2. Computer Organization and Architecture

Introduction

Computer Technology after its evolution, has made incredible improvement. In early stage of evolution, the computers were not able to store data, the size of the computer was a very huge one with very less computational power.

Now a days, a personal computer has more computational power, more main memory, more disk storage, smaller in size and is available in affordable cost. The advances in the technology brought a rapid rate of improvement in building computers and innovation in computer design.

2.1 Organization of a Small Computer

The model of a computer consists of four basic units, shown in following figure (Figure 2.1). The basic units are:

I. Central Processing Unit (CPU)

II. Input and Output Devices

III. Memory Unit

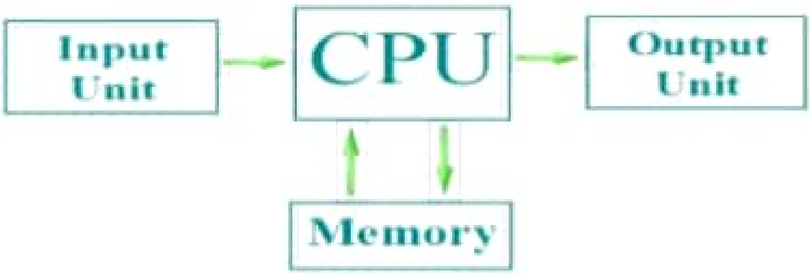

Figure 2.1 : Basic unit of a Computer

2.1.1 Central Processing Unit (CPU)

Central processing unit is also known as a brain of computer. It is a hardware in a computer that carries out the instructions of a computer program by performing the logical, arithmetical and input/output operations of the system.

CPU consist of three basic blocks (refer Figure 2.2):

a. Arithmetic Logical Unit

b. Control Unit

c. Registers

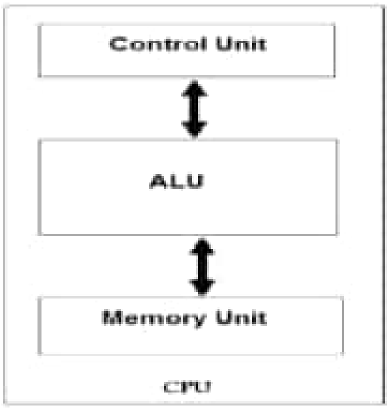

Figure 2.2 : Central Processing Unit

a. Arithmetic Logic Unit (ALU)

ALU performs mathematical, logical and decision operations in a computer. After the processing is done by the ALU, the result is sent to the computer memory. In some computer processors, ALU is divided into two distinct parts:

i. Arithmetic Unit

ii. Logical Unit

i. Arithmetic Unit (AU)

Arithmetic Unit performs arithmetic operations.

- The ALU has got two input parameters named as A and B and one output storage parameter, named as C.
- It perform the operation as: C = A op B (where op stands for operand which can be either of the given operations)

  I. Addition(C=A + B)

  II. Subtraction(C=A - B)

  III. Multiplication(C = A * B)

  IV. Division(C = A / B)

The input data are stored in parameters A and B, and according to the operation specified, the ALU perform the operation and put the result in parameter C.

ii. Logical Unit(LU)

Logical Unit performs the logical operations. ALU is having three logical operands:

I. AND

II. OR

III. NOT

I. AND

The AND operator is represented by '&' and performs logical conjunction on two Boolean expressions.

- If both expressions are True, then the 'AND' returns True.
- If either or both expressions are False, then 'AND' returns False.

As shown in the given table (Figure 2.3) : C = A & B

A	B	C
0	0	0
0	1	0
1	0	0
1	1	1

Figure 2.3 : Logical Unit - AND

II. OR

The OR operator represented by '|' performs logical disjunction on two Boolean expressions.

- If either expression is True, 'Or' returns True.
- If neither expression is True, 'Or' returns False.

As shown in the given table (Figure 2.4) : C = A | B

A	B	C
0	0	0
0	1	1
1	0	1
1	1	1

Figure 2.4 : Logical Unit - OR

III. NOT

The NOT operator is represented by '!' and performs the NOT operation on a single operand.

