# UNIT 2

## PROCESSOR ARCHITECTURE

# COURSE CONTAINS

2.1 Uniprocessor Architecture

2.2 CISC and RISC Architecture

2.3 Parallel processing mechanism for uniprocessor

- 2.3.1 Parallelism and pipelining within CPU

2.4 Multiprocessor and Multicomputer Model

- 2.4.1 UMA Model
- 2.4.2 NUMA Model
- 2.4.3 COMA Model
- 2.4.4 NORMA Model

2.5 Flynn's Taxonomy

2.6 Feng's classification

2.7 Distributed Memory Multicomputer

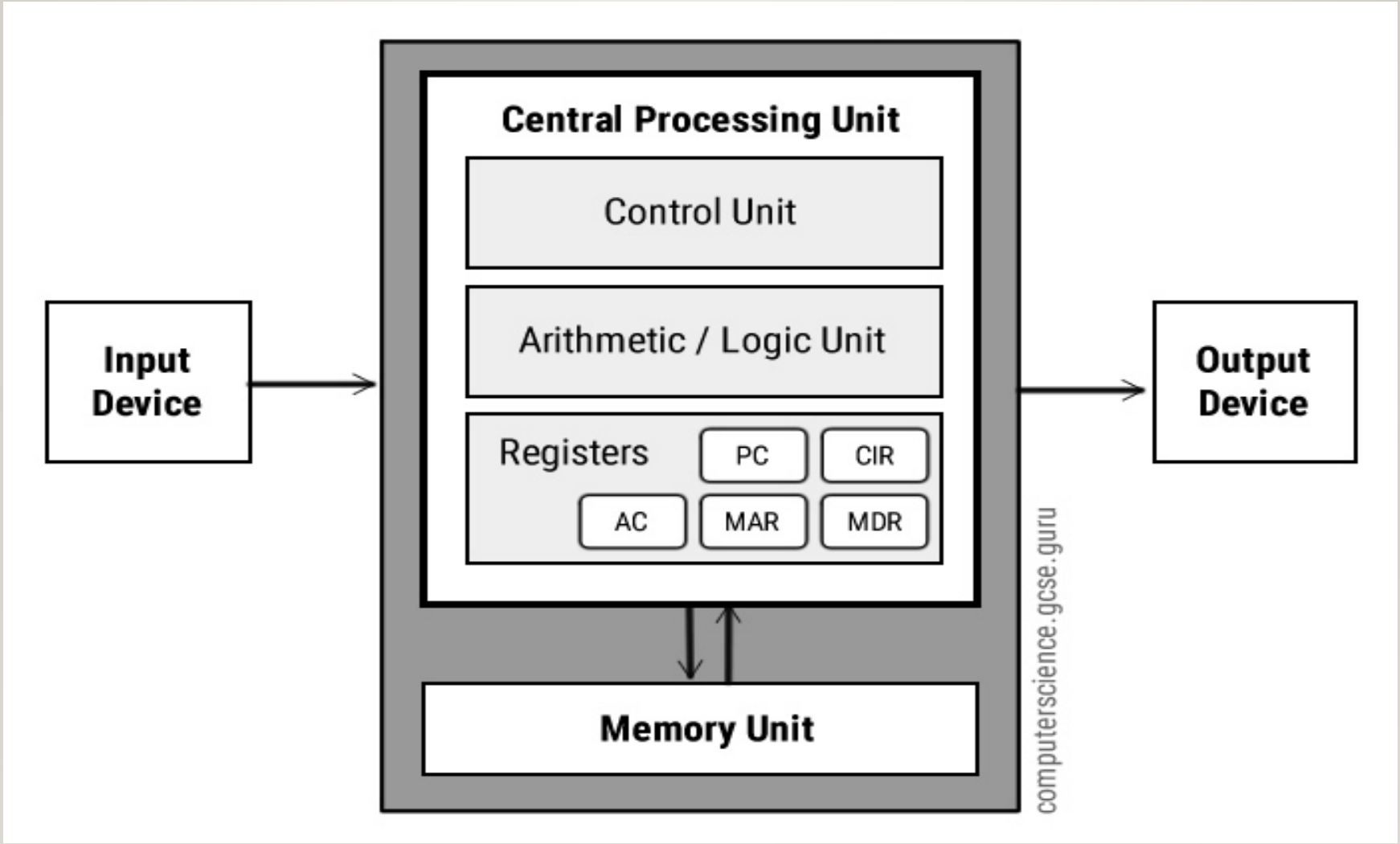
2.8 Shared Memory Multiprocessor

# 2.1 UNIPROCESSOR ARCHITECTURE

A **uniprocessor** (or a **uniprocessor system**) is defined as a computer system that has a single central processing unit that is used to execute computer tasks.

**Von Neumann Architecture**

- Von Neumann Architecture was first published by John Von Neumann in 1945.
- Von Neumann architecture is based on the stored-program computer concept, where instruction data and program data are stored in the same memory. This design is still used in most computers produced today.
- It consists of a Control Unit, Arthimetic and Logic Unit(ALU),Memory Unit, Registers and Input/Output Devices.


Central Processing Unit
Control Unit
Arithmetic / Logic Unit
Registers
PC
CIR
AC
MAR
MDR
Input Device
Output Device
Memory Unit
computerscience.gcse.guru

**1.Central Processing Unit(CPU)**

- The CPU is the electronic circuit responsible for executing the instructions of a computer program.
- The CPU contains the ALU,CU and variety of registers.

**a. Arithmetic and Logic Unit(ALU):** The ALU allows arithmetic (add, subtract etc) and logic (AND, OR, NOT etc) operations to be carried out.

**b. Control Unit(CU):**The control unit controls the operation of the computer's ALU, memory and input/output devices, telling them how to respond to the program instructions it has just read and interpreted from the memory unit. CU also provides the timing and control signals required by other computer components.

**c. Registers**: Registers are high speed storage areas in the CPU. All data must be stored in a register before it can be processed. Registers contains

i. MAR(Memory Address Register): Holds the memory location of data that needs to be accessed.

ii. MDR(Memory Data Register): Holds data that is being transferred to or from memory

iii. AC(Accumulator): Where intermediate arithmetic and logic results are stored

iv. PC(Program Counter): Contains the address of the next instruction to be executed

v. CIR(Current Instruction Register): Contains the current instruction during processing

**Buses:** Buses are the means by which data is transmitted from one part of a computer to another, connecting all major internal components to the CPU and memory. A standard CPU system bus is comprised of a control bus, data bus and address bus.

i. Address Bus: Carries the addresses of data (but not the data) between the processor and memory

ii. Data Bus: Carries data between the processor, the memory unit and the input/output devices

iii. Control Bus: Carries control signals/commands from the CPU (and status signals from other devices) in order to control and coordinate all the activities within the computer

**2. Memory Unit**

- The memory unit consists of RAM, sometimes referred to as primary or main memory.
- Main memory is fast and also directly accessible by the CPU. Main memory stores both data and instruction.

**3.Input/Output Devices:**

- It is operated by the control unit.
- Program or data is read into main memory from the *input device* or secondary storage under the control of CPU input instruction.
- *Output devices* are used to output the information from a computer. If some results are evaluated by computer and it is stored in the computer, then with the help of output devices, we can present it to the user.

# VON NEUMANN ARCHITECTURE VS HARVARD ARCHITECTURE

## VON NEUMANN ARCHITECTURE

1. This has a single common memory space where both program instructions and data are stored.
2. There is a Single data bus which fetches data and instructions
3. Execution time taken by the CPU to fetch a program is high.
4. The advantage is its simplicity and economy.

