Semnan University

بسم الله الرحمن الرحیم

Semnan University

Advanced Computer Architecture

Parallel processing

Teacher

Fatemeh Daraee

f_daraei@semnan.ac.ir

https://fdaraei.profile.semnan.ac.ir

An Overview of Parallel Processing

Semnan University

- **What is parallel processing?**
  - Parallel processing is a method to improve computer system performance by executing two or more instructions simultaneously.
- **The goals of parallel processing:**
  - One goal is to reduce the "wall-clock" time or the amount of real time that you need to wait for a problem to be solved.
  - Another goal is to solve bigger problems that might not fit in the limited memory of a single CPU.

Parallelism in a Uniprocessor System

- A reconfigurable arithmetic pipeline is an example of parallelism in a uniprocessor system.
- Each stage of a reconfigurable arithmetic pipeline has a multiplexer at its input. The multiplexer may pass input data, or the data output from other stages, to the stage inputs. The control unit of the CPU sets the select signals of the multiplexer to control the flow of data, thus configuring the pipeline

Vector Arithmetic Unit

A vector arithmetic unit contains multiple functional units that perform addition, subtraction, and other functions. The control unit routes input values to the different functional units to allow the CPU to execute multiple instructions simultaneously.

- For the operations **A ← B + C** and **D ← E - F**, the CPU would route B and C to an adder and then route E and F to a subtractor for simultaneous execution.

Organization of Multiprocessor Systems

- **Flynn's Classification**
  - Was proposed by researcher Michael J. Flynn in 1966.
  - It is the most commonly accepted taxonomy of computer organization.
  - In this classification, computers are classified by whether it processes a single instruction at a time or multiple instructions simultaneously, and whether it operates on one or multiple data sets

Semnan University

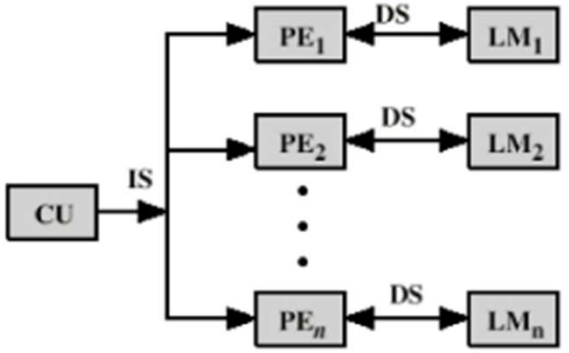
Single Instruction, Single Data (SISD)

- SISD machines execute a single instruction on individual data values using a single processor.
- Based on traditional Von Neumann uniprocessor architecture, instructions are executed sequentially or serially, one step after the next.
- Until most recently, most computers are of SISD type.

Semnan University

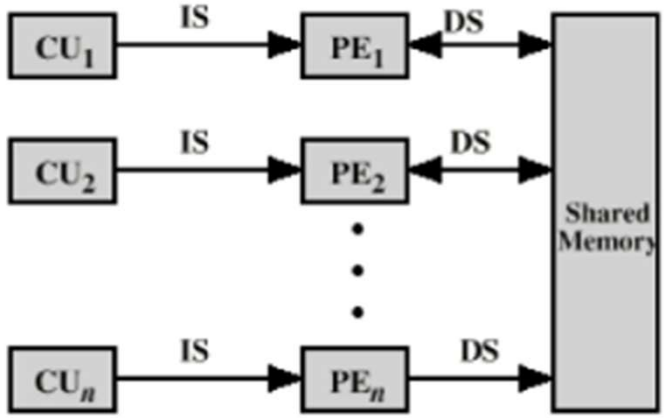
Single Instruction, Multiple Data (SIMD)

- An SIMD machine executes a single instruction on multiple data values simultaneously using many processors.
- Since there is only one instruction, each processor does not have to fetch and decode each instruction. Instead, a single control unit does the fetch and decoding for all processors.
- SIMD architectures include array processors.



Multiple Instruction, Multiple Data (MIMD)

- MIMD machines are usually referred to as multiprocessors or multicomputers.
- It may execute multiple instructions simultaneously, contrary to SIMD machines.
- Each processor must include its own control unit that will assign to the processors parts of a task or a separate task.
- It has two subclasses: Shared memory and distributed memory.



Multiple Instruction, Single Data (MISD)

- This category does not actually exist. This category was included in the taxonomy for the sake of completeness.

System Topologies

- A system may also be classified by its topology.
- A topology is the pattern of connections between processors.
- The cost-performance tradeoff determines which topologies to use for a multiprocessor system.

Topology Classification

- A topology is characterized by its diameter, total bandwidth, and bisection bandwidth
  - Diameter – the maximum distance between two processors in the computer system.
  - Total bandwidth – the capacity of a communications link multiplied by the number of such links in the system.
  - Bisection bandwidth – represents the maximum data transfer that could occur at the bottleneck in the topology.

Uniform memory access (UMA)

- The UMA is a type of symmetric multiprocessor, or SMP, that has two or more processors that perform symmetric functions. UMA gives all CPUs equal (uniform) access to all memory locations in shared memory.
- They interact with shared memory by some communication mechanism like a simple bus or a complex multistage interconnection network.

Nonuniform memory access (NUMA)

- NUMA architectures, unlike UMA architectures do not allow uniform access to all shared memory locations. This architecture still allows all processors to access all shared memory locations but in a nonuniform way, each processor can access its local shared memory more quickly than the other memory modules not next to it.