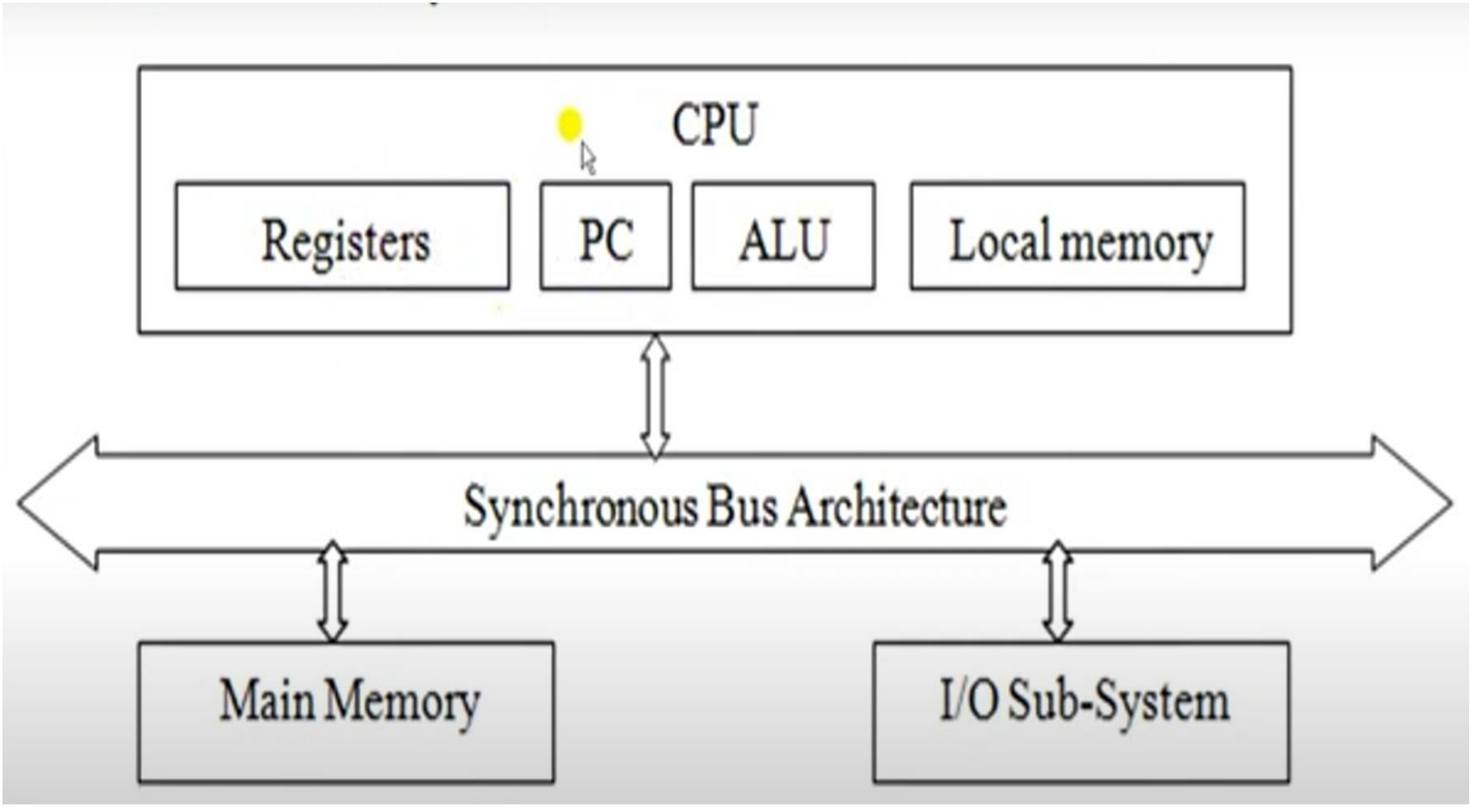
Unit 6
Parallel Processing Paradigm

Basic Uniprocessor Architecture

Uniprocessor : A uniprocessor is a system with a single processor which has three major components that are main memory i.e. the central storage unit, the central processing unit i.e. CPU, and an input-output unit like monitor, keyboard, mouse, etc.



Parallelism in Uniprocessor

- **Parallelism in a uniprocessor** means a system with a single processor performing two or more than two tasks simultaneously.
- Parallelism can be achieved by two means hardware and software.

