BACHELOR OF COMPUTER APPLICATIONS

SEMESTER 3

DCA2103
COMPUTER ORGANIZATION

Unit 1

Computer Evolution

Table of Contents

1. INTRODUCTION

We are all familiar with what a computer is in a specific, contemporary sense. Personal computers are found in most aspects of daily life, and for some, it is hard to even imagine a world without them. A computer is capableof computing and processing data. Computers run instructions that aregiven to it. Instructions are low-level commands like add, subtract, or compare two numbers, run some instruction if the comparison is true, and run some other instruction if it's false. In this unit, you will be introduced to the definition, history, and other fundamental concepts of computers.

1.1 Objectives:

After studying this unit, you should be able to:

- ❖ *Define computer*
- ❖ *Give the history of computers*
- ❖ *Give the classification of computers*
- ❖ *Explain the basic structure of the computer*
- ❖ *Discuss the central processing unit of the computer*
- ❖ *Explain the bus or interconnection system*
- ❖ *Describe the von neumann architecture.*

2. BRIEF HISTORY OF COMPUTER

A computer is a digital device or a machine for manipulating data according to a list of instructions. Earlier the term "computer" referred to a person who performed numerical calculations often with the aid of a mechanicalcalculating device. The group of instructions is called the **program** and internal storage is called **memory**. Information fed to a computer can be categorized as either *instruction* or *data*. **Instructions** *are* the commands given to the computer which tells what operation that computer should do. In other words, *the Instructions* are the explicit commands that govern the transfer of information within the computer as well as between the computer and its I/O devices and specify the operations to be performed.

2.1 Mechanical and Electromechanical Ancestors

The first machine to attract widespread attention was built in 1642, by a French philosopher and scientist B. Pascal. This machine was a mechanicalcounter for carrying operations like addition and subtraction. It consisted of two sets of gears with teeth (counter wheels) for representing decimalnumbers. Each gear had 10 decimal digits engraved on it. The position ofthe dial indicated the decimal value. Each set of dials was used to temporarily hold a number like a register. One special register was considered like an accumulator which held the running total. There was one more register which was used to enter a number to be added or subtracted from the accumulator. Thus when the machine was set in motion, the numbers in the two sets of dials were added, with the result appearing in theaccumulator. By means of complement number representation, the machine could accomplish both addition and subtraction.

In 1671, the German mathematician and philosopher **Gottfried Leibniz** constructed a calculator that could perform multiplication and division along with addition and subtraction. Leibniz's machine could perform additional operations in a repetitive step-by-step fashion using chains and pulleys. Thismachine was the forerunner of many machines that were also called as four function calculators.

In 1750, punch cards were developed to specify the pattern in thetechnology of weaving. Symbols stored on punch cards were used to represent instructions. There were a series of modifications and finally in 1801,Joseph Marie Jacquard made an improvement to the

textile loom. The machine used a series of punched paper cards. He produced a very successful loom in which all the power was supplied mechanically and all the control through the punch card. And the cards were moved through loomapparatus. The presence or absence of a hole dictated the movements of parts of the loom to create the desired pattern. Thus this loom was aprogrammable process control machine with the 'program' supplied on punch cards that were used as a template to allow his loom to weave intricate patterns automatically. The resulting Jacquard loom was an important step in the development of computers because the use of punched cards to define woven patterns can be viewed as an early, albeit limited, form of programmability.

The ideas behind the general-purpose programmable computer were developed by **Charles Babbage** in the 19th century. He designed two machines: **Difference Engine and Analytical Engine**.

Difference Engine: The difference engine was designed to calculate the entries of a table automatically and transfer them via steel punches to an engineer's plate from which the tables could be printed. The only arithmetic operation performed by this machine was addition. However using a mathematical technique called a *method of finite difference*, a large numberof useful formulas could be computed either exactly or approximately using only additions. The formulas included polynomials and trigonometric functions like sine functions. The structure of Babbage's Difference Engine is shown in figure 1.1 below.
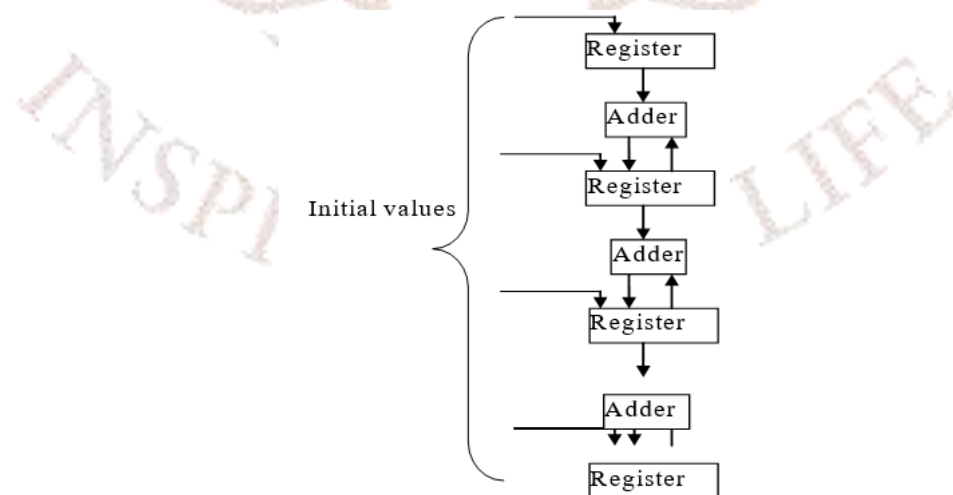

Fig 1.1: Structure of Babbage's Difference Engine

The difference engine consisted of a number of mechanical registers. Each register consisted of a set of counter wheels and could store a decimal number. Pairs of adjacent registers were connected by an addingmechanism. To compute results, initial values were loaded into the registers.When the difference engine was driven by a suitable motor it could then perform a series of steps to give an answer. This machine could accommodate third-degree polynomials and 15-digit numbers. Babbage proposed to build a machine that would accommodate sixth-degree polynomials and 20-digit numbers but could not succeed as the British government withdrew its support. In the meantime, he came up with a more powerful and ambitious machine which he called an **Analytical engine**.