## HARVARD ARCHITECTURE

1. Harvard architecture computers have separate memory areas for program instructions and data.
2. There are two or more internal data buses which allow simultaneous access to both instructions and data.
3. Time for execution of program is low.
4. Its high-speed execution time is at the cost of more hardware complexity.

# 2.2 CISC AND RISC ARCHITECTURE

## CISC (Complex Instruction Set Computer) Architecture

- Complex instruction set computing is a CPU design where single instructions can execute multiple low-level operations(such as a load from memory, perform  an arithmetic operation, and store the result in memory ).
- Called “complex” because of the complex work performed per instruction.
- Example : add X and Y and put the result in Z(for X,Y,Z memory address)
- A CISC instruction set typically includes many instructions with different sizes and execution cycles , which makes CISC instructions harder to pipeline.

**The basic Characteristics of CISC are:**

1. A large number of instructions –typically from 100 to 250 instructions.
2. Some instructions that perform specialized tasks and are used infrequently.
3. A large variety of addressing modes-typically 5 to 20 different modes.
4. Variable-length instruction formats.

**Disadvantage**

- CISC include a complex instruction decoding scheme, an increased size of the control unit, and increased logic delays.
- Some instruction take more time then other.

# RISC (REDUCED INSTRUCTION SET COMPUTER) ARCHITECTURE

- RISC is a type of microprocessor architecture that is designed to perform smaller number of types of computer instruction so that it can operate at a higher speed.
- It utilizes a small, highly-optimized set of instructions, rather than a more specialized set of instruction often found in other types of architectures.
- Called “reduced” because of the reduction of work performed by an instructions.
- Break operation into simpler sub operation. Example : instruction: load X , load Y , add X and Y , store Z.
- Computers typically use CISC while tablets , smartphones and other devices use RISC.

**The basic characteristics of RISC are:**

1. RISC simplifies the instruction set and  has less number of instruction.
2. The use of only a few addressing modes.
3. By using a relatively simple instruction format, the instruction length can be fixed.
4. Shorter instructions – Breaking the complex instruction into several short simpler instructions and can run several instruction simultaneously.
5. It has the ability to execute one instruction per clock cycle.
   - This is done by overlapping the fetch, decode and execute phases of two or three instructions by using a procedure referred to as **pipelining.**

# RISC 5 STAGE PIPELINING

- Instruction Fetch(IF)

  Get instruction from memory

- Instruction Decode (ID)

  Translate opcode into control signals and read registers

- Execute(EX)

  Perform ALU operation

- Memory(MEM)

  Access memory if needed

- Write Back(WB)

  Update register file

| Instr. No. | Pipeline Stage | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | IF | ID | EX | MEM | WB | | |
| 2 | | IF | ID | EX | MEM | WB | |
| 3 | | | IF | ID | EX | MEM | WB |
| 4 | | | | IF | ID | EX | MEM |
| 5 | | | | | IF | ID | EX |
| Clock Cycle | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

## Advantage of RISC processor over CISC

- **Speed:** perform smaller number of set of instruction so that it can operate at a higher speed.
- **Better performance** : The performance of RISC processors is often two to four times than that of CISC processors because of simplified instruction set.
- **Simpler hardware:** Because the instruction set of a RISC processor is so simple , it uses up much less chip space and simple hardware requirements.
- **Shorter design cycle :** RISC processors can be designed more quickly than CISC processors due to its simple architecture and also can complete there work in 1 clock cycle.

# CISC VS RISC ARCHITECTURE

## CISC

- Complex Instruction Set Computer
- CISC has large number of instruction
- Variable length instruction format.
- Large number of addressing modes.
- Slower and more powerful.
- Instruction takes more than one clock cycle.

## RISC

- Reduced Instruction Set Computer
- RISC has less number of instruction
- RISC has fixed length instruction format.
- Few number of addressing modes.
- Faster and less powerful.
- An instruction execute in a single clock cycle

# CISC VS RISC ARCHITECTURE

## CISC

- CISC is high in Code size .
- Cost is high.
- Requires less number of registers
- CISC uses memory to memory operation
- CISC does not support pipelining.
- CISC uses micro-programming

## RISC

- RISC is less in Code size .
- Cost is less
- Requires more number of registers
- RISC uses register to register operation.
- Pipelining is the major feature of RISC.
- RISC uses hardwired programming

# 2.3 PARALLEL PROCESSING MECHANISM FOR UNIPROCESSOR

Parallelism can be achieved in Uni-Processor systems using two main methods:

1. Hardware method
2. Software method

**1. Hardware method of parallelism**

a. Multiplicity of functional units

b. Parallelism and pipelining within the CPU

c. Overlapped CPU and I/O operations

d. Use of a hierarchical memory system

e. Balancing of subsystem Bandwidth

2. **Software method of parallelism**

a. Multiprogramming

b. Time sharing

# HARDWARE METHOD OF PARALLELISM

**a. Multiplicity of functional units**

- Many of the functions of the ALU can be distributed to multiple and specialized Functional Units which can operate in parallel. Functional Units may be ADD,DIVIDE,MULTIPLY,BOOLEAN.
- Example: The CDC-6600 has 10 functional units built into its CPU. These 10 units are independent of each other and may operate simultaneously.

## b. Parallelism and Pipelining within the CPU

- In addition with multiple functional units, to explore parallelism now almost all CPUs make use of **parallel adders** like carry-look ahead and carry –save adders. This is in contrast to the bit serial adders used in the first generation machines.
- **High speed multiplier recoding and convergence division** are the techniques for exploring parallelism and the sharing of hardware resources for the functions of multiply and divide.
- The use of multiple function units is a form of parallelism within the CPU.
- Various phases of instructions executions such as instruction fetch, instruction decode, instruction execution and result storage can be **pipelined.**

- Instruction pipelining where the execution of multiple instructions can be partially overlapped.

**Basic five-stage pipeline**

| Instr. No. \ Clock cycle | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | IF | ID | EX | MEM | WB | | |
| 2 | | IF | ID | EX | MEM | WB | |
| 3 | | | IF | ID | EX | MEM | WB |
| 4 | | | | IF | ID | EX | MEM |
| 5 | | | | | IF | ID | EX |

(IF = Instruction Fetch, ID = Instruction Decode, EX = Execute, MEM = Memory access, WB = Register write back).

In the fourth clock cycle (the green column), the earliest instruction is in MEM stage, and the latest instruction has not yet entered the pipeline.

# PIPELINING CONCEPT

- Pipelining is an implementation technique where multiple expression are **overlap** in execution. Pipelining is used to improve processing time.

Example : Performing laundary works

W→Wash→30 minutes

D→Dry→40 minutes

F→Fold→40 minutes

- Applying pipelining approach

| | 30 | 40 | 40 | 30 | 40 | 40 |
|---|---|---|---|---|---|---|
| A | W | D | F | | | |
| B | | | | W | D | F |
| C | | | | | | |
| D | | | | | | |

| | 30 | 40 | 40 | 30 | 40 | 40 |
|---|---|---|---|---|---|---|
| A | W | D | F | | | |
| B | | W | D | F | | |
| C | | | W | D | F | |
| D | | | | W | D | F |

## c. Overlapped CPU and I/O operations

- I/O operations can be performed simultaneously with the CPU computations by using separate I/O controllers, channels, or I/O processors.
- The direct memory access (DMA) channel can be used to provide direct information transfer between the I/O devices and the main memory .

