POST-GRADUATE DIPLOMA IN COMPUTER APPLICATIONS

PAPER : PGDCA-I INTRODUCTION TO INFORMATION TECHNOLOGY

SECTION-A

LESSON NO:

Department of Distance Education

Punjabi University, Patiala

LESSON NO. 1.1 **AUTHOR: ARUN SHARMA**

HISTORY OF COMPUTERS

1.1.0 Objectives

After completing this lesson, you will be able to :

- *Know the historical evolution of Computer*
- *Discover the development stages of the System*\

1.1.1 Introduction

A computer is an information processing machine. It can perform arithmetic operations (addition, subtraction, multiplication and division) and take logical decisions. It has a memory and can store lot of information. The stored information may be retrieved, moved and operated upon as desired. Computations are done at an extremely fast speed with complete reliability and accuracy. The speed of execution of operations by modern computers ranges from several million operations per second for fast computers to tens of thousands of operations per second for slow computers.

1.1.2 Computers

There are basically two types of computers--Analog and Digital.

The *analog computer* accepts, processes and generates continuous (unbroken) data. Computations are carried out with physical quantities, such as length, voltage, current, etc. Slide rule, voltmeter, ammeter, potentiometer are examples of analog devices. You know that when current is passed through an ammeter, it gives information about current passing through it. The deflection of the needle indicates the amount of current passing. Deflection is more for higher currents and less for lower currents. Now current and deflection are both continuous quantities. The ammeter receives current (input) and gives deflection (output) after detecting current (processing of input). Thus, ammeter is an analog device. Analog computers use this principle, though they are much more complicated and can perform sophisticated processing of data.

Analog computers are comparatively *slow* and less accurate. They are designed for special applications only. We cannot use a given analog computer for all purposes.

The *digital computer* accepts, processes and produces discrete (discontinuous) data. Computations are done with discrete quantities, such as numerical digits. Usual facit machines, electronic calculators are simple examples of digital devices.

Digital computers are much faster than analog computers and the computations are far more accurate. They come in various sizes - starting from pocket size to large systems which occupy few normal-sized rooms. Digital computers can be designed either for special or for general purposes.

Another type of computer which combines the features of both analog and digital computers is referred to as *Hybrid* computer. It is faster than analog but much slower than digital computer. Hybrid computers find applications in special areas only.

Normally, when we speak of a computer, it is understood as a digital computer. Nowadays these are the most widely used machines.

The present state of computer design and technology has evolved over a period of several hundred years due to efforts of a large number of scientists, mathematicians and engineers. A brief history of development of calculating machines and electronic computers will give you an overview of this.

1.1.3 Early Developments

Man has been interested in counting since ages. The earliest counting device was the Abacus. It was developed sometime between 1000 and 400 B.C.

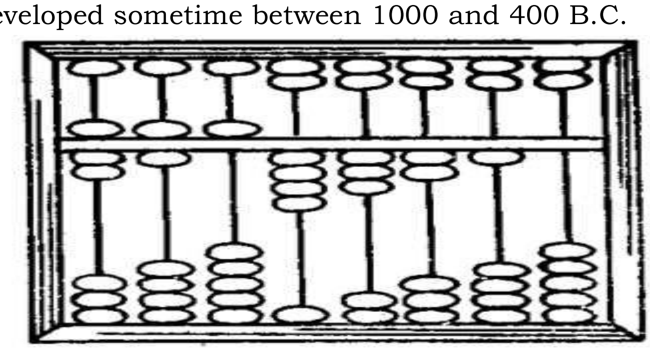


Fig. 1.1 The abacus

The abacus is typically constructed of various types of hardwoods and comes in varying sizes. The frame of the abacus has a series of vertical *rods* on which a number of wooden *beads* are allowed to slide freely. A horizontal *beam* separates the frame into two sections, known as the *upper deck* and the *lower deck*. The beads in the upper portion count five each and those in the lower portion count one each. The standard abacus can be used to perform addition, subtraction, division and multiplication; the abacus can also be used to extract square-roots and cubic roots.

Significant developments in counting devices took place around 1640, when Blaise Pascal, a young French mathematician, invented a simple adding machine (Fig. 1.2).

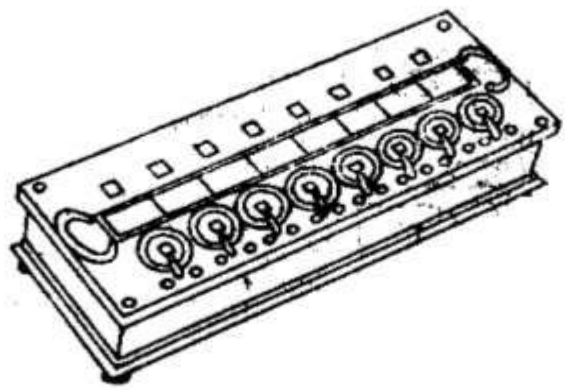
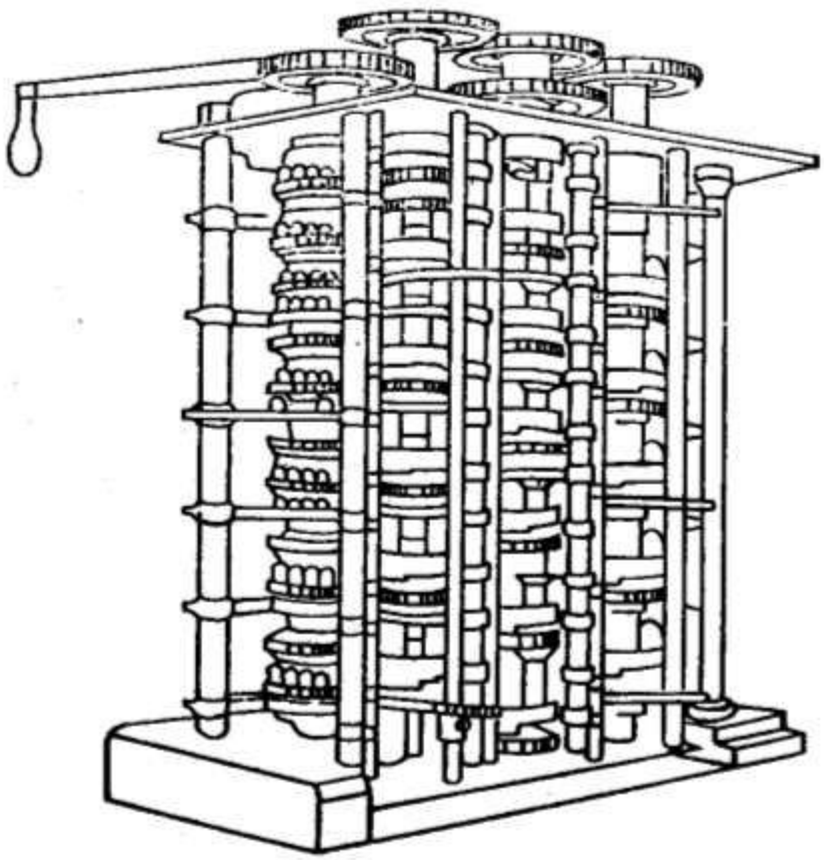


Fig. 1.2 Adding machine

It consisted of a series of toothed wheels. There were ten teeth on each wheel. These represented digits from 0 to 9. When a wheel was rotated past 9, a small indicator on one wheel caused the next wheel to rotate automatically. This way, when one wheel turned past digit 9, *it* exchanged *its* ten teeth for one tooth on the next wheel. Pascal's machine could perform additions. It was later improved upon by Leibnitz, a German mathematician whose machine could perform multiplications and divisions as well.

Charles Babbage, an English mathematician, developed a mechanical calculating device, called **Difference Engine** (Fig. 1.3), for automatic computation of mathematical tables around 1830. Babbage, was also involved in the design of another calculating machine which could perform many general functions in an automatic way. After much effort, he constructed a machine called **Analytical Engine**. This machine had a memory device, an arithmetic device, a punched card input system and an external memory store. Thus, Babbage's analytical engine had many of the same fundamental features as the MODERN computers. However, the analytical and difference engines were never produced commercially. This was because of the lack of manufacturing technology at that time. But Babbage is considered one of the great pioneers of the field of computation for his foresight to design the analytical engine.

These were all mechanical machines and their reliability was rather poor. Nevertheless, they laid the basis for the development of advanced calculating machines for the future.



1.1.3.1 Some Early Computers

Mark I Computer-The design and development of the modern computing machines started around 1935. Howard Aiken, Professor at the Harvard University, began to build a machine for automatic computation using the concepts of Babbage and punched cards which were invented by Herman Hollerith. After much efforts by Aiken and IBM engineers, the project was completed and a machine was designed in 1944 (IBM stands for-international Business Machines, Inc. It is the name of a company which is known world-wide for its contributions in the computer field). It was named MARK I. This was considered to be the first digital computer. MARK I could accept data from punched cards, store them in memory, make calculations by means of automatically controlled electromagnetic relays and arithmetic counters which were mechanical. Thus, it was an electromechanical computer, extremely slow compared with modern computer. It could be programmed, that is, given a set of commands to carry out certain operations. It performed arithmetic and logical operations and solved scientific problems.

The ENIAC Computer-Efforts were going on to increase the speed for automatic computations. MARK I, being an electromechanical machine, could not be improved upon much. Around 1940, vacuum tubes (diodes and triodes) were being used in electronic circuits. J.W. Mauchly and J.P. Eckert, Professors at the University of Pennsylvania, used these tubes to design a computer which was completed in 1946. It was called Electronic Numerical *Integrator And* Calculator, abbreviated as ENIAC. This was the first electronic computer, faster than MARK I but much slower than the modern machines.

The EDSAC Computer- Von Neumann, a noted mathematician at the University of Princeton, suggested the concept of stored program and the use of binary number system in computers around 1946. Earlier machines could be programmed (in a complicated way, though), but the idea of storing instructions in the computer memory was not there. The suggestion of Neumann brought a revolution in the development of computers. Instructions could be stored internally by coding them in numerical form, like data. The idea of computer program got initiated from now: onwards. These concepts could not be incorporated into the ENIAC, as they came late. The first computer to use the stored program concept was designed and completed in 1949 at the Cambridge University, England. It was named Electronic Delayed Storage Automatic Computer called EDSAC. The program was fed into the storage unit by means of paper tape. EDSAC also used vacuum tubes and was a little faster than the ENIAC.

Universal Automatic Computer-Further developments took place in the input/output media. The first computer to use magnetic tapes for data input and output was UNIVAC I *(UNIVersal* Automatic Computer). It was produced by Univac Division of Remington Rand Company, USA, in 1951. It was the first computer to be produced commercially. It could process numeric as well as alphabetic data. Vacuum tubes were used in this system also.

Around the same time, IBM came up with another computer, called IBM 650 This also used vacuum tubes for internal operations.

The disadvantages with the use of vacuum tubes have been:

(i) Size of the computer becomes very large, so lot of space is required for their storage.

(ii) Life time of the tubes is small.

(iii) Tubes are expensive and less reliable.

(iv) Tubes generate considerable amount of heat, thus making it necessary to place the computer in air-conditioned environments.

(v) Switching times of the tubes are very high, so the speed of the computer is also small. (Switching time is defined as the time taken by the tube to go from one state to another state, say for example, from the conducting state to non-conducting state or vice versa. Conducting state means when current is flowing through a tube.)

After the discovery of the transistor, vacuum tubes were replaced by transistors. A transistor is a solid state device whose size is much smaller than a vacuum tube.

Many of the drawbacks of the vacuum tubes were either eliminated or minimized with the use of transistors. However, with further research in electronics technology, transistors have been replaced by extremely superior solid state devices. Complicated circuits can be designed on very small *silicon chips.* This has led to the tremendous reduction in the size, cost of computers, improvement in quality and enhancement of speed. Latest computers can perform several million operations per second.

Self-Check Exercise

Q: Write short notes on the following :

- **Analog Computer**
- **Digital Computer**

1.1.4 Generations of Computers

The history of computer development is often referred to in reference to the different generations of computing devices. In computer terminology, the word 'generation' is described as a stage of technological development or innovation. A major technological development that fundamentally changed the way computers operate, resulting in increasingly smaller, cheaper, more powerful and more efficient and reliable devices characterize each generation of computer.Each feveration is defined by a significant technological developments that changes fundmentally how computers operate. According to the kind of 'processor' installed in a machine, there are five generations of computers, which are discussed in the next few sections.

1.1.4.1 First Generation (1940-56): Vacuum Tubes

First generation computers were vacuum tubes/thermionic valves based machines. These computers used vacuum tubes for circuitry and magnetic drums for memory. A magnetic drum is a metal cylinder coated with magnetic iron-oxide material on which data and programs can be stored. Input was based on punched cards and paper tape and output was displayed on printouts.

Figure 1.4 Vacuum Tube

First generation computers relied on binary-coded language (language of 0s and 1s) to perform operations and were able to solve only one problem at a time. Each machine was fed with different binary codes and hence were difficult to program. This resulted in lack of versatility and speed. In addition, to run on different types of computers, instructions must be rewritten or recompiled.

Examples: ENIAC, EDVAC and UNIVAC were some of the computers of this generation.

Characteristics of first generation computers

- These computers were based on the vacuum tube technology.
- They were the fastest computing devices of their times (computation time was milliseconds)
- These computers were very large and required a lot of space for installation.
- Since thousands of vacuum tubes were used, they generated a large amount of heat. Therefore, air conditioning was essential.
- These were non-portable and very slow equipments.
- They lacked in versatility and speed.
- They were very expensive to operate and used a large amount of electricity.
- These machines were unreliable and prone to frequent hardware failures. Hence constant maintenance was required.
- Since machine language was used, these computers were difficult to program and use.
- Each individual component had to be assembled manually. Hence, commercial appeal of these computers was poor.

1.1.4.2 Second Generation (1956-63): Transistors

Second generation computers used transistors instead of vacuum tubes, which were superior to vacuum tubes. A transistor is made up of semiconductor material like germanium and silicon. It usually had three leads and performed electrical functions such as voltage, current or power amplification with low power requirements. Since transistor is a small device, the physical size of computers was greatly reduced. Computers became smaller, faster, cheaper, energy-efficient and more reliable than their predecessors. In second generation computers, magnetic cores were used as primary memory and magnetic disks as secondary storage devices. However, they still relied on punched cards for input and printouts for output.

Figure 1.5 Transistors

One of the major developments of this generation includes the progress of machine language to assembly language. Assembly language used mnemonics (abbreviations) for instructions rather than numbers, for example, ADD for addition and MULT for multiplication. As a result, programming became less cumbersome. Early high-level programming languages such as COBOL and FORTRAN also came into existence in this period.

Examples: PDP-8, IBM 1401, and IBM 7090.

Characteristics of second generation computers

- These machines were based on the transistor technology.
- They were smaller as compared to first generation computers.
- The computational time of these computers was reduced to microseconds from milliseconds.
- They were more reliable and less prone to hardware failure. Hence, required less frequent maintenance.
- They had better portability and generated less amount of heat.
- Assembly language was used to program computers. Hence, programming became more time efficient and less cumbersome.
- Second generation computers still required air conditioning.
- Manual assembly of individual components into a functioning unit was still required.

1.1.4.3 Third Generation (1964-Early 1970s): Integrated Circuits

The development of the integrated circuit was the trait of the third generation computers. Also called an IC, an integrated circuit consists of a single chip (usually silicon) with many components such as transistors and resistors fabricated on it. Integrated circuits replaced several individually wired transistors. This development made computers smaller in size, reliable and efficient.

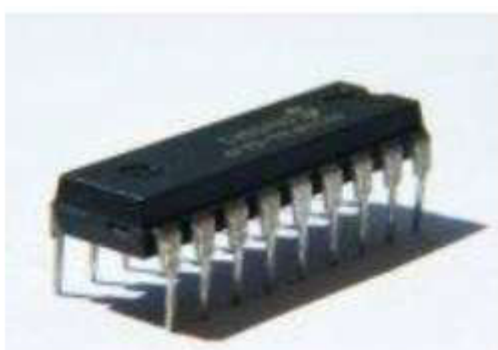


Figure1.6 Integrated Circuit

Instead of punched cards and printouts, users interacted with third generation computers through keyboards and monitors and interfaced with an operating system.

This allowed the device to run many different applications at one time with a central program that monitored the memory. For the first time, computers became accessible to mass audience because they were smaller and cheaper than their predecessors. **Examples:** NCR 395 and B6500.

Characteristics of third generation computers

- These computers were based on integrated circuit (IC) technology.
- They were able to reduce computational time from microseconds to nanoseconds.
- They were easily portable and more reliable than the second generation.
- These devices consumed less power and generated less heat. In some cases, air conditioning was still required.
- The size of these computers was smaller as compared to previous computers.
- Since, failing of hardware occurred very rarely, its maintenance cost was very low.
- Extensive use of high-level languages became possible.
- Manual assembling of individual components was not required, so it reduced the large requirement of labor and cost. However, highly sophisticated technologies were required for the manufacture of IC chips.
- Commercial production became easier and cheaper.

1.1.4.4 Fourth Generation (Early 1970s-Till Date): Microprocessors

The fourth generation is an extension of third generation technology. Although, the technology of this generation was still based on the integrated circuit, these have been made readily available to us because of the development of the microprocessor (circuits containing millions of transistors). The Intel 4004(10) chip, which was developed in 1971, took the integrated circuit one step further by locating all the components of a computer (central processing unit, memory and input and output controls) on a minuscule chip. A microprocessor is built onto a single piece of silicon, known as chip. It is about 0.5 cm along one side and no more than 0.05 cm thick.

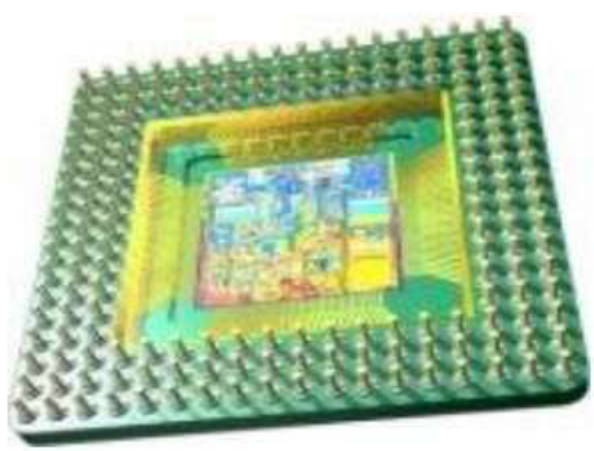


Figure 1.7 Microprocessor

The fourth generation computers led to an era of **large scale integration (LSI)** and **very large scale integration (VLSI)** technology. LSI technology allowed thousands of transistors to be constructed on one small slice of silicon material whereas VLSI squeezed hundreds of thousands of components on to a single chip. Ultra-large scale integration (ULSI) increased that number into the millions. This way computers became smaller and cheaper than ever before.