- It checks the status of a current operand and reverses the value of a resultant operand.
- If the value of an operand is true then it is reversed to false.
- If the value of an operand is false then it is reversed to true.

As shown in the given table (Figure 2.5) : B = !A

A	B
0	1
1	0

Figure 2.5 : Logical Unit - NOT

b. Control Unit (CU)

The Control Unit extract instructions from memory, decodes and executes them, and then sends the necessary signals to the ALU to perform the operations required.

It is a circuitry that directs operations within the computer's processor by directing the input and output of a computer system. The processor then controls how the rest of the computer operates (giving directions to the other parts and systems). It works by gathering input through a series of commands it receives from instructions in a running programs and then outputs those commands into control signals that the computer and other hardware attached to the computer carry out. The control unit basically controls the operations inside the CPU and "directing traffic" in a sense.

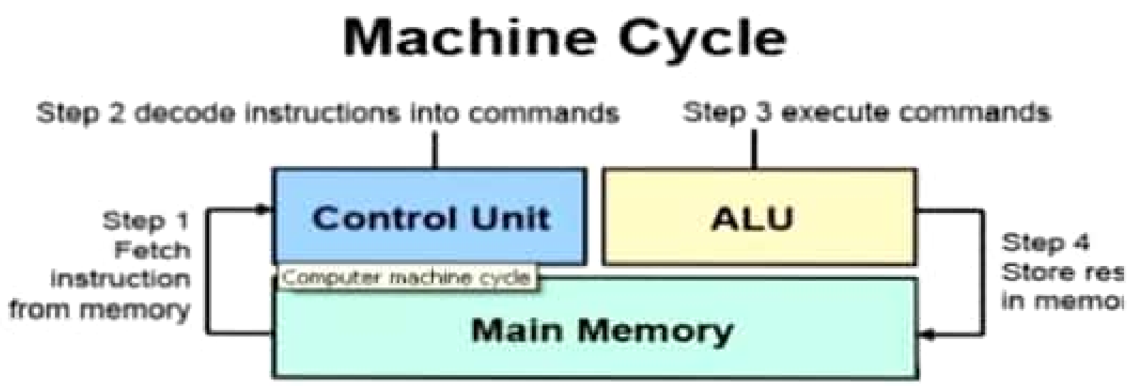

Figure 2.6 : Working of Control Unit

The control unit fetches instruction from main memory then decodes those instructions into commands. ALU then executes those commands and it stores the result in main memory (As shown in Figure 2.6).

c. Registers

Registers are contained within the central processing unit, and holds the data involved in the computation being performed. The size of the register determines the processing capabilities of the CPU. The amount of information that can be held in a register at a time for processing is known as Register Size. The larger the register size, the faster is the speed of processing.

Registers are the fastest and most costly storage units in a computer. Computer instructions are stored in consecutive memory locations and are executed sequentially one at a time.

The control unit reads an instruction from a specific address in memory and executes it. Then it continues by reading the next instruction in sequence and executes it. In this type of instruction sequencing, a counter is needed to calculate the address of the next instruction after execution of the current instruction.

It is necessary to provide a register in the control unit for storing the instruction code after it is read from memory. The computer needs processor registers (refer to Glossary) for manipulating data and a register for holding a memory address.

Registers Performs following operations:-

- Fetch - The Fetch Operation is used for taking the instructions given by the user.
- Decode - The Decode Operation is used for interpreting the meaning of the instructions fetched by registers. Decoding instructions means the CPU will find out which operation is to be performed on the instructions.
- Execute - The Execute Operation is performed by the CPU and the results produced by the CPU are then stored into the memory and after that they are displayed on the user screen.

In given table (Figure 2.7), registers are listed in table with a brief description of their function and the number of bits that they contain.