**d. Use of a hierarchical memory system**

- The CPU is 1000 times faster than memory access. A hierarchical memory system can be used to close up the speed gap between CPU and Memory. The hierarchical order listed is

- Registers
- Cache
- Main Memory
- Magnetic Disk
- Magnetic Tape

- This hierarchy is used to broaden the memory bandwidth of CPU.
- The inner most level is the register files directly addressable by ALU.
- Cache memory can be used to serve as a buffer between the CPU and the main memory. Virtual memory space can be established with the use of disks and tapes at the outer levels.

## e. Balancing of subsystem Bandwidth

- CPU is the fastest unit in computer.
- The **bandwidth** of a system is defined as the number of operations performed per unit time. In case of main memory the memory bandwidth is measured by the number of words that can be accessed per unit time.
- **Bandwidth balancing between CPU and Memory**

The speed gap between CPU and memory can be closed-up by using fast **cache memory** between them.

- **Bandwidth balancing between Memory and I/O Devices**

To balance the speed gap of the I/O devices and memory, I/O channels with different speeds are used between the slow I/O devices and main memory.

# SOFTWARE METHOD OF PARALLELISM

**a. Multiprogramming**

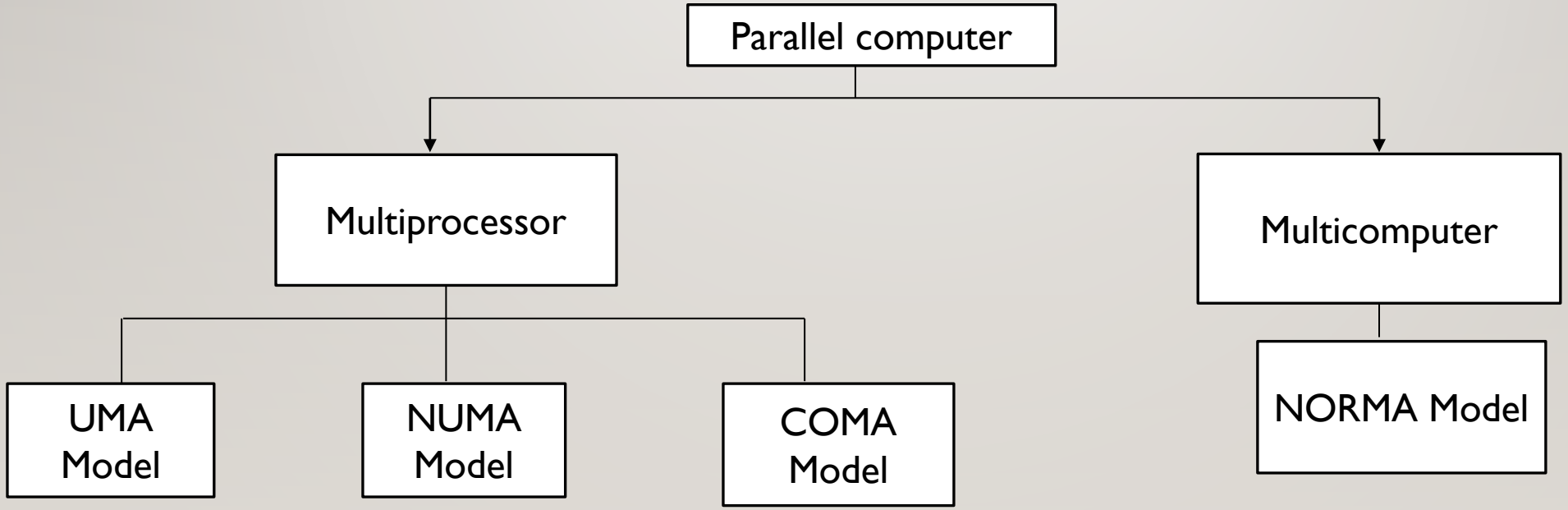
- As we know that some computer programs are CPU bound and some are I/O bound.
- Whenever a process P1 is tied up with I/O operations. The system schedular can switch the CPU to process P2.This allows simultaneous execution of several programs in the system. This interleaving of CPU and I/O operations among several programs is called **multiprogramming**, so the total execution time is reduced.
- Here processes do not have to wait for each other to complete and hence can execute simultaneously(parallel).

## b. Time sharing

- Multiprogramming on a uniprocessor is centered around the sharing of the CPU by many programs. Sometimes high priority programs may occupy the CPU for long time and other programs are put up in queue.
- This problem can be overcome by a concept called as Time sharing in which every process is allotted a time slice of CPU time and thereafter after its respective time slice is over CPU is allotted to the next program if the process is not completed it will be in queue waiting for the second chance to receive the CPU time.

# 2.4 MULTIPROCESSOR AND MULTICOMPUTER MODEL

- Multiprocessor and multicomputer are types of parallel computer.



**1. Multiprocessor or Shared memory multiprocessor**

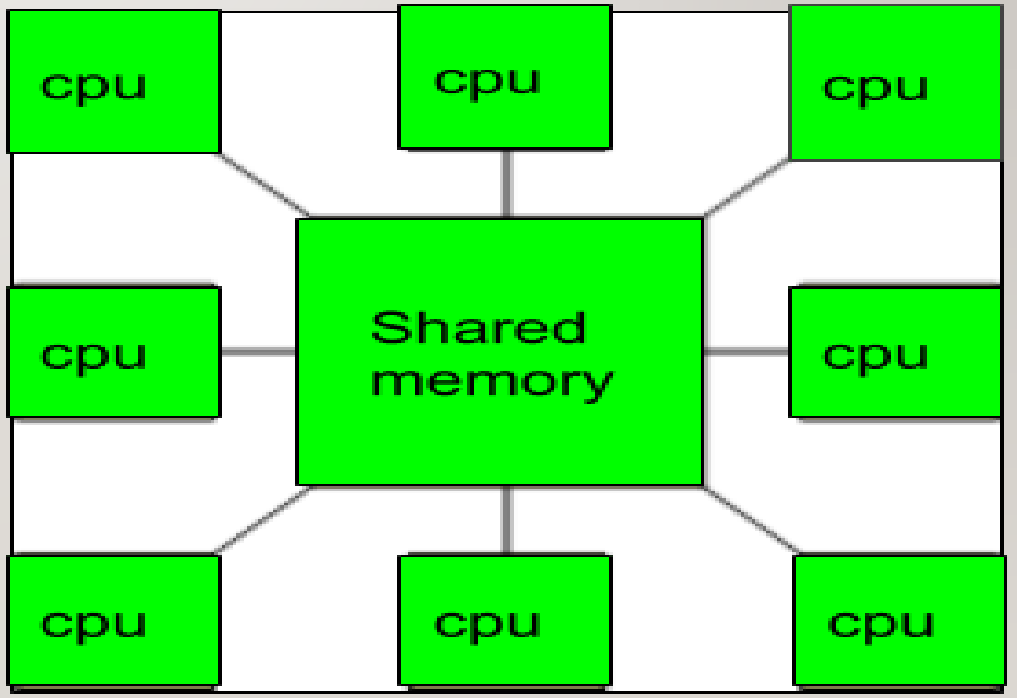
A Multiprocessor is a computer system with two or more CPUs or processor , all the CPUs shares the common memory also called **shared memory multiprocessor.**

**2. Multicomputer or distributed memory multicomputer**

Multicomputer is a computer made up of several computers , in which each processor has its own memory. No shared memory occurs also known as **distributed memory multicomputer.**