Fourth generation computers became more powerful, compact, reliable and affordable. As a result, it gave rise to *personal computer (PC)* revolution. During this period, magnetic core memories were substituted by semiconductor memories, which resulted in faster random access main memories. Moreover, secondary memories such as hard disks became economical, smaller and bigger in capacity. The other significant development of this era was that these computers could be linked together to form networks, which eventually led to the development of the Internet. This generation also saw the development of the GUls (Graphical User Interfaces) mouse and handheld devices. Despite many advantages, there is only one disadvantage of this generation, that is this generation required complex and sophisticated technology for the manufacturing of CPU and the other components.

Examples: Apple II, Altair 8800 and CRAY-1.

Characteristics of fourth generation computers

- Fourth generation computers are microprocessor-based systems.
- These computers are very small in size.
- Fourth generation computers are the cheapest among all the other generations.
- They are portable and very reliable.
- These machines generate negligible amount of heat, hence they do not require air conditioning.
- Hardware failure is negligible so minimum maintenance required.
- The production cost was very less. In addition, labour and cost involved at assembly stage is also minimal.
- Graphical user interface and pointing devices enable users to learn to use the computer quickly.
- Interconnection of computers leads to better communication and resource sharing.

1.1.4.5 Fifth Generation (Present and Beyond): Artificial Intelligence

The fifth generation computers are still in the development stage. **These computers will use 'Super Large Scale Integrated' chips, which will result in production of microprocessor having millions of electronic components on a single chip**. The goal of fifth-generation computing is to develop devices that respond to natural language input and will have the capability of learning and self-organization. These computers will use intelligent programming (artificial intelligence or AI) and knowledge-based problem solving techniques. The intelligent program will tell the computer what to do and not how to do. The conventional data processing is based on processing information whereas artificial intelligence deals with processing ideas and knowledge. Moreover, they will be capable of performing multiple, simultaneous instructions using more than one micro-processing chip. The input and output for these machines will be in the form of graphic images or speeches. Such machines will be able to interpret natural languages (NLP-Natural Language Processing).

The ongoing development and application of artificial intelligence has also enabled a field of *expert systems*. It consists of a knowledge base and a set of rules for manipulating the knowledge. With the advent of fifth generation computers, expert systems will become much more sophisticated and popular. Japan and USA have undertaken projects to design and develop such computers. Presently, they are used in the field of medicine, treatment planning, monitoring and so on, on a very small scale.

The characteristic features of fifth-generation computers are as follows:

- Portable PCs (called notebook computers) are much smaller and handy than the PCs of the fourth generation, allowing users to use computing facility even while travelling.
- The desktop PCs, mainframes and workstations are several times more powerful than the PCs of the fourth generation.
- Although the mainframes require proper air-conditioning of the rooms/areas in which they are located, no air-conditioning is normally required for the notebook computers, desktop PCs, and workstations.
- They consume much less power than their predecessors.
- They are more reliable and less prone to hardware failures than their predecessors. Hence, the maintenance cost is negligible.
- They have faster and larger primary and secondary storage as compared to their predecessors.
- Use of standard high-level programming languages allows programs written for one computer to be easily ported to and executed on another computer.
- More user-friendly interfaces with multimedia features make the systems easier to learn and use by anyone, including small children.
- The explosion in the size of the Internet, coupled with Internet-based tools and applications, have made these systems influence the life of even common men and women.

1.1.5 Summary

Many devices, which mankind developed for their computing requirements, preceded computers. Charles Babbage was the first scientist who proposed the idea of a programmable computer. Thus, he is considered as the father of modem digital computers. Some of the well-known early computers, which are considered the stepping-stones of modem computers, are MARK - I, ENIAC, EDSAC and UNIVAC.

Computer development is divided into five main generations. With every generation, computer technology has fundamentally changed, resulting in increasingly smaller, cheaper, more powerful, more efficient and reliable devices. First generation computers were vacuum tube-based machines. These computers were very large, required a lot of space for installation, generated a large amount of heat, and were non-portable and very slow equipments. In addition, these machines were unreliable and prone to frequent hardware failures. Second generation computers used transistors in place of vacuum tubes. Since transistor was a small device, the physical size of computers was greatly reduced. Computers became smaller, faster, cheaper, energy-efficient and more reliable than their predecessors. Third generation of computers was integrated circuit (IC) based machines. Integrated circuits replaced several individually wired transistors. This development made computers smaller in size, reliable, and efficient. Fourth generation computers use microprocessor (circuits containing millions of transistors) as their basic processing device. These computers are the most powerful, compact, reliable and affordable as compared to their predecessors. Fifth generation computers are still in the development stage. These computers will use 'Super Large Scale Integrated' chips, which will result in the production of microprocessors having millions of electronic components on a single chip. They will use intelligent programming (artificial intelligence or AI) and knowledge-based problem solving techniques.

Main Characteristics of First four Generations have been shown in the following Table

Generation	Years	Switching device	Storage device	Switching time	MTBF*	Software	Applications
First	1940-55	Vacuum tubes	Acoustic delay lines and later magnetic drum. 1 kbyte memory.	0.1 to 1 millisecond	30 mts. to 1 hour	Machine languages and Simple monitoring software	Mostly scientific. Later simple business applications
Second	1956-63	Transistors	Magnetic core main memory, tapes and disk peripheral memory. 100 Kbyte main memory.	1 to 10 microseconds	About 10 hrs.	Symbolic and assembly languages. Early version of rs High level languages FORTRAN, COBOL, Algol, Batch operating systems.	Extensive business applications. Engineering design optimization, scientific research.
Third	1964-75	Integrated Circuits (IC)	High speed magnetic cores. Large disks (100 MB). 1 Mbyte main memory.	0.1 to 1 micro-second	About 100 hrs.	FORTRAN IV, COBOL 68. PL/1, Timeshared operating system.	Data base management systems. On line systems.
Fourth–First decade	1976-85	Large scale inte-grated circuits. Micro-processors (LSI)	Semiconductor memory. Winchchester disk. 10Mbyte main memory. 1000 Mbyte disks.	10 to 100 nano-seconds	About 1000 hrs.	FORTRAN 77, Pascal, ADA, COBOL-74, Concur-rent Pascal.	Personal computers. Distributed systems. Integrated CAD/CAM Real time control. Graphics oriented systems.
Fourth–Second decade	1986-1995	Very large scale integrated circuits. Over 3 million transistors per chip	Semiconductor memory. IGB main memory. 100 GB disk.	1 to 10 nano-seconds.	About 10,000 hrs.	C, C++,. JAVA, PROLOG	Simulation, Visual-ization, Parallel computing, Virtual reality, Multimedia.

1.1.6 Keywords:

Computer : A device that computes, especially a programmable electronic machine that performs high-speed mathematical or logical operations or that assembles, stores, correlates, or otherwise processes information.

Analog Computer : A Computer in which numerical data are represented by measurable physical variables, such as electrical voltage.

Digital Computer : A computer that performs calculations and logical operations with quantities represented as digits, usually in the binary number system.

Vacuum Tubes : An electronic device that controls the flow of electrons in a vacuum. It is used as a switch, amplifier or display screen. Used as on/off switches, they allowed the first generation computers to perform digital computations.

Transistors : In the binary block of computing, a transistor is mostly a switch and is the fundamental building block of computer circuitry.

Microprocessor : An integrated circuit that contains the entire central processing unit of a computer on a single chip.

AI (Artificial Intelligence) : The ability of a computer or other machine to perform those activities that are normally thought to require intelligence.

1.1.7 Short Answer Type Questions:

- What is a Computer?
- What are the various types of computers?
- How are third generation computers superior to second generation computers?
- What were the disadvantages of using vacuum tubes? How have they been overcome?
- What do the following names stand for -ENIAC, EDSAC, UNIVAC, IBM.

1.1.8 Long Answer Type Questions :

- What is a Computer? Explain briefly the working principle of various types of computers. Give examples.

- Trace the history of development of modern computers.
- What are the various computer generations? Explain them.

1.1.9 Suggested Readings:

1. Computer Fundamentals By Pradeep K. Sinha and Priti Sinha (BPB publications)
2. Introduction to Information Technology by V.Rajaraman, PHI, New Delhi.
3. Fundamentals of Information Technology, by Alexis Leon and Matthews Leon Vikas Publishing House
4. Fundamentals of Information Technology by Deepak Bharihoke, Pentagon Press.
5. Computer : A history of the Information Machine by Martin Campbell-Kelly and William Aspray.
6. Media Technology & Society : A History : From the Telegraph to the Internet by Brian Winston.

LESSON NO. 1.2 **AUTHOR: ARUN SHARMA**

INTRODUCTION TO COMPUTERS

1.2.0 Objectives:

After completing this lesson, you will be able to :

- Describe the elements of the computer System
- State the characteristics of Computer System
- Define types of Computers along with their features
- Discuss applications of Computers in daily life

1.2.1 Introduction

A computer is an electronic device that gets input from the user, process it and gives the output in a meaningful manner. Computer system is an information-processing machine doing arithmetic and logical manipulations (calculations), with a capacity to store data, temporarily and or permanently. In early days of computer systems, they were treated as computing machines but in present days they are treated as an information processor for compution as well as word processing systems. Present computer systems are electronic having incredibly high speed microprocessors. Pentium Core-2-Duo computer system is now available in high-tech market. The speed of Pentium Core -2-Duo processor is the highest ever in the present day systems.

The Intel's new research chip i.e. "Wireless Internet on the Chip" feature logic (microprocessor), flash memory and analog communications circuits on a piece of silicon using a single manufacturing process. Each of these types of circuits is traditionally manufactured on separate process technology in different factories. Chips produced by the new process may be up to five times more powerful than those used in today' s Hand Held Computers or cell phones, capable of operating at speeds of up to 1 GHz and providing up to a month of battery life.

1.2.2 Structure of a Computer

Every computer system has essentially four important units:

- Input device
- Central processing unit
- Memory
- Output device.

The central processing unit itself has three units, namely, control unit, arithmetic logic unit (ALU) and registers. The functional blocks of computer system are illustrated in the figure 2.1.

Memory is the separate unit connected to the CPU by bus. In addition to the Memory unit, computers also employ secondary storage devices for holding data or instructions on a permanent basis.

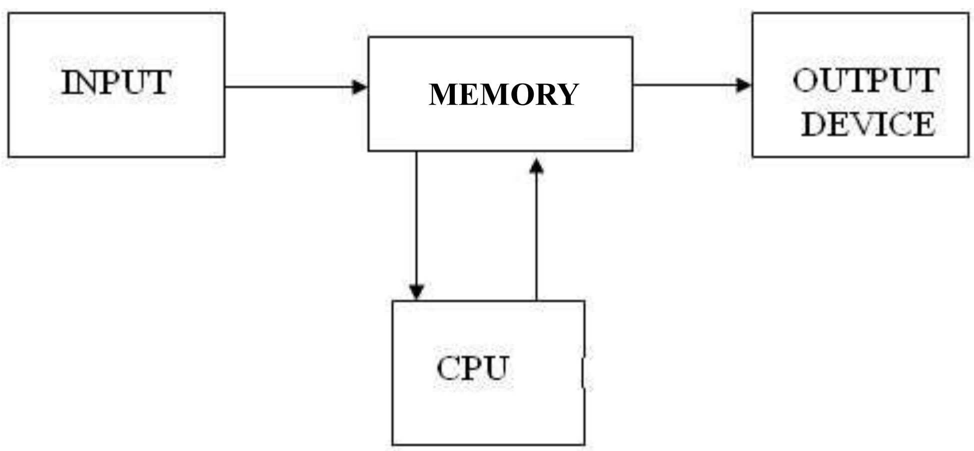


Figure 2.1 : Structure of a Computer

1.2.2.1 Input Device

The function of the input device is to transfer / feed the information into memory unit of the computer. Various input devices are used with computers system like mark and character recognition, hand held and hand print terminal, teletypewriter, keyboard, mouse etc.

1.2.2.2 Central Processing Unit

The function of CPU (Central Processing Unit) is to process the information as per the instruction given to it. The information given to the CPU is of two types.

i) Data: It is the raw material on which processing is done. It may be alphabetic, alphanumeric and numeric.

ii) Program: It is the set of instructions to perform a given task.

CPU further consists of three main parts.

(a) Control Unit

Like an activity of a brain in case of the human beings, control unit accepts instructions held in main memory, interprets these instructions and processes them through appropriate units of the system in the correct sequence. This unit ensures that according to the stored instructions, the right operation is done on the right data at the right time. It also informs A.L.U. for the precise operation to be done and directs for the transfer of data items. It consists of electronic circuits designed to decode program instructions from the memory.

(b) Registers

Information is read from the memory by a register. Register are a type of computer memory used to quickly accpet, store and transferend and instructions that are being used immediatly by the CPU. It also interconnects the Memory Cells. The identification code of the each register in the Memory is known as its address. It is usually specified as a binary number and placed in a register called Memory Address Register (MAR). MAR is a register to which the address location of the memory is stored from/to which the information is read from/written into by the decoders. The data read from the memory is placed in a register called Memory Data Register (MDR).

A set of connectors that carries bits in parallel and has associated with control scheme is known as Bus. A bus that carries a word (bits) to or from memory is known as Data Bus (DB). In order to retrieve the word from memory it is necessary to specify its address. Address space in general will consist of registers and main memory. This address is carried by Memory Address Bus (MAB). Control Bus (CB) is used to pass the instructions. Let us see the interconnection of CPU and Memory in the Figure 2.2. by using the buses.

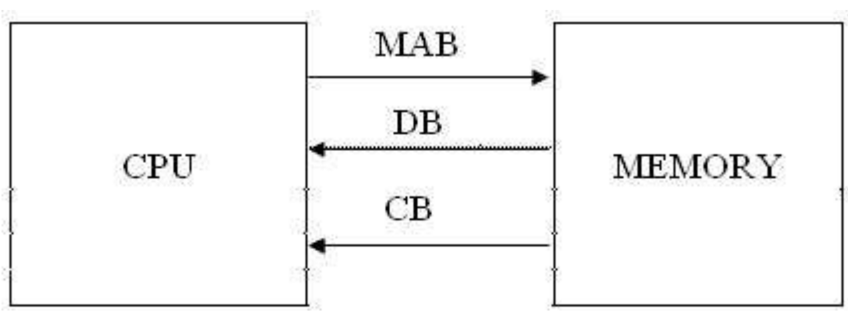


Figure 2.2 : Interconnection of CPU and Memory

(c) Arithmetic Logic Unit (A.L.U.)

Can you think about logical and arithmetical operations? Look at the A.L.U. The function of this unit is to perform these operations on the numeric and alphanumeric information as per the directives of control unit. As you know there are five basic arithmetic operations i.e. addition, subtraction, multiplication, division and exponentiation and three main logical operations i.e. AND, OR, NOT.

Special purpose memory locations are used in A.L.U. called registers. The main register of A.L.U. is accumulator. It is a memory space in which arithmetic operations are performed. After completing the operation, the result will be also stored in the accumulator. Working Registers and Status registers are the other registers used in A.L.U. Working registers are used to store temporarily intermediate results obtained during the computation. Status Registers are used to indicate the results of ALU operations.

1.2.2.3 Memory

The function of memory is to store the information. The internal memory is also called as main memory. There are two types of memory units i.e. main memory and secondary memory.

The main memory unit is used to store the program, the data and the results. With the capacity of the memory, it holds the program and data totally or partially as RAM.

The secondary memory is provided to supplement the main storage and is used to store a large amount of data or instructions on permanent basis. Secondary storage is much cheaper than main memory of the computer system e.g. Hard Disk.

1.2.2.4 Output Device

The function of the output device is to display the results from the processing and calculations. Various output devices are used with the computer systems i.e. visual display unit, printers, plotters etc.

1.2.2.5 Operation of the Computer System

As you have learnt, Input device gives instructions to the memory of the CPU. If it contains arithmetical and/or logical operations, information is given to ALU to perform. Processed information and results are given back to the memory for storage. This total operation is under the supervision of Control unit. The Output Device takes out the results and required information.

1.2.3 Characteristics of Computers

As you are aware that the computerization of the every sector is possible only because of the following characteristics. Every characteristic plays a vital role to increase the utility of computer system. There are six characteristics.

a. **Speed:** The speed of computer systems is consistently increasing. High-speed characteristic has led to many scientific and engineering projects being carried out which were previously impossible. A powerful computer is capable of adding two IS-digit numbers in 300 to 350 nanoseconds only. The ability to get answers fast enough so that one has time to take action on them. It makes real-time computing possible. The internal speed of computer is virtually instantaneous because of incredible speeds of electrical pulses.

b. **Storage:** Computer systems are called Information Explorers. They generate vast amount of information after processing. It becomes necessary to store it with the memory devices. Such storage devices are of two types i.e. internal and external. The internal memory of CPU is mainly used to retain system information and program software and hence majority of data is stored outside the memory of computer. Also there is inverse relationship between size of internal memory and speed of the computer. Hence, size of internal memory is limited. Secondary storage devices are therefore used to store the information. However, computers very often work on a vast amount of data. Secondary storage devices are used to retain the data on a permanent basis. Several different devices can provide this additional storage space. Floppy disk, magnetic disk, Winchester disk, magnetic tapes are the examples of secondary storage devices.

c. **Accuracy:** The accuracy of computer system is 100% provided that the given instructions to it are right. Errors occurred in the result are only because of wrong instructions given to it. It means errors are due to the human rather that the technological weakness of the machine i.e. Imprecise thinking of the programmer, inaccurate data feeding or poorly designed systems.

d. **Versatility:** Computer system is a versatile instrument as it is capable of performing almost any task. It performs only three basic operations.

1. It transfers data within CPU
2. It performs basic arithmetical and logical operations.
3. It exchanges information through input / output devices.

e. **Automation:** Once a program is in the computer memory, the individual instructions are transferred sequentially to the control unit for execution. The CPU (Central Processing Unit) follows these instructions until it meets a last instruction that says 'stop program execution' i.e. human assistance required for execution is almost nil. A computer system works much more than the calculator and executes the program instructions automatically.

f. **Diligence:** A computer system does not suffer from the human traits of tiredness and lack of concentration. It performs calculations with exactly the same accuracy and speed as the first.

Self Check Exercise

Q: Writer short notes on the following :

- **Input device**
- **Output device**
- **CPU**

1.2.4 Classification of Computers

These days, computers are available in many sizes and types. You can have a computer that can fit in the palm of your hand to those that can occupy the entire room; multi user computers can be used by hundreds of users simultaneously. Computers also differ based on their data processing abilities. Hence, computers can be classified according to purpose, data handling, and functionality.

