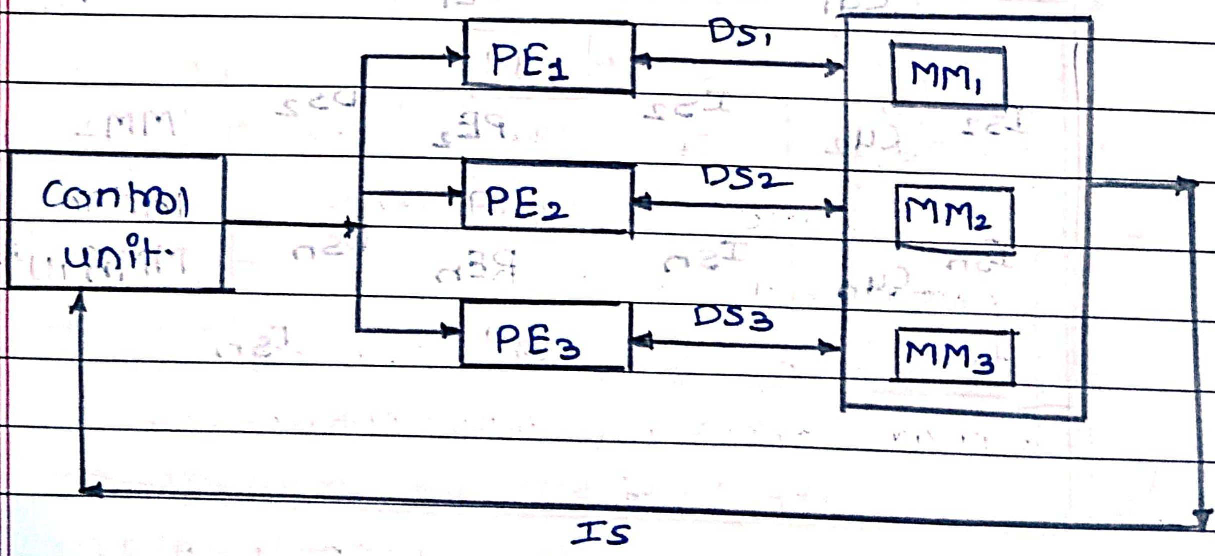
Assignment No. 01.

Q.1. Explain with suitable diagrams SIMD, MIMD Architecture?

→ SIMD (Single Instruction Stream, Multiple Data Stream).

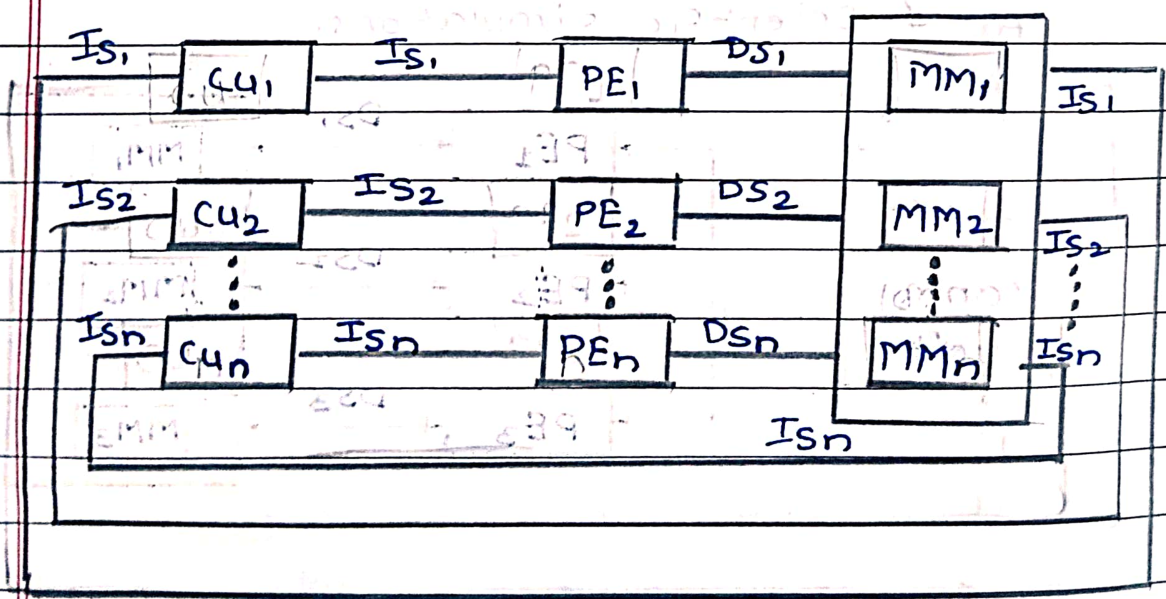
- In this case a single control unit that dispatches the same instruction to multiple processing elements, but work on different data.
- This kind of system is mainly used when many data (array of data) have to be operated with same operation.
- Vector processors & array processors fall into this category.
- Used in application like graphics processing AI & scientific simulations.