# MULTIPROCESSOR

- A Multiprocessor is a computer system with two or more CPUs or processor that is capable of performing multiple tasks simultaneously.
- All the CPUs shares the common memory also called **shared memory multiprocessor.**
- The main objective of using a multiprocessor is to boost the system's execution speed and enhance performance.
- Types multiprocessor(UMA,NUMA,COMA)

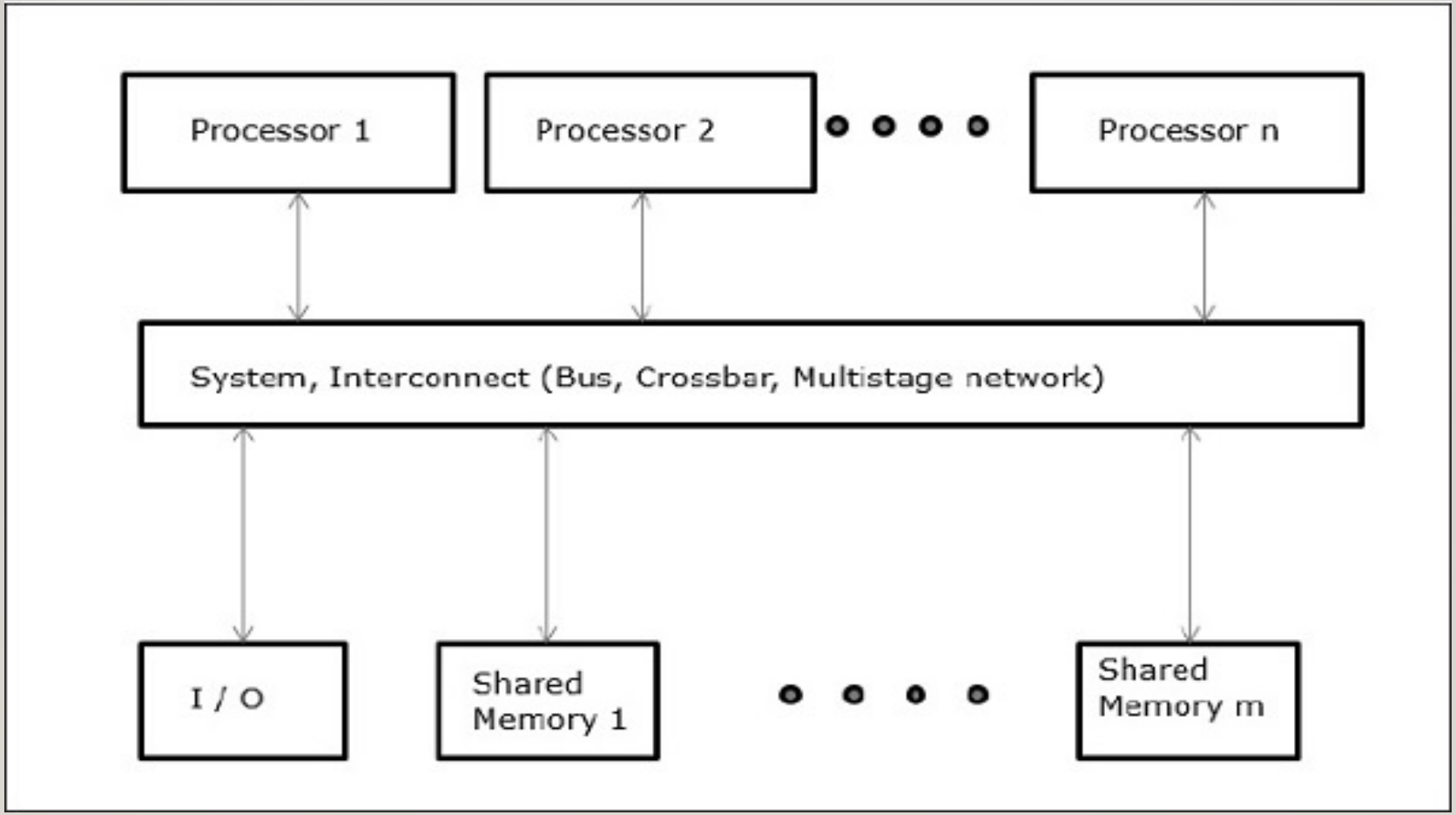


**Types of multiprocessor or three most common shared memory multiprocessors models are-**

1. UMA Model
2. NUMA Model
3. COMA Model

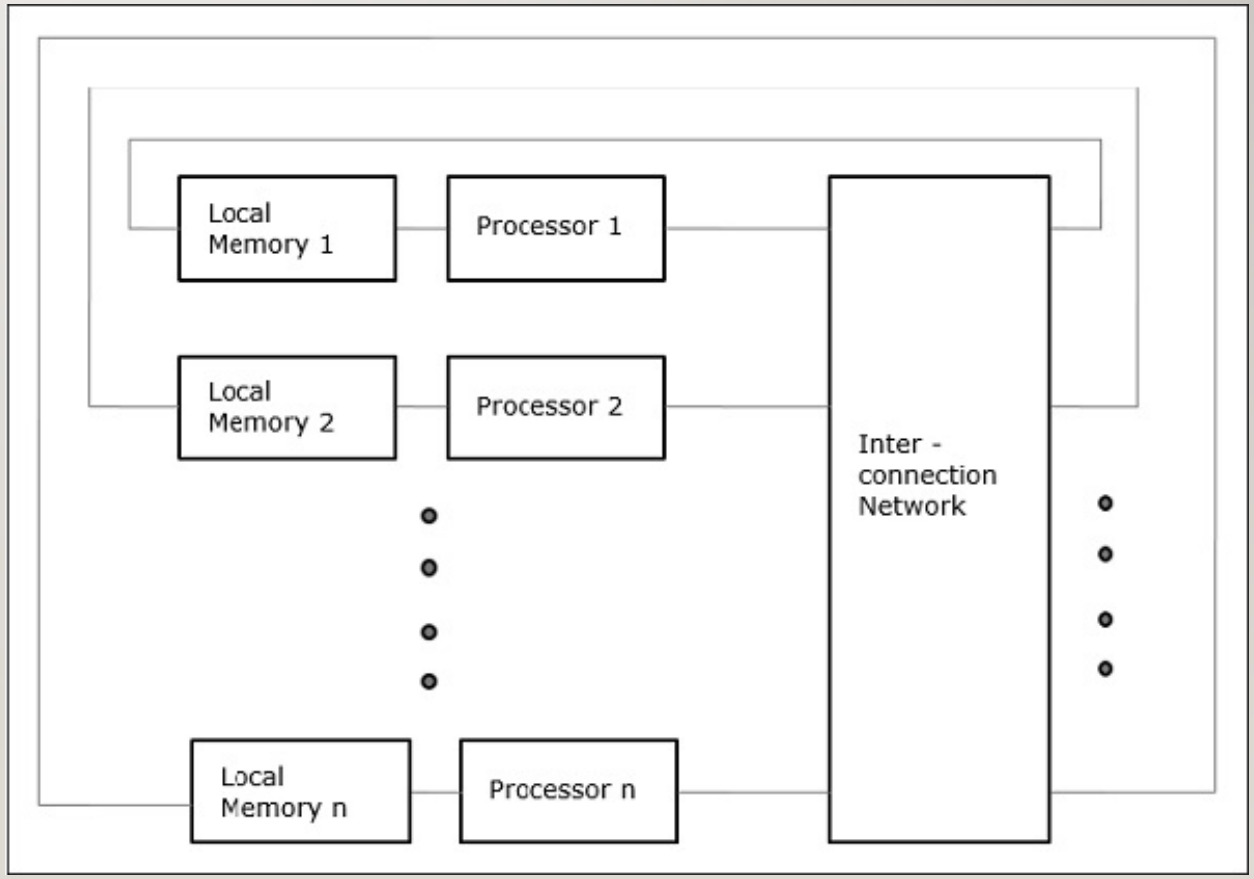
# 1. UMA MODEL

- **UMA: Uniform Memory Access Model**
- In this model, all the processors share the physical memory uniformly
- All the processor have equal access time to all the memory words.
- Each processor may have a private cache memory. Same rule is followed for peripheral devices.
- When all the processors have equal access to all the peripheral devices, the system is called a **symmetric multiprocessor**.
- When only one or a few processors can access the peripheral devices, the system is called an **asymmetric multiprocessor**.
- Tightly coupled systems use a common bus , crossbar , or multistage network to connect processors , peripherals, and memories.