1.2.4.1 Classification According to Purpose

Computers are designed for different purposes. They can be used either for general purposes or for specific purposes.

- **General purpose computers** : A general purpose computer, as the name suggests, is designed to perform a range of tasks. These computers have the ability to store numerous programs. These machines can be used for various applications, ranging from scientific as well as business purpose applications. Even though such computers are versatile, they generally lack in speed and efficiency. The computer that you use in your schools and homes are general purpose computers.
- **Specific purpose computers** : These computers are designed to handle a specific problem or to perform a single specific task. A set of instructions for the specific task is built into the machine. Hence, they cannot be used for other applications unless their circuits are redesigned, that is, they lacked versatility. However, being designed for specific tasks, they can provide the result very quickly and efficiently. These computers are used for airline reservations, satellite tracking, and air traffic control.

1.2.4.2 Classification According to Type of Data-Handled Techniques

Different types of computers process the data in a different manner. According to the basic data handling principle, computers can be classified into three categories: **analog, digital, and hybrid.**

- **Analog computers :** A computing machine that works on the principle of measuring, in which the measurements obtained are translated into desired data is known as analog computer. Modern analog computers usually employ electrical parameters, such as voltages, resistances or currents, to represent the quantities being manipulated. Such computers do not deal directly with the numbers. They measure continuous physical magnitudes (such as temperature, pressure, and voltage), which are analogous to the numbers under consideration. For example, the petrol pump may have an analog computer that converts the flow of pumped petrol into two measurements - the quantity of petrol and the price for that quantity.

  Analog computers are used for scientific and engineering purposes. One of the characteristics of these computers is that they give approximate results since they deal with quantities that vary continuously. The main feature of analog computers is that they are very fast in operation as all the calculations are done in 'parallel mode'. It is very easy to get graphical results directly using analog computer. However, the accuracy of analog computers is less.
- **Digital computers :** A computer that process information, numerical or otherwise, represented in a digital form is known as digital computer. Such

computers process data (including text, sound, graphics and video) into a digital value (in 0's and 1's). In digital computers, analog quantities must be converted into digital quantity before processing. In this case, the output will also be digital. If analog output is desired, the digital output has to be converted into analog quantity. The components, which are performing these conversions, are the essential parts or peripherals of the digital computer.

Digital computers can give the results with more accuracy and at a faster rate. The accuracy of such computers is limited only by the size of their registers and memory. The desktop PC at your home is a classic example of digital computer.

- **Hybrid computers :** Hybrid computer incorporated the measuring feature of an analog computer and counting feature of a digital computer. For computational purposes, these computers use the analog components and for the storage of intermediate results, digital memories are used. In order to bind the powers of analog and digital techniques, analog to digital and digital to analog, the hybrid computers comprehensively use converters. Such computers are broadly used for scientific applications, various fields of engineering and in industrial control processes.

1.2.4.3 Classification According to Functionality

Based on physical size, performance and application areas, we can divide computers generally into four major categories: micro, mini, mainframe and super computers.

1 **Micro computers :** A micro computer is a small, low cost digital computer, which usually consists of a microprocessor, a storage unit, an input channel and an output channel, all of which may be on one chip inserted into one or several PC boards. The addition of a power supply and connecting cables, appropriate peripherals (keyboard, monitor, printer, disk drives, etc.), an operating system and other software programs can provide a complete micro computer system. The micro computer is generally the smallest of the computer family. Originally, they were designed for individual users only, but nowadays they have become powerful tools for many businesses that, when networked together, can serve more than one user. IBM-PC Pentium 100, IBM-PC Pentium 200 and Apple Macintosh are some of the examples of micro computers. Micro computers include desktop, laptop and hand-held models such as PDAs (Personal Digital Assistants).

Desktop Computer: Desktop computer or PC (Personal Computer) is the most common micro computer. It is principally intended for stand-alone use by an individual. These micro computers typically consist of a system unit, a display monitor, a keyboard, internal hard disk storage and other peripheral devices. The major criterion behind the importance of the PCs is that they are not very expensive to purchase by the individuals or the small businesses. Some of the major personal computer manufacturers are APPLE, IBM, Dell and Hewlett-Packard.

Laptop: A laptop is a portable computer, that is, a user can carry it around. Since the laptop computer resembles a notebook, they are also known as notebooks. Laptops are small computers enclosing all the basic features of a normal desktop computer. The biggest advantage of this computer is that one can use this computer anywhere and at anytime, specially when one is travelling and does not have a proper place to keep it. Moreover, these computers do not need any external power supply as a rechargeable battery is completely self-contained by them. These computers are expensive as compared to desktop computers.

Fig 2.3 Desktop Computer Fig 2.4 Laptop Computer

Hand-held Computers: A hand-held, also called Personal Digital Assistant (PDA), is a computer that can conveniently be stored in a pocket (of sufficient size) and used while the user is holding it. PDAs are essentially small portable computers and are slightly bigger than the common calculators. A PDA user generally uses a pen or electronic stylus, instead of a keyboard for input. As shown in Figure, the monitor is very small and is the only apparent form of output. Since, these computers can be easily fitted on the palmtop, they are also known as palmtop computers. Hand-held computers usually have no disk drives rather they use small cards to store programs and data. However, they can be connected to printer or a disk drive to generate output or store data. They have limited memory and are less powerful as compared to desktop computers. Some examples of PDAs are Apple Newton, Casio Cassiopeia and Franklin eBookMan.

Fig 2.5 Personal Digital Assistant

2. **Mini computers :** In the early 1960s, Digital Equipment Corporation (DEC) started shipping its PDP series computer, which the press described and referred as mini computers. A mini computer is a small digital computer, which normally is able to process and store less data than a mainframe but more than a micro computer, while doing so less rapidly than a mainframe but more rapidly than a micro computer. They are about the size of a two-drawer filing cabinet. Generally, they are used as desktop devices that are often connected to a mainframe in order to perform the auxiliary operations. Mini computer (sometimes called a mid-range computer) is designed to meet the computing needs for several people simultaneously in a small to medium size business environment. It is capable of supporting from 4 to about 200 simultaneous users. It serves as a centralized storehouse for a cluster of workstations or as a network server. Mini computers are usually multi-user systems so these are used in interactive applications in industries, research organizations, colleges and universities. These are also used for real-time controls and engineering design work. High-performance workstations with graphics I/O capability use mini computer. Some of the widely used mini computers are PDP 11, IBM (8000 series) and VAX 7500.

3. **Mainframe computers** : A mainframe is an ultra-high performance computer made for high-volume, processor-intensive computing. It consists of a high-end computer processor, with related peripheral devices, capable of supporting large volumes of data processing, high performance on-line transaction processing systems and extensive data storage and retrieval. Normally, it is able to process and store more data than a mini computer and far more than a micro computer. Moreover, it is designed to perform faster than a mini computer and much faster than a micro computer. Mainframes are the second largest (in capability and size) of the computer family. However, a mainframe can usually execute many programs simultaneously at a high speed, whereas super computers are designed for single processes.

   Mainframe allows its user to maintain large information storage at a centralised location and be able to access and process this data from different computers located at different locations. They are typically used by large businesses and for scientific purposes. Examples of mainframe computers are IBM's ES000, VAX 8000 and CDC 6600.

4. **Super computers :** Super computers are the special purpose machines, which are specially designed to maximize the numbers of FLOPS (Floating Point Operation Per Second). Any computer below one gigaflop/sec is not considered a super computer. A super computer has the highest processing speed at a given time for solving scientific and engineering problems. It basically contains

a number of CPUs that operate in parallel to make it faster. Its processing speed lies in the range of 400-10,000 MFLOPS (Millions of Floating Point Operation Per Second). Due to this feature, super computers help in many applications such as information retrieval computer-aided design.

A super computer can process a great deal of information and make extensive calculations very, very quickly. They can resolve complex mathematical equations in a few hours, which would have taken a scientist with paper and pencil a lifetime or years, using a hand calculator. They are the fastest, costliest and most powerful computers available today. Typically, super computers are used to solve multi-variant mathematical problems of existent physical processes, such as aerodynamics, metrology and plasma physics. These are also required by the military strategists to simulate defence scenarios. Cinematic specialists use them to produce sophisticated movie animations. Scientists build complex models and simulate them in a super computer. Here, it is used to model the actions and reactions of literally millions of atoms as they interact. Super computer has limited use because of its price tag and limited market. The largest commercial uses of super computers are in the entertainment/ advertising industry. Examples of super computers are CRAY-3, Cyber 205 and PARAM.

Self Check Exercise :

Q: Write short notes on the following :

- **Microcomputers**
- **Minicomputers**
- **Mainframe Computers**

1.2.5 Applications of Computers

For the last few decades, computer technology has revolutionized the businesses and other aspects of human life all over the world. Practically, every company, large or small is now directly or indirectly dependent on computers for information processing. Computer systems also help in the efficient operations of railway and airway reservation, hospital records, accounts, electronic banking and so on. Computers not only save time, but also save paper work. Some of the areas where computers are being used are listed below.

- **Science:** Scientists have been using computers to develop theories to analyze, and test the data. The fast speed and the accuracy of the computer allow different scientific analyses to be carried out. They can be used to generate detailed studies of how earthquakes affect buildings or pollution affects weather pattern. Satellite-based applications have not been possible without the use of computers. It would also not be possible to get the information of our solar system and the cosmos without computers.
- **Education:** Computers have also revolutionized the whole process of education. Currently, the classrooms, libraries and museums are efficiently utilizing computers to make the education much more interesting. Unlike recorded

television shows, computer-aided education (CAE) and Computer Based Training (CBT) packages are making learning much more interactive. It also helps in research where it has facilitated the search of new research papers and journals.

- **Medicine and Health Care:** There has been an increasing use of computers in the field of medicine. Now, doctors are using computers right from diagnosing the illness to monitoring a patient's status during complex surgery. By using automated imaging techniques, doctors are able to look inside a person's body and can study each organ in detail (such as CAT scans or MRI. scans), which was not possible few years ago. There are several examples of special purpose computers that can operate within the human body such as a cochlear implant, a special kind of hearing aid that makes it possible for deaf people to hear.
- **Engineering/Architecture/Manufacturing:** The architects and engineers are extensively using computers in designing and drawings. Computers can create objects that can be viewed from all the three dimensions. By using the techniques like virtual reality, architects can explore houses that have been designed but not built. The manufacturing factories are using computerized robotic arms in order to perform hazardous jobs. Besides, Computer Aided Manufacturing (CAM) can be used in designing the product, ordering the parts and planning production. Thus, computers help in co-coordinating the entire manufacturing processes.
- **Entertainment:** Computers are finding greater use in entertainment industry. Computers are used to control the images and sounds. The special effects, which mesmerize the audience, would not have been possible without the computers. In addition, computerized animation and colorful graphics have modernized the film industry.
- **Communication:** E-Mail or Electronic Mail is one of the communication media in which computer is used. Through e-mail the messages and reports are passed from one person to one or more persons by the aid of computer and the telephone line. The advantage of this service is that while transferring the messages it saves time, avoids wastage of paper and so on. Moreover, the person who is receiving the messages can read the messages whenever he is free and can save it, reply it, forward it or delete it from the computer.
- **Business Application:** This is one of the important uses of the computer. Initially computers were used for batch processing jobs, where one does not require the immediate response from the computer. Currently computers are mainly used for real time applications (like at the sales counter) that require immediate response from the computer. There are various concerns for which computers are used such as in business forecasting, to prepare pay bills and personal records, in banking operations and data storage, in the various types

of Life Insurance Business and as an aid to management for managing employee database. Businesses are also using the networking of computers, where a number of computers are connected together to share the data and the information. Use of E-mail and Internet has changed the ways of business.

- **Publishing:** Computers have created a field known as DTP or Desk Top Publishing. In DTP, with the help of computer and a laser printer one can perform the publishing job all by oneself. Many of the tasks requiring long manual hours such as making table of contents and index can be automatically done by the application of computers and DTP software.
- **Banking:** In the field of banking and finance, computers are extensively used. People can use the ATM (Automated Teller Machine) services 24 hours of the day in order to deposit and withdraw cash. When the different branches of the bank are connected through the computer networks, then the inter branch transaction such as cheque and draft can be done by the computers without any delay.
- Mass Media Industry : Computer is widely used in mass media industry for editing, refixing and broadcasting news, serials, movies etc.

1.2.6 Summary

A computer can be viewed as a system that comprises several units (central processing unit, input unit, and output unit). These individual units work together to convert data into information. Central Processing Unit (CPU) has three parts: arithmetic logic unit (ALU), control unit and memory unit. ALU performs arithmetical (add, subtract) and logical operations (and, or) on the stored numbers. Control unit controls the input/ output devices and directs the overall functioning of the other units of the computer. Memory unit holds the intermediate results during the course of calculations and provides the data as and when required. The computer takes in the data through various input devices like keyboard, mouse and light pens. The output unit supplies the resulting data through, monitors, printers and plotters etc.

Computer is an electronic device that performs mathematical and non-mathematical operations with the help of instruction to process the information to achieve desired result. Speed, accuracy, reliability, versatility, diligence, lack of intelligence and feelings characterizes the computer.

Computers are classified according to purpose, data handling capability and functionality. Based on purpose, computers can be classified as general purpose computers and special purpose computers. According to the basic data handling principle, computers can be classified as analog, digital and hybrid. Based on physical size, performance, and application areas, we can divide computers generally into four major categories: micro, mini, mainframe and super computers. A micro computer is a small, low cost digital computer. Micro computers include desktop, laptop, and hand-held models such as PDAs (Personal Digital Assistants).A mini computer is a small digital computer, which normally is able to process and store less data than a mainframe but more than a micro computer. A mainframe is an ultra-high performance computer made for high-volume, processor-intensive computing. Super computers are the special purpose machines. These are unique and parallel architectures in order to achieve high speeds.

Computers have entered in every sphere of human life and found applications in various fields such as medicine and health care, business, education, science, technology, government, entertainment, engineering, and architecture.

1.2.7 Keywords :

CPU : The part of a computer that interprets and executes instructions.

Registers : A part of the central processing unit used as a storage location.

Control Unit : Within the computer, hardware that performs the physical data transfers between memory and a peripheral device, such as a disk or screen.

ALU : The arithmetic logic unit (ALU) is a digital circuit that calculates an arithmetic operation (like an additions, subtraction, etc.) and logic operations (like an Exclusive or) between two numbers.

1.2.8 Short Answer Type Questions :

Q:1 What is Central Processing Unit?

Q:2 What is a Register?

Q:3 What is Memory?

Q:4 What is ALU?

1.2.9 Long Answer Type Questions:

Q:1 Draw a block diagram of computer system and explain the functions of various units.

Q:2 Discuss the various characteristics of computers.

Q:3 List out the various applications of computers.

Q:4 Explain the various classifications of Computers?

1.2.10 Suggested Readings:

1. "Computer Fundamentals" By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Introduction to Information Technology by V.Rajaraman, PHI, New Delhi.
3. Fundamentals of Information Technology, by Alexis Leon and Matthews Leon Vikas Publishing House
4. Fundamentals of Information Technology by Deepak Bharihoke, Pentagon Press.
5. Computer Fundamentals by Anita Goel, Pearson Publication

LESSON NO. 1.3 **AUTHOR: ARUN SHARMA**

PRIMARY MEMORY

1.3.0 Objectives

After completing this lesson, you will be able to :

- *State the relevance of Memory*
- *Explain how memory can be represented*
- *Define the types of Memory*
- *Differentiate between RAM and ROM*
- *Explain the types of RAM and ROM*

1.3.1 Introduction

The computer can be used to solve a mathematical problem, to type a letter, to draw some figure etc., for all these applications you will have to execute some program or set of instructions. Before executing a program the program is stored into the computers memory and from the memory, computer executes these instructions one by one to carry out the required work.

The data required to operate on or the result after the operation are also kept in the computer's memory - before transferring them onto some other permanent storage device. When we talk about memory, we generally refer to the primary memory only and when we talk of storage, secondary memory is referred. In this lesson we will discuss primary memory, various facets of primary memory such as its types and what purpose it solve. The secondary memory will be discussed in the next lesson in detail.

1.3.2 Memory Hierarchy

The processor is the 'brain' of the computer where all the essential computing takes place. Unlike a human brain, which combines memory with processing power, a computer processor has very little memory. It must rely on other kinds of memories to hold data and instructions and to save results. The memory in a computer system is of three fundamental types:

- Internal Processor Memory
- Primary Memory
- Secondary Memory

1.3.2.1 Internal Processor Memory

This memory is placed in the CPU (processor) or is attached to a special fast bus. Internal memory usually includes cache memory and special registers, both of which can be directly accessed by the processor. This memory is used for temporary storage of data and instructions on which the CPU is currently working. Processor memory is the fastest among all the memory but is also the most expensive. Therefore, a very diminutive part is used in the computer system. It is generally used to compensate for the speed gap between the primary memory and the processor.

1.3.2.2 Primary Memory

The main memory or the memory on the motherboard of the computer is called the primary memory. primary memory is directly accssed by the CPU. This is also called the on-line memory because it is always directly available to the processor.

This memory can be further divided into two types of memory :

- **RAM**
- **ROM**

The RAM is a type of memory which is used by the computer to store temporary values, programs etc., the RAM is also called a volatile memory.

The volatile memory is a memory which loses its content when the power supply to this memory is switched off. Any information in the volatile memory or the RAM is transferred to some permanent storage device such as floppy disk or hard disk before switching off the machine.

Non-volatile memory is a permanent memory. It does not lose its contents even when the power is not available to this type of memory. A non-volatile memory such as ROM is used inside the computer to keep permanent information such as boot program which is required every time the machine is switched on.

1.3.2.3 Secondary Memory

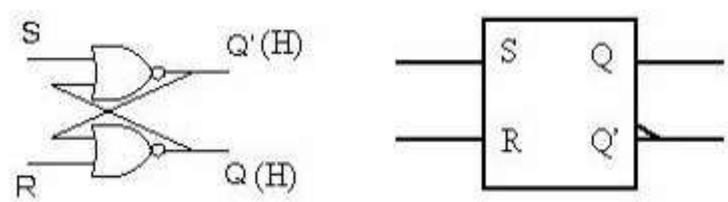
Secondary memory is not a memory in conventional terms, **it is actually the storage media used to store the program or the data, the floppy disk, magnetic disks, magnetic tapes, CD ROMs and the hard disk are example of secondary memory.** These are the least expensive among all the memories, but they are slow in comparison. This is also called off-line memory because it is not directly accessible to the processor, to use any information stored on the secondary memory, first it needs to be transferred to the primary memory, from where the processor can access it.