1. Hardware Method

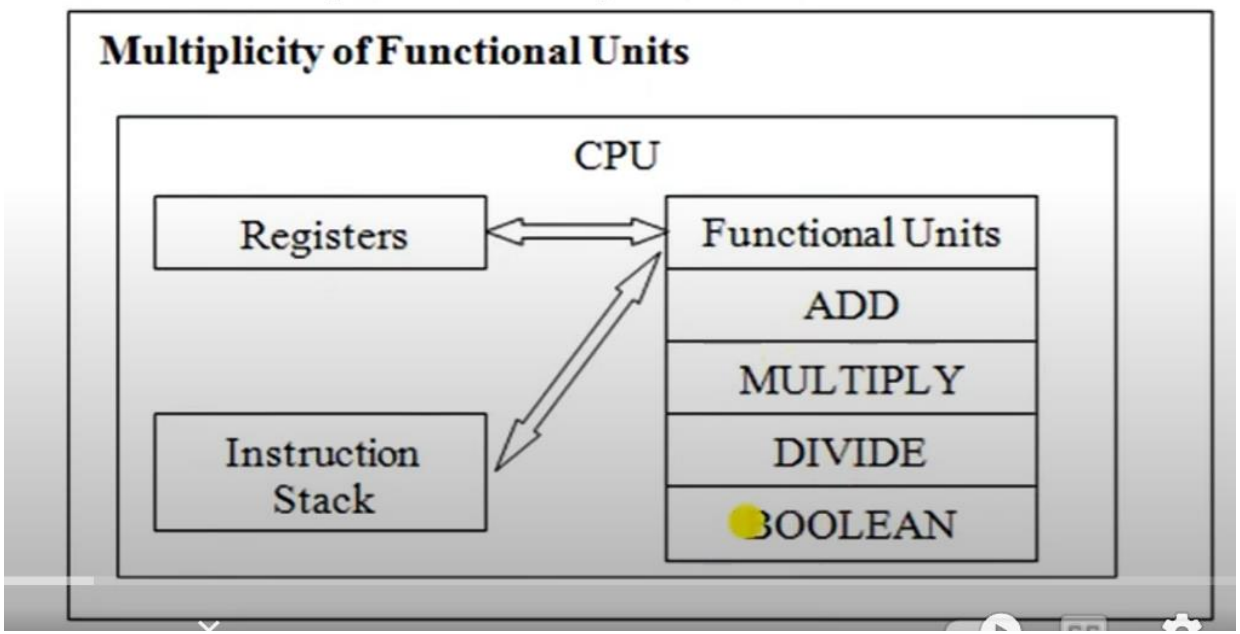
- Multiplicity of functional unit.
- Parallelism and pipelining within the CPU.
- Overlapped CPU and I/O operation.
- Use Hierarchical Memory System.
- Balancing of subsystem bandwidth.

2. Software Method

- Multiprogramming
- Time Sharing

1. **Multiplicity of Functional Unit:**

- In earlier computers, the CPU consists of only one arithmetic logic unit which used to perform only one function at a time.
- This slows down the execution of the long sequence of arithmetic instructions.
- To overcome this the functional units of the CPU can be increased to perform parallel and simultaneous arithmetic operations.
- Many of the functions of ALU can be distributed among multiple and specialized functional units which can operate parallel.



2. **Parallelism and Pipelining within CPU:**

- **Parallel adders** can be implemented using techniques such as carry-lookahead and carry-save. A parallel adder is a digital circuit that adds two binary numbers, where the length of one bit is larger as compared to the length of another bit and the adder operates on equivalent pairs of bits parallelly.

- The **multiplier** can be recoded to eliminate more complex calculations.

- **Various instruction execution phases are pipelined** and to overcome the situation of overlapped instruction execution the techniques like instruction prefetch and data buffers are used.

3. Overlapped CPU and I/O Operation: To execute I/O operation parallel to the CPU operation we can use I/O controllers or I/O processors. For direct information transfer between the I/O device and the main memory, direct memory access (DMA) can be used.

4. Use Hierarchical Memory System:

- We all are aware of the fact that the processing speed of the CPU is 1000 times faster than the memory accessing speed which results in slowing the processing speed.
- To overcome this speed gap hierarchical memory system can be used.
- The faster accessible memory structure is registered in CPU, then cache memory which buffers the data between CPU and main memory.

5. Balancing of Subsystem Bandwidth

The processing and accessing time of CPU, main memory, and I/O devices are different. If arrange the processing time of these units in descending order the order would be:

- $t_d > t_m > t_p$

where t_d is the processing time of the device, t_m is the processing time of the main memory, and t_p is the processing time of the central processing unit. The processing time of the I/O devices is greater as compared to the main memory and processing unit. CPU is the fastest unit.

1. Cache Memory (Processing speed of Main Memory and Processor)
2. I/O Channels (Processing speed of I/O devices and Main Memory)

Software Approach for Parallelism in Uniprocessor

- **1. Multiprogramming:**
- There may be multiple processes active in computers and some of them may be competing for **memory some for I/O devices and some for CPU**. So, to establish a balance between these processes, program interleaving must be practiced. This will boost resource utilization by overlapping the I/O and CPU operations.