Processor 1
Processor 2
Processor n
System, Interconnect (Bus, Crossbar, Multistage network)
I / O
Shared Memory 1
Shared Memory m

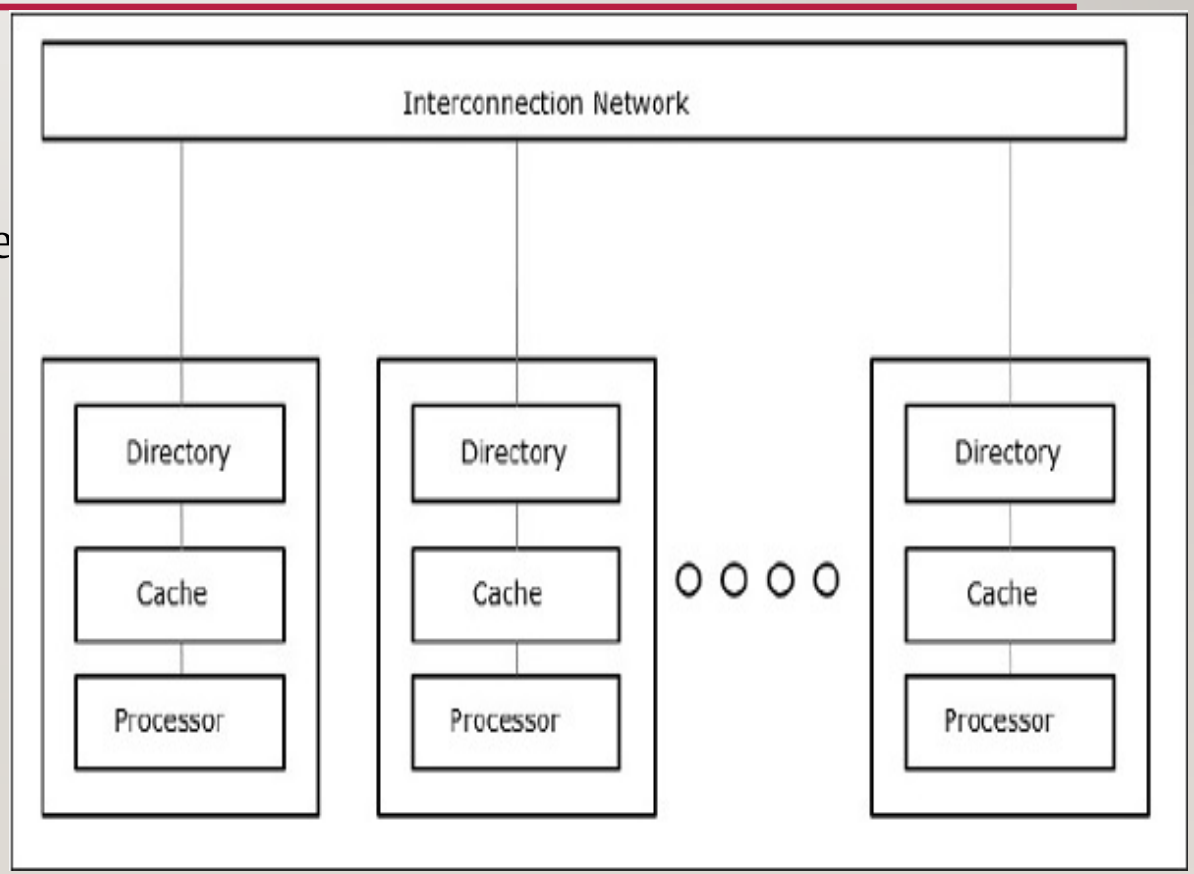
# 2. NUMA MODEL

- **NUMA :Non-Uniform Memory Access Model**
- In NUMA multiprocessor model, the access time varies with the location of the memory word.
- The memory is shared among all the processors, called local memories.
- The collection of all local memories forms a global address space which can be accessed by all the processors.



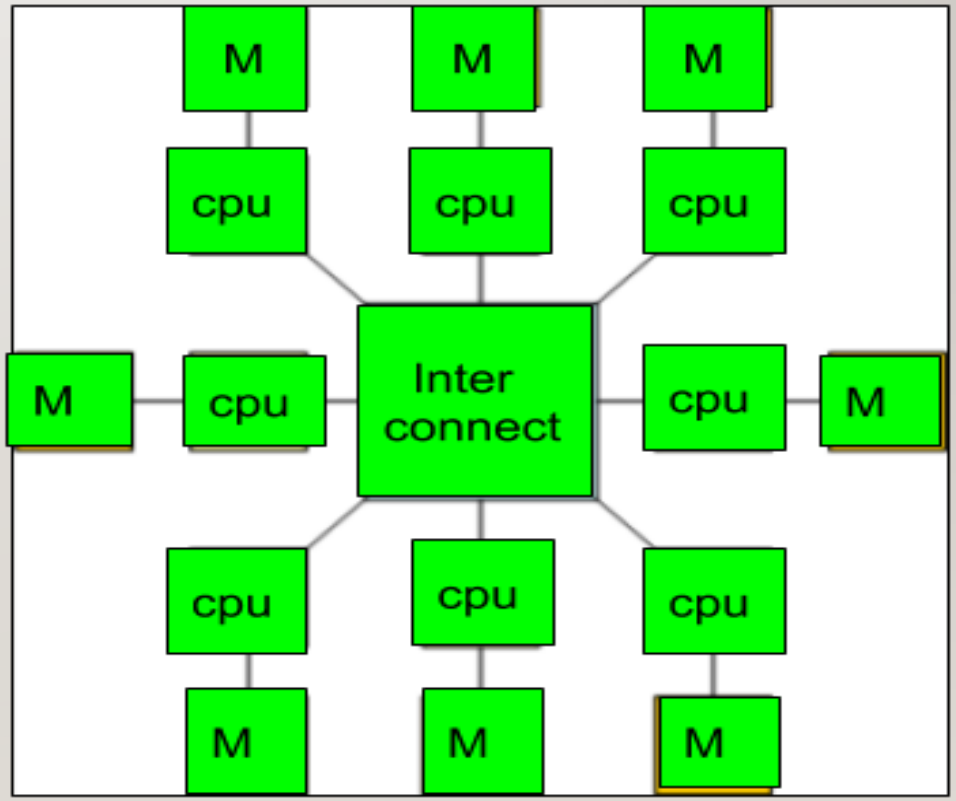
# 3. COMA MODEL

- **COMA: Cache Only Memory Access Model**
- The COMA model is a special case of the NUMA model.
- In COMA Model all the distributed main memories are converted to cache memories.
- Remote cache access is assigned by means of cache directories.