They have much larger storage capacity than primary memory. Because of its size, it is also called mass storage device.

1.3.3 Memory Representation

Bit or the Binary digit i.e. number 0 and number 1 are the most basic information that one can store inside the computer.

A single bit can be stored inside the computer as a charged capacitor, or by using any bistable multivibrator such as flip-flops, or by using the transistor as a switch, or by using a relay etc.



The output Q is called the ***state*** of the flip-flop

S=1, R=0 **Sets** the state to 1

S=0, R=1 **Resets** (or Clears) the state to 0

S=0, R=0 is the **Hold** state

S=1, R=1 is not allowed

Figure *S-R Flip-Flop using NOR gates, used to store 1 bit.*

A S-R Flip-Flop made using the NOR gates is shown in the figure, this circuit can be made using about 5 to 6 transistor and is used to store single bit of information. As shown in the table,

- When the input S and R both are 0, the flip-flop will remain in the last set stage i.e. it will remember the last set state.

- When the input S is 1 and R is 0, the flip-flop will get set or will store a binary number 1
- When the input S is 0 and R is 1, the flip-flop will get reset or will store a binary number 0
- When the input S and R both are 1, it will be treated as an invalid input by the flip-flop.

Bit	1

Nibble	1	0	0	1

Byte	1	1	0	1	0	1	1	0

Figure *Bit, Nibble and Byte*

A large memory can be constructed by arranging this type of many flip-flops in an array. As a single bit alone does not provide much information, a combination of these bits are used to convey different information. When two bits are used they can be combined in four different forms 00, 01, 10, and 11 and each combination can convey a different information. A combination of 4 bit is called a “nibble” and is used to store information inside the calculators, as the 4 bit can have $2^4 = 16$ different combinations, it is enough to convey 10 different digits (0 to 9) and some additional information such as decimal point etc.. The computer can manipulate characters as well as numbers so a combination of 8 bits (called a “byte”) is used inside the computer to store a character. A combination of 8 bits can convey $2^8 = 256$ different combinations, which is enough for all the alphabet, lower case, upper case, numbers and some special text and graphic symbols used by the computer. Different combination of 0s and 1s of these 8 bits is used to store different information inside the computer’s memory for example

Unit	Symbol	Bytes
Byte	B	2^0 = 1 byte
Kilobyte	KB	2^{10} = 1024 bytes
Megabyte	MB	2^{20} = 1024 KB
Gigabyte	GB	2^{30} = 1024 MB
Terabyte	TB	2^{40} = 1024 GB
Petabyte	PB	2^{50} = 1024 TB
Exabyte	EB	2^{60} = 1024 PB

Table: Memory Units

01000001 is used to store character "A"

00110001 is used to store the number "1"

1 byte can store 1 character, a character can be any alphabet, number or special symbol character such as @, #, $, * etc.

1.3.4 Random Access Memory (RAM)

RAM or Random Access Memory is the main memory used inside the computer to store program, data and results. The term Random access is used because this memory is organized in such a way that any part of the memory can be accessed without going through all the previous parts. For example, to read the memory location 5, one need not pass over the preceding locations 1, 2, 3 and 4, instead the location 5 can be directly accessed.

As most of the other primary memories such as ROM, EPROM etc. are also random access memories, a better term to describe the RAM would be Read /Write Memory, because one can read the stored information from the RAM as well as write or store some information into it.

The RAM can be further divided into two types

- **DRAM**
- **SRAM**

1.3.4.1 Dynamic RAM (DRAM)

A dynamic RAM (DRAM) is cost wise cheaper than the other type i.e. the static RAM (SRAM), and this is the main reason of using the DRAM in the computer as the main memory. As the DRAM utilizes capacitors to store the information, the problem with the DRAM is once some data is stored into it, it can't retain the data for a long time unless the data is again written back or refreshed after some time.

Refreshing of a DRAM memory is done by rewriting the content of the memory every few millisecond. Unless the DRAM is refreshed every few millisecond it loses its content.

Designing memory refresh circuitry for DRAM memory modules is one of the main job for motherboard designers. The DRAM is refreshed by reading its contents and then rewriting it back into the same location.

The positive point about the DRAM is it can be made in higher density or higher capacity compared to the SRAM chips. Because of its cost being lower than the SRAM memory, DRAM is the most common memory used in today's computers.

1.3.4.2 Static RAM (SRAM)

Unlike the DRAM, information stored inside any SRAM memory remains as it is as long as the power supply is provide to the SRAM chip. However, data lost when the power gets down, which makes SRAM a volatile memory. One need not refresh the information stored into the SRAM. Because the SRAM does not require any refresh circuits the circuitry required to interface SRAM is very simple compared to the interface circuitry for the DRAM chips.

SRAM basically use flip flops to store the information.

Self Check Exercise :

Q: What is the difference between Volatile and Non-Volatile memory?

Q: Write short notes on the following:

- **Primary Memory**
- **DRAM**
- **SRAM**

1.3.5 Cache Memory

Cache memory is random access memory (random access memory that a computer microprocessor can access more quickly than it can access regular RAM).It is an extermely fast memory that arts as buffer between RAM and CPU. As the microprocessor processes data, it looks first in the cache memory and if it finds the data there, it does not have to do the more time-consuming reading of data from larger memory.

Some memory caches are built into the architecture of microprocessors. The Intel 80486 microprocessor, for example, contains an 8K-memory cache and the Pentium has a 16K cache. Such internal caches are often called Level 1 (L1) caches. The Pentium® III processor includes two separate 16 KB level 1 (L1) caches, one for instruction and one for data. Most modern PCs also come with external cache memory, called Level 2 (L2) caches. These caches sit between the CPU and the DRAM. Like L1 caches, L2 caches are composed of SRAM but they are much larger. Certain versions of the Pentium III processor include a Discrete, off-die level 2 (L2) cache. This L2 cache consists of a 512 KB unified, non-blocking cache that improves performance over cache-on-motherboard solutions by reducing the average memory access time and by providing fast access to recently used instructions and data. Performance is also enhanced over cache-on-motherboard implementations through a dedicated 64-bit cache bus. In such system, L1 and L2 are on chip and the off-chip cache is referred to as L3.

Disk caching works under the same principle as memory caching, but instead of using high-speed SRAM, a disk cache uses conventional main memory. The most recently accessed data from the disk (as well as adjacent sectors) is stored in a memory buffer. When a programme needs to access data from the disk, it first checks the disk cache to see if the data is there. Disk caching can dramatically improve the performance of applications, because accessing a byte of data in RAM can be thousands of times faster than accessing a byte on a hard disk. When data is found in the cache, it is called a *cache hit* and the effectiveness of a cache is judged by its hit rate. Many cache systems use a technique known as smart caching, in which the system can recognize certain types of frequently used data. The strategies for determining which information should be kept in the cache constitute some of the more interesting problems in computer science.

1.3.6 ROM

ROM or Read Only Memory is as its name suggest is a memory that can be read only, one can not write any information into it. One big plus point about the ROM

is, it does not lose its content when the power supply is cut off and this is a non-volatile type memory.
Like a RAM, a ROM is also random access memory i.e. one can directly access any part of the ROM without serially going through the complete ROM.
There are many types of ROM available in the market, some of them are

1.3.6.1 Mask ROM

A mask ROM is the basic ROM chip. In this type of chips the information is fabricated into the chip at the time of manufacturing itself. The mask is the master pattern used during the chip manufacturing to make various circuit elements on the chip.
Major drawbacks of this type of ROM is it is not easy to make any change in them, as the information is written at the time of manufacturing itself.
Number of chips to be manufactured should be very high when this type of ROM is made, to justify the high investment required in their making.

1.3.6.2 PROM

Creating a ROM chip from scratch is a time-consuming and expensive process. For this reason, developers created a type of ROM known as Programmable Read-only Memory (PROM), which can be programmed. PROM is the ROM memory with a small difference. Unlike the ROM, the information on this chip is written after they are manufactured.
At the time of manufacturing this chip is made as a blank ROM chip and later using special PROM programmers the information is stored into them.

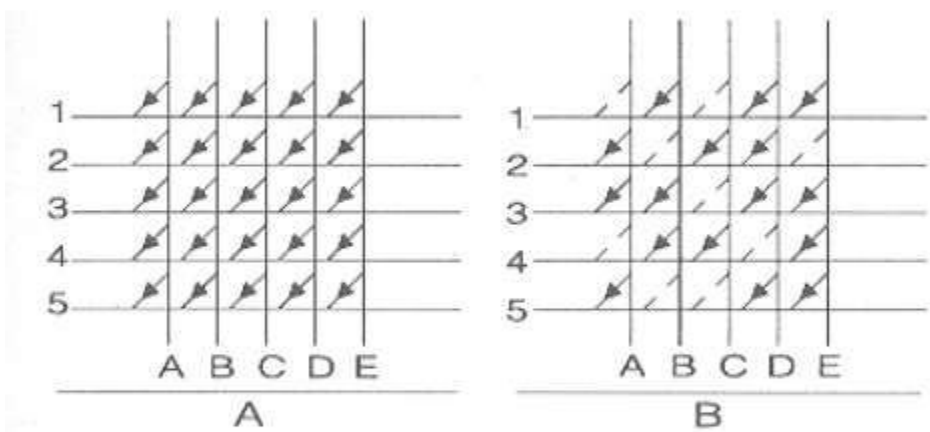


Figure *PROM Memory Before and After the Programming*

Initially when the PROM is made it contains row and address connections **as shown in the figure A i.**e. all the locations contain a binary 1. Later to store any information, the connection between the row and column is broken **as shown in the figure B** to store a binary 0 and the connection is made to remain as it is to store a binary number 1.
Once some information is stored into the PROM by "burning" the information into it, it becomes equal to a ROM i.e. now the information into it can not be changed or removed

and removal of the power to the chip will not destroy the data stored into it.
As the information once written can not be changed or removed to update or revise it with new information, the PROM are not in common use in the computer industry.

1.3.6.3 EPROM

EPROM or Erasable Programmable Read Only Memory, is a PROM but with an option to be able to remove or erase it contents if the user wants to write some new information into it.

As shown in the figure, the EPROM is easily distinguishable from other chips because of a window in the middle of this chip. The window is covered with a transparent quartz glass. As the quartz is transparent to the UV radiation, it is used in place of the ordinary glass. Through this window, one can see the internal circuitry of the EPROM chip. This windows is used to erase the content of the EPROM chip.

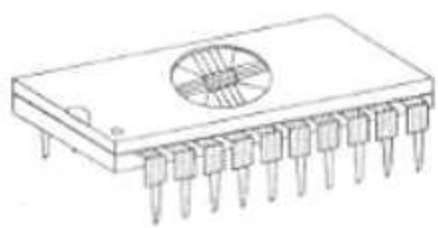


Figure *The EPROM Memory Chip*

To erase the content stored in the EPROM, the EPROM is put under a short wave ultra-violet light source inside an EPROM erasure device. Once the content of the EPROM is erased it can be reprogrammed or rewritten using the EPROM programmer. After the EPROM is programmed the top window should be closed with some opaque label, because the sun rays which contains the ultra violet light can slowly delete/ erase the content of the EPROM chip.

In EPROM one single memory location can not be erased or changed to change the content of even a single location the entire EPROM's content should be erased and then the complete EPROM should be re-written with the new value.

EPROMs are one of the most common ROMs used inside the computers

1.3.6.4 EEPROM

EEPROM or Electrically Erasable Programmable Read Only Memory (pronounced "ee ee prom" or "double e prom") is another type of EPROM. The difference between a EPROM and EEPROM is in the way its content is erased. In an EPROM the content is removed by shining ultraviolet rays through the window directly to the circuits, whereas in the EEPROM the content is removed by using a higher than normal voltage. For example if the "EEPROM is normally used with +5 Volt current for the read operation than by using a voltage higher than +5 Volt for example + 12 Volt, the content of the

EEPROM can be erased..

One big advantage of the EEPROM is its content can be deleted without removing it form the circuit, whereas to erase the content of the EPROM one has to remove it from the circuit and put under the ultraviolet rays for some time By providing the erase voltage in the circuit itself the content of the EEPROM can be changed without removing it from the circuit.

Even though the EEPROM does not lose its content when the power supply to it is switched off and it can be erased and programmed without removing it from the circuit, it can not be used in place of RAM memory in your computer.

Two main reasons behind this are the following :

- The EEPROM has one big drawback that it can be erased and programmed for a limited number of times only, the number of time is a big number i.e. around tens or hundreds of thousand erase and reprogramming can be done, but even this many not be enough for the life time of computer. During the life of a computer the RAM may be written many many thousands of time.
- The second problem with the EEPROM is that to change even one bit of information stored in the EEPROM, the entire content is to be first removed and then everything is to be written back with the new values, this will be almost impossible to do if EEPROM is used as the main memory inside the computer, as the main memory's content are changed many times in a second of operation.

1.3.6.5 EAROM

EAROM or Electrically Alterable Read only Memory is another type of ROM which can be written and read in somewhat the same fashion as the RAM. The EAROM is made in such a way that by applying a high voltage to a particular bit of the memory the content of that location can be changed. To change the content of one location the entire content need not be erased and rewritten as it is done in the EPROM and EEPROM devices. Once some information is stored into the EAROM it works just like a normal ROM, it does not lose its contents when the power supply is cut off.

The only drawback of the EAROM is the write operation is very slow compared to the RAM memory.

EAROMs are useful in areas when small amount of information is to be stored permanently, which may require changes from time to time and where the battery backup may not be available or possible.

1.3.6.6 Flash ROM

A new type of EEPROM that can be erased and reprogrammed using the normal operating voltage is called flash memory. The flash memory has the same limitation as the EEPROM that it can be erased and programmed only for a fixed number of times and it must be erased and programmed in blocks.

Because of the above two reasons the "flash memory" has not gained acceptance as the main memory in the computer Industry. Some manufacturers use it as disk emulators to provide a very fast disk drive, but it requires special software or drivers to make sure that the number of erase and write operation remain minimum.

Flash memory gets its name because the micro chip is organized so that a section of memory cells are erased in a single action or 'flash'. Flash memory is used in digital cellular phones, digital cameras, LAN switches, PC Cards for notebook computers and other devices.

1.3.7 Summary

Table 1 summarizes the features of each type of memory, but keep in mind that different memory types serve different purposes. Each memory type has its strengths and weaknesses.

Type	*Volatile?*	*Writeable?*	*Erase Size*	*Max Erase Cycles*	*Cost (per Byte)*	*Speed*
SRAM	Yes	Yes	Byte	Unlimited	Expensive	Fast
DRAM	Yes	Yes	Byte	Unlimited	Moderate	Moderate
Masked ROM	No	No	n/a	n/a	Inexpensive	Fast
PROM	No	Once, with a device programmer	n/a	n/a	Moderate	Fast
EPROM	No	Yes, with a device programmer	Entire Chip	Limited	Moderate	Fast
EEPROM	No	Yes	Byte	Limited	Expensive	Fast to read, slow to erase/write
Flash	No	Yes	Sector	Limited	Moderate	Fast to read, slow to erase/write

Table 1. Characteristics of the various memory types

Computer memory refers to the electronic holding area for instructions and data, where computer's microprocessor can quickly reach. The word *bit* is a contraction of a *binary digit.* It is the smallest unit of information in a computer, either ON or OFF, represented in binary as either 1 or 0.The various units that are used to measure computer memory are bit, byte, kilobyte, megabyte, gigabyte and terabyte. Computer's memory hierarchy can be defined into three fundamental types, namely, internal processor memory, primary memory and secondary memory.

Internal Processor memory is placed in the CPU and it includes cache memory and special registers that are directly accessed by the processor. Primary memory refers to RAM (Random Access Memory), which is volatile in nature and ROM (Read Only Memory), which is non-volatile in nature. Secondary memory also called as auxiliary memory, is used to provide backup storage for instructions and data. It has much larger capacity than primary memory.

RAM is a kind of memory that allows computer to store data for the immediate manipulation and to keep track of what is currently being processed. Computer RAM is categorized into two types, namely, SRAM and DRAM. *SRAM* stands for Static Random Access Memory. It retains data as long as power is provided to the memory chip and does not require constant periodic refreshing. *DRAM* stands for Dynamic Random Access Memory. It retains data as long as power is provided to the memory chip but does require constant periodic refreshing. They are slower and less expensive than SRAM.

The non-volatile memory chip, where initial startup instructions (special boot program) are stored is called *ROM. Masked ROM* was the first hard-wired device that contained a pre-programmed set of instructions. *PROM* chips are non-volatile but more fragile than other ROMs. Erasing of data in *EPROM chips* is done by exposing the chip under the ultraviolet radiations. Process of re-programming these chips is much flexible than PROM chips. Erasing of data in *EEPROM chips* is done by applying the electrical charges on the chip. These chips are much faster than the *EPROM* chips. Flash ROM is also called as *flash BIOS.* It is a non-volatile memory that can be erased and reprogramed in blocks of data.

1.3.8 Keywords :

Primary Memory : The computer's internal memory, which is typically made up of dynamic RAM chips. Until non-volatile RAM, such as magnetic RAM (MRAM), becomes commonplace, the computer's primary storage is temporary. When the power is turned off, the data in primary storage are lost.

Secondary Memory : Secondary memory (or Secondary Strorage) is the slowest and cheapest form of memory. It cannot be processed directly by the CPU. It must first be copied into primary storage (also known as RAM). Secondary memory devices include magnetic disks like hard drives and floppy disks; optical disks such as CDs and CDROMs; and magnetic tapes, which were the first form of secondary memory.

RAM : A type of computer memory that can be accessed randomly; that is, any byte of memory can be accessed without touching the preceding bytes. RAM is the most common type of memory found in computers.

ROM : A type of computer memory on which data has been prerecorded. Once data has been written onto a ROM chip, it cannot be removed and can only be read.

Cache Memory : A memory cache, sometimes called a cache store or RAM cache, is a portion of memory made of high-speed static RAM (SRAM) instead of the slower and cheaper dynamic RAM (DRAM) used for main memory. Memory caching is effective because most programs access the same data or instructions over and over. By keeping as much of this information as possible in SRAM, the computer avoids accessing the slower DRAM.

1.3.9 Short Answer Type Questions :

Q:1 What is Primary Memory?

Q:2 What is Cache Memory?

Q:3 Explain the following terms:

- RAM
- ROM
- EPROM
- EEPROM

Q:4 Distinguish between the following:

- RAM and ROM
- Static RAM and Dynamic RAM
- Primary memory and Internal processor memory

1.3.10 Long Answer Type Questions :

Q:1 What are the types of memories available in the computer system? Write down the various units, which are used in measuring the computer memory.

Q:2 What is RAM? How it is different from ROM? Explain the various types of RAM.

Q:3 What is ROM? Explain the various types of ROM.