- Program interleaving can be understood as when a process P_1 is engaged in I/O operation the process schedular can switch CPU to operate process P_2. This led process P_1 and P_2 to execute simultaneously. This interleaving between CPU and I/O devices is called multiprogramming.

- **2. Time-Sharing:** Multiprogramming is based on the concept of time-sharing. The CPU time is shared among multiple programs. Sometimes a high-priority program can engage the CPU for a long period starving the other processes in the computer.

- The concept of timesharing assigns a fixed or variable time slice of CPUs to multiple processes in the computer. **This provides an equal opportunity to all the processes in the computer.**

- So, in this way, we can effectively introduce parallelism in the uniprocessor. Parallelism increases the efficiency of the system and makes computation even faster. We have seen both the approaches hardware and software to introduce parallelism in a system with a single processor.

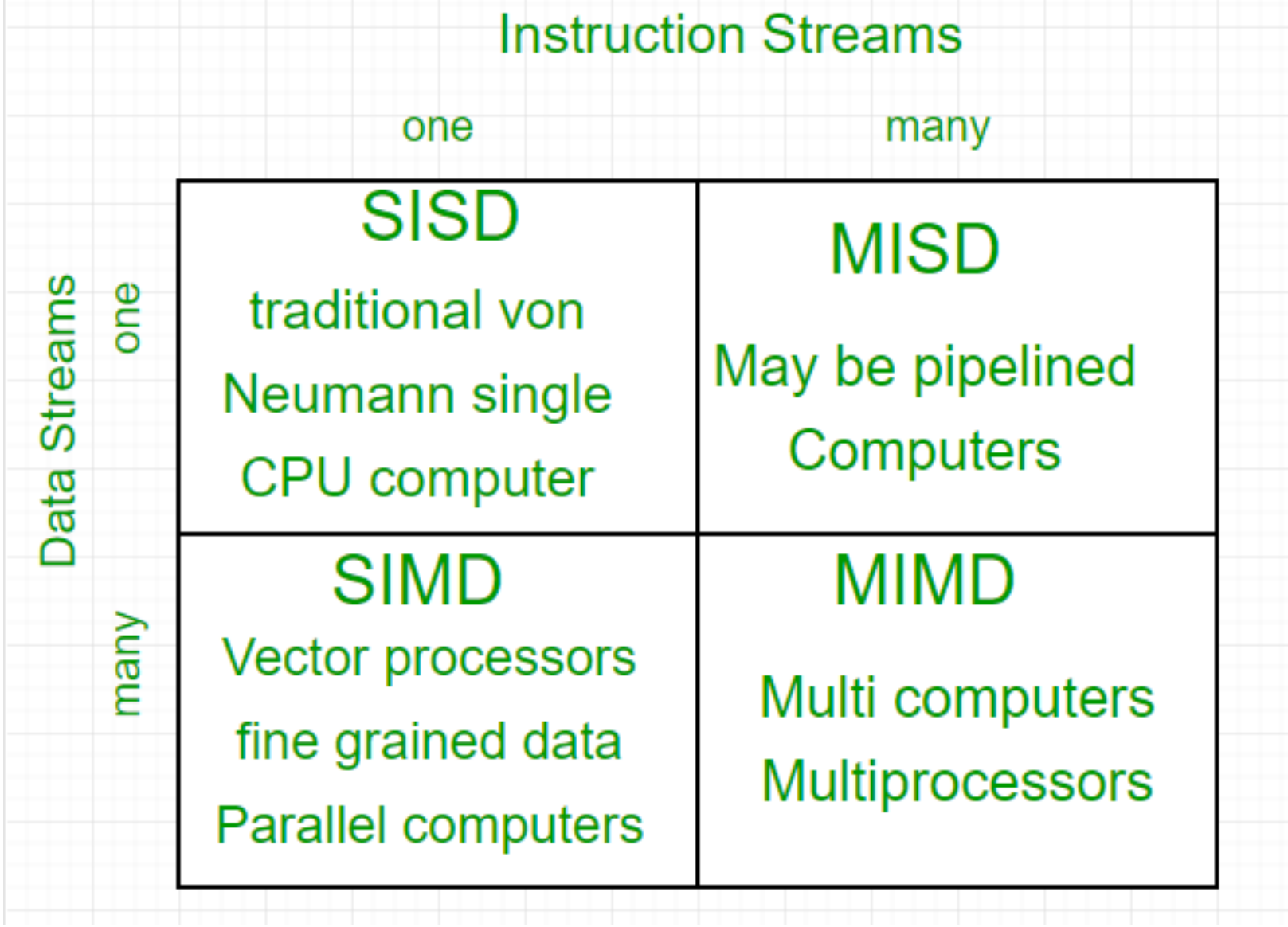
Computer Architecture | Flynn’s taxonomy

- **Stream**→ Stream refers to sequence of objects (may be data or instructions) that is executed by a single CPU.

