UNIT I – Chapter 1

1. BASIC STRUCTURE OF COMPUTERS

1.1 What is a Computer (Digital Computer)?

It is a fast electronic calculating device that accepts digitized information, processes it according to the list of internally stored instructions and produces the result in the form of output.

Computer Types:

1. Personal Computer:

These are the widely used types of computers in homes, offices and colleges. There are basically two types of this computer:

i. Desktop Computers:

Computer having a processing or a storage unit, visual display unit, audio unit, input units that can all be located on a desk of a home or an office.

ii. Portable Notebook Computers:

These are compact version of personal computer with all kinds of devices of the desktop that is being packaged in a single unit.

2. Workstation:

A computer with high resolution graphic input and output capability which has the dimension of the desktop but with a higher computational power.

3. Enterprise systems and servers:

Also called as mainframes used for large business processing that has high computation power and storage capacity than workstation/

Example: ATM

Servers:

These are the one which can handle large volume of requests to access data. Here request and response both are transferred through internet.

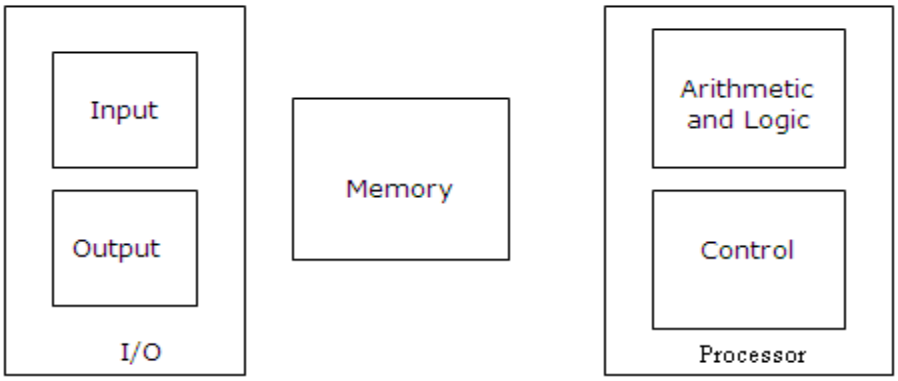
4. Supercomputers:

Computers used for large scale numerical calculations and R & D areas.

1.2 Functional Units of Computer

A computer system has 5 functionally independent units for instruction processing:

1. Input Unit
2. Memory Unit
3. Arithmetic & Logic Unit (ALU)
4. Output Unit
5. Control Unit (CU)



Basic functional units of a computer

All the functional units that is being represented in the above schematic handles information for their functionality and those information's are called as **instruction.** Set of these instructions constitute to what is called as **Program** is used to perform a specific task.

Next we will discuss the overall properties and functionalities of the different units

1. Input Unit:

→Main functionality of this unit is to accept coded information from the human operators or any other external systems.

→One of the well known input devices is **Keyboard.** In keyboard whenever any key corresponding to letter or a digit is pressed, it is automatically converted to the binary code and is transmitted over the cable into the memory system or the processor.

2. Memory Unit:

→Main functionality of this unit is to store programs and data that has to be executed. The memory unit is broadly classified into two classes:

a. Primary Storage
b. Secondary storage

a. Primary Storage:

→Often known as a fast memory that operates at a very high speed.

→Any programs must be stored in this memory while they are being executed. The memory system is made of a large number of storage cells which are capable of storing one bit of information.

→While processing (accessing) the information, the information is accessed in terms of a group of **n** bits which are called as a **word.** To provide easy access to the word, each word is associated with a unique address called **word location.** Number of bits associated with a word is called as **word length** and a word length ranges from 16 to 64 bits.

→When memory is accessed usually one word of data is read or written from or into the memory.

→Primary memory components are called as a **RAM (Random Access Memory)** components. RAM is the memory in which any word location can be reached in a short and fixed amount of time. Time required access its word location is called as **Memory Access Time** and it is constant irrespective of word location. The RAM units again have the hierarchy in the memory systems which are classified as:

i. Cache memory: It is a very fast and smaller RAM unit that is built on top of the processor chip.

ii. Main Memory: Largest and slowest RAM unit.

b. Secondary Storage:

A slower memory unit when compared to the primary memory but storage area is larger and permanent and are usually **magnetic disks** and **optical disks**

3. Arithmetic & Logic Unit (ALU):

Most of the computer operations are being performed in this unit of the processor. Any ALU operation is first initiated by bringing the required operands into the processor and is stored in the high speed storage elements of the processor called **registers** which usually have one word capacity**.**

4. Output Unit:

Main functionality is to send the processed result to the outside world through any of the output device.