1.3.11 Suggested Readings:

1. Computer Fundamentals” By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Introduction to Information Technology by V.Rajaraman, PHI, New Delhi.
3. Fundamentals of Information Technology, by Alexis Leon and Matthews Leon Vikas Publishing House
4. Fundamentals of Information Technology by Deepak Bharihoke, Pentagon Press.
5. Memory Systems : Cache, DRAM, Disk by Bruce Jacob

LESSON NO. 1.4 **AUTHOR: ARUN SHARMA**

SECONDARY STORAGE

1.4.0 Objectives

After completing this lesson, you will be able to :

- State the relevance of Secondary Memory
- Define methods of accessing data from Secondary Memory
- List the different Secondary Storage Devices
- Define how data can store in these devices
- Learn the advantages and disadvantages of these devices

1.4.1 Introduction

In the previous lesson, we discussed about primary memory, which is volatile in nature and has a very limited storage capacity. This kind of memory is mainly used for processing the data. Being volatile, primary memory cannot hold data or instructions once the computer is switched off. Therefore, a computer requires more stable (non-volatile) type of memory so that it can store all the data (files) and instructions (software programs) even after the computer is turned off. This kind of memory is known as secondary memory or auxiliary memory or peripheral storage or secondary storage. As opposed to primary storage, which the Central Processing Unit (CPU) uses for processing data and instructions, secondary storage is used to store data and programs when they are not being processed. Secondary storage includes devices like hard disks, floppies, CD-ROMs and magnetic tapes. These devices have a larger (and more permanent) storage capacity and they are less expensive as compared to primary storage devices, but they are slow in comparison.

In this lessson, different types of secondary storage devices have been discussed.

1.4.2 Benefits of Secondary Storage

Modern computing systems consist of four hardware components: central processing unit, memory, secondary storage and input-output devices that interact with end users. The input device is used for providing data to the computer. The CPU processes this data and provides the output via the output device. The secondary storage is used to store the data and instructions permanently. These data (files) and instructions (software programs) are then loaded in primary memory (RAM) so that the computer can process the data efficiently. Once the output is presented by the computer, if the user wants to, it is stored on the secondary storage.

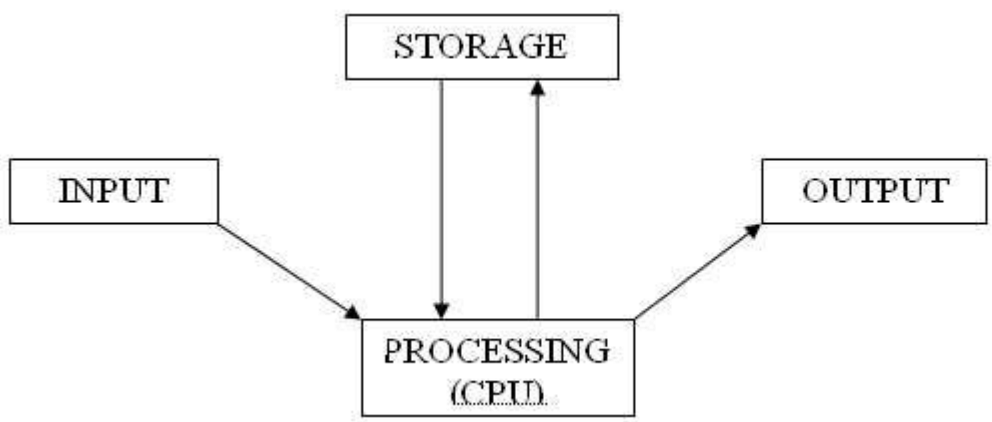


Figure: Basic Computer Operation

Imagine how many filing-cabinet drawers would be required to hold the millions of files of, say, tax records kept by a big corporation. The record storage rooms would have to be enormous. Computers, in contrast, permit storage on a tape or disk in extremely compressed form. Storage capacity is unquestionably one of the most valuable assets of the computer. Secondary storage is the storage space where one can store software and data on a semi-permanent basis. It is necessary because primary storage can be used only temporarily. The user probably will like to store the data that he derives from processing; that is why secondary storage is needed. Furthermore, main memory is limited in size, whereas secondary storage media can store as much data as necessary.

The benefits of secondary storage can be summarised as follows:

- **Non-Volatility:** By nature, a secondary storage device is non-volatile, that is, it does not lose its contents even when its power is cut off. Hence, such devices do not require a continuous supply of electricity as the primary memory does.
- **Capacity:** Secondary storage devices are used by organisations so that they

can store large volumes of data (equivalent of a room full of data on paper) in sets of disks that take up less space than a breadbox. A simple diskette holds the equivalent of 500 printed pages or 1 book. An optical disk can hold the equivalent of approximately 500 books.

- **Reliability:** Data in secondary storage is safe because secondary storage is physically reliable. Sometimes, however, disks may fail but the overall reliability of secondary storage is excellent. Moreover, it is more difficult to tamper or illegally access the data on secondary storage as compared to data stored in traditional file cabinets.
- **Convenience:** With the help of a computer, authorised people can locate and access data quickly.
- **Cost:** It is less expensive to store data on a tape or disk (the principal means of secondary storage) than to buy and house filing cabinets. Data that is reliable and safe is less expensive to maintain than data subject to errors. Nevertheless, the greatest savings can be found in the speed and convenience of filing and retrieving data.
- **Reusability:** The data remains in the secondary storage as long as it is not overwritten or deleted by the user. When the data is placed in the RAM, the computer actually makes a copy of that data in the memory. The user can change the data as per requirement, and in case he wants to revert to the original copy, he can simply close the data file as long as he has not saved the modified work. In that case, CPU will replace the original copy with the modified copy.
- **Portability:** Modern day storage devices like CD-ROMs and floppy disks are so small that they can be easily ported from one computer to another. In addition, since most of the storage devices are standardised, they can be used with almost every computer irrespective of their construct.

These benefits apply to all the secondary storage devices but some devices are better than the others. In the following sections, we will begin with a look at the various storage media, including those used for persona! computers and then consider what it takes to get data organised and processed.

Difference between Primary and Secondary Memory

Sr. No.	Primary Memory	Secondary Memory
1.	Fast	Slower than Primary Memory
2.	Expensive	Cheaper
3.	Volatile	Non-volatile
4.	It supports ongoing CPU activities	Acts as a storage device

1.4.3 Classification of Secondary Storage Devices

Secondary storage devices allow us to store the information and programs permanently. The information in a secondary storage device can be accessed, depending upon how the information is stored on the storage medium. Primarily, there are two methods of accessing data from the secondary storage devices:

1. **Sequential:** Sequential access means the computer system must search the storage device from the beginning until it finds the required piece of data. The most common sequential access storage device would be a magnetic tape where data is stored sequentially and can be processed only sequentially. Suppose, a tape contains information regarding employees of an organisation. Here, to look for employee number 100's information, the computer will have to start with 1, then go past 2, 3, and so on, until it finally comes to 100. This data access method is less expensive than other methods because it uses magnetic tape, which is cheaper than disks. The disadvantage of sequential organisation is that searching for data is slow.
2. **Direct:** Direct access, also known as random access, means that the computer can go directly to the information that the user wants. The most common direct access storage is the disk and the most popular types of disks today are

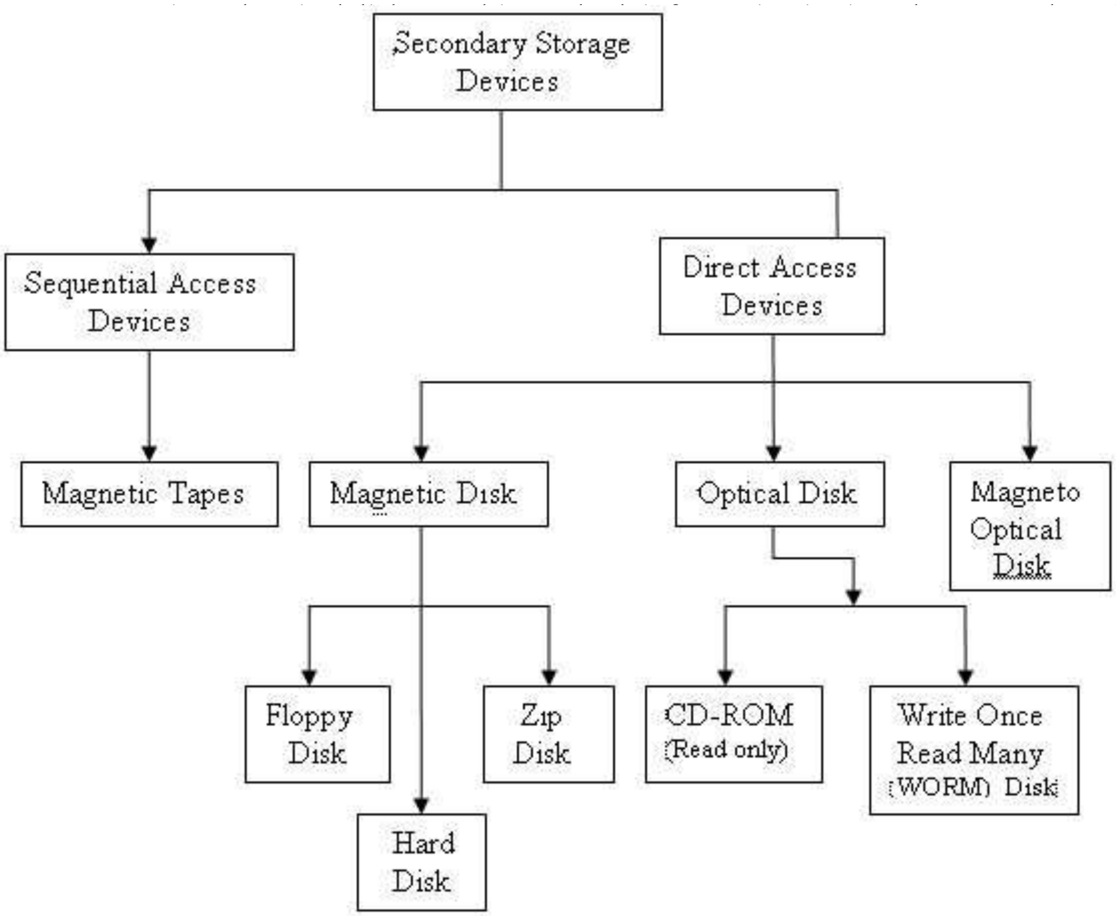


Figure Classification of Secondary Storage Devices

Self Check Exercise

Q: What is Secondary Memory?

Q: What are the benefits of Secondary Storage devices?

1.4.4 Magnetic Tape

It is oldest type of secondary storage device. Magnetic tape mechanism is similar to the commonly used audio tape recorders. It is a plastic tape with magnetic coating. The data is stored in the form of tiny segments of magnetised and de-magnetised portion on the surface of material. Magnetised portion of the surface refers to the bit value '1' where as demagnetised portion refers to the bit value '0'. Tapes come in a number of forms, including ½-inch wide tape wound on a reel, ¼-inch wide tape in data cartridges and cassettes and tapes that look like ordinary music cassettes but are designed to store data instead of music.

The major differences between magnetic tape units are the speed at which the tape is moved past the read/write head and the density of the recorded information. The amount of data or the number of binary digits that can be stored on a linear inch of tape is known as the tape's recording density.

Magnetic tapes are very durable and can be erased as well as reused. These tapes are the least expensive and reliable storage medium for organising archives and taking backup. However, magnetic tapes are not suitable for data files that need to be revised or updated often because it stores data in a sequential manner. Being sequential access devices means that the user must advance or rewind the tape to the position where the requested data starts. Tapes are also slow due to the nature of the media. While data transfer can be increased by increasing the speed of the tape, this can lead to two problems. One, the heads are more likely to miswrite due to not having enough time to align the polarities of the magnetic particles, or may misread due to the tape not being under the head long enough. The other problem has to do with the durability of the tape media itself - the faster the tape is started, pulled and stopped, the more it will stretch. If the tape stretches too much, then it will render it unusable for data storage and data loss may result. The tape now has a limited role because disk has proved to be a superior storage medium.

Advantages:

(1) Data-entry to CPU from magnetic tape is 100 times faster than the card readers.

(2) Magnetic tape is used for input, output and very large amount of data storage.

(3) Magnetic tape cassettes are easily transportable.

(4) Magnetic tapes are available in varying densities.

(5) Magnetic tapes hold high data recording density, resulting in low cost per bit of storage.

Disadvantages:

(1) Loading and unloading of tape require much time.

(2) Handling of the tape must be done carefully. Tapes should be kept in

dust free environment and away from corrosive gases or chemicals.

(3) Effect of magnetism and heat may damage the tapes.

(4) It is necessary to check the accuracy of data stored on the tape.

(5) They are not flexible when file updating requires record insertion and deletion.

1.4.5 Magnetic Disk

Magnetic disks are the most widely used and popular storage medium for direct access secondary storage. They offer high storage capacity, reliability and have the capability to access the stored data directly. A magnetic disk comprises a thin piece of plastic/metal circular plate/platter, which is coated with magnetic oxide layer. Data is represented as magnetised spots on a disk. A magnetised spot represents as 1 (bit) and the absence of a magnetised spot represents a 0 (bit). To read the data, the magnetised data on the disk is converted into electrical impulses, which is transferred to the processor. Writing data onto the disk is accomplished by converting the electrical impulses from the processor into magnetic spots on the disk. The data in a magnetic disk can be erased and reused virtually infinitely. The disk is designed to reside in a protective case or cartridge to shield it from the dust and other external interference.

The surface of a disk is divided into imaginary tracks and sectors. Tracks are concentric circles where the data is stored. Disk sectors refer to the number of fixed size areas that can be accessed by one of the disk drive's read/write heads, in one rotation of the disk, without the head having to change its position. An intersection of a track and a disk sector is known as track sector. Generally, a disk has eight or more disk sectors per track. However, different types of magnetic disk may have different number of tracks. Modern day disks are marked (tracks and sectors) on both surfaces, hence they are also known as double-sided disks.

In order to make the disk usable, first it must be formatted to create tracks and sectors. The formatting of a magnetic disk refers to the assignment of addresses to the different areas in the disk. Without formatting, there would be no way to store data into the disk. If a sector is corrupted after formatting, it will be marked as damaged, and the data saved on it will not be readable.

Frequently, multiple disks are maintained and used together to create a large disk storage system. Typically, two, three or more platters are stacked on top of each (disk pack) other with a common spindle, which turns the whole assembly. There is a gap between the platters, making room for magnetic read/write head. There is a read/write head for each side of each platter, mounted on arms, which can move them towards the central spindle or towards the edge. This concept of stacking the disk is known as cylinder.

1.4.5.1 Types Of Magnetic Disk

Broadly, magnetic disks can be classified into three types: floppy disk, hard drive and zip disk.

Floppy Disk

It is the most popular secondary storage device used to transfer the files from one PC to another Pc. Floppy disk is made of plastic material. It is coated with magnetic oxide. It is normally covered with a plastic or cardboard sleeve for protection. A hole at the center is provided for mounting disk drive hub and an another small hole for index mark sensing. The data is stored along concentric circles called tracks. A notch at the right side of the disk is used to protect floppy from writing over it. There are two standard sizes, 5½ inch and 8 inch. Now 8-inch floppy is outdated. A more recent development is 3½ inch. The capacity of floppy disk is varied (e.g. 360 KB, 1.2 MB, 2 MB, 1.44 MB etc.)

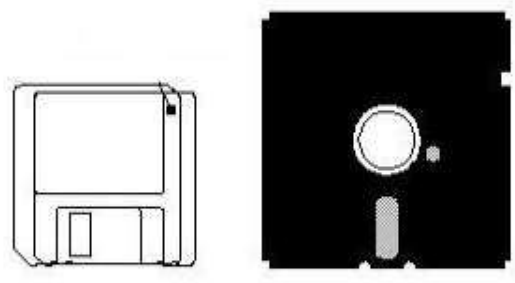


Figure: Floppy disks

The floppy disk is slipped into the drive mechanism for read/write operation. The mechanism holds the envelope and the flexible disk are rotated inside the envelope by the drive mechanism. The read/write head is in contact with a slit for operation. The head is moved radically. i.e. track to track. The floppy disk is a low cost device. Floppies have a speed limitation of about 360-RPM. A low-density floppy disk has 40 tracks, 9 sector per track and 515 bytes per sector. Floppy disks are used for data preparation, small business applications and word processing systems.

Hard Disk:

The hard disk, also called the hard drive or fixed disk is the primary storage unit of the computer. A hard disk consists of a stack of disk platters that are made up of aluminum alloy glass substrate coated with a magnetic material and protective layers. The hard drive's speed is discussed in terms of access time (the speed at which the hard drive finds data), which is measured in milliseconds. The average drives have an access time of 9 to 14 ms. The lower the access time, the faster the hard drive. The capacity, or amount of information that a hard drive can store, is measured in bytes. Hard disk plays a significant role in the following important aspects of a computer system:

- **Performance:** The hard disk plays a very important role in overall system performance. The speeds at which the computer boots up and programs load are directly related to the hard disk speed. The hard disk's performance is also critical when multitasking is exercised.
- **Storage Capacity:** A bigger hard disk lets one store more software and data into it, thereby permitting the user to store large software related to complex processes such as graphics and multimedia.

- **Software Support:** Nowadays almost all software needs large storage space and faster hard disks to load them efficiently. Usually, older hard disks can barely hold the operating system and few uncomplicated software that does not require much space.
- **Reliability:** One way to assess the importance of any hardware component is to consider how much damage is caused in case it fails. By this standard, the hard disk is considered the successful storage component by a long way. A good quality hard disk, combined with smart maintenance and backup habits, can help ensure that the nightmare of data loss doesn't become part of daily life.

Zip disk

Zip disk is a removable storage device and it has a capacity to store 100-250 MB of data Zip disk's drive unit is measured at 18 x 13 x 4 cm and weighs about half kilogram. It has rubber feet to stabilise the unit in either vertical or horizontal position. The substrate for the disk is made up of plastic material on which magnetic oxide particles are coated. There are two indicator lights, green for power and amber for disk access. It has an eject button, but no ON/OFF switch. It comes complete with drive, connection cable, power supply, operating software and a starter 100 MB disk. After installation, you can transfer files to and from the zip drive as if you are accessing hard disk or floppy drive.

The zip disk is similar in diameter to that of the 3½-inch floppy diskette. However, one should keep in mind that although they may look similar to floppy disks, but a zip drive cannot read or write floppy disks and similarly a floppy drive cannot read/write data onto a zip disk. Either the zip drive can be built into the computer or it can be connected to the computer via a parallel or SCSI cable. There are different versions of zip disks and drives available in the market. These versions can hold 100, 250 and 110 MB of data. As a zip disk can hold as much as 70 times of information as compared to a floppy disk, it can be used to store heavy graphics, music or presentation files.