- **Instruction Stream**→ Instruction stream refers to sequence of instructions that is executed by the CPU.

- **Data Stream**→ Data stream refers to the sequence of data including input, partial or temporary results called for by instruction stream.

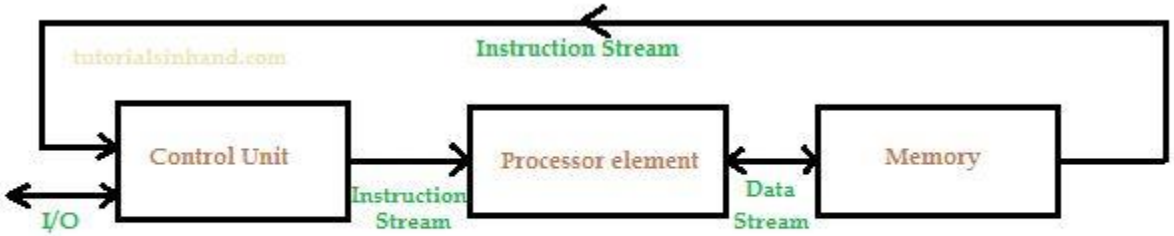
Flynn's classification is based on multiplicity of instructions and data streams in a computer system.


Instruction Streams
one
many
Data Streams
one
many
SISD
traditional von Neumann single CPU computer
MISD
May be pipelined Computers
SIMD
Vector processors
fine grained data
Parallel computers
MIMD
Multi computers
Multiprocessors

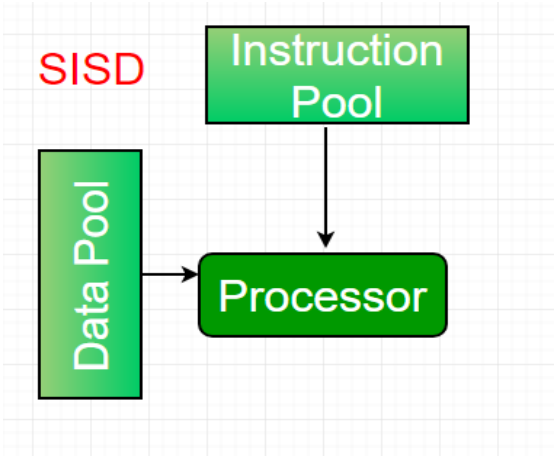
Flynn's classification –

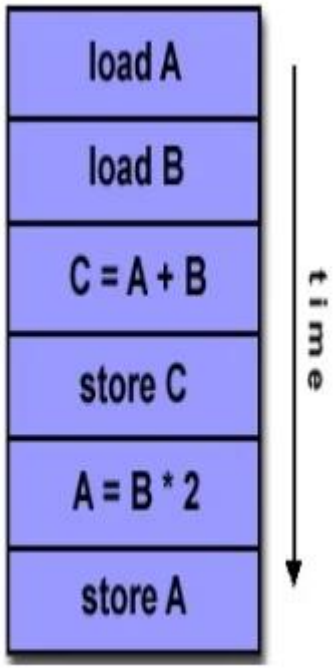
1. Single-instruction, single-data (SISD) systems –

An SISD computing system is a uniprocessor machine which is capable of executing a single instruction, operating on a single data stream.