5. Control Unit (CU):

→Operations of all other units described earlier are controlled and co ordinate by this dedicated unit.

→Control unit controls and co ordinates all the functionalities and communication between the other units by sending the control signals to other units in the form of timing signals (clock pulses). Timing signals are the signals that determine when a given action has to take place.

1.3 Basic operational Concepts

→In a computer system, to perform a given task, appropriate program consisting of list of instructions is stored in the memory. Individual instruction from the memory is loaded into the processor and is executed one by one by the processor to accomplish a given task. Data to be used as operands are also stored in the memory.

Example of typical instruction:

ADD LOCA, R0

Above instruction adds the operand at memory location A to operand in register R0 and stores the result in register R0 as shown:

R0 = [LOCA] + [R0]

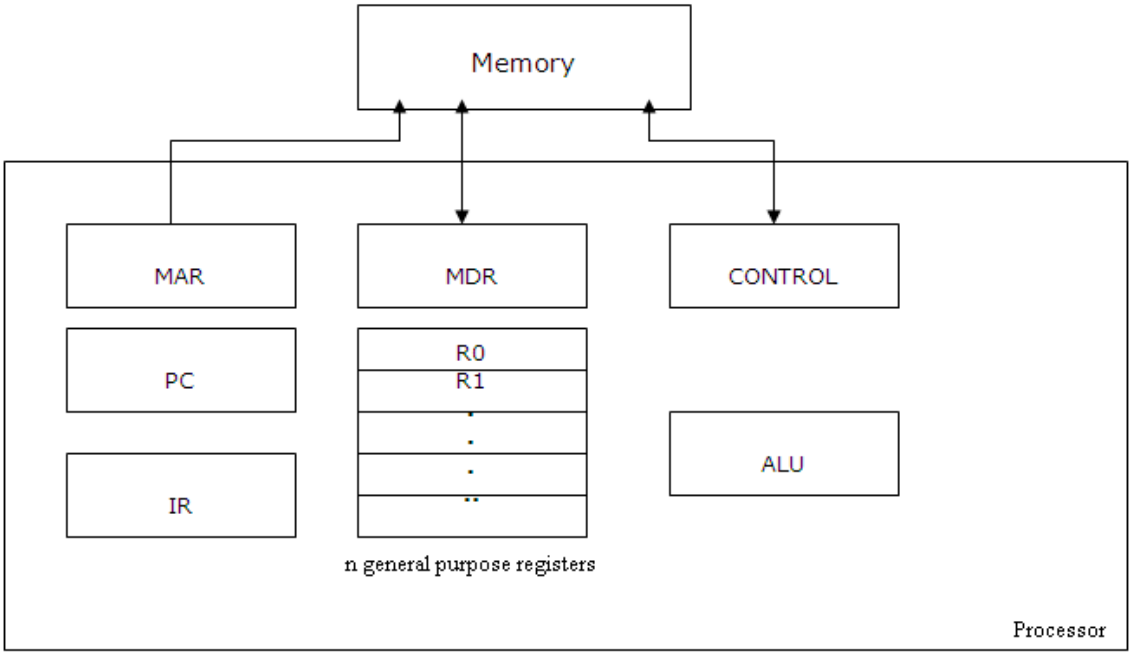
Execution Steps:

For the time being we will mention the steps of execution very abstractly, later we will see the detailed execution. For the execution of any mathematical operations, it has to be performed by bringing the operands from the memory into the processor registers. Steps of operation are:

1. First instruction is fetched from the memory in to the processor.
2. Then the execution of the instruction is done as shown below:

```
Load    LOCA, R1        (R1 = [LOCA])
Add     R1, R0          (R0 = [R1] + [R0])
```

Connection between processor and memory



Transfer between memory and processor starts by sending address of memory location to be accessed into the memory unit along with the appropriate control signal for performing the required action.

In addition to ALU and Control Unit, processor has a number of other registers which are classified as dedicated registers and general purpose register.

1. Instruction Register (IR): A dedicated holds the instruction that is currently being executed.

2. Program Counter (PC): A dedicated register that keeps track of the execution of the program. It points to the next instruction to be executed.