Register Symbol	Number of Bits	Register Name	Function
DR	16	Data Register	Holds memory operand
AR	12	Address Register	Holds address for memory
AC	16	Accumulator	Processor Register

IR	16	Instruction Register	Holds instruction code
PC	12	Program Counter	Holds address for instruction
TR	16	Temporary Register	Holds temporary data
INPR	8	Input Register	Holds input character
OUTR	8	Output Register	Holds output character

Figure 2.7 : Register Details

- 12 bits instruction word are needed to specify the address of an operand.
- This leaves 3 bits for operation part of the instruction and bit to specify a direct or indirect address.
- The data register (DR) holds the operand read from memory.
- The accumulator register (AC) is a general purpose processing register.
- The instructions read from memory are placed in the instruction register (IR).
- The temporary register (TR) is used for holding temporary data during the processing.
- The input register (INPR) receives an 8-bit character from an input device.
- The output register (OUTR) holds an 8-bit character for an output device.
- A program counter (PC) is a register in a computer processor that contains the address of the instruction being executed at current point of time. As each instruction is fetched, the program counter increases its stored value by 1 and points to the next instruction in sequence. When the computer is reset or restarts, the program counter normally reverts to 0.
- An address register (AR) is a high-speed circuit that holds the addresses of data to be processed or of the next instruction to be executed.

2.1.2 Input and Output Devices

Input and Output devices can be of three types

a. Input Unit/Devices

An input device is a hardware device that sends data to the computer.

- Without input device, a computer would only be a display device and would not allow users to interact with it.

- With the help of input device, data from outside can be supplied to the computer.
- Program or data is read into main storage using input device or secondary storage under the control of CPU.

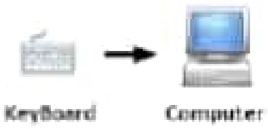

Figure 2.8 : Input Unit

Following are some of input devices:

- Keyboard
- Mouse, touchpad or any other pointing device
- Barcode reader
- Keyboard
- Biometrics (e.g. fingerprint scanner)
- Digital camera and Digital Camcorder
- Touch screen
- Webcam

The drives such as a CD-ROM, DVD are capable of sending information to computer, but they are not input devices. These devices are considered as storage devices.

b. Output Unit/Devices

Any information that has been processed and comes from a computer is considered as an output.

- When someone views the output, they're viewing it on an output device such as a computer monitor or a hard copy print out.
- The figure (Figure 2.9), is an example of output being sent from the computer to a printer.

Figure 2.9 : Output Unit

- The computer results are provided to the user with the help of output unit, the results can also be stored in storage device permanently for future use.

- Output data from main storage is sent to output device under the control of CPU output instructions.
- Example of output devices:
    - Printer
    - Monitor etc.

c. Input/Output Devices

Also referred to as an IO device, is a hardware device that accepts inputted information and also has the capability to output that information. Examples of input/output devices are

- Floppy diskette drive(Figure 2.10)
- CD-ROM drives(Figure 2.11)
- Modem(Figure 2.12)
- USB thumb drives etc.(Figure 2.13)




Figure 2.10 : Floppy diskette *Figure 2.11 : CD-ROM* *Figure 2.12 : Modem* *Figure 2.13 : USB thumb drives*

2.1.3 Memory

Memory unit is an essential component in any computer, needed to store the data and program. CPU works with information stored in memory unit. It is also termed as primary memory or main memory.

A very small computer with a limited application may be able to fulfill its intended task without the need of additional storage capacity. Most general purpose computer is run more efficiently if it is equipped with additional storage beyond the capacity of main memory.

There is just not enough in one memory unit to accommodate all the programs used in typical computer, users accumulate and continue to accumulate large amounts of data processing software. There, it is more economical to use low cost storage devices to serve as a backup for storing the information that CPU is not using currently.

The unit that communicates directly with CPU is called the main memory. The most common memory device used are magnetic disks and tapes, used for storing large data files, system programs and other backup information. Only data currently needed by the processor reside in main memory. Rest of the information is stored in auxiliary memory and transferred to main memory when required.