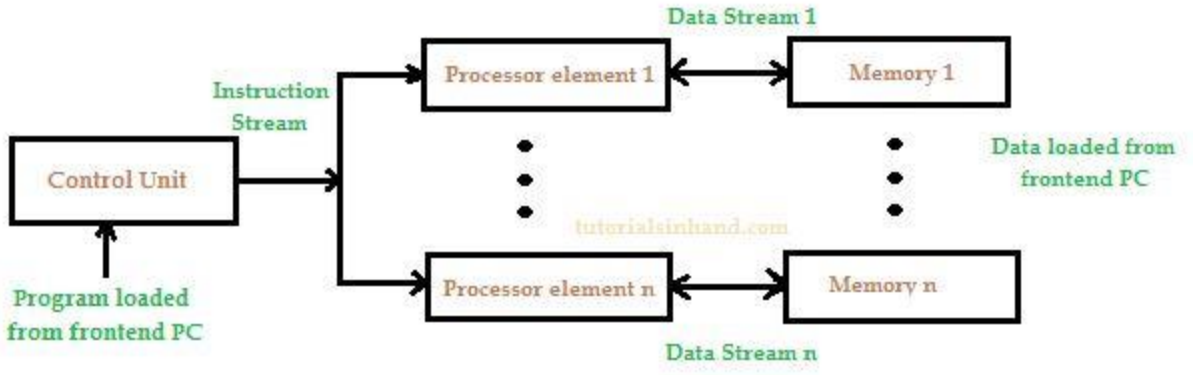


SISD Computer

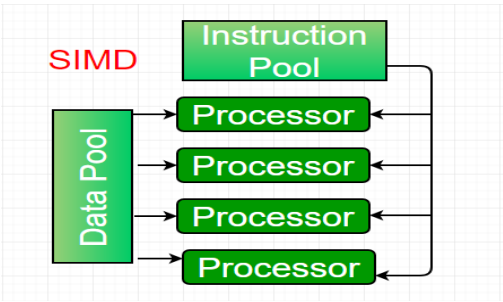


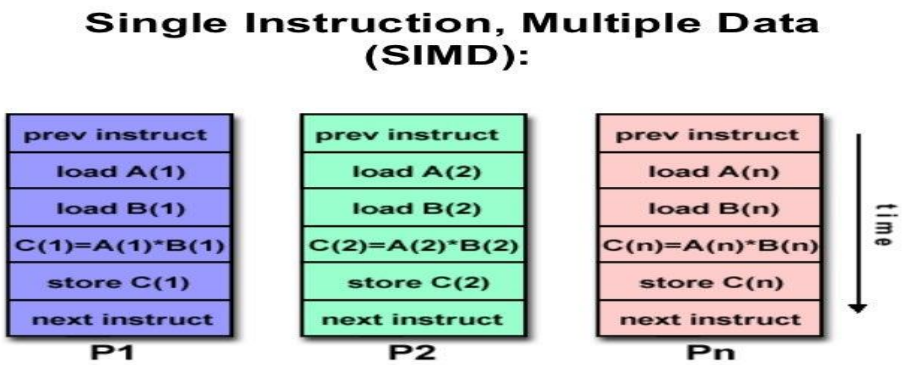


- **Single-instruction, multiple-data (SIMD) systems –**
  An SIMD system is a multiprocessor machine capable of executing the same instruction on all the CPUs but operating on different data streams.