Analytical engine: This machine was designed to be a general-purpose device that is capable of performing any mathematical operation automatically. The Analytical engine's structure is shown in Figure 1.2 below.
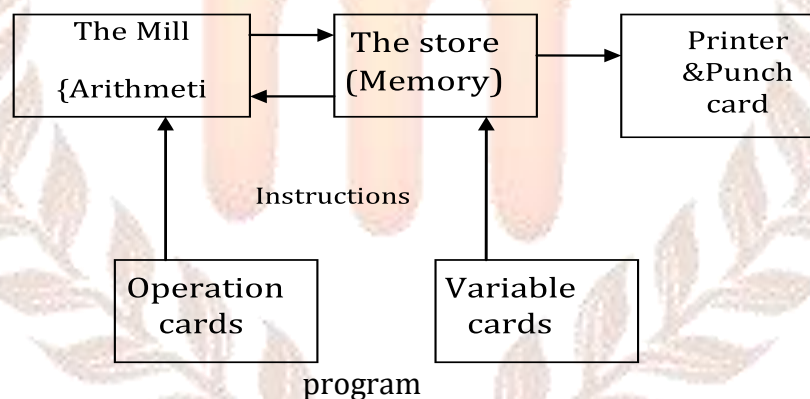

Fig 1.2: Structure of Babbage's analytical engine

The Analytical engine consists of five units of which two main units are the **Store** and the **Mill**. To control the sequence of operations, punch cards were used which were of the type developed earlier for Jacquard's loom. Each of these components are described below:

- **The Store:** Is a memory unit comprising sets of counter wheels
- **The Mill:** It corresponds to a modern Arithmetic Logic Unit. It was capable of performing four basic arithmetic operations. It operated on pairs of mechanical registers and produced a result stored in another mechanical register, all of which were stored in the store (a memory as defined above)

- **The punch cards:** It constituted a computer program and were divided into two groups:
  - o **90 Operation cards:** These cards were used to control the operation of the Mill. Each operation card selected one of the four possible operations (that is addition, subtraction, multiplication, and division) at each step in the program.
  - o **Variable cards:** It is used to select the memory locations to be used by a particular operation, that is the source of the input operands andthe destination of the results.
- **Output:** Is a printer or a card punch device so that output data is either printed on a printer or punched on cards.

Large-scale automated data processing of punched cards was performedfor the US Census in 1890 by tabulating machines designed by Herman Hollerith and manufactured by the Computing Tabulating Recording Corporation, which later became IBM.

During the first half of the 20th century, many scientific computing needs were met by increasingly sophisticated analog computers. However, these were not programmable and generally lacked the versatility and accuracy of modern digital computers.

Computers take numerous physical forms. Early electronic computers were the size of a large room, consuming as much power as several hundred modern personal computers. Today, computers can be made small enough to fit into a wristwatch and be powered by a watch battery. Society has come to recognize personal computers and, the laptop computer as icons ofthe information age. However, the most common form of computer in use today is by far the embedded computer. Embedded computers are small, simple devices that are often used to control other devices. For example, they may be found in machines ranging from fighter aircraft to industrial robots, digital cameras, and even children's toys.

The ability to store and execute programs makes computers extremelyversatile. Computers can range from personal computers to supercomputerswhich are all able to perform the same computational tasks so long as time and storage capacity are not considerations. The use of digital electronics and more flexible programmability were important steps, in the development of the digital electronic computer.

First Generation Electronic Computers (1937-1953)

These devices used electronic switches, in the form of vacuum tubes, instead of electromechanical relays. In 1941, J. V. Atanasoff, a professor of physics and mathematics at Iowa State, and graduate student Clifford Berry had succeeded in building a machine that could solve 29 simultaneous equations with 29 unknowns The first electronic computer *ENIAC (Electronic Numerical Integrator and Computer)* was designed and constructed by at the Moore school of engineering (university of Pennsylvania). This was able to perform 5000 additions or subtractions per second.

EDSAC *(Electronic Delay Storage Automatic Calculator)* was one of the first computers to implement the stored program (von Neumann) architecture. Several developers came up with a far more flexible and elegant design, which came to be known as the *stored program architecture* or *von Neumann architecture*. This design was first formally described by John von Neumann in the paper "First Draft of a Report on the EDVAC", published in 1945. Nearly all modern computers implement some form of stored program architecture.

Second Generation (1954-1962)

The second-generation witnessed several important developments at all levels of computer system design. Electronic switches in this era werebased on discrete diode and transistor technology and the first machines to be built with this technology include TRADIC at Bell Laboratories in 1954 and TX-0 at MIT's Lincoln Laboratory. Important commercial machines of this era include the IBM 704 and its successors, the 709 and 7094. In the 1950s the first two supercomputers were designed specifically for numeric processing in scientific applications.

Third Generation (1963-1972)

Technology changes in this generation include the use of integrated circuits, or ICs (semiconductor devices with several transistors built into one physical component), semiconductor memories, microprogramming as a technique for efficiently designing complex processors, and the introduction of operating systems and time-sharing. Early in this third generation, Cambridge University and the University of London cooperated in the development of CPL (Combined Programming Language, 1963). The first ICs were based on small-scale integration (SSI) circuits, which had around 10 devices per circuit (or 'chip'), and

evolved to the use of medium-scale integrated (MSI) circuits, which had up to 100 devices per chip.

Fourth Generation (1972-1984)

Large scale integration (LSI - 1000 devices per chip) and very large scale integration (VLSI - 100,000 devices per chip) were used in the construction of the fourth generation computers. In 1972, Dennis Ritchie developed the Clanguage from the design of the CPL and Thompson's B. IBM worked with Microsoft during the 1980s to start what we can really call PC (Personal Computer) life today.

Fifth Generation (1984-1990)

Fifth-generation is based on parallel processing hardware and ArtificialIntelligence (AI) software. In this generation, the higher-level languages like C and C++, Java, .Net, etc. are used. The ULSI (Ultra Large Scale Integration) technology became popular and this resulted in the productionof microprocessor chips having ten million electronic components. Parallel Processing is coming and showing the possibility that the power of many CPU's can be used side by side, and computers will be more powerful than those under central processing. The main features of fifth-generation are:

- ULSI technology
- Development of artificial intelligence
- Advancement in Parallel Processing
- Advancement in Superconductor technology
- More user-friendly interfaces with multimedia features
- Availability of very powerful and compact computers at cheaper rates

Laptop, Notebook, UltraBook and Chrome Book are some of the computertypes of the fifth generation.