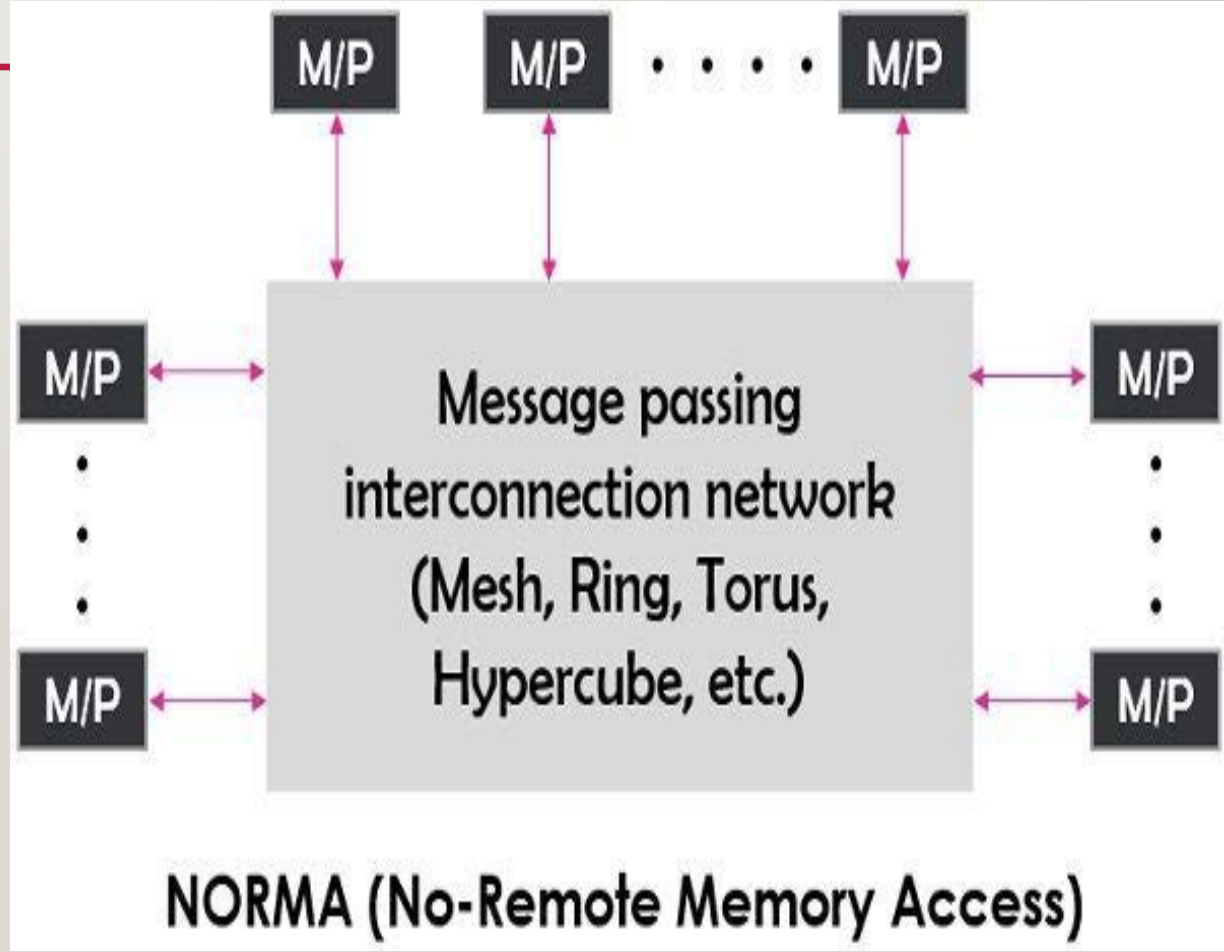
# MULTICOMPUTER

- A computer made up of several computers . The term generally refers to an architecture in which each processor has its own memory.
- A multicomputer system is a computer system with multiple processors that are connected together to solve a problem.
- Each processors can communicate with each other via an interconnection network. No shared memory occurs also known as **distributed memory multicomputer.**
- Types of Multicomputer(NORMA)



# NORMA MODEL(NO-REMOTE MEMORY ACCESS MODEL )

- The system consists of multiple computers ,known as **nodes.**
- Nodes are inter-connected by message passing network. This network of connecting element may be in mesh, ring, tours or hypercube.
- Each node acts as an autonomous computer consisting of a processor, a local memory and sometimes I/O devices.
- All local memories are private and are accessible only by local processors. So the traditional machines are called **no-remote-memory-access (NORMA)** machines. Example : MIT J-machine



# MULTIPROCESSOR VS MULTICOMPUTER MODEL

## MULTIPROCESSOR

1. A multiprocessor system is a single computer that operates with multiple CPUs
2. They are also called tightly coupled due to the higher degree of resource sharing.
3. More difficult and costly to build.
4. Support parallel computing.
5. Program tends to be easier.
6. Also known as shared memory multiprocessor.

## MULTICOMPUTER

1. multicomputer system is a cluster of computers that operate as a singular computer.
2. They are also called loosely coupled as there is no resource sharing.
3. Easier and cost effective to build.
4. Supports distributed computing.
5. Program tends to be more difficult.
6. Also known as distributed memory multicomputer.

# ARCHITECTURAL CLASSIFICATION SCHEMES

**1. Flynn's taxonomy or Classification**

It is based on the multiplicity of instruction stream and data streams in a computer system.

SISD,SIMD,MISD,MIMD

**2. Feng's Classification**

It is based on serial versus parallel processing.