BE-B-14311-Mayuresh Karnavat

* MIMD (Multiple Instruction Stream, Multiple Data Stream)-

- Multiple processors execute different instructions on different Data independently.
- This is a complete parallel processing examples
- Here each processor has its own control unit; each processing element can execute different instructions on set of different data.
- Examples of this kind of systems are SMPs (symmetric Multiprocessors), clusters & NUMA (Non-uniform memory Access).



BE-B-14211 - Mayuresh Karnavat

Q.2. Explain the impact of memory Latency & memory Bandwidth on system performance

→ Memory Performance plays a crucial role in determining the overall speed & efficiency of a computer system. Two imp factors affecting memory performance are memory latency & memory Bandwidth.

1. Memory Latency:

refer to the time delay between a processor requesting data from memory & the data being available for use. It is measured in nanoseconds (ns) & impacts systems performance significantly.

Impact on System Performance:

- High latency slows down the CPU, as it must wait for data to arrive before processing.
- Increased ~~Rdctas~~ idle cycle, reducing overall system efficiency.
- Affects real-time application, such as gaming & financial trading.

2. Memory Bandwidth: is the rate of which data can be read from or written to memory by the CPU. It is measured in gigabytes per second (GB/s) & depends on factors such as memory bus width & clock speed.

Impact on system performance:

classmate

Date
Page

BE-B-14211 - Mayuresh Karnavat

- Higher bandwidth improves data transfer rates, reducing CPU wasting time.
- Essential for high-Performance Computing gaming & AI workloads.
- Low bandwidth creates bottlenecks, limiting the system's ability to process large datasets.

Q.3. Explain Message Passing Costs in Parallel Computers in Parallel machines.

→ Platforms To Transfer a message over a network, total time required is calculated based on the following factors:

(i) Startup time (t_s) : Time spent at Sending & receiving nodes.

(ii) per-hop time (t_h) : It is a function of number of hopes & includes factors Such as network delays, switch latency, etc.

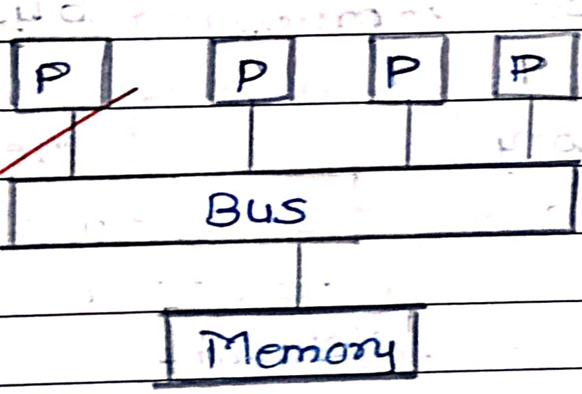
(iii) Per word transfer time (t_w) : It includes all overhead that are determined by the length of the message This includes error checking & correction, bandwidth, etc

Q.4. Describe uniform-memory-Access & Non-uniform-memory Access with diagrammatic representation.

BE-B-14211- mayuresh Karnavar

→ Uniform Memory Access (UMA):

- All processors share a unique centralized primary memory, so each CPU has the same memory access time
- Each processor gets equal priority to access the main memory of the machine.
- These systems are also called symmetric shared-memory multiprocessors (SMP.
- used in shared memory multiprocessors.