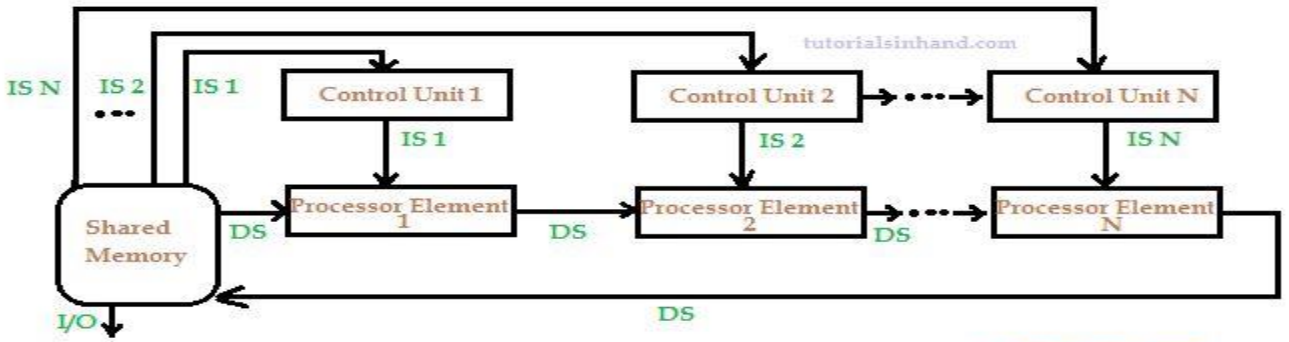
SIMD Computer

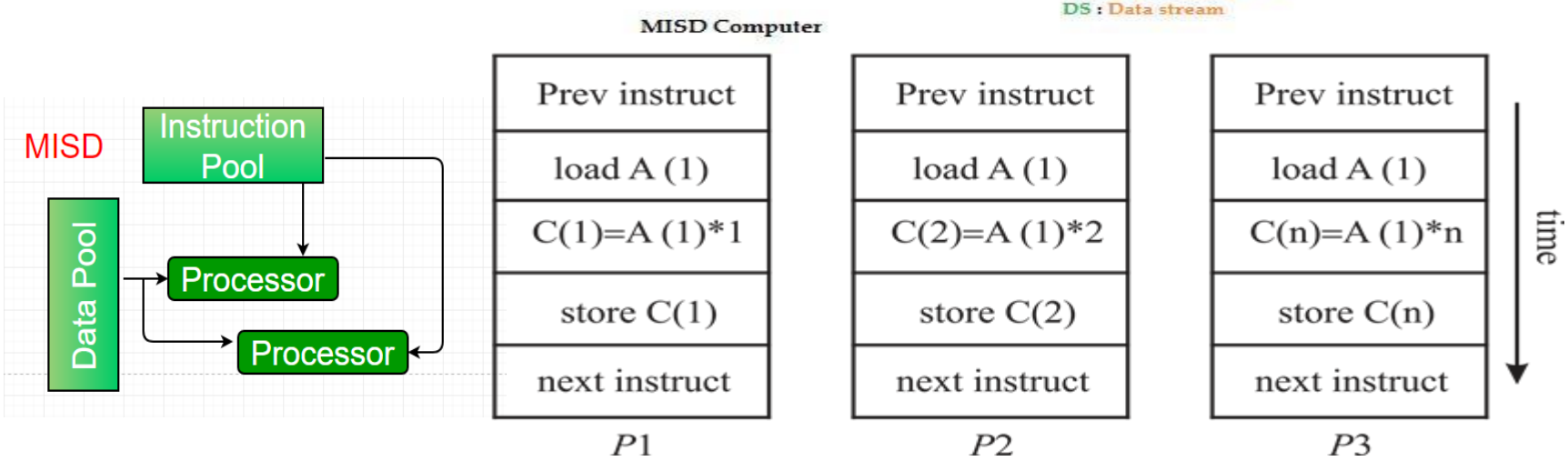




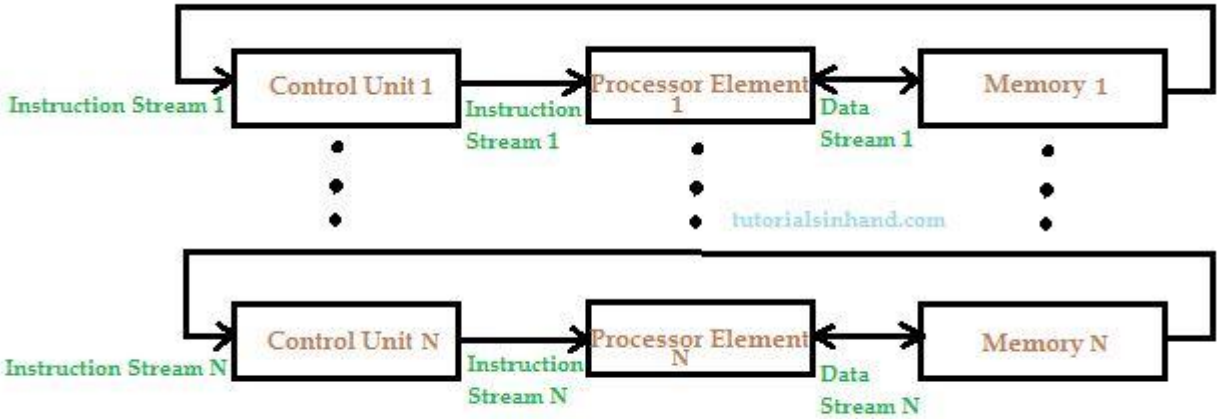
- **Multiple-instruction, single-data (MISD) systems –**
  An MISD computing system is a multiprocessor machine capable of executing different instructions on different PEs but all of them operating on the same dataset .

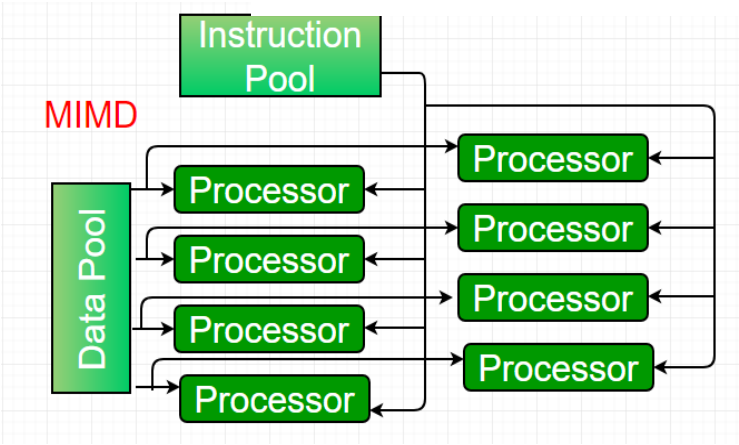
  Example Z = sin(x)+cos(x)+tan(x)