Sixth Generation (1990- onwards)

This sixth generation of computers differs from previous generations interms of processing speed, size, and the complexity of tasks that computers can now perform. One of the most important changes that you can observe in this sixth generation is the explosive growth of wide-area networking. Network bandwidth and speed of operation and networking capabilities havekept developing tremendously. Also, this generation resulted in many gains

in parallel computing, both in the hardware area and in developing algorithms to exploit parallel architectures. Now the processors are being constructed with a combination of RISC, pipelining, and parallel processing. Personal computers (PCs) can now operate with Gigabit per second processors, multi-gigabyte disks, hundreds of Mbytes of RAM, color printers, high-resolution graphic monitors, and graphical user interfaces. Also, the parallel systems now compete with vector processors in terms of total computing power, and combinations of parallel/vector architectures are well established.You can even expect parallel systems to dominate the future.

Microprocessors

As the density of elements on memory chips has continued to rise, t h e density of elements on processor chips also increased. A breakthrough was achieved in 1971 when Intel developed its 4004. The 4004 was the first chipto contain all of the components of a CPU on a single chip: It was called the microprocessor. The next microprocessor that was introduced by Intel was Intel 8008 in 1972. This was the first 8-bit microprocessor and was almost twice as complex as the 4004. Table 1.1 shows the evolution of Intel Microprocessors.

Table 1.1: Evolution of Intel Microprocessors.

(a) 1970s Processors

	4004	8008	8080	8086	8088
Introduced	1971	1972	1974	1978	1979
Clock speeds	108 kHz	108 kHz	2 MHz	5 MHz, 8 MHz, 10 MHz	5 MHz, 8 MHz
Bus width	4 bits	8 bits	8 bits	16 bits	8 bits
Number of transistors	2,300	3,500	6,000	29,000	29,000
Feature size (μm)	10		6	3	6
Addressable memory	640 Bytes	16 KB	64 KB	1 MB	1 MB

(b) 1980s Processors

	80286	386TM DX	386TM SX	486TM DX CPU
Introduced	1982	1985	1988	1989
Clock speeds	6 MHz–12.5 MHz	16 MHz–33 MHz	16 MHz–33 MHz	25 MHz–50 MHz
Bus width	16 bits	32 bits	16 bits	32 bits
Number of transistors	134,000	275,000	275,000	1.2 million
Feature size (μm)	1.5	1	1	0.8–1
Addressable memory	16 MB	4 GB	16 MB	4 GB
Virtual memory	1 GB	64 TB	64 TB	64 TB
Cache	--	--	--	8 kB

(c) 1990s Processors

	486TM SX	Pentium	Pentium Pro	Pentium II
Introduced	1991	1993	1995	1997
Clock speeds	16 MHz–33 MHz	60 MHz–166 MHz,	150 MHz–200 MHz	200 MHz–300 MHz
Bus width	32 bits	32 bits	64 bits	64 bits
Number of transistors	1.185 million	3.1 million	5.5 million	7.5 million
Feature size (μm)	1	0.8	0.6	0.35
Addressable memory	4 GB	4 GB	64 GB	64 GB
Virtual memory	64 TB	64 TB	64 TB	64 TB
Cache	8 kB	8 kB	512 kB L1 and 1 MB L2	512 kB L2

(d) Recent Processors

	Pentium III	Pentium 4	Core 2 Duo	Core 2 Quad
Introduced	1999	2000	2006	2008
Clock speeds	450–660 MHz	1.3–1.8 GHz	1.06–1.2 GHz	3 GHz
Bus sidth	64 bits	64 bits	64 bits	64 bits
Number of transistors	9.5 million	42 million	167 million	820 million
Feature size (nm)	250	180	65	45
Addressable memory	64 GB	64 GB	64 GB	64 GB
Virtual memory	64 TB	64 TB	64 TB	64 TB
Cache	512 kB L2	256 kB L2	2 MB L2	6 MB L2

Sixth Generation (1990 -)

This generation brought about gains in parallel computing in both the hardware and in improved understanding of how to develop algorithms to exploit parallel architectures. Workstation technology continued to improve, with processor designs now using a combination of RISC, pipelining, and parallel processing. Personal computers (PCs) now operate with Gigabit persecond processors, multi-gigabyte disks, hundreds of Mbytes of RAM, color printers, high-resolution graphic monitors, stereo sound cards, and graphical user interfaces. This generation has brought about microcontrollertechnology.

Modern Definition of computer

The modern definition of a *computer* is an electronic device that performs calculations on data, presenting the results to humans or other computers ina variety of useful ways. A computer is a complex system that contains millions of elementary electronic components. Hence computers can be considered to have hierarchic nature. A hierarchic system is a set of interrelated subsystems each again hierarchical in nature until we reach the lowest level of the elementary subsystem. A hierarchic nature of a complex system is essential to both its

design and its description. At each level, the designer is concerned with the structure and function.

Definitions of few terms

- **Structure:** It defines the way in which the components of a computer are interrelated.
- **Function:** It defines the operation of each individual component as apart of the structure.
- **Architecture:** It refers to those attributes of a computer system whichare visible to a programmer.
- **Organization:** It refers to the operational units of a computer and their interconnections, and how they implement the architecture of the system.

2.2 Classification Of Computers

Computers can be classified into many types based on their **size, speed, and cost.**

- The smallest machines are called *microcomputers*. For example a personal computer, that is widely used in schools, homes, and business offices.
- The next higher-level machine is called a *minicomputer*. Minicomputersare widely used in payroll, scientific computing applications, etc.
- *Mainframes* are used for business data processing when computingand storage capacity is larger than what the minicomputers can handle.
- *Supercomputers* are used for large-scale numerical calculations suchas weather forecasting, missile launching, aircraft, and simulation.

2.3 Designing For Performance

Year by year, the cost of computer systems continues to drop dramatically, while the performance and capacity of those systems continue to rise equally dramatically. For example, desktop applications that require thegreat power of today's microprocessor-based systems include

- Image processing
- Speech recognition
- Videoconferencing

- Multimedia authoring
- Simulation modeling

Workstation systems now support highly sophisticated engineering and scientific applications, as well as simulation systems, and have the ability to support image and video applications.

What gives Intel x86 processors or IBM mainframe computers such mind-boggling power is the relentless pursuit of speed by processor chip manufacturers. While processor power has raced ahead at breakneck speed, other critical components of the computer have not kept up. The result is a need to look for performance balance: an adjusting of the organization and architecture to compensate for the mismatch among the capabilities of the various components. Designers constantly strive to balance the throughput and processing demands of the processorcomponents, main memory, I/O devices, and the interconnection structures.