WSBS,WPBS,WSBP,WPBS

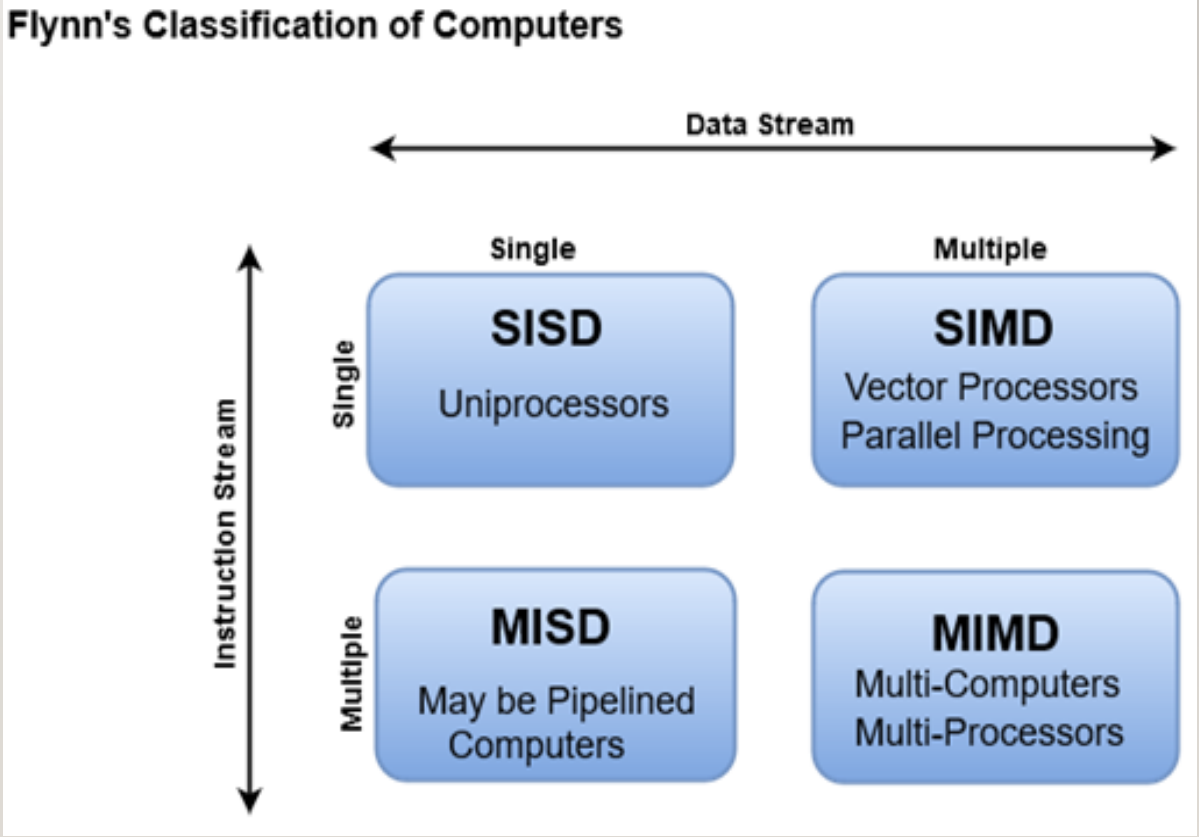
# 1. FLYNN'S TAXONOMY

- This **taxonomy** proposed by Michael J. Flynn in 1966.
- Flynn's taxonomy is a "**classification of parallel computer architectures based on number of instruction stream and data stream".**
- The sequence of instructions read from memory is called an **instruction stream.**
- The operation performed on the data in the processor is called **data stream.**

- Note: The term 'Stream' refers to the flow of instructions or data

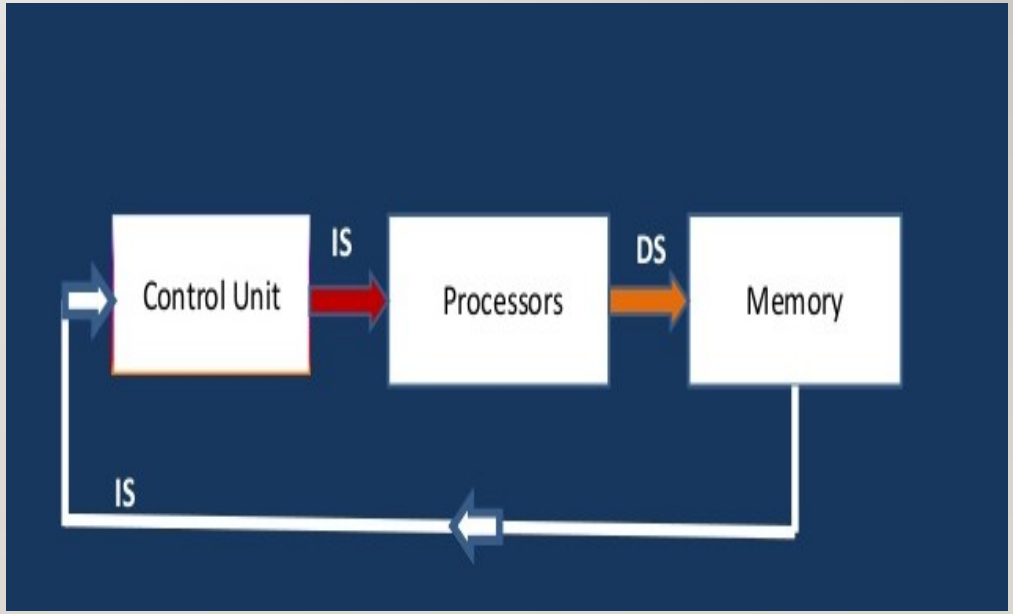
Flynn's classification divides computers into four major groups that are:

1. Single instruction stream, single data stream (SISD)
2. Single instruction stream, multiple data stream (SIMD)
3. Multiple instruction stream, single data stream (MISD)
4. Multiple instruction stream, multiple data stream (MIMD)



# 1. SINGLE INSTRUCTION STREAM, SINGLE DATA STREAM (SISD)

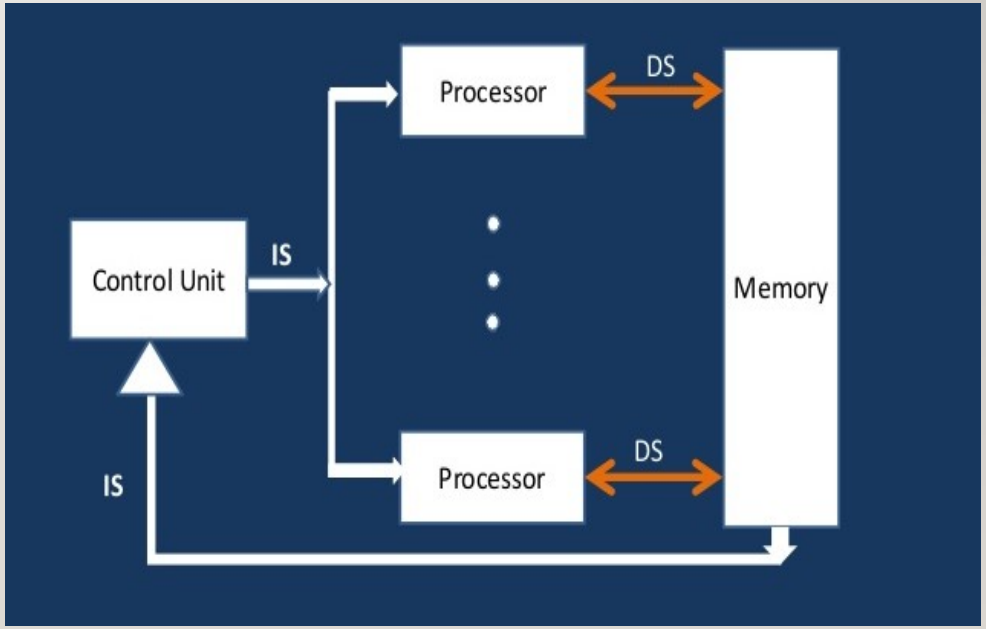
- An SISD computing system is a uniprocessor machine which is capable of executing a single instruction, operating on a single data stream.
- SISD is a serial or sequential computer.
- SISD represents the organization of a single computer containing a control unit, a processor unit, and a memory unit.
- Instructions are executed sequentially and the system may or may not have internal parallel processing capabilities.
- In SISD computer instructions are executed but may overlap in their execution stages.
- Example: IBM 7001, VAX 11, CDC-6600 .