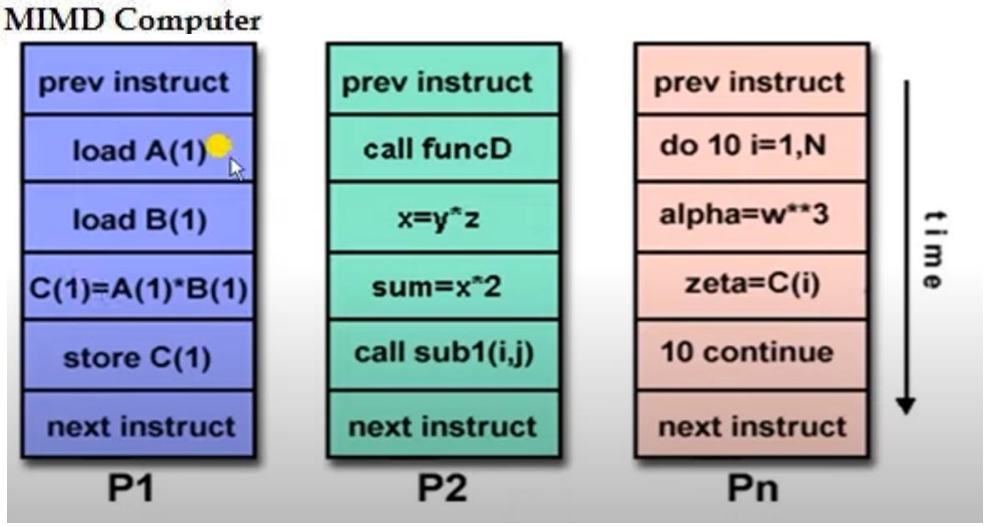




- **Multiple-instruction, multiple-data (MIMD) systems –**
  An MIMD system is a multiprocessor machine which is capable of executing multiple instructions on multiple data sets.