> **Self-Assessment Questions – 1**
>
> 1. The list of instructions is called_____.
> 2. _____is was a machine designed to be a general purpose device that is capable of performing any mathematical operation automatically.
> 3. _____ are widely used in payroll, scientific computingapplications etc.

3. STRUCTURE OF A COMPUTER SYSTEM

A computer is an entity that interacts in some or the other way with its external environment. All of its linkages to the external environment are classified as peripheral devices or communication lines. In particular, the basic model of a computer is shown in figure 1.3. It consists of four main components, **Central Processing Unit (CPU), Memory, Input / Output,** and a **Bus**. The bus can also be a wire or a communication line or in general, it can be referred to as a system interconnection.
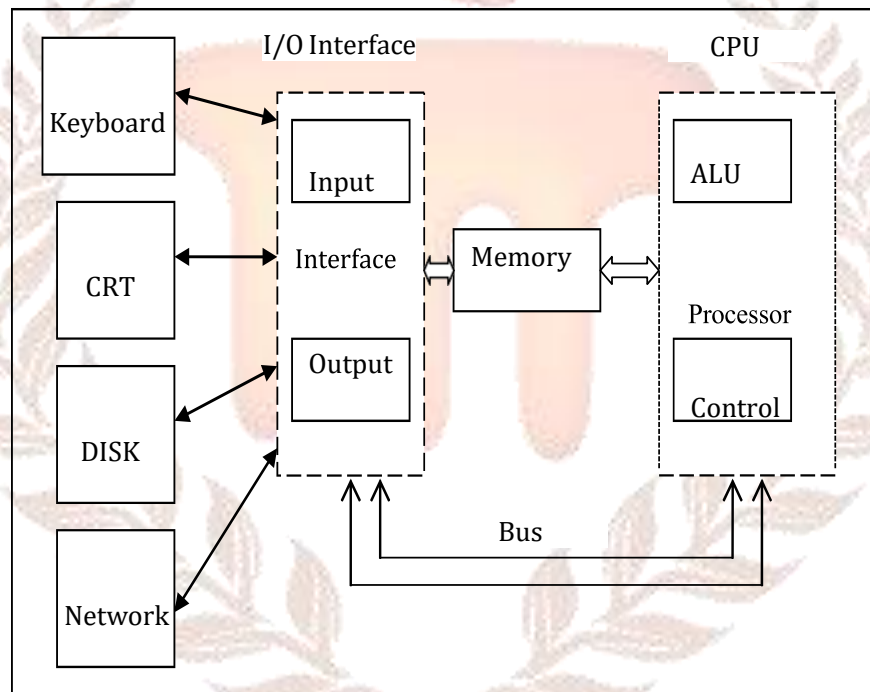
Fig 1.3: Functional units of a computer

- **CPU:** This is the computational unit and is the computer's heart. This entity controls the operations of the computer and performs its data processing functions. It is usually referred to as a processor. All the actions of all components are controlled by the control unit of the CPU. CPU comprises two units called **Arithmetic Logic Unit (ALU),** where all the arithmetic and logical operations are performed and the **Control Unit (CU)** which coordinates with all other units for proper system operation.

- **Memory:** Memory is used to store the instructions, data, and result as well. The memory unit is an integral part of a computer system. The main function of a memory unit is to store the information needed by the system.

- **Input/output interface:** They are used to move data from the computer and from its external environment. The external environment may be an input or an output device like Printer, displays, keyboard etc.

- **System interconnection:** This constitutes some mechanism that provides for communication among CPU, Memory & I/O. These can be referred to as a system BUS.

Traditionally the computer system consists of a single CPU. But some machines like multiprocessing involve the use of multiple CPU's and share a single memory.

Central Processing Unit

It is the heart or core component of CPU. Figure 1.4 shows the basic functional components of the Central Processing Unit.
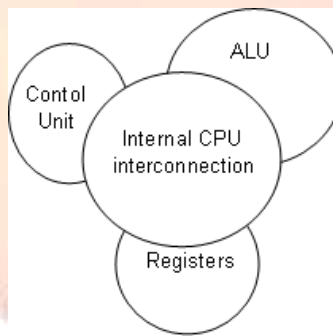

Fig 1.4: Basic Block diagram of CPU

Its major structural components are the Control unit, ALU, Registers, and CPU interconnections.

- **Control Unit:** It controls the operations of the CPU and hence the computer system.

- **Arithmetic logic unit (ALU):** It performs the computers data processing functions

- **Registers:** These form the internal memory for CPU. It provides storage, internal to the CPU.

- **CPU interconnections:** It provides means for communication among the control unit, ALU, and registers of the CPU.

The CPU must carry out the tasks given below:

1. Read the given instructions

2. Decode them

3. Get operands for execution

4. Process the instruction
5. Give out / store the result

To carry out these tasks the CPU needs to temporarily store some data. It must remember the location of the last instruction so that it can know where to get the next instruction. It needs to store instruction and data temporarily while an instruction is being executed. In general, the CPU needs an internalmemory for all stores, either instruction or data. The CPU contains a handful of registers which act as local variables. The CPU runs instructions and performs computations, mostly by the ALU. The registers are the only memory the CPU has. Register memory is very fast for the CPU to access since it resides in the CPU itself. However, the CPU has rather limited memory. All the local memory it uses is in registers. It has very fast accessto registers, which are onboard on the CPU. It has much slower access to RAM.

- **Arithmetic logic unit (ALU)**

  The ALU is a collection of logic circuits designed to perform arithmetic (addition, subtraction, multiplication, and division) and logical operations (not, and, or, and exclusive-or). It's basically the calculator of the CPU. When arithmetic or logical operation is required, the values andcommand are sent to the ALU for processing.

- **Control Unit:**

  The purpose of the control unit is to control the system operations by routingthe selected data items to the selected processing hardware at the right time. The Control unit acts as nerve center for the other units.

- **Instruction decoder:**

  All instructions are stored as binary values. The instruction decoder receives the instruction from memory, interprets the value to see what instruction is to be performed, and tells the ALU and the registers which circuits to energize in order to perform the function.

- **Registers:**

  The registers are used to store the data, addresses, and flags that are inuse, by the CPU.