There are following types of Memories:

a. Main memory

   i. RAM (Random Access Memory)

   ii. ROM (Read Only Memory)

b. Auxiliary Memory

   i. Magnetic Disks

   ii. Magnetic tapes etc.

a. Main Memory

The main memory is central storage unit in a computer system. It is used to store programs and data being used during computer operation.

Following comes under Main Memory:

- RAM (Random Access Memory):

RAM, also known as Volatile memory that loses its contents when the computer or hardware device loses power.

- ROM(Read Only Memory):

ROM is used for storing programs that are permanently in a computer and for tables of constants that do not change its value once the production of computer is completed.

b. Auxiliary Memory

An Auxiliary Memory, also known as high-speed memory bank is used in mainframes and supercomputers (refer to Glossary). The major difference between main memory and auxiliary memory is that CPU directly access the main memory but the auxiliary memory is not accessed by the CPU directly. For this the data is first transferred from auxiliary memory to main memory and then from main memory the data is transferred to the CPU for further processing.

The most commonly used auxiliary memory devices in the computer systems are magnetic disks and magnetic tapes.
There are following auxiliary memories:

- Magnetic Disk
  - A magnetic disk is a circular plate made up of metal or plastic coated with magnetized material.
  - Both the sides of disk are used and several disks may be stacked on one spindle with read/write heads available on each surface.
  - Bits are stored on magnetized surface in spots along concentric circles called tracks.

- Tracks are divided into sections called sectors.
- Disk that are attached and cannot removed by occasional user are called Hard disk.
- A disk drive with the removable disks are called a floppy disks.

- Magnetic Tapes
    - A magnetic tape is also known as Removable Media consists of electric, mechanical and electronic components to provide the parts and control mechanism for a magnetic tape unit. The major difference between magnetic disk and magnetic tape is that, magnetic tapes support sequential access.
    - The tape is a plastic coated strip with a magnetic recording medium.
    - The Bits are recorded as magnetic spots on tape along several tracks.
    - Tracks are divided into sections called sectors.
    - Disk that are attached and cannot removed by occasional user are called Hard disk.
    - A disk drive with the removable disks are called a floppy disks.

2.2 Bus

A bus enables a computer processor to communicate with the memory or a video card.

- The processor, main memory and Input/Output devices can be interconnected through common data communication lines which are termed as a common bus, as shown in Figure 2.10.
- The primary function of common bus is to provide a communication path between devices for transfer of data.
- The bus include control lines needed to support interrupts and arbitration.

Basically, all computers utilize two types of bus:

I. Internal Bus : also known as local bus,enables a communication between internal components such as a computer video card and memory.

II. External Bus : also known as expansion bus, communicates with external components such as a USB or SCSI device(refer to Glossary).

The bus lines which are used to transfer data are grouped into three categories:

I. Data Bus

II. Address Bus

III. Control Bus

- Data Bus
    - Data bus is a computer subsystem that allows transferring of data from one component to another on a motherboard, or between two computers.
    - This includes transferring data to and from the memory, or from the central processing unit (CPU) to other components.
- Address Bus
    - An address bus is used to specify a physical address.
    - When a processor or Direct memory access (DMA) enabled device needs to read or write to a memory location, it specifies that memory location on the address bus.
- Control Bus
    - A control bus is a part of a computer bus, used by CPUs for communicating with other devices within the computer.
    - It carries commands from the CPU and returns status signals received from the devices.