3. Memory Address Register (MAR): A dedicated register that holds the address of any location of the memory unit that has to be accessed.

4. Memory Data Register (MDR): A dedicated register that contains the data to be written into or read out from the memory unit.

5. General purpose Registers: There are n general purpose registers for holding the operands that is decoded from the instruction.

Steps of Execution:

1. Execution of the program starts when **PC** is set to point to the first instruction of the program that is to be executed.

2. Contents of **PC** are then transferred to **MAR** and a **read** control signal is sent to the memory unit to fetch the instruction from the memory location. Addressed word is read out of the memory location and is loaded into the register **MDR.**
3. Next contents of **MDR** are transferred into **IR** for the execution purpose. If the instruction involves any operation that needs operand from the memory location, operands are fetched from the memory location by sending their address along with the **read** signal, Value will be transferred into **MDR** and then into the **general purpose register.**
4. After completing the operation, result may be kept in the general purpose register or may be sent into the memory location by sending the computed data into **MDR** along with write signal by the control unit

1.4 Bus structures

→These are the one that defines how to connect different functional units of the computer.

→**Bus:** Group of lines (wires) that serves as a connecting path for the several devices connected across the system is called as **Bus.**

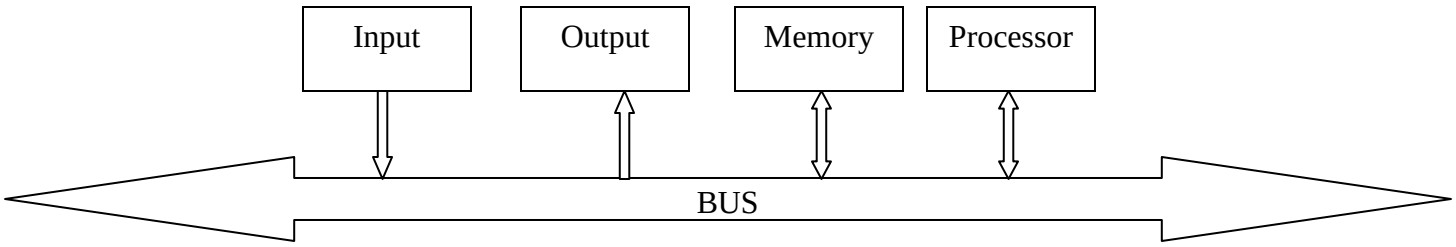
Lines basically constitute of:

1. Address lines → Carries address to be accessed from the memory location.
2. Data lines → Carries data from or into the memory unit.
3. Control lines → Carries control signals required performing the operation.

Bus structures are broadly classified into two types:

1. Single Bus structure
2. Multiple Bus structure

Single Bus structure:



Here all the units are connected to a single bus. At a time bus can be used for only one transfer i.e. two units can actively use bus at any given time. Bus control lines are used to arbitrate multiple bus requests which are called as **Bus arbitration.** Main advantage of this bus is low cost and high flexibility.

1.5 Some basic definitions:

Software: Collection of programs are called as software.

System Software: Software that is responsible for co ordination of all the activities of the computing system.

Operating System: Large program that is used to control the interaction among the various computer units as they execute application program.

Performance: Computer performance is measured in terms of how quickly it can execute programs to accomplish a given task successfully.

Factors affecting Computer performance:

1. Compiler
2. Machine Instruction set
3. Hardware Design

Processor Clock: These are the timing signals that control the processor circuits.

Above shown is the representation of the timing signals were one positive and negative edge constitutes to **clock cycle.** A timing signal is made of several such clock cycles and the length of the clock cycle affects the processor speed greatly. During the execution of an instruction each and every step of the execution is made to complete in one clock cycle. Length of clock cycle is nothing but the clock period (**P)** and is inversely proportional to the clock rate (**R**).

$$R = 1/P$$

Clock rate is mentioned in terms of cycles per second and a standard unit called **Hertz** is used to represent clock rate.

For example: 500 MHz means 500 million cycles per second and 500 GHz means 500 billion cycles per second were Mega and Giga is used to represent million and billion respectively.

If clock rate of a processor is 500 MHz then clock period **P** is given by

$$\mathbf{P = 1/R = 1/500X10^{6} = 0.002X10^{-9} = 2ns}$$

1.6 Basic Performance Equation (Processor Time Equation)

→The Equation is also called T,N,S,R equation.