# 2. SINGLE INSTRUCTION STREAM, MULTIPLE DATA STREAM (SIMD)

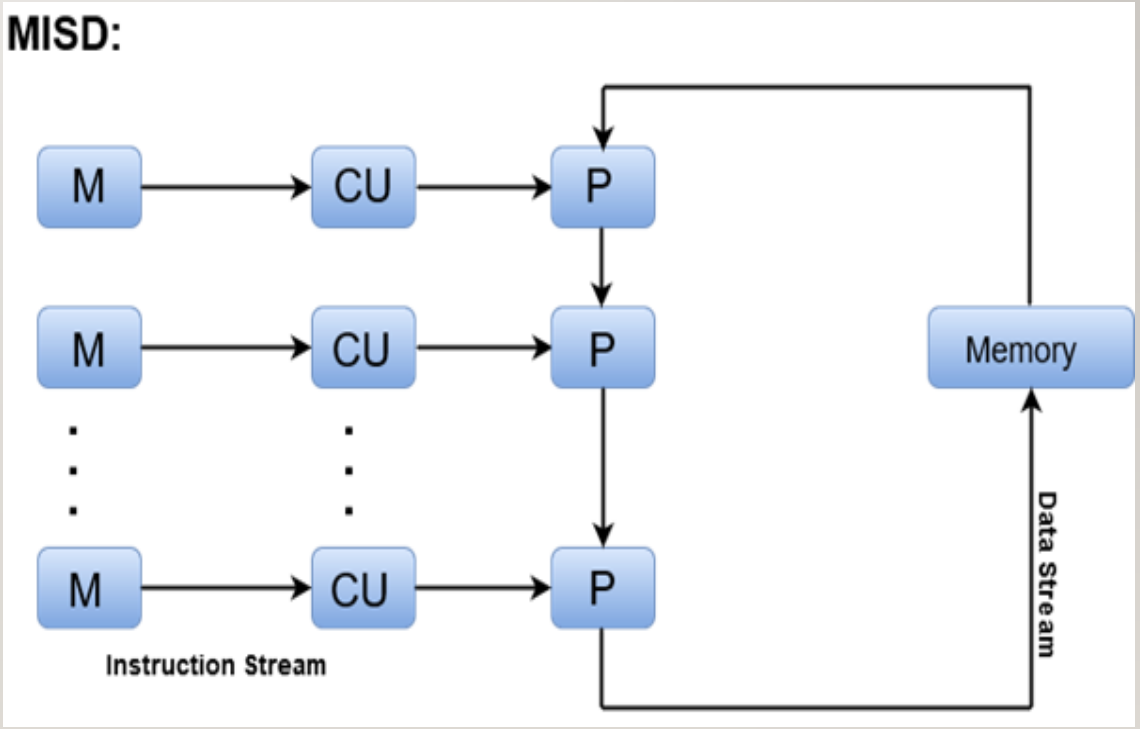
- This category corresponds to array processor and vector processor.
- SIMD represent an organization that includes many processing units(processor) under the supervision of a common control unit.
- All processor receive the same instruction from the control unit but operate on different items of data.
- Example: Array Processor : MPP

  Vector Processor: IBM 9000,C90



# 3. MULTIPLE INSTRUCTION STREAM, SINGLE DATA STREAM (MISD)

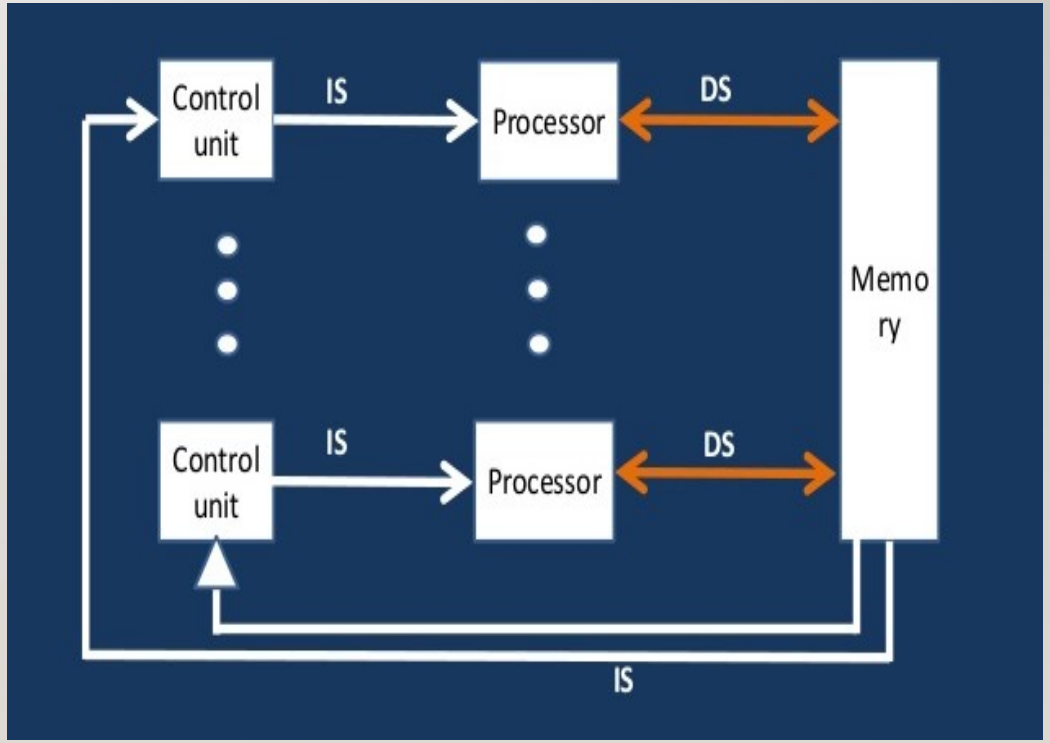
- In MISD, multiple processing units operate on one single-data stream. Each processing unit operates on the data independently via separate instruction stream.
- The result of one processor become the input of the next processor in the micro-pipe. So it is mostly used in pipelined computers.
- MISD structure is only of theoretical interest since no practical system has been constructed using this organization.



# 4. MULTIPLE INSTRUCTION STREAM, MULTIPLE DATA STREAM (MIMD)

- In MIMD , there are more than one processor unit having the ability to execute several program simultaneously.
- MIMD organization refers to a computer system capable of processing several programs at a same time.
- Execution of multiple instruction on multiple data.
- In MIMD, each processor has a separate program and an instruction stream is generated from each program.
- Most multiprocessor and multicomputer system can be classified in this category.

Example: Cray T90,IBM-SP2



# FENG'S CLASSIFICATION

- Tse-yun Feng [1972] proposed a scheme on the basis of degree of parallelism to classify various computer architecture.
- Feng's classification is based on serial versus parallel processing.
- The maximum number of binary digits that can be processed within a unit time by a computer system is called the maximum parallelism degree P.
- In Feng's classification parallelism is based on word length (n) and a bit-slice length (m)

❑word length is the number of bits in a word.

❑A bit-slice is a string of bit one from each of the words at the same vertical position.

Feng’s Computer system classification:

1. Word Serial Bit Serial (WSBS)
2. Word Parallel Bit Serial (WPBS)
3. Word Serial Bit Parallel (WSBP)
4. Word Parallel Bit Parallel (WPBP)

**1. Word Serial Bit Serial (WSBS)**

It is called bit serial processing because one bit is processed at a time. Slow

**2. Word Parallel Bit Serial (WPBS)**

It is called bit slice processing because m-bit slice is processed at a time.

**3. Word Serial Bit Parallel (WSBP)**

It is found in most existing computer and has been called as word slice processing because one word of n bit is processed at a time.

**4. Word Parallel Bit Parallel (WPBP)**

It is known as fully parallel processing in which an array on n x m bits is processed at one time. Maximum parallelism is achieved here. Fastest

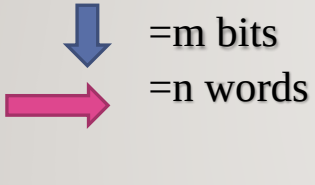



figure 1 :WSBS
figure 2 : WPBS
figure 3 : WSBP
figure 4: WPSP