Memory Units

Memory is basically a large array of bytes. The main function of a memory unit is to store the information needed by the system. Information stored can be data, an instruction that is nothing but programs and maybe some garbage. Memory locations that do not contain any valid data may store some arbitrary values and hence they are termed garbage data. The memory unit is an integral part of a computer system. The system performance is largely dependent on the organization, storage capacity, and speed of operation of the memory system. The CPU can read or write to the memory, but it's much slower than accessing registers. Nevertheless, you need memory because registers simply hold too little information. Most of the memory is in RAM, which can be thought of as a large array of bytes as shown in figure 1.5. In an array, we can refer to individual elements using an index. In computer organization, indexes are more commonly referred to as addresses. Addresses are the numbers used to identify successive locations. Specifying its address and command that performs the storage or retrieval process can access a word.
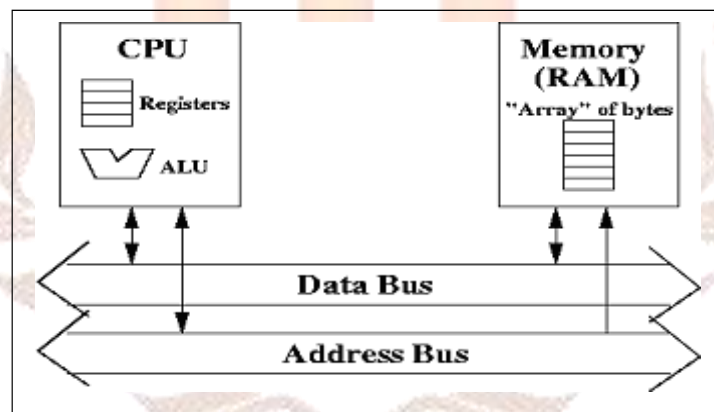

Fig 1.5: System showing the registers and memory as array of bytes

The number of bits in each word is called as *word length* of the computer. Large computers usually have 32 or more bits in a word. The word length of microcomputers ranges from 8 to 32 bits. The capacity of the memory is one

factor that decides the size of the computer. Data are usually manipulated within the machine in units of words, multiples of words, or parts of words. During execution, the program must reside in the main memory. Instructions and data are written into the memory or read out from the memory under the control of a processor. Most memory is byte-addressable, meaning that each address refers to one byte of memory.

The bulk of the memory is stored in a separate device called RAM usually called ***physical memory***. RAM stores programs as well as data. The CPU fetches an instruction in RAM to a register, which is referred to as an instruction register. This register then determines what instruction it has, andexecutes the instruction. Executing the instruction may require loading data from RAM to the CPU or storing data from the CPU to RAM. The time required to access one word is called the ***memory access time***.

Classification of Memory system of a Computer

The memory system of a computer can be broadly classified into four groups.

- **Internal Memory**

  Internal memory refers to a set of CPU registers. These serve as working memory, storing temporary results during the computation process. They form a general-purpose register file for storing the data asit is processed. Since the cost of these registers is very high only a few registers can be used in the CPU.

- **Primary Memory**

  Primary memory is also called ***main memory***, which operates at electronic speeds. CPU can directly access the program stored in the primary memory. Main memory consists of a large number ofsemiconductor storage cells. Each cell is capable of storing one bit of information. *Word* is a group of these cells. Main memory is organizedso that the contents of one word, containing n bits, can be stored or retrieved in one basic operation.

- **Secondary Memory**

  This memory type is much larger in capacity and also much slower than the main memory. Secondary memory stores system programs, large data files, and information which is not regularly used by the CPU. When the capacity of the main memory is exceeded, the additional information is stored in the secondary memory. Information from the secondary memory is accessed indirectly through the I/O programs that transfer the information between the main memory and secondary memory. Examples of secondary memory devices are magnetic hard disks and CD-ROMs.

- **Cache Memory**

  The performance of a computer system will be severely affected if the speed disparity between the processor and the main memory is significant. The system performance can be improved by placing a small,fast-acting buffer memory between the processor and the main memory. This buffer memory is called *cache memory*. The cost of this memory is very high.

Input / Output and I/O Interface

Any movement of information from or to the computer system is considered as Input/Output. The CPU and its supporting circuitry provide I/O methods. Input and output units are usually combined under the term input-output unit (I/O). For example consider the keyboard of a video terminal, which consistsof the key- board for input and a cathode ray tube display for output.

There are a wide variety of peripherals that deliver different amounts of data, run at different speeds, and present data in different formats. All the I/Operipherals are slower than the CPU and RAM. Hence I/O units need proper I/Ointerfaces.

Input Devices:

The computer accepts the coded information through the input unit. It has the capability of reading the instructions and data to be processed. The most commonly used input device is the keyboard of a video terminal. This is electronically connected to the processing part of a computer. The keyboardis wired such that whenever a key is pressed the corresponding letter ordigit is automatically translated into its corresponding code and is directly sent either to memory or the processor.

Output Devices

The output unit displays the processed results. Examples are video terminals and graphic displays.

I/O devices do not alter the information content or the meaning of the data. Some devices can be used as output only e.g. graphic displays.

Following are the Input / Output techniques

- o Programmed
- o Interrupt driven
- o Direct Memory Access (DMA)

System Interconnection / BUS

"A bus is a communication pathway connecting two or more devices." A key characteristic of a bus is that it is a **shared transmission medium**. Multipledevices connect to the bus, and a signal transmitted by any one device is available for reception by all other devices attached to the bus. If twodevices transmit during the same time period, their signals will overlap and become garbled. Thus, only one device at a time can successfully transmit. The communication between the external environment and CPU is established through the System Bus. The system bus is classified into three different types, depending on whether it carries Data, Control, or Address information, and is indicated in figure 1.5.