Advantages of Magnetic Disks

- Magnetic disks follow direct access mode for reading and writing onto the data files, thereby making an ideal device for accessing frequently accessed data.
- Magnetic disks are easily moveable from one place to another because of their small size and lightweight.
- The data transfer rate of disks is much higher than magnetic tapes.
- Due to their low cost and high data recording densities, the cost per bit in magnetic disks is minimum.
- The storage capacity of these disks is virtually unlimited as numbers of such disks can be added to store data.
- Magnetic disks are less prone to the corruption of data as they can withstand the temperature and humidity change in a much better way

as compared to magnetic tapes.

Disadvantages of Magnetic Disks

- Magnetic disks must be stored in a dust-free environment in order to protect them from crashing down.
- The cost of Magnetic disks storage is more costly than the cost of magnetic tapes.
- Magnetic disks are less secure than magnetic tapes.

Self Check Exercise:

Q Name the various types of Magnetic Disk.

Q What is Hard Disk?

1.4.6 Optical Disk

As compared to magnetic tape and magnetic disk, optical disk is a relatively new secondary storage medium. During the last few years, it has proved to be a promising random access medium for high capacity secondary storage, because it can store extremely large amounts of data in a limited space.

An optical-disk storage system consists of a rotating disk, which is coated with a thin metal or some other material that is highly reflective. Laser beam technology is used for recording/reading of data on the disk. Due to the use of laser beam technology, optical disks are also known as laser disks or optical laser disks.

Unlike magnetic disks, which have several concentric tracks, an optical disk has one long track, which starts at the outer edge and spirals inward to the center. This spiral track is ideal for reading large blocks of sequential data, such as music. However, it makes for slower random access time than the concentric tracks used by magnetic disks, whose sectors can be located faster because they are always found on a given track at a fixed distance from the center.

(a) Track pattern on an optical disk (b) Track pattern on a magnetic disk

Difference in track patterns on optical and magnetic disks.

Like a track on a magnetic disk, the track of an optical disk is split up into sectors, but with optical disks, each sector has the same length, regardless of whether it is located near the disk's center or away from the center. This type of data organization allows data to be packed at maximum density over the entire disk. However, it also requires a more complicated drive mechanism, because the rotation speed of the disk must vary inversely with the radius; the drive must slow down the disk's rotation speed to read sectors towards the outside of the disk and speed it up to read sectors towards the center of the disk.

The cost-per-bit of storage is very low for optical disks, because of their low cost and enormous storage density. They come in various sizes, ranging from 12-inch to 4.7-inch diameter. The most popular one is of 5.25 inch diameter, whose capacity is around 650 Megabytes. This storage capacity is equivalent to about 2,50,000 pages of printed text.

1.4.6.1 Types of Optical Disks

All optical disks are round platters. They come in different sizes and capacities. The two most popular types of optical disks in use are CD-ROM and WORM disks.

- **CD-ROM:** CD-ROM stands for Compact Disk-Read-Only Memory. It is a spin-off of music CD technology, and works much like the music CDs used in music systems. In fact, if you have a soundboard and speakers connected to your computer, you can play music CDs with your computer.

  The CD-ROM disk is a shiny, silver color metal disk of 5½ inch (12 cm) diameter. It has a storage capacity of about 650 Megabytes. It is so called, because of its enormous storage capacity on a compact-size disk and because it is a read-only storage medium. That is, these disks come pre-recorded and the information stored on them cannot be altered.
- **WORM Disk:** WORM stands for write-once, read-many. WORM disks allow the users to create their own CD-ROM disks by using a CD-recordable (CD-R) drive, which can be attached to a computer as a regular peripheral device. WORM disks, which look like standard CD-ROM disks, are purchased blank and encoded using a CD-R drive. The information recorded on a WORM disk by a CD-R drive can be read by any ordinary CD-ROM drive. As the name implies, data can be written only once on a WORM disk, but can be read many times. That is, as with a CD-ROM disk, once data has been etched on to the surface of a WORM disk, it becomes permanent, which can be read, but never altered.

Advantages:

1. The cost-per-bit of storage for optical disks is very low, because of their low cost and enormous storage density.
2. The use of a single spiral track makes optical disks an ideal storage medium for reading large blocks of sequential data, such as music.
3. Optical disk drives are more reliable storage medium than magnetic tapes or magnetic disks.
4. Optical disks have a data storage life in excess of 30 years.
5. Since data once stored on an optical disk becomes permanent, the danger of stored data getting inadvertently erased/overwritten is not there with optical disks.
6. Due to their compact size and light weight, optical disks are easy to handle, store, and port from one place to another.

7. Music CDs can be played on a computer having a CD-ROM drive along with a sound board and speakers. This allows computer systems to be also used as music systems, whenever desired.

Limitations:

1. It is a read-only (permanent) storage medium. Data once recorded, cannot be erased, and hence, the optical disks cannot be reused.
2. It is not easy to copy an optical disk, as it is to a floppy disk. One needs to have software and hardware for writing disks.
3. The data access speed for optical disks is slower than magnetic disks.
4. Optical disks require a more complicated drive mechanism than magnetic disks.

Uses of Optical Disks

Optical disks are typically used for one or more of the following purposes:

1. For distributing large amounts of data at low cost. For example, a complete encyclopedia, dictionary, world atlas, dictionary of quotations, biographies of great people, information about all educational institutions of a particular type in a country, etc. are often distributed on CD-ROM disks.
2. For distribution of electronic version of conference proceedings, journals, magazines, books, product catalogs, etc.
3. For distribution of new or upgraded versions of software products by software vendors.
4. For storage and distribution of a wide variety of multimedia applications, such as video games.
5. For archiving of data, which are not used frequently, but which may be used once in a while.
6. WORM disks are often used by end-user companies to make permanent storage of their own proprietary information. For example, many banks use them for making a permanent record of their daily transactions.

1.4.7 Magneto-Optical Disk

As implied by the name, these drives use a hybrid of magnetic and optical technologies, employing Laser to read data on the disk, while additionally needing magnetic field to write data. A magneto-optical disk drive is so designed that an inserted disk will be exposed to a magnet on the label side and to the light (laser beam) on the opposite side. The disks, which come in 3½-inch and 5½-inch formats, have a special alloy layer that has the property of reflecting laser light at slightly different angles depending on which way it is magnetised, and data can be stored on it as north and south magnetic spots, just like on a hard disk.

Magneto-optical drives use a laser to target and heat specific regions of magnetic particles. This accurate technique enables magneto-optical media to pack in a lot more information than other magnetic devices. Once heated, the magnetic particles can easily have their direction changed by a magnetic field generated by the read/write head. Information is read using a less powerful laser, making use of the Kerr effect, where the polarity of the reflected light is altered depending on the orientation of the magnetic particles. Where the laser/magnetic head has not touched the disk, the spot represents a '0', and the spots where the disk has been heated up and magnetically written will be seen as data' 1 '. However, this is a 'two-pass' process, which, coupled with ill tendency for magneto-optical heads to be heavy, resulted in early implementations being relatively slow. Nevertheless, magneto-optical disks can offer very high capacity and cheap media as well as top archival properties, often being rated with an average life of 30 years, which is far longer than any magnetic media.

1.4.7.1 Mass Storage Devices

In order to get vast amount of storage capacity of several bytes (trillions and more) in a computer system, a different kind of storage system is used. In such type of system, multiple units of similar kinds of storage media are associated together to form a chain of mass storage devices. These storage media may include multiple magnetic tape reels or cartridges, multiple arrays of magnetic disks or multiple CD-ROMs as a secondary storage device.

Broadly speaking, we can categorise mass storage devices into three types:

- Redundant Array of Inexpensive Disks (RAID): The basic idea of RAID is to combine multiple hard disks into an array of disk drives to obtain high performance, large capacity, and reliability that exceeds that of a single large drive. These arrays of drives appear to the host computer as a single logical drive.
- Automated Tape Library: An automated tape library comprises numerous set magnetic tapes along with their drives and controllers mounted in a single unit. The unit comprises one or more tape drives to perform read/write operations on the tapes in the tape library. In the multiple tape drive environment, these tapes can be simultaneously read or write, thus, resulting in the speedy rate of data transfer. Multiple drives lead to the reliability of the storage unit because if one of the drives fails, then the unit can continue to operate with other tape drives. These tape libraries, can store up to several terabytes of data.
- CD-ROM Jukebox: A CD-ROM jukebox comprises numerous set of CD-ROM disks along with their drives and controllers mounted in a single unit. The unit comprises one or more CD-ROM drives to perform read/write operations on the CD-ROM in the jukebox. In the multiply CD-ROM drive environment, these CD-ROMs can be simultaneously read or write, thus resulting in the speedy rate of data transfer. Multiple drives lead to the reliability of the storage unit because if one of the drives fails, then the unit can continue to operate with other CD-ROM drives. These jukeboxes can store up to several terabytes of data.

1.4.8 Summary

Secondary memory refers to the memory, which can store all the data (files) and instructions (software programs) even after the computer is turned off. It includes devices like hard disk, floppy disk, CD-ROM and tape drives. Secondary storage comes with many benefits, that is, non-volatility, capacity, reliability, convenience, cost, reusability and portability.

Secondary storage devices are classified into two types according to data access, that is, sequential access and direct access. Sequential access refers to the searching of the storage device from the beginning until it finds the required piece of data. The most common sequential access storage device is the magnetic tape where data is stored sequentially and can be accessed only sequentially. Direct access means going directly to the information that the user wants. The most common direct access storage is the disk and the most popular types of disks today are magnetic and optical disks.

Magnetic tape is like a plastic tape with a magnetic coating on it. They are very durable and can be erased as well as reused. These tapes are the least expensive and reliable storage medium. Magnetic disks are the most widely used and popular storage medium for direct access secondary storage. They offer high storage capacity, reliability and have the capability to access the stored data directly.

A floppy disk is a removable round, flat piece of mylar plastic, coated with ferric oxide and encased in a protective plastic. This kind of disk is read and written by a floppy disk drive. Zip disk is a removable storage device having a capacity to store 100-250 MB of data. Zip disk's drive unit is measured at 18 x 13 x 4 cm and weighs about half kilogram.

An optical disk is a flat, circular, plastic disk coated with material on which bits may be stored in the form of highly reflective areas and significantly less reflective areas, from which the stored data may be read when illuminated with a narrow-beam of laser diode. A compact disk is an optical disk that is used to store music, in the form of digital audio. These disks are also capable of being used as a data storage device. WORM (Write Once-Read Many) disk is an optical disk that can be used to read as well as to write (usually only once) data onto its surface. Data on these disks can be written, but cannot be erased.

MO (Magneto-optical) disk is a hybrid of magnetic and optical technologies. It employs laser beam to read data on the disk, while additionally needing electromagnetic head to write data. Redundant Array of Inexpensive Disks (RAID) is a type of mass storage device that combines multiple inexpensive hard disk drives into an array of disk drives to obtain high performance, large capacity, and reliability that exceed that of a single large drive. An automated tape library comprises numerous sets of magnetic tapes along with their drives and controllers mounted in a single unit. A CD-ROM jukebox comprises numerous set of CD-ROM disks along with their drives and controllers mounted in a single unit.

1.4.9 Keywords :

Secondary Memory : Secondary memory (or Secondary Strorage) is the slowest and cheapest form of memory. It cannot be processed directly by the CPU. It must first be copied into primary storage (also known as RAM). Secondary memory devices include magnetic disks like hard drives and floppy disks; optical disks such as CDs and CDROMs; and magnetic tapes, which were the first forms of secondary memory.

Magnetic Tape: Tape on which computer-readable data is electronically stored via magnetic particles embedded in the tape.

Magnetic Disk : A memory device, such as a floppy disk, a hard disk, or a removabel cartriodge, that is covered with a magnetic coating on which digital information is stored in the form of microscopically small, magnetized needles.

Floppy Disk : A flexible plastic disk coated with magnetic material and covered by a protective jacket, used primarily by computers to store data magnetically. Also called diskette.

Hard Disk : A rigid magnetic disk fixed permanently within a drive unit and used for storing computer data. hard disks generally offer more strorage and quicker access to data than floppy disks do.

Optical Disk : A plastic-coated disk that stores digital data, such as music or text, as tiny pits etched into the surface and is read with a laser scanning the surface. Also called laser disk.

1.4.10 Short Answer Type Questions

Q:1 What is auxiliary memory? State at least four uses of it.

Q:2 Discuss advantages and disadvantages of:
- Magnetic disks
- Optical disks

Q:3 Explain the following:
- Fixed Disk
- Zip Disk
- Hard Disk
- Floppy Disk

Q:4 Difference between Primary and Secondary Memory

1.4.11 Long Answer Type Questions

Q:1 What are secondary storage devices? How are they useful in storing large amount of data? What are the benefits of secondary storage devices?

Q:2 Write a short note on different types of magnetic tapes and magnetic disks.

Q:3 How is sequential access method different from the direct access method?

1.4.12 Suggested Readings:

1. Computer Fundamentals” By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Introduction to Information Technology by V.Rajaraman, PHI, New Delhi.
3. Fundamentals of Information Technology, by Alexis Leon and Matthews Leon Vikas Publishing House
4. Fundamentals of Information Technology by Deepak Bharihoke, Pentagon Press.
5. Memory System : Cache, DRAM, Dush by Bruce Jacob

LESSON NO. 1.5 **AUTHOR: ARUN SHARMA**

INPUT AND OUTPUT DEVICES

1.5.0 Objectives

After completing this lesson you will be able to :

- Defining basic Input and Output Concepts
- State the role of peripheral device
- List the input-output devices
- Show the operation of peripheral devices

1.5.1 Introduction

Many traditional ways to input and output data were designed more with the machines in mind than the humans who used them. However, the trend in input and output (I/O) devices is to move toward a more natural means of communication. The most natural way for humans to communicate with a computer is to use the same techniques that people use to communicate with each other. For example, entering data into the computer might be accomplished by speaking, handwriting or physically pointing at specific objects on the screen. Speech, written text, and visual images such as color graphs are forms of output that humans find easiest to interpret and use. Advances such as voice recognition and speech synthesis have only begun to reach their potential. This lesson will present some basic I/O concepts as well as look at the more common I/O devices.

1.5.2 Input Concepts

For computer processing, data needs to be entered into the computer (Input) and the result of processing needs to be communicated to the user (output). Input is the process of entering and translating incoming data into machine readable form. The data to be entered are often referred to as input. Output is the process of translating data in machine - readable form into a form understandable by humans or readable by other machines. The information that is the result of processing is also often referred to as output. Before going into detail, let us recall the meanings of some terms. Hardware is all the physical components of a computer system. Any hardware item which is attached to the main unit of a computer that houses the CPU is referred to as a peripheral device. An input device is a peripheral device through which data are entered and transformed into machine-readable form.

Three things to be concerned with during the input process are:

1) The method of data preparation
2) The type of processing to be used
3) The accuracy of the data being entered

1.5.2.1 Data Preparation Methods

There are two concepts related to data preparation methods:

(a) On-line
(b) Off-line

(a) On-Line refers to an operation or device in which data are transferred directly to or from a computer.
(b) Off-Line refers to an operation or device in which data are not immediately transferred to or from a computer.

The end product of an off-line input operation may be used as data for an on-line input operation. Now-a-days on-line devices and operations are most commonly used.

Data for input can be prepared in two ways:

- Transcriptive data entry

- Source data entry

(i) Transcriptive Data Entry

The data are prepared on documents at the source of the data. The data must then be transcribed to another medium that can be read and interpreted by a computer.

(ii) Source Data Entry

The data are prepared at the source in a machine readable form that can be used by a computer without a separate intermediate data transcription step. This type of data preparation reduces the number of errors during input process.

1.5.2.2 Batch and Transaction Processing

The two main data processing methods are:

(a) Batch Processing

(b) Transaction Processing

(a) Batch Processing

Batch processing is a method in which data are gathered and batched through either an on-line or off-line data gathering process and processed periodically without user intervention.

(b) Transaction Processing

Transaction processing is an on-line method in which data are processed immediately, and files are updated as a transaction takes place. This allows the user to communicate with the computer during processing through interactive programs which supply the results of processing so that they can be used when needed to control an operation or modify a process. This procedure is called real time processing. It is used in airlines offices to book and prepare a ticket.

1.5.2.3 Data Accuracy

There are two most commonly used methods which are used to help eliminate errors at the time of data entry.

a) Verification which is performed by the user

b) Validation which is programmed into the computer

(a) Verification

This can be conducted by corroborating the data entered by checking them against a known source.

(b) Validation

Computer can be programmed to accept only a certain range of data. Checking for data items that deviate from this range is called validation.

1.5.3 INPUT DEVICES

Before computer processing data must be entered into the computer by an input device so that they can be translated into a machine readable form. Let us have a look at various input devices available:

1.5.3.1 Punched Cards and Card Readers

It is a rigid piece of paper containing rows and columns of numbers. These cards are put into a keypunch machine which places holes in the card to represent the data. This pattern of punched holes used to represent characters on the card is called the Hollerith code. These holes are interpreted by a card reader that translates the punched information into a machine readable code and then sends it to the CPU for processing. Now-a-days this system is totally replaced by newer, more efficient input device.

Thus the punched card has numerous uses in industry. It is a widely used file in small-scale computer installations.

Its basic advantages are the following:

1. Machine and user accessibility. i.e. people as well as machines, can read card data
2. Relatively low cost
3. Can be used as a turn around document

Its basic disadvantages are the following:

1. Large volumes of cards becomes difficult to control
2. Mechanical parts of many card devices cause frequent breakdown
3. Low speed of card reader and punch devices
4. Not reusable
5. Inflexibility of the card format-either 80 or 96 columns per card

While cost is minimal efficiency is limited with punched card processing. Not only can cards be misplaced or missorted rather easily, but machines that read cards are subject to mechanical breakdowns. Moreover, card reader and punch devices are notoriously slow as compared to computer processing speed. The average speed of card devices is far slower than the billionths of seconds (nanoseconds) required for actually processing the data. Similarly, the physical limitation of 80 or 96 characters per record is often severely restrictive.

