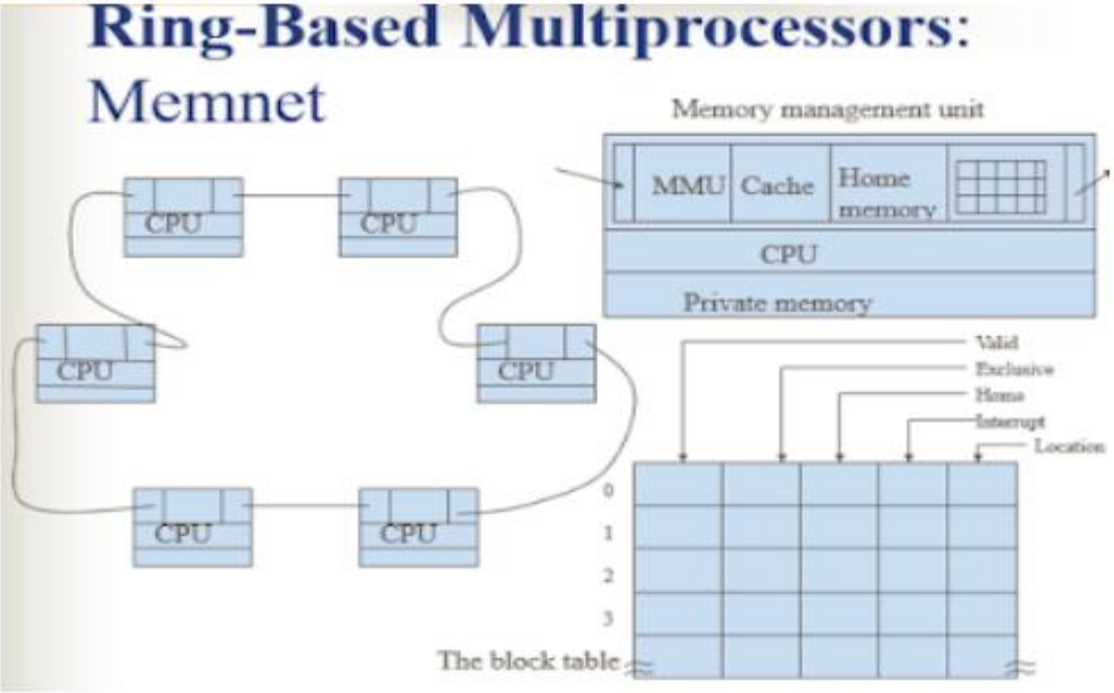
- Such chips are widely used in appliances cars and even toys.
- CPU portion of the chip has data.
- Address lines directly connect to memory portion.
- it is expensive and complicated to construct chip like this.

DSM Architecture



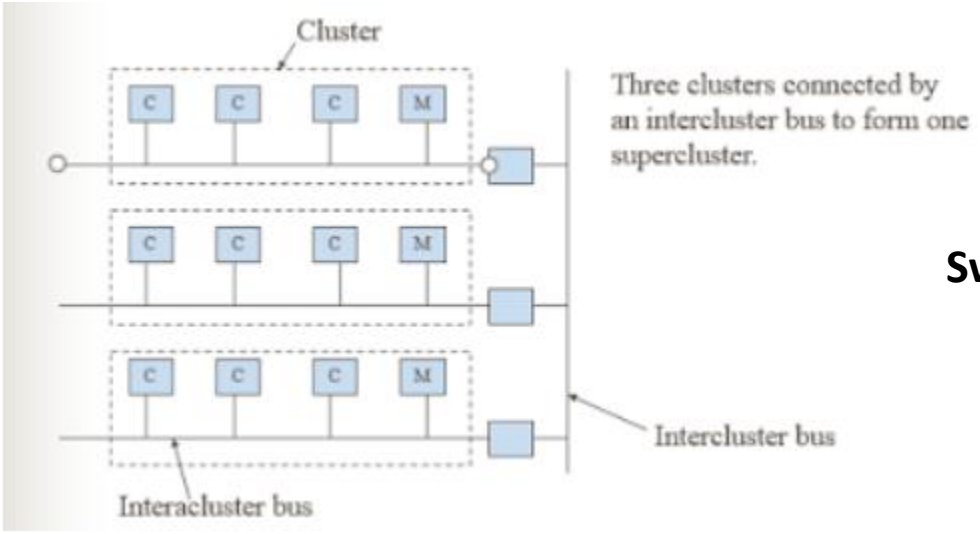
• Connection between CPU and memory is set of parallel wires some holding address of CPU wants to read or write or for sending or receiving messages
• Network traffic is reduced by using caches with each cpu.
•Algorithms are used to prevent two CPU trying to access same memory simultaneously.
•Having single bus makes it overloaded.