Self-Assessment Questions – 2

4. ALU stands for_____.

5. The number of bits in each word is called as_____of the computer.

6. *Primary memory is also called as_____.*

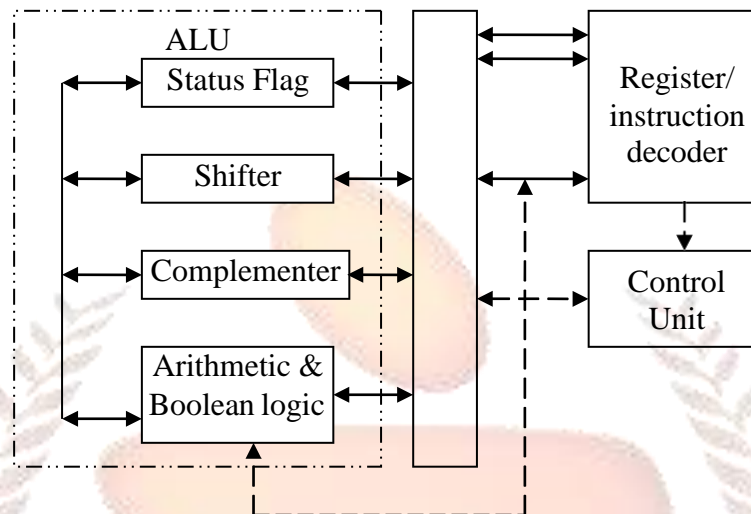
4. ARITHMETIC LOGIC UNIT (ALU)

Fig 1.6: CPU showing internal components of ALU

The basic components of ALU are as shown in figure 1.6. The Arithmetic and Logic Unit is the core of any processor. It performs the calculations on the input data given. ALU is needed to transfer data between the various registers. Always ALU operates only on data in the internal CPU memory. ALU is capable of performing any arithmetic and Boolean operations. An Arithmetic-Logic Unit or ALU can be considered as a combination of various circuits in a single circuit that are used to execute data processing instructions. The complexity of ALU is determined by the way in which its arithmetic instructions are realized. Simple ALUs that perform fixed-point addition and subtraction, as well as logical operations, can be realized by combinational circuits.

The ALU realized using a combinational logic that are basically constructed from AND, OR, and NOT gates. It is basically an implementation of a Boolean function. A generic diagram for a combinational logic circuit of ALU is shown in figure 1.7. In general, a combinational logic circuit has inputs, which are divided into data inputs and control inputs, and outputs. Control inputs tell the circuit what to do with the data inputs.
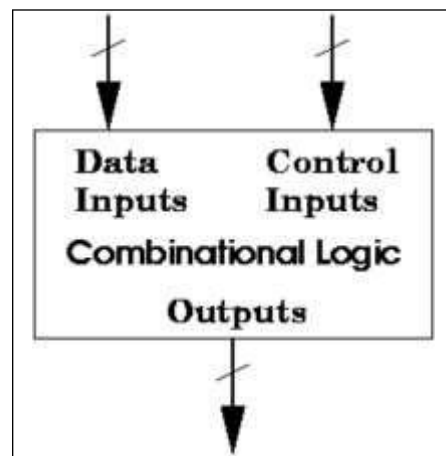
Fig 1.7: A generic ALU that has 2 inputs and 1 output

A typical ALU will have two input ports and a result port. It will also have a control input telling it which operation to perform. For example add, subtract,and, or, etc. It also consists of additional output bits for condition codes. Basically, these bits indicate some facts about the computation. For example carry, overflow, negative, zero results. Additional output bits are together called the *status bits*. The status bits are used for branching operations.

ALUs may be simple and perform only a few operations: Integer arithmetic like add and subtract and Boolean logic like and, or, complement and left Shift, right Shift, rotate. Such simple ALUs may be found in small 4- and 8- bit processors.

Example: Consider a 32-bit ALU as shown in figure 1.8. It consists of source 1 labeled as SRC1 and source 2 labeled as SRC2 as the two 32-bit data inputs. It also has a control input, labeled C which is a signal for addition. These control bits tell the ALU to perform an addition operation on the data inputs.
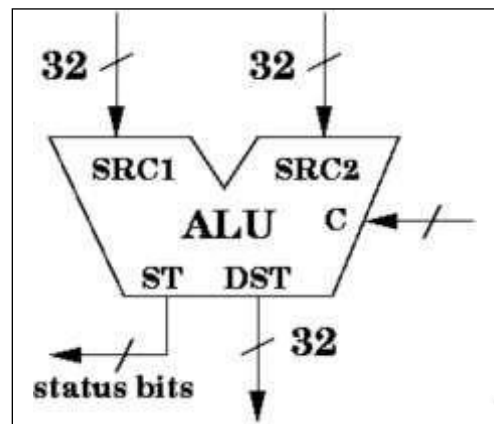
Fig 1.8: A 32 bit ALU

The result of the computation is sent to the output labeled **DST**. It is also 32 bits. There are some additional output bits labeled **ST**. In our example, they may indicate whether the output is zero, or has overflowed.

More complex ALUs will support a wider range of integer operations like multiplication and division, floating-point operations like add, subtract, multiply, and divide. It can even compute mathematical functions like square root,sine, cosine, log, *etc.*

To perform arithmetic and logic operations necessary operands are transferred from the memory location to ALU where one of the operands is stored temporarily in some register. This register is called a ***temporary register***. Each register stores one word of data.

5. CONTROL UNIT

The control unit is the portion of the processor that actually causes things to happen. The purpose of the control unit is to control the system operations by routing the selected data items to the selected processing hardware at the right time. The Control unit acts as a nerve centre for the other units. This unit decodes and translates each instruction and generates the necessary enable signals for ALU and other units. The Control unit has two responsibilities i.e., *instruction interpretation and instruction sequencing*.

In *instruction interpretation,* the control unit reads an instruction from the memory and recognizes the instruction type, gets the necessary operand, and sends them to the appropriate functional unit. The signals necessary to perform the desired operation are taken to the processing unit and resultsobtained are sent to the specified destination.

In *instruction sequencing* control unit determines the address of the next instruction to be executed and loads it into the program counter.

In general, the I/O transfers are controlled by the software instructions that identify both the devices involved and the type of transfer. But the actual timing signals that govern the transfers are generated by the control circuits. Similarly, the data transfer between a processor and the memory is controlled by the control circuits.

The operation of the computer can be summarized as below:
  ➢ The computer accepts information through the input unit and transfers itto the memory.
  ➢ Information stored in the memory is fetched into arithmetic and logic unitsto perform the desired operations.
  ➢ The processed information is transferred to the output unit.
  ➢ All activities inside the machine are controlled by a control unit.

Self-Assessment Questions – 3

7. _____is capable of performing any arithmetic and Booleanoperations.
8. In_____control unit determines the address of thenext instruction be executed and loads it into the program counter.

---

6. BUS STRUCTURE

A bus consists of 1 or more wires. There's usually a bus that connects the CPU to memory and to disk and I/O devices. Real computers usually have several busses, even though the simple computer we have modeled only has one bus where we consider the data bus, the address bus, and the control bus as part of one larger bus.