1.5.3.2 Keyboards

One of the most common and familiar input device is the standard keyboard. The traditional QWERTY typewriter comprises the basic portion of today's computer keyboard. A typical computer keyboard contains all letters, numbers, symbols of a regular typewriter, plus other keys which may include:

(i) a numeric keypad which functions like calculator

(ii) function keys whose operation can be determined by the user or preprogrammed by the software being used

(iii) special keys such as those used to control movement of the cursor (special character or symbol that indicates the user's position on the screen or focuses attention to a specific area, to allow communication and interaction between the user and the program) on the computer screen

A cursor is a special character or symbol that indicates the user's position on the screen or focuses attention to a specific area to allow communication and interaction between the user and the program. Cursor movement keys typically include four directional arrow keys which move the cursor one space at a time in either an up, down, left, or right direction. Other keys, typically named Home, End, PgUp (page up), and PgDn (page down), move the cursor over longer distances quickly. The cursor movement keys are often part of the numeric keypad. If this is the case, a key, such as the NumLock, acts as a toggle switching the function of these keys between cursor movement and numeric operations. Other special keys common to many keyboards are ESC (escape), CTRL (control), INS (insert), DEL (delete), and PRT SC(print screen). Many of these keys were added to increase efficiency in programming and in using applications such as word processing packages.

Alphanumeric keyboard

1.5.3.3 Mouse

A mouse is a pointing device. It usually contains one or three buttons; as the user rolls it on a flat surface. The mouse controls cursor movement on the screen. When the user presses one of the buttons the mouse either marks a place on the screen or makes selections from data on the screen. It can be used for many applications ranging from games to designing products with graphics. It can be used as an alternative to keyboard or it can be used in combination with a keyboard to enhance input operations. Doug Engelbart developed the mouse as an alternative to the keyboard in the 1960s.

Mouse

1.5.3.4 Joystick

A joystick, is a most popular input device used to play video games, for other applications, which includes flight simulator, training simulator, CAD/CAM system and for controlling industrial robots. A joystick uses a level to control the position and speed with which the joystick is moved into digital signals that are sent to the computer to control the cursor movement.

1.5.3.5 Track Ball

A track ball uses a hard sphere to control cursor movement. The ball can be rotated by hand in any direction. The track ball translated the sphere's direction and speed of rotation into a digital signal, used to control the cursor.

Joystick

Track Ball

1.5.3.6 Touch Screen

A touch screen registers input when a finger or other object comes in beams and ultrasonic acoustic waves. Infrared beams crisscross the surface of the screen and when a light beam is broken, that location is recorded.

1.5.3.7 Light pen

A light pen is also a pointing device like mouse. It can be used to choose a displayed menu option. The pen contains a photocell placed in a small tube. As the user moves the tip of the pen over the surface of the screen, it detects the light coming from a limited field of view. The light from the screen causes the photocell to respond and this electric response is transmitted to the process which can identify the menu option that is triggering the photocell. Light pen is useful for graphic work, especially for Computer Aided Design (CAD) purposes.

Recall that the data entered directly into a computer system without a transcription process is referred to as source data. Methods of source data input are:

1.5.3.8 Scanners

An scanner is an input device, which translates paper documents into an electronic format, which can be stored in a computer. The input documents may be typed text, pictures, graphics or even handwritten material. This input device has been found to be very useful in preserving paper documents in electronic form. The copy of document stored in a computer in this manner will never deteriorate in quality or become yellow with are, and be displayed or printed, whenever desired. If the computer in which the scanned document is stored has the right kind of software (called image-processing software), the stored images can be altered and manipulated in interesting ways.

Scanners come in various shapes and sizes. The two commonly used types are:

1. **Flatbed scanner:** A flatbed scanner is like a copy machine, which consists of a box having a glass plate on its top and a lid that covers the glass plate. The document to be scanned is placed upside down on the glass plate. The light source is situated below the glass plate and moves horizontally from left to right when activated. After scanning one line, the light beam moves up a little, and scans the next line. The process is repeated for all the lines. It takes about 20 seconds to scan a document of size 21 cm x 28 cm.
2. **Hand-held scanner:** A hand-held scanner has a set of light emitting diodes encased in a small case, which can be conveniently held in hand during operation. To scan a document, the scanner is slowly dragged from one end of the document to its other end, with its light on. The scanner has to be dragged very steadily and carefully over the document, otherwise the conversion of the document into its equivalent bit map will not be correct. Due to this reason, hand-held scanners are used only in cases where high accuracy is not needed. They are also used when the volume of documents to be scanned is low. They are also much cheaper as compared to flatbed scanners.

1.5.3.9 Voice Input/Recognition Devices

The human need for easy and faster way of input is giving rise to a very friendly input systems, which take input using human voice, i.e. when one speaks it takes the input. These devices are in stage of experimenting and development, where various problems like speech should be independent of persons, where even the individual's voice is different, it should take input. This input device uses a microphone which converts human speech into electric signals, which are matched with already existing dictionary of patterns when match is found it gives appropriate output. The great limitation is the vocabulary and modulation, which will surely be overcome and we can get voice response systems.

1.5.3.10 Magnetic Ink Character Recognition (MICR)

MICR is the interpretation by a computer of a line of characters written in a special magnetic ink. These characters can be read by humans as well. For instance, there are a line of numbers and some odd-shaped characters on the bottom of bank's cheques. These magnetic ink characters are bank processing symbols representing the cheque number, customer account number and bank identification number. When the bank receives the cheque, the amount of the cheque is also printed in magnetic ink. These cheques are sent through a MICR reader to interpret the information, update the appropriate accounts and in some cases, sort the cheques afterwards.

There are several advantages associated with the use of MICR:

- Checks may be roughly handled, folded, smeared, and stamped, but they can still be read with a high degree of accuracy
- Processing is speeded because checks can be fed directly into the input device
- People can easily read the magnetic ink characters

The main limitation of MICR is that only the 10 digits and 4 special characters needed for bank processing are used. No alphabetic characters are available.

1.5.3.11 Magnetic Strips

They are thin bands of magnetically encoded data that are found on the back of credit cards. The data stored on the card vary from one application to another but they include account numbers or special access codes. Data in the form of magnetic strips cannot be seen or interpreted by simply looking at the card and so it can be highly sensitive or personal.

1.5.3.12 Optical Recognitions

Optical recognition occurs when a device scans a printed surface and translates the image the scanner sees into a machine readable format that is understandable by the computer. Optical recognition can be of the following types:

(a) Optical Mark Recognition (OMR)

It employs mark sensing to scan and translate, based on its location which is a series of pen or pencil marks into computer readable form. For instance, the objective type multiple choice question paper we get in the bank recruitment exam. A computerized optical mark reader scores the tests by identifying the position of the mark.

(b) Optical Bar Recognition (OBR)

This is a slightly more sophisticated type of optical recognition. Have you seen a series of thin black bars and spaces on the back of few books? They are bar codes or product codes which are arranged to represent data, such as the name of the manufacturer, and the type of product etc. An optical bar reader recognizes and interprets them on the basis of the width of the lines. The bar code is read by a scanner, and the price and product are then matched by the computer.

(c) Optical Character Recognition (OCR)

This is the most sophisticated type of optical recognition. An optical card reader works in much the same way as the human eye. It recognizes specially shaped numeric and alphabetic characters.

A set of optical characters may be used to print merchandise tags that can be read using on OCR Reader. To process the sales transaction, the information on the tag, such as item price and the inventory number can be automatically entered at a point of sale terminal. Pertinent data can be saved and transferred to the company's main computer system to be used in activities such as managing inventories and analyzing sales.

The primary advantage of OCR is that it eliminates some of the duplication of human effort required to get data into the computer. This reduction in effort can improve data accuracy and can increase the timeliness of the Information processed. However, difficulties in using OCR equipment may be encountered when documents to be read are poorly typed or have strikeovers or erasures. Also, form design and ink specifications may become more critical than is the case when people key the data from the forms.

Self Check Exercise

Q Write short notes on the following:

- **Batch Processing**
- **Transaction Processing**
- **OMR**
- **OCR**

1.5.4 Output Concepts

An output device is a peripheral device that allows a computer to communicate Information to humans or another machine by accepting data from the computer and transforming them into a usable form.

Output is divided into two general categories:

a) Output that can be readily understood and used by humans

b) Output to secondary storage devices that hold the data to be used as input for further processing by computer

Output that can be understood by humans can be in the form of

1. Hard copy
2. Soft copy

Hard Copy

It is output on paper and can be read immediately or stored and read later. This is a relatively stable and permanent form of output.

Soft Copy

It is usually a screen displayed output. It is a transient form of output and is lost when the computer is turned off.

1.5.5 Output Devices

1.5.5.1 Hard Copy Devices

Hard copy devices can be broadly grouped into two categories:

(a) Printer

(b) Plotter

(A) Printer

A printer produces output usually in the form of text on paper. Before getting into details about the types of printers, let us have an idea of the print quality. an important feature of hard copy devices.

Print quality available from hard copy output can vary considerably.

1. Near - typeset Quality This print is similar in quality to that produced by a type set machine such as print found in a magazine.
2. Letter Quality This print is made using fully formed (solid line) characters.
3. Near - letter Quality Print Printers that don't produce the fully formed characters can still print high quality documents using this type of print. On some printers this is done when the print head makes multiple passes over the same letters filling in the spaces between the dots or lines.
4. Standard - Quality Print It is produced when characters composed of dot lines are formed by a single pass of the print head.
5. Draft - Quality Print The characters are formed with a minimum number of dots or lines and are smaller than the standard quality characters.

Types of Printers

(i) Dot-Matrix Printer

It uses printer heads containing 9 to 24 pins. These pins produce patterns of dots on the paper to form the individual characters. Dot matrix printers are very popular as they are relatively inexpensive and typically print at speed of 100-600 characters per second. These printers can give us draft quality, standard quality and near letter quality prints.

(ii) Daisy Wheel Printer

This printer is used in order to get the quality of type found on type writers. The daisy wheel printer is so called because the print mechanism looks like a daisy. This printer gives us letter quality print but is slow, typically 25-55 characters per second.

The speed of printers, we have discussed above is incredibly slow compared with the processing speed of the computers to which they are connected. The processor while it is sending texts to the printers cannot normally be used for other productive work until printing is done. There are two approaches which, if used by computer systems allow users to continue to operate their computers while printing is in progress.

a) **Spooler program (software) -** allows the processor to alternate between processing a user's ongoing activity and controlling the printing process. But for using this approach, disk space or dedicated block of primary storage should be available to store the text to be printed by the spooler.

b) **Printer buffer (hardware) -** It is an additional storage device that can accept text to be printed as fast as the computer can send it. The buffer then slowly releases the text data to match the printer's speed while the computer is free to do other things.

(iii) Line Printer

Line printers or line-at-a-time printers use special mechanisms mentioned above that can print a whole line at once. They are very fast with speed ranging from 1200 to 6000 lines per minute.

(iv) Ink-Jet Printers

Ink-Jet printers form characters on paper by spraying ink from tiny nozzles through an electrical field that arranges the charged ink particles into characters at the rate of approximately 250 characters per second. The ink is absorbed into the paper and dries instantly. Various colors of ink can also be used. Although this might sound like a messy way of printing. Ink-jet printers are reliable but expensive.

(v) Thermal- Transfer Printers/Thermal Printers

An inexpensive alternative to the ink-jet printer is the thermal-transfer printer which uses heat to transfer ink to paper. These printers bond the ink onto the paper by heating pins which press against a special ink ribbon. Thermal-transfer printers can produce color printouts by using a color ribbon.

(vi) Laser Printer

When speed and quality comparable to typeset material are required and cost is no factor. a laser printer is the solution. They produce images on paper by directing a laser beam at a mirror which bounces the beam on to a drum. The laser leaves a negative charge on the drum to which positively charged black toner powder will stick. As the paper rolls by the drum, the toner is transferred to the paper. A hot roller bonds the toner to the paper.

Laser printers use buffers to store an entire page at a time and that is why they print very fast at the rate of approximately 21000 lines or 437 pages per minute.

Features of Printers

(a) Impact versus Non impact

An impact printer is one that generally operates by using a hammer to strike a character against an inked ribbon: the impact then causes an image of the character to be printed.

Impact printers function just as typewriters do. Their main disadvantage is that they are relatively slow noisy and subject to mechanical breakdowns. Despite the disadvantages, most printers we use today are impact devices.

Type of impact printers are:

- Dot matrix
- Daisy wheel
- Line

Non impact printers were recently developed to meet some of the needs that were not being satisfied by impact technology. Most high-speed page printers, for example, use some form of non impact device; but high speed is not the only benefit of nonimpact technology.

Among the non impact printers available are:

- Laser-xerographic - high speed
- Electro sensitive - high speed
- Electro photographic - high speed
- Thermal - ideal for lightweight portable terminals; quiet
- Ink-jet - for letter - quality printing

(b) Print Mechanisms for Impact Printers

1. **Wire-Matrix:** Some impact printers print character images out of a series of dots. The pattern of each character is created by extended wires that strike a carbon ribbon leaving the desired image.
2. **Character Printers**
    a) **Print wheel:** The print wheel rotates until each character moves into the required print position on the line. When all of the characters are in their proper position a hammer strikes the paper against the wheel and a line is printed.
    b) **Print Chain:** With a chain mechanism there is one print hammer for each print position on the line. The characters are printed when the hammer presses the paper against the linked ribbon.
    c) **Print drum:** The print drum is a metallic cylinder. Each column on the drum contains all the characters where each array of characters corresponds to a single print position on the line. As the drum rotates each character that is to print reaches the appropriate position and a hammer then strikes that character.

(B) Plotter

A plotter reproduces drawings using pens that are attached to movable arms. The pens are directed across the surface of a stationary piece of paper. High quality bar graphs, pie charts created with a plotter give a very good quality output. Plotters are ideal for making bar charts, maps, architectural drawings, technical drawings, and three-dimensional illustrations.

1.5.5.2 Soft Copy Devices

An output on soft copy devices enables viewing of the work which allows corrections or rearrangement of material to suit specific needs. The most commonly used soft copy device is a monitor.

(A) Visual Display Terminal or Terminal

A monitor is a television like device used to display data or information. allowing users to view the results of processing. The combination of monitors with the keyboard is called a terminal. Computer industry has divided VDT offerings into three categories.

(1) Dumb Terminal

The dumb terminal is a stand alone keyboard and display screen that can send or receive data. but cannot process that data. They are simply devices that immediately transmit each keyed data character to the processor.

(2) Smart Terminal

They are designed with a microprocessor chip that contains the circuits needed to perform arithmetic logic and control functions and they also have an internal storage capability. They can move the cursor in all directions to add, delete and change keyed characters. Smart terminals can store and consolidate input data prior to sending them to the processor. But they cannot normally be programmed by users.

(3) Intelligent Terminal

It combines VDT hardware with built-in microprocessors that can be programmed by the user. They can process small jobs without the need to interact with a larger computer. The intelligent Video Display Terminal is designed with circuits that permit it to communicate with other terminals and larger processors.

The difference between an intelligent terminal and a personal computer is very little. The intelligent terminal is designed with circuits that allow it to communicate with other terminals and larger processor, while most personal computers have been designed primarily to operate alone. But plug in hardware and needed software permit Interaction between main frames and personal computer. And other plug in modules are available to convert existing dumb and smart terminals into full fledged microcomputers.

(B) Visual Display Unit (Monitor)

A monitor is a television like device used to display input data or information. allowing users to view the results of processing. Quality of the monitor is often judged in terms of resolution which is a measure of the number of picture elements or pixels. a screen contains. A pixel is the smallest increment of a display screen that can be controlled individually. The higher the number of pixels the clearer and sharper is the image. Screens for monitor are of two types:

(i) Cathode Ray Tube (CRT)

On this type of screen a data image can be produced by moving an electron beam across a phosphor-coated screen. By intensifying the strength of the beam the phosphor-coating glows in certain places, forming the characters. The CRT has a display screen of 25 lines of 80 characters each. it is used in most desk top monitors. Some CRT screens are monochrome (one color) while others produce many colors.

Two types of color monitors are commonly used: composite and red-green-blue. A composite color monitor uses only one electron gun to control the intensity of all three phosphor dots in each pixel. In a red-green-blue (RGB) monitor, three electron guns, one for each dot in a pixel are used to control the intensity of the phosphor dots. A much sharper picture is produced with the RGB monitor because the three electron guns allow finer control over the intensities of the phosphor dots.

Artists may wonder why red, green, and blue are used as the primary colors instead of red, yellow, and blue. Simply stated, the difference between electronic waves and color pigments requires that green be used instead of yellow to get the appropriate color mixes.

(ii) Flat-Panel Display

A CRT screen is reliable but it is bulky and consumes a lot of power and that is why it is cumbersome to use it with portable computers. For small computers, flat panel display is used. It does not have a picture tube. The most common type of flat panel display is the Liquid Crystal Display (LCD), which produces images by aligning molecular crystals. When a voltage is applied, the crystals line up in a way that blocks light from passing through them and his absence of light is seen as characters on the screen. LCDs are better than CRTs as LCDs don't flicker, thereby avoiding eye strain and fatigue caused by CRTs during prolonged sessions at the computer.

The flat-panel display is still in its infancy; better resolution and contrast. as well as other new features are being developed. One advance is the gas-plasma screen. These displays offer flicker-free viewing and have a much higher contrast than LCDs. Gas-plasma displays contain an ionized gas (plasma) between two glass plates. One glass plate contains a set of horizontal wires; the other has a set of vertical wires. The intersection points of each horizontal and vertical wire identifies a pixel. A pixel is turned on when current is sent through the appropriate vertical and horizontal wires. The plasma at the pixel emits light as the current is applied. Characters are formed by lighting the appropriate pixels.

1.5.6 Summary

Input is the process of entering and translating incoming data into machine readable form. During the input process three things to be concerned with are method of data preparation, type of processing to be used and the accuracy of the data being entered.

An input device is a peripheral device through which data is entered and transformed into the machine-readable form. Examples: keyboard, mouse, trackball, and scanner. Keyboards are the most commonly used data entry devices. By pressing down the keys of the keyboard, data are entered into the computer. Mouse is a small hand-held pointing device, which can be used to input commands or information. It contains two or three buttons. A trackball is an upturned mouse, with a movable ball on the top of a stationary base. It is used to control the cursor movements and the actions on a computer screen. Joystick is a device that moves in all directions and controls the movement of a pointer. Using a light pen, one can select the objects on the monitor by directly pointing to the objects. A touch screen is a type of display screen device that is placed on the computer monitor in order to allow the direct selection or activation of the computer when somebody touches the screen. Speech recognition is the technology by which sounds, words or phrases spoken by humans are converted into digital signals. A scanner is a device that scans an image and transforms the image to ASCII codes. The common types of the scanners are: hand-held scanner and flat-bed scanner. An optical character recognition (OCR) is a procedure that involves reading text from paper

and translating the images into a form that the computer could manipulate.
The computer output generated by output devices is of two types: Hard Copy and Soft Copy. The printed form of output is referred as Hard Copy while the form of output, which is shown on a display screen or is in audio or voice form is referred as Soft Copy. Based on the hard copy and soft copy outputs, the output devices are classified into two types: Hard copy output devices and Soft copy output devices. Hard copy output devices are very slow in operation as compared to the soft copy output devices. Printers and plotters are the most commonly used hard copy output devices.
A printer is a device that prints information and data from the computer on to paper. The two major categories of printer technologies are: Impact printers and Non-impact printers. Plotters are special-purpose drawing devices, which reproduce graphic images on paper using a pen whose movements are controlled by the computer.
Cathode Ray Tubes (CRTs) are most common types of monitors for the office and the home. Liquid Crystal Display (LCD) is a type of projection technology used to create digitized images.