DSM Architecture



- A single address line is partitioned into a private area and shared area.
- Private area is divided up into regions so each machine has a piece for its stack
- Shared area is divided into 32 byte blocks.
- All machines are connected via token passing ring. All components are interconnected via Memnet device.
- No centralized global memory.

DSM Architecture

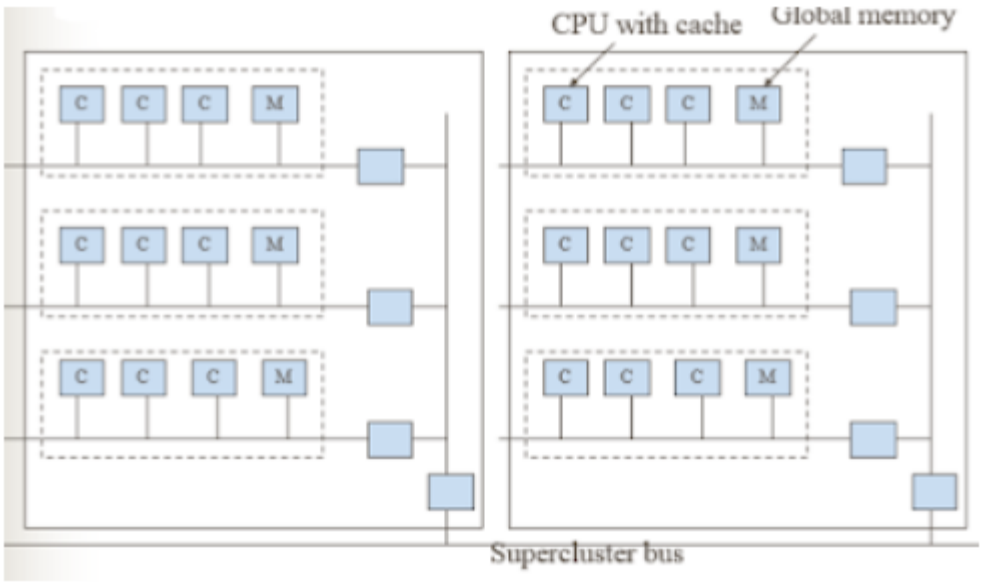


Switched Multiprocessors

Two approaches can be taken to attack the problem of not enough bandwidth.

- Reduce the amount of communication. E.g. Caching.
- Increase the communication capacity. E.g. Changing topology.
- One method is to build the system as a hierarchy. Build the system as multiple clusters and connect the clusters using an intercluster bus. As long as most CPUs communicate primarily within their own cluster, there will be relatively little intercluster traffic. If still more bandwidth is needed, collect a bus, tree, or grid of clusters together into a supercluster, and break the system into multiple superclusters.

DSM Architecture



Switched Multiprocessors

One method is to build the system as a hierarchy. Build the system as multiple clusters and connect the clusters using an intercluster bus. As long as most CPUs communicate primarily within their own cluster, there will be relatively little intercluster traffic. If still more bandwidth is needed, collect a bus, tree, or grid of clusters together into a supercluster, and break the system into multiple superclusters.