The size of the bus is the number of wires in the bus. We can refer to individual wires or a group of adjacent wires with subscripts. A bus can be drawn as a line with a dash across it to indicate there's more than one wire. The dash in it is then labeled with the number of wires and the designationof those wires.
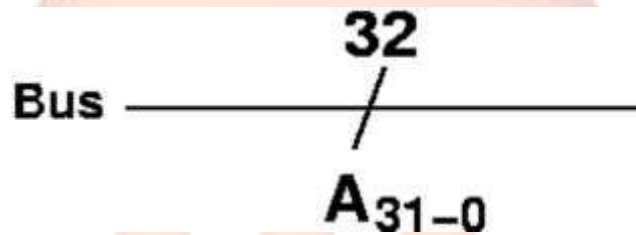

Fig 1.9: Representation of a 32-bit Bus

For example, consider a bus as shown in figure 1.9. It consists of a slanting dash on the horizontal line that represents it is a bus that carries more wires.Also, the slant dash is labeled 32 which indicates that the number of wires in that bus is 32 and the dash is also labeled A_{31-0} which indicates individual 32 wires from A_0 to A_{31}. We can then refer to, say A_{10-0} or A_{15-9} to refer to some subset of the wires.

A bus allows any number of devices to hook up to the bus. Devices connected to the bus must share the bus. Only one device can write to it ata time. One alternative to using a bus is to connect each pair of devices directly. Unfortunately, for N devices, this requires about N^2 connections, which may be too many. Most devices have a fixed number of connections which doesn't permit dedicated connections to other devices. A bus doesn't have this problem.

Data, Address, and Control Busses

There are usually 3 kinds of buses. There's a 32-bit data bus, which is used to write or read 32 bits of data to or from memory. There's a 32-bit address bus for the CPU to specify which

address to read or write from or to memory.Finally, there's a control bus which may consist of a single wire or multiple wires to allow the CPU and memory to communicate.

For example, a control signal is required to indicate when and whether a read or write is to be performed. To support two 32-bit busses, both the CPU and memory require 64 pins or connections 32 for data and 32 for address. Earlier there was a shortage of pins and hence it was necessary to multiplex the address and data bus. Multiplexing uses the same bus as both address and data bus.

There are other kinds of buses that are used primarily for I/O devices like USB. These are mostly high-speed buses for external devices.
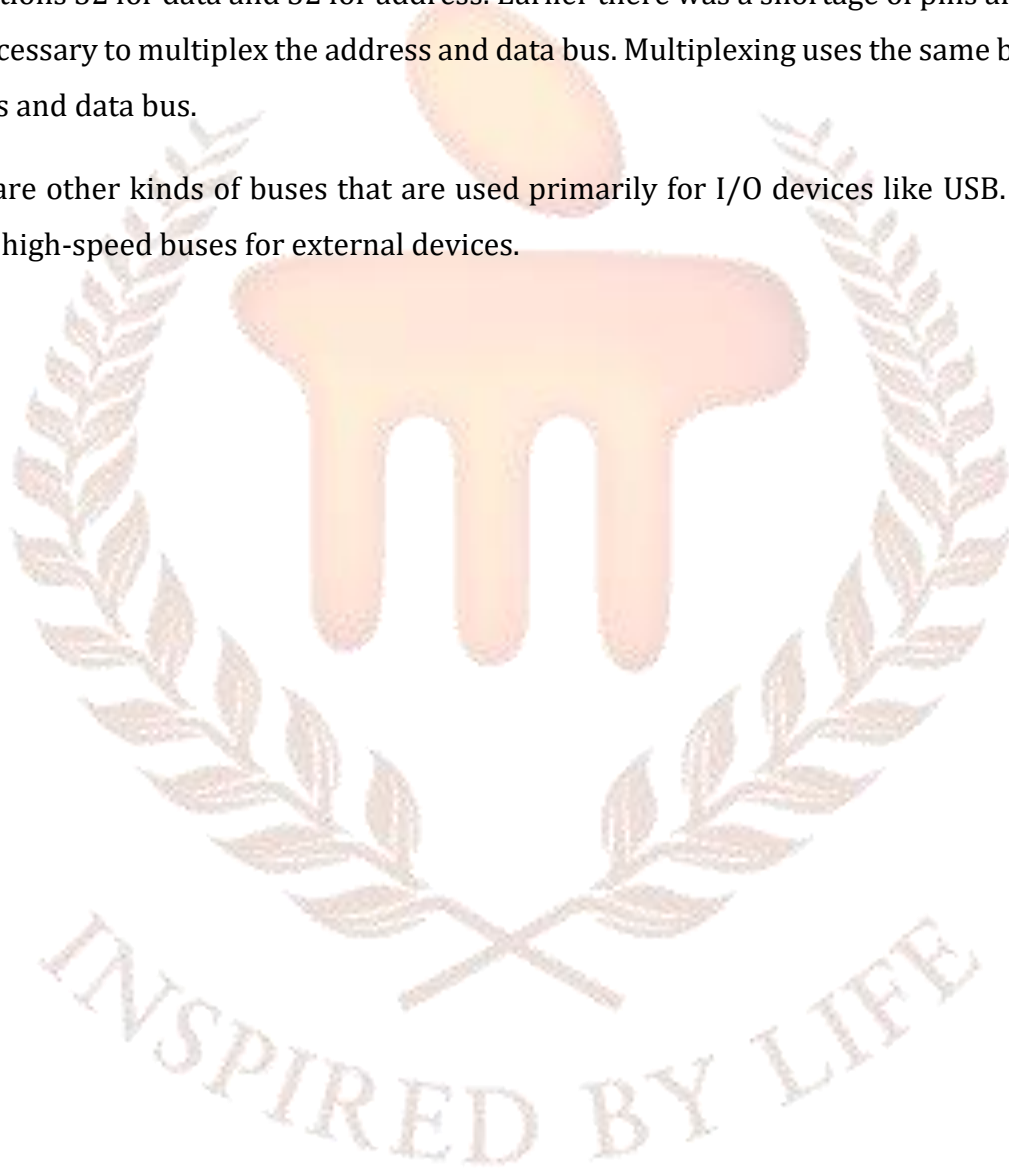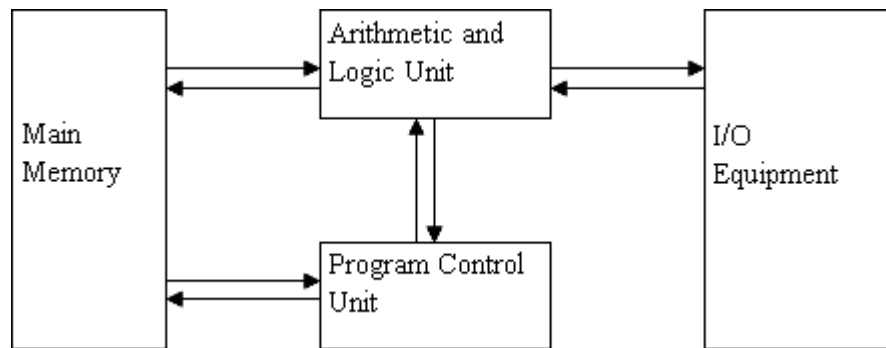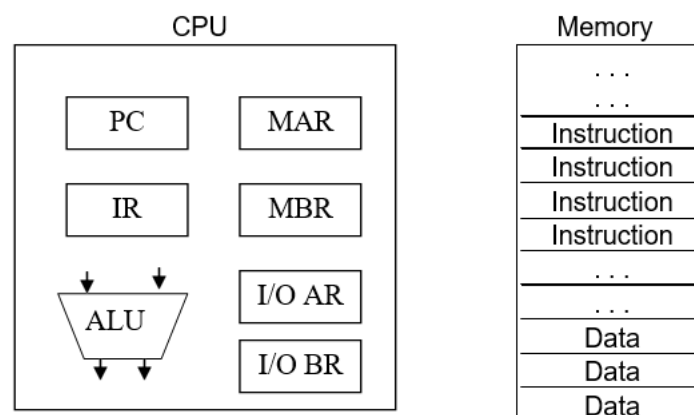
7. VON NEUMANN ARCHITECTURE