- Let **T** be the processor time required to execute a program that has been prepared in some high level language.
- Compiler generates machine level language for these programs. Assume that complete execution requires **N** machine languages to be executed.
- Let **S** be the average basic step needed to execute one instruction where each step is completed in one clock cycle
- Let **R** be the clock rate of the processor clock for a system.

Program Execution Time

$$T = \frac{NXS}{R}$$

1.7 Pipelining & Superscalar Operations

Pipelining

Overlapping the execution of successive instruction in order to improve performance is called as **Pipelining.**

Normally when executing any instruction, execution of the instructions is carried out in different steps such as **Fetch, Decode & Execute**. Each step of the execution is done by a separate functional unit. In case of serial execution new instruction is fetched only when the previous instruction has completed all the steps of execution. But in case of pipelining if any step of the execution is left idle, new instruction can be sent into that unit.

Example: ADD R1, R2, R3

Move LOC, R1

In the above example mentioned, according to the technique, execution of the instructions is done in sequence one by one. i.e. execution of the instruction Move starts only after the completion of execution of instruction ADD. But in the technique of pipelining once after completing the fetching of operands R1 and R2 of Add instruction, next instruction will be fetched into the processor when fetching unit is idle.

Superscalar operation

Multiple instruction pipelines are implemented in order to achieve higher degree of concurrency and this approach is called **superscalar operation.**

1.8 Performance Measurement

Computer community adopted the idea of measuring computer performance using standard programs which are independent of particular system platform and these programs are called as **Benchmark programs.**

Here performance measure is the time taken by computer to execute a given **Benchmark Suite** with respect to Benchmark systems. **Benchmark Suite** is a package of various benchmark programs that is being developed for various application domain such as database, gaming, compilers etc.

To select and publish benchmark programs, a nonprofit organization called **System Performance Evaluation Corporation (SPEC)** is working. To rate the system they have a standard unit which is called as **SPEC rating** which is calculated for each and every benchmark programs individually and is given by the formula:

$$\text{SPEC Rating} = \frac{\text{Running time on refernce computer}}{\text{Running time on computer under test}}$$

If SPEC rating is 50, it means Computer under test is 50 times faster than the standard Benchmark system in terms of performance.

The overall **SPEC rating** is the product of individual rating and is given by the formula:

$$\text{SPEC rating} = \left(\prod_{i=1}^{n} SPECi \right)^{\frac{1}{n}}$$

1.9 Differences between RISC and CISC

(Not mentioned in the syllabus but a commonly asked question)

RISC means Reduced Instruction Set Computer and CISC means Complex Instruction Set Computer which differentiates two major computer machine types based on the structure of the instructions.

RISC	CISC
1. Simple Instruction taking one cycle	Complex instructions taking multiple cycles
2. Only load/store instructions refer memory	Any instruction may refer memory
3. Highly pipelined	Very less or not pipelined
4. Instructions are executed by hardware	Instructions are executed by a special unit called **Microprogram**
5. Fixed format instruction	Variable format instruction
6. Few instruction set and modes	Many instruction set and modes
7. Complexity of the instruction execution is handled by compiler	Handled by Microprogarm.
8. Multiple register sets	Single register set

1.10 Generations of Computer

There are four generations of computer that is discussed very briefly in this section:

1. First Generation

→Stored program concept was introduced by **John Von Neumann** in this generation of the computers

→Programs and data both resided in the same memory space

→Programs were written in the **Assembly level language** which later was translated into **Machine level language.**

→Basic ALU operations were performed by **Vacuum tubes.**

2. Second Generation

→In this generation vacuum tubes were replaced by **Transistors.**

→Application programs were written in High level languages like Java, C, C++ etc

→Compilers were developed to convert High Level Languages into Machine Level Languages.

→Separate I/O processors that could run in parallel with processor was developed

→Even Magnetic core memories were introduced in this generation

3. Third Generation

→Integrated Circuits (IC) technology was introduced to produce high speed processors.

→IC memory replaced magnetic memory.

→Microprogramming, parallelism and pipelining were introduced in this generation.

→Cache and virtual memories were developed.

4. Fourth Generation

→VLSI technology was introduced to build a millions of IC on a small area.

→This lead to the existence of high speed processors called **Microprocessor.**

→Network technology got invented in this generation

☺ **"If you have only a day to live; fill it with study"** ☺