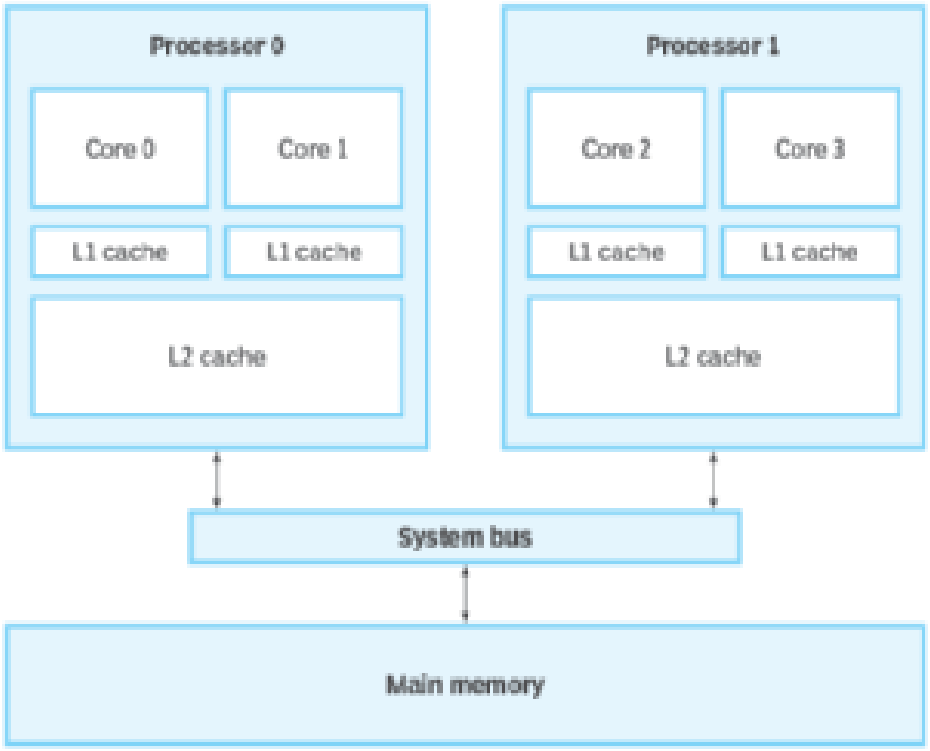




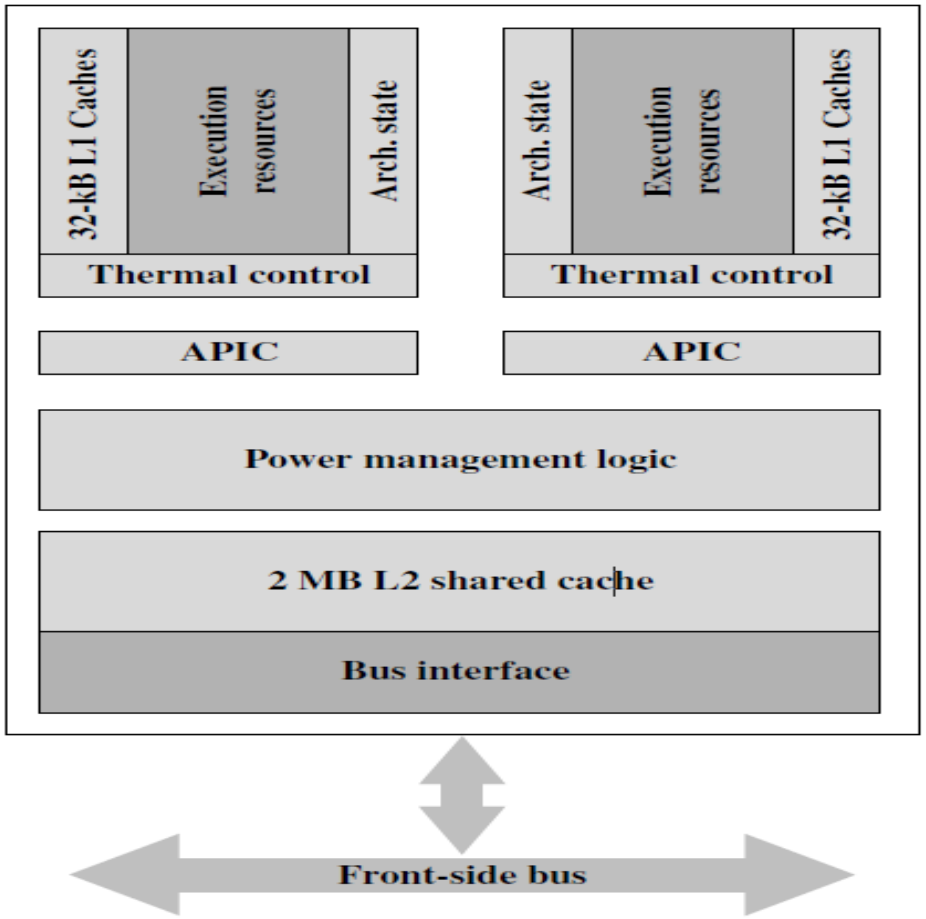
What is a multicore processor?

- A multicore processor is an integrated circuit that has two or more processor cores attached for enhanced performance and reduced power consumption.
- These processors also enable more efficient simultaneous processing of multiple tasks, such as with parallel processing and multithreading.
- A dual core setup is similar to having multiple, separate processors installed on a computer.
- However, because the two processors are plugged into the same socket, the connection between them is faster.
- The use of multicore processors or microprocessors is one approach to boost processor performance without exceeding the practical limitations of semiconductor design and fabrication.

Architecture of multicore processors



Intel Core Duo



Intel Core Duo Block Diagram

- The general structure of the Intel Core Duo is shown in Figure.
- As is common in multicore systems, each core has its own dedicated **L1 cache**. In this case, each core has a 32-KB instruction cache and a 32-KB data cache.

A. Thermal Control Unit:

- Each core has an independent **thermal control unit**.
- The Core Duo thermal control unit is designed to manage chip heat dissipation to maximize processor performance within thermal constraints.
- Thermal management also improves ergonomics with a cooler system and lower fan a acoustic noise.
- If the temperature in a core exceeds a threshold, the thermal control unit reduces the clock rate for that core to reduce heat generation.

B . Advanced Programmable Interrupt Controller (APIC):

The APIC performs a number of functions, including the following:

1. The APIC can provide inter processor interrupts, which allow any process to interrupt any other processor or set of processors. In the case of the Core Duo, a thread in one core can generate an interrupt, which is accepted by the local APIC, routed to the APIC of the other core, and communicated as an interrupt to the other core.

2. The APIC accepts I/O interrupts and routes these to the appropriate core.

3. Each APIC includes a timer, which can be set by the OS to generate an interrupt to the local core.

C. Power management logic:

1. Theis is responsible for reducing power consumption when possible, thus increasing battery life for mobile platforms, such as laptops.

2. In essence, the power management logic monitors thermal conditions and CPU activity and adjusts voltage levels and power consumption appropriately.

3. It includes an advanced power-gating capability that allows for an ultra fine grained logic control that turns on individual processor logic subsystems only if and when they are needed.

4. Additionally, many buses and arrays are split so that data required in some modes of operation can be put in a low power state when not needed.

D. Shared 2-MB L2 cache:

- The Core Duo chip includes a shared 2-MB **L2 cache**.
- The cache logic allows for a dynamic allocation of cache space based on current core needs, so that one core can be assigned up to 100% of the L2 cache
- The L2 cache includes logic to support the MESI protocol for the attached L1 caches.

E. Bus interface .

The **bus interface** connects to the external bus, known as the Front Side Bus, which connects to main memory, I/O controllers, and other processor chips.