Fig 1.10: Structure of the IAS Computer

IAS is the first digital computer in which the von Neumann Architecture was employed. The general structure of the IAS computer is as shown in figure 1.10:

- The main memory, which stores both instructions and data
- An arithmetic and logic unit (ALU) capable of operating on binary data
- A control unit, which interprets the instructions in memory and causes them to be executed
- Input and Output (I/O) equipment operated by the control unit

The **von Neumann Architecture** is based on three key concepts:

1. Data and instructions are stored in a single read-write memory.
2. The content of this memory is addressable by location, without regard tothe type of data contained therein.
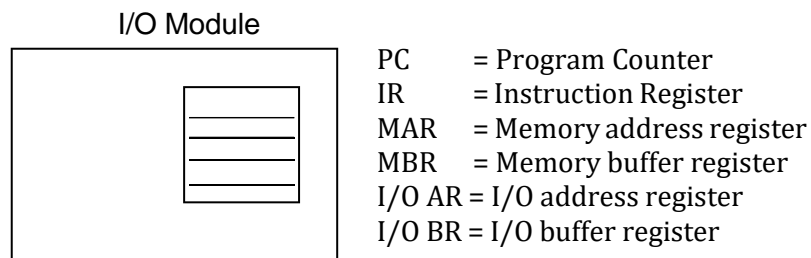3. Execution occurs in a sequential fashion unless explicitly modified from one instruction to the next.

I/O Module

PC   = Program Counter
IR   = Instruction Register
MAR   = Memory address register
MBR   = Memory buffer register
I/O AR = I/O address register
I/O BR = I/O buffer register

Fig 1.11: Computer components Von Neumann architecture

The CPU consists of various registers as listed in figure 1.11. They are

1. **Program Counter (PC):** It contains an address of an instruction to befetched.

2. **Instruction Register (IR):** It contains the instruction most recentlyfetched.

3. **Memory Address Registers (MAR):** Contains the address of a locationin memory.

4. **Memory Buffer Register (MBR):** It contains a word of data to be writtento memory or the word most recently used.

5. **I/O Address Register (I/O AR):** Contains the address of an I/O.

6. **I/O Buffer Register (I/O BR):** It contains a word of data to be written toI/O device.

Any instruction to be executed must be present in the System Memory. The instruction is read from a location pointed by the PC of the memory and then transferred to IR through the data bus. The instruction is decoded andthen the data is brought to the ALU either from memory or register etc. ThenALU computes the required operation on the data and stores the result in a special register called Accumulator. All the sequence of actions is controlledby the control signals generated by the control unit. Thus *Accumulator* is a special purpose register designated to hold the result of an operation performed by the ALU.

Self-Assessment Questions – 4

9. The size of the bus is the number of_____in the bus.

10. _____ contains an address of an instruction to befetched.

11. MAR stands for_____.

12. Accumulator is a special purpose register designated to hold the resultof an operation performed by the ALU.(True / False)

8. SUMMARY

In this unit we provided a historical perspective of a computer with the different generations which explained as to how the computers have evolved. We also introduced to you the classification of computers, various components of the computers, followed by memory which is considered to be the crucial component. Finally, this unit concludes with a brief discussion on ALU and Bus Architecture. *A computer* is a digital device or a machine for manipulating data according to a list of instructions. Large scale integration (LSI - 1000 devices per chip) and very large scale integration (VLSI -100,000 devices per chip) were used in the construction of the fourth generation computers. The purpose of the control unit is to control the system operations by routing the selected data items to the selected processing hardware at the right time.

9. TERMINAL QUESTIONS

1. Give and explain the classification of computers
2. Explain the functional units of a basic computer with a neat diagram
3. Explain the functions of ALU.
4. Write a note on the Bus structure.
5. Discuss on von Neumann Architecture.
6. Explain the various types of registers that CPU consists of.

10. ANSWERS

Self-Assessment Questions

1. Program
2. Analytical engine
3. Minicomputers
4. Arithmetic logic unit
5. word length
6. main memory
7. ALU
8. instruction sequencing
9. wires
10. Program Counter (PC)
11. Memory Address Register
12. True

Terminal Questions:

1. Computers can be classified into micro, mini, main frame and supercomputers. Refer Section 1.2.2.
2. Computer consists of four main components, Central Processing Unit(CPU), Memory, Input / Output and a Bus. Refer Section 1.3
3. The Arithmetic and Logic Unit is the core of any processor. It performsthe calculations on the input data given. Refer Section 1.4.
4. A bus consists of 1 or more wires. There's usually a bus that connectsthe CPU to memory and to disk and I/O devices. Refer Section 1.6.
5. IAS is the first digital computer in which the von Neumann Architecture was employed. Refer Section 1.7.
6. CPU consists of various registers. They are Program Counter (PC),Instruction Register (IR) etc. Refer Section 1.7 for details.