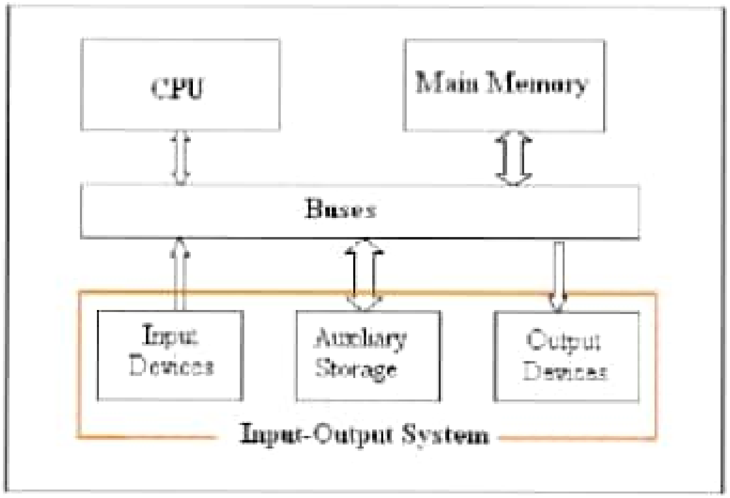

Figure 2.14 : Working of Bus

2.3 Addressing Modes

Addressing modes are the ways how the address of an object/operand is being accessed.

- In other words, 'The strategy to identify WHERE the operand is located.'
- In computer architecture every tasks user attempts to perform is converted into an instruction (i.e Machine codes) and that instruction consists of some operands to work on it.

Ex: Consider this machine code ADD A, B here ADD is opcode and A,B are operands. This means add A and B and store the result in A. Here A and B refers to some locations where the numbers are actually stored. This a kind of indirect addressing mode.

There are several types of addressing modes present. Each addressing mode has a unique formula to locate the operand.
Some of them are mentioned below:

I. Immediate Addressing Mode (Refer to glossary for detail).

II. Direct Addressing Mode (Refer to glossary for detail).

III. Indirect Addressing Mode (Refer to glossary for detail)

Glossary

- **Byte** - A byte is a unit of measurement used to measure data. 1 byte contains 8 binary bits, or a series of eight zeros and ones. So, each byte can be used to represent 2^8 or 256 different values.

- **Processor Registers** - A processor register is a local storage space on a processor that holds data that is being processed by CPU. They generally occupy the top-most position in the memory hierarchy and provide high-speed storage space and fast access to data. A register may include the address of memory location instead of the real data itself.

- **Immediate Addressing Mode** - In immediate addressing, the instruction contains the value to be used. In many cases, an instruction requires a constant quantity and a bit pattern which will never change no matter when or how often the instruction is executed. This mode of bit pattern is called the immediate addressing mode. In the given Figure 2.15, the op-word for instruction includes a group of bits which identifies this mode of addressing. Since the instruction is located in program memory the constant is itself in program memory.
  In Immediate addressing mode, the instruction does not explicitly state the location of operand, rather, it explicitly states the operand itself. This addressing mode is used when a particular constant value is to be fixed within a program itself. The value is found in memory "immediately" after the instruction code word that may never change at any time.

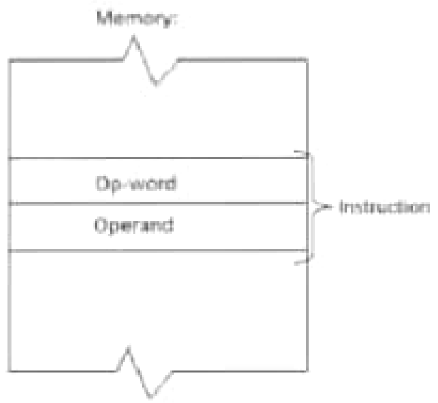

Figure 2.15 : Immediate Addressing Mode

- **Direct Addressing Mode** - In direct addressing, the instruction tells where the value can be found, but the value itself is out in the memory. Here the instruction explicitly states the location of an operand or destination that is either in memory or in a processor register.
  There are two sub-classification of direct addressing mode-

  - **Absolute Addressing** - The mode is referred to as Absolute addressing, when the location is in memory.

- **Register Direct Addressing** - The mode is referred to as Register Direct Addressing, when the location is a processor register.

In Figure 2.16 , the instruction specifies the address of the operand. In Figure 2.17, the instruction specifies the register containing the operand.