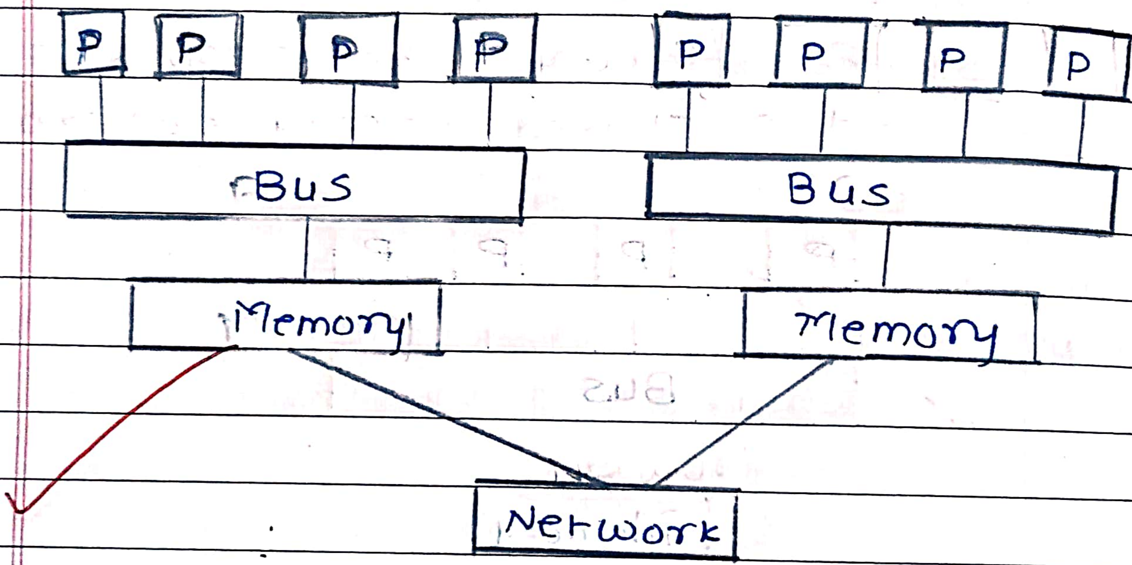
• Non-uniform memory Access (NUMA):

In each NUMA shared memory architecture each processor has its own local memory module that it can access directly.

- At the same time, it can also access any memory module belonging to another processor using shared bus or any type of inter connect.
- These systems have a shared logical address space, but physical memory is distributed among CPUs, so that access time to data depends on data position.

BE-B-K211-mayuresh Kornavak

- All processor has direct path to the block of its local memory attached to it.
- All processor can see all memory.
- Access to the memory of other processors is slower.



Q.5 Describe the scope of parallel computing Give applications of parallel computing.

→ Scope of parallel computing:

1) Faster processing & Efficiency.
- Split large tasks into smaller ones to be processed simultaneously.
- Reduces execution time significantly compared to sequential computing.

2) Scalability & High performance Computing.
- Parallel computing allows systems to scale up by adding more processors.

BE-B-14211 - Mayuresh Karnavar

3) Large scale data processing:
- Essential for Big data analysis where massive datasets requires fast processing.

4) Artificial Intelligence & Machine Learning.
AI models, such as deep Learning, requires immense computational power

5) Real time systems & simulations.
- used in real-time weather forecasting financial modeling & scientific simulations.

* Applications of Parallel computing.
1) Scientific application
2) Commercial application
3) Applications in Engineering & Design.
4) Application in computer system.
5) Weather-climate Research.
6) Application of parallel processing in Medical.

Q.6. Write short note on.
a. Data flow model
b. Demand Driven computation
c. cache memory.

classmate

Date ____
Page ____

BE - B - 14211 - Mayuresh Karnavat

a) Data flow model:

is a computational model where the execution of operations depends on the availability of data, rather than a predefined sequence of instructions.

- In this models. It defines the function of the internal processes in the system with the aid of DFDs (Data Flow Diagrams)
- Functional Modeling is represented through a hierarchy of DFDs.
- The DFD is a graphical representation of a systems that shows the inputs to the system, the processing upon the inputs, the outputs of the system as well as the internal data stores.
- DFD illustrate the series of transformation or computations performed on the object or the systems, & the external controls & object that affect the transformation
- The four main parts of a DFD are -

1. Processes.
2. Dataflows.
3. Actors.
4. Data stores.

- The other parts of a DFD are constraints & ~~[illegible]~~ control flows.

classmate

Date
Page

BE-B-14211 - Mayuresh Karnavat

b. DEMAND-DRIVEN Computation:

- represents a general framework for deriving demand-driven algorithms for interprocedural data flow analysis of imperative programs.
- The goal of demand-driven analysis is to reduce the time and/or space overhead of conventional exhaustive analysis by avoiding the collection of information that is not needed.
- The derived demand-driven algorithm find responses to these queries through a partial reversal of the respective data flow analysis.
- Depending on whether minimizing time or space is of primary concern, result caching may be incorporated in the derived algorithm
- Framework is applicable to interprocedural data flow problems with a finite domain set.

c) Cache memory:

- cache memory is a high-speed storage located between the CPU & main-memory
- stores frequently accessed data & instruction

35
25, 09:37

BE - B - 14211 - mayuresh - Karnavar

reducing memory latency.

- speeds up CPU processing by minimizing delays in fetch data.
- organized into L1 (fastest, small), L2 & L3 (slower, Larger) cache levels.
- It is a faster & smaller ~~or~~ segment of memory whose access time is as close registers.
- Data in primary memory can be accessed faster than secondary memory but still access times of primary memory are generally in a few microseconds, whereas the CPU is capable of ~~[illegible]~~ Performing operations in nanoseconds.