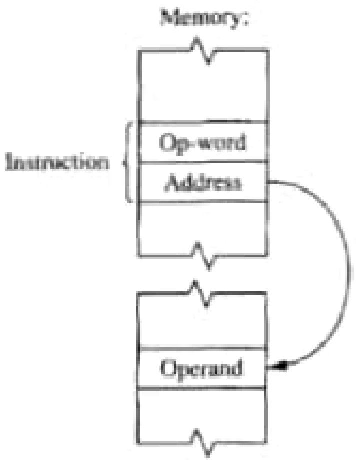

Figure 2.16 : Absolute Addressing

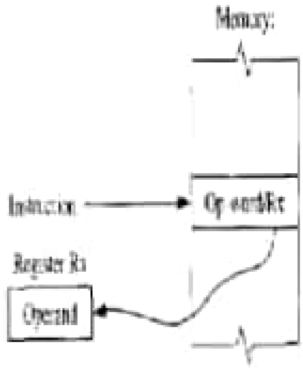

Figure 2.17 : Register Direct Addressing

- **Indirect Addressing Mode** - Indirect addressing uses address register to point at the location of the operand in memory, that is, the address of an operand is obtained indirectly via an address register. In the Indirect Addressing Mode the instruction tells the processor neither the address of the operand nor the operand itself. The instruction may either state explicitly the address of a location in a memory or the name of a processor register. In the given Figure 2.18, the op-word for the instruction includes a group of bits which identifies this mode of addressing and the indirect address is specified in one or more additional post-words. Indirect Addressing is used when a program operates upon different data values under different circumstances.

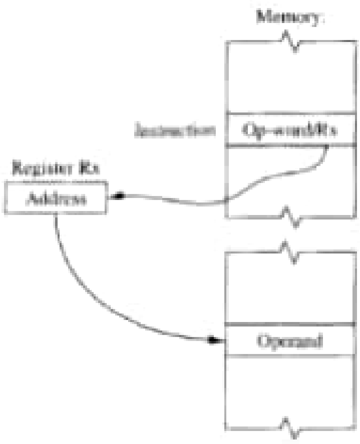

Figure 2.18 : Indirect Addressing Mode

- **SCSI device** - A SCSI device is connected to the SCSI (Small Computer System Interface). The SCSI device functions on its own and uses the SCSI interface to communicate with the rest of the computer. Figure 2.19 and Figure 2.20 are example of SCSI devices.

Figure 2.19 : SCSI Device

Figure 2.20 : SCSI Device

- **Mainframes** – A Mainframe computer is very large and powerful computer, dedicated to lengthy and complex calculations. Mainframe is also referred as the hub of a system, capable of supporting hundreds or even thousands of users simultaneously. A mainframe is an ultra-high-performance computer made for processor-intensive and high-volume computing. They are used for scientific purposes and large businesses.
- **Supercomputers** - A supercomputer is a high performance computing machine. It is the fastest type of computer having extremely fast processing speeds. Supercomputers are very costly and are used for specialized applications, where immense amounts of mathematical calculations are required. For example, animated graphics, nuclear energy research, weather forecasting and petroleum exploration works on supercomputer.

The major difference between a supercomputer and a mainframe is that a supercomputer uses its power in executing few programs as fast as possible, whereas a mainframe uses its power to execute many programs concurrently.

Summary

In given document, we have presented a brief overview of Computer organization and architecture.

- **Computer Architecture** - refer to the relationship between different hardware components of a computer system.
- **Computer Organization** - refer to how operational attributes are linked together to realized the architecture specifications.

We have covered basic units of a computer:

- **Central Processing Unit (CPU)** - Controls the operation of the computer and performs its data processing functions.
    - **Arithmetic Logic Unit (ALU)** - It performs computer's data processing functions.
    - **Control Unit(CU)** - It performs computer's data processing functions.
    - **Registers** - Registers provides internal storage to the CPU.
- **Input Unit/Devices** - They provide input to the CPU.
- **Output Unit/Devices** - They provide output from CPU to the user.
- **Memory Unit** - Memory Unit is needed to store the data and program.
- **Bus** - It consists of set of wires that allow data to be passed back and forth.
- **Addressing Mode** - Addressing modes are the ways how the address of an object/operand is being accessed.