1.5.7 Keywords :

Hard Copy : A printed copy, especially of the output of a computer or word processor.

Soft Copy : Information that is displayed on a screen, given by voice, or stored in a form that cannot be read directly by a person, as on magnetic tape, disk, or microfilm.

Input device : A device, such as a keyboard, used to enter information into a computer.

Output device : A device, such as a printer, video display, or speaker that presents data from a computer to a user.

1.5.8 Short Answer Type Questions :

Q:1 What is the purpose of an output device?

Q:2 Output can be divided into two categories. What are they?

Q:3 List three types of optical recognition.

Q:4 Distinguish between :

- Impact printer and Non-Impact printers.
- Batch and transaction processing
- Serial and parallel interfaces
- Dumb terminal and a smart terminal
- Intelligent terminal and a personal Computer.

1.5.9 Long Answer Type Questions :

Q:1 What is an Input device? What is the purpose of an Input device? List the various types of input devices.

Q:2 What is an output device? What do you mean by hard copy and soft copy? List the various types of Output devices.

1.5.10 Suggested Readings:

1. Computer Fundamentals" By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Introduction to Information Technology by V.Rajaraman, PHI, New Delhi.
3. Fundamentals of Information Technology, by Alexis Leon and Matthews Leon Vikas Publishing House
4. Fundamentals of Information Technology by Deepak Bharihoke, Pentagon Press.
5. FYSOS : Input and output Devices by Benjamin David Lunt
6. Input/output devices and Interaction Techniques by K. Kinckley

Lesson No. 1.6 **Author : Arun Sharma**

NUMBER SYSTEMS

Chapter Outline:

1.6.0 Objectives

- *What is Number System?*
- *Categories of number system*
- Types of Number Systems
- *Method of conversion between number bases*

Since the early days of human civilisation, people have been using their fingers, sticks and other things for counting. It all started perhaps, with the need to figure out the assets a person had. As daily activities became more complex, numbers became more important in trade, time, distance and in all other spheres of human life. It became apparent that we needed more than our fingers and toes to keep track of the number in our daily routine. Furthermore, ever since people discovered that it was necessary to count objects, they have been looking for easier ways to count them. Signs and symbols gained popularity for number representation. The early forms were straight lines or groups of lines.

In 3400 BC, the ancient Egyptians started using special symbols for writing the numbers. This was a major advancement because it reduced the number of symbols required. However, it was difficult to represent large or small numbers by using such a graphical approach.

1.6.1 Number Systems

A number system defines a set of values used to represent *quantity.* We talk about the number of people attending a class, the number of modules taken by each student and use numbers to represent grades achieved by students in tests. Quantifying values and items in relation to each other is helpful for us to make sense of our environment.

Number systems have been around for thousands of years. We can see the remnants of several systems in our present day civilisation. The common system is the existing system based on number ten. Although, today the most common number system in use is the Arabic system, the number systems can be categorised in two broad categories:

- Non-Positional Number Systems
- Positional Number Systems

1.6.1.1 Non-Positional Number Systems

In ancient times, people used to count on their fingers. When the fingers became insufficient for counting, stones, pebbles or sticks were used to indicate the values. This method of counting is called the *non-positional number system.* It was very difficult to perform arithmetic with such a number system, as it had no symbol for zero. The most common non-positional number system is the Roman Number System. In this number system, only a few characters are used to represent the numbers. The characters, which are used in this number system are I, V, X, L (for fifty), C (for hundred), etc. Moreover, since it is very difficult to perform the addition or any other arithmetic operations in this system, as a result no logical or positional techniques are used in this system.

1.6.1.2 Positional Number Systems

In positional number systems, the value of each digit in a number is defined not only by the symbol but also by the symbol's position. Positional number systems have a base or radix. The first positional number system was invented by the Babylonians. They used a base 60 system. The positional number system, which is being used nowadays is called the *decimal number system.* This system is base 10 system, that is, it contains 10 digits (0, 1, 2, 3... 8, 9). Apart from the decimal number system, there are some other positional number systems such as *binary number system, octal number system* and *hexadecimal number system* each having a radix of 2, 8 and 16, respectively. However, the principles which are applied to the decimal number system are also applicable for the other positional number systems.

1.6.1.3 Base (or Radix) of System

In the number system, the base or radix tells the number of symbols used in the system. In the earlier days, different civilisations were using different radixes. The Egyptians used the radix 2, the Babylonians used the radix 60 and Mayans used 18 and 20. In contrast, modern computers use the radix 2 because they recognise only two symbols, which are represented in digital circuits as 0s and 1s.

Radix of the system is always expressed in decimal numbers. The base or radix of the decimal system is 10. This implies that there are 10 symbols: 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. Similarly, the system using three symbols 0, 1, 2 will be of base 3; four symbols will be of base 4 and so forth.

The base of a number system is indicated by a subscript (decimal number) and this will be followed by the value of the number.

For example:

$(7592)_{10}$ is of base 10 number system.
$(214)_8$ is of base 8 number system.
$(123)_{16}$ is of base 16 number system.

1.6.2 Computer and Numbers

We apply numbers everyday, and knowing how numbers work enables us to know how a computer manipulates and stores numbers.

For a computer, everything is a number whether it may be numbers, alphabets, punctuation marks, its own instructions, etc. Let us understand with the help of an example. Consider the word 'words' which always appears on the computer screen (for us) as a series of alphabetic characters. However, for the computer, it is a combination of numbers. To the computer it appears as:

0111 0111 0110 1111 0111 0010 0110 0100 0111 0011
(w o r d s)

Eventually, the number systems that are generally used by the computers are:

- Decimal System
- Binary System
- Octal System
- Hexadecimal System

Table 1.6.1 Types of Number Systems

Number System	Radix Value	Set of Digits.	Example
Decimal	r = 10	(0, 1,2,3,4,5,6,7,8,9)	$(25)_{10}$
Binary	r = 2	(0, 1)	$(11001)_2$
Octal	r = 8	(0, 1,2,3,4,5,6, 7)	$(31)_8$

Hexadecimal	r = 16	(0, 1,2,3,4,5,6,7,8,9, A, B, C, D, E, F)	$(19)_{16}$

The important thing about the number systems is that each system is just a different method for representing the quantities. Moreover, the quantities do not change but the symbols used to represent those quantities are changed in each number system.

1.6.2.1 Decimal Number System

The primary number system used is a base ten number system or *decimal number system.* The decimal system is the system which we use everyday while counting. The name is derived from the Latin word *Decem,* which means ten. This number system includes the ten digits from 0 through 9. These digits are recognized as the symbols of the decimal system. Each digit in a base ten number represents units ten times the units of the digit to its right.

Starting at the decimal point and moving to the left, each position is represented by the base (radix) value (10 for decimal) raised to a power. The power starts at 0 for the position just to the left of the decimal point. The power is incremented for each position that continues to the left.

10^3 10^2 10^1 10^0

where,

10^3	=	10	x	10	x	10	=	1000
10^2	=	10	x	10			=	100
10^1	=	10					=	10
10^0							=	1

Moving to the right of the decimal point is just like moving to the left except that we will need to place a minus sign in front of each power.

$.10^{-1}$ 10^{-2} 10^{-3}

Consider the number 9735. In the first column of the following table, we write 9735 in the expanded notation. In the second column we write the same sum but express 9000 as 9 x 1000, 700 as 7 x 100, 30 as 3 x 10, and 5 as 5 x 1. In the third column, again we write the same numbers, but express 1000, 100, 10 and 1 as powers of 10.

9735	9000 + 700 + 30 + 5	Is Equivalent to	9 x 1000 7 x 100 3 x 10 5 x 1	Is Equivalent to	9×10^3 7×10^2 3×10^1 5×10^0

So, $9735 = (9 \times 10^3) + (7 \times 10^2) + (3 \times 10^1) + (5 \times 10^0)$.

1.6.2.2 Binary Number System

In the early stages of computer development, the problem of storing data was the most difficult problem. Consequently, before organising a device that could hold data with the available technology, it was necessary to reduce the data to its most fundamental state.

Computers do not use the ten digits of the decimal system for counting and arithmetic. Their CPU and memory are made up of millions of tiny switches that can be either in the ON or OFF states. **Two digits, 0 and 1 are used to refer for the two states of ON and OFF, respectively.**

Suppose we have two tiny switches, they can represent the following four patterns:

Switch1	Switch2	Pattern
OFF	OFF	1
OFF	ON	2
ON	OFF	3
ON	ON	4

The pattern shown in the above table is not drawn randomly. They have some logical order. According to the above table, if we replace each 'ON' switch with 'l' and each 'OFF' with '0' then we get a number system called *binary number system.* With this kind of system, it is very easier for the hardware to represent the data since it has to deal with only two numbers (0 and 1). Accordingly, most of the modem computer systems are operating by using this system.

The place value of the binary number system is based on the number two. In this system, we have the one's place, the two's place, the four's place, the eight's place, the sixteen's place and so on. Each place in the number represents two times (2x's) the place to its right.

The weight of each binary bit of a binary number depends on its relative position within the number. In other words, the weight of a digit in any positional number system depends on its relative position within the number and the base of the number system.

In the binary number system with base 2, the weight of nth bit of the number from Right Hand Side (RHS) is n^{th} bit x $^{2n-1}$

The weighted values for each position is determined as follows:

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	2^{-1}	2^{-2}
125	64	32	16	8	4	2	1	.5	.25

Table 1.6.2 Decimal Binary Comparison

Decimal	Binary
0	0
1	01
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010

The problem with binary system is that it takes a large number of digits to represent numerical values. Binary is not efficient in representing fractional values. It cannot represent these values accurately and needs many digits to even come near to approximation.

1.6.2.3 Octal Number System

The octal number system with its 8 digits, '0', '1', '2', '3', '4', '5', '6' and '7' is a base-eight system. The table below shows the weighting for the octal number system up to 3 decimal places before and 2 decimal places after the *octal* point (**.**).

Octal Weights	8^3	8^2	8^1	8^0	·	8^{-1}	8^{-2}
Values	512	64	8	1	·	0.125	0.015625

The octal or base 8 number system is commonly used with computers. With reference to the above table, we find that one octal digit is the equivalent value of three binary digits. The following example of the conversion of octal $(225)_8$ to binary and vice versa will further illustrate this conversion.

Binary and Octal Comparison

Octal to Binary

2	2	5
010	010	101

Binary to Octal

010	010	101
2	2	5

This system is a positional notation number system. Just as the decimal system that uses powers of 10 and the binary system uses powers of 2, the octal system uses powers of 8 to determine the digit of a number's position.

Table 1.6.3 Octal Number System

Binary Number	**Decimal Number**	**Octal Number**
000	0	0 (0×8^0)
001	1	1 (1×8^0)
010	2	2 (2×8^0)
011	3	3 (3×8^0)
100	4	4 (4×8^0)
101	5	5 (5×8^0)
110	6	6 (6×8^0)
111	7	7 (7×8^0)
1000	8	10 ($1 X 8^1 + 0 \times 8^0$)
1001	9	11 ($1 X 8^1 + 1 \times 8^0$)
1010	10	12 ($1 \times 8^1 + 2 \times 8^0$)

1.6.2.4 Hexadecimal Number System

Hexadecimal is another number system that works exactly like the decimal and binary number systems except that the base is 16. Just as the decimal number represents a power of 10, each hexadecimal number represents a power of 16. To represent the decimal numbers, this system uses 0 to 9 numbers and A to F characters to represent 10 to 15, respectively.

The largest hexadecimal digit F is equivalent to binary 1111. Thus, in other words, a single hexadecimal can represent a combination of 4 bits. Since, a byte consists of 8 bits, so a byte can be represented by exactly two hexadecimal digits. For example, consider a binary number 01101111.

Now, split the above number into two parts as shown below:

0110 1111

We see that,

0110 (binary) = 6 (hex)

1111 (binary) = F (hex)

Thus, this number is $6F_{hex}$ or $6F_{16}$

Table 1.6.4 Decimal-Hexadecimal-Binary Comparisons

Decimal	Hex	Binary
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111
8	8	1000
9	9	1001
10	A	1010
11	B	1011
12	C	1100
13	D	1101
14	E	1110
15	F	1111

1.6.3 Conversion Between Number Bases

We have discussed earlier that internally computer uses binary numbers for data representation whereas externally it uses decimal numbers. However, any number in one number system can be represented in any other number system. Let us see the various methods which can be used to convert numbers from one base to another.

1.6.3.1 Conversion of Decimal to Binary

The method, which is used for the conversion of decimal into binary, is often called as the remainder method. This method involves the following steps:

1. Begin by dividing the decimal number by 2 (the base of binary number system).
2. Note the remainder separately as the rightmost digit of the binary equivalent
3. Continually repeat the process of dividing by 2 until the quotient is zero and keep writing the remainders after each step of division (these remainders will either be 1 or 0).
4. Finally, when no more division can occur, write down the remainders in reverse order (last remainder written first).

Example 1: Determine the binary equivalent of $(36)_{10}$.

2	36	**Remainder**	
2	18	0	**Least Significant Bit (LSB)** ↑
2	9	0	
2	4	1	
2	2	0	
2	1	0	
	0	1	**Most Significant Bit (MSB)**

Taking remainders in reverse order, we have 100100. Thus, the binary equivalent of $(36)_{10}$ is $(100100)_2$.

Example 2: Determine the binary equivalent of $(671)_{10}$.

2	671	**Remainder**	
2	335	1	**Least Significant Bit (LSB)** ↑
2	167	1	
2	83	1	
2	41	1	
2	20	1	
2	10	0	
2	5	0	
2	2	1	
2	1	0	
	0	1	**Most Significant Bit (MSB)**

Taking remainders in reverse order, we have 1010011111. Thus, the binary equivalent of $(671)_{10}$ is $(1010011111)_2$.

In every number system, we will number each bit as follows:

- The first bit from the right in a binary number system is bit position zero.
- Each bit to the left is given as the next successive bit number.

Here, bit at position zero is usually referred to as the LSB (least significant bit). The first bit from the left is typically called the MSB (most significant bit). In the above examples 1 and 2, the LSB and the MSB are indicated. The intermediate bits are referred by their respective bit numbers.

1.6.3.2 Conversion of Binary to Decimal

In the binary to decimal conversion, each digit of the binary number is multiplied by its weighted position, and each of the weighted values is added together to get the decimal number. Consider the following examples:

Example 1: Determine the decimal equivalent of $(11010)_2$.

Binary Number	1	1	0	1	0
Weight of Each Bit	2^4	2^3	2^2	2^1	2^0
Weighted Value	2^4 x 1	2^3 x 1	2^2 x 0	2^1 x 1	2^0 x 0
Solved Multiplication	16	8	0	2	0

Sum of weight of all bits = 16 + 8 + 0 + 2 + 0 = 26

Thus, the decimal equivalent of $(11010)_2$ is $(26)_{10}$.

Example 2: Determine the decimal equivalent of $(10110011)_2$.

Binary Number	1	0	1	1	0	0	1	1
Weight of Each Bit	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
Weighted Value	$2^7 \times 1$	$2^6 \times 0$	$2^5 \times 1$	$2^4 \times 1$	$2^3 \times 0$	$2^2 \times 0$	$2^1 \times 1$	$2^0 \times 1$
Solved Multiplication	128	0	32	16	0	0	2	1

Sum of weight of all bits = 128 + 0 + 32 + 16 + 0 + 0 + 2 + 1
= 179

Thus, the decimal equivalent of $(10110011)_2$ is $(179)_{10}$.

1.6.3.3 Conversion of Decimal to Octal

To convert a decimal number into its octal equivalent, the same procedure is adopted as in decimal to binary conversion but the decimal number is divided by 8 (the base of the octal number system).

Example 1: Determine the octal equivalent of $(359)_{10}$.

8	359	**Remainder**	
8	44	7	**Least Significant Bit (LSB)**
8	5	4	↑
8	0	5	**Most Significant Bit (MSB)**

Taking remainders in reverse order, we get 547. Thus, the octal equivalent of $(359)_{10}$ is $(547)_8$.

Note: *Here, in the octal base conversion, the concept of LSB and MSB is similar to that of the binary conversions.*

Example 2: Determine the octal equivalent of $(432267)_{10}$.

8	432267	**Remainder**	
8	54033	3	**Least Significant Bit (LSB)**
8	6754	1	
8	844	2	
8	105	4	
8	13	1	
8	1	5	
8	0	1	**Most Significant Bit (MSB)**

Taking remainders in reverse order, we get 1514213. Thus, the octal equivalent of $(432267)_{10}$ is $(1514213)_8$.

1.6.3.4 Conversion of Octal to Decimal

In the octal to decimal conversion, each digit of the octal number is multiplied by its weighted position and each of the weighted values is added together to get the decimal number.

Example 1: Determine the decimal equivalent of $(456)_8$.

Octal Number	4	5	6
Weight of Each Bit	8^2	8^1	8^0
Weighted Value	$8^2 \times 4$	$8^1 \times 5$	$8^0 \times 6$
Solved Multiplication	256	40	6

Sum of weight of all bits = 256 + 40 + 6
= 302

Thus, the decimal equivalent of $(456)_8$ is $(302)_{10}$.

Example 2: Determine the decimal equivalent of $(127662)_8$.

Octal Number	1	2	7	6	6	2
Weight of Each Bit	8^5	8^4	8^3	8^2	8^1	8^0
Weighted Value	8^5 x 1	8^4 x 2	8^3 x 7	8^2 x 6	8^1 x 6	8^0 x 2
Solved Multiplication	32768	8192	3584	384	48	2

Sum of weight of all bits = 32768 + 8192 + 3584 + 384 + 48 + 2
= 44978

Thus, the decimal equivalent of $(127662)_8$ is $(44978)_{10}$.

1.6.3.5 Conversion of Binary to Octal

The conversion of an integer binary number to octal is accomplished by the following steps:

1. Break the binary number into 3-bit sections starting from the LSB to the MSB.
2. Convert the 3-bit binary number to its octal equivalent.

For whole numbers, it may be necessary to add a zero as the MSB in order to complete a grouping of three bits.

Note: *By adding a zero, the MSB will not change the value of the binary number.*

Example 1: Determine the octal equivalent of $(010111)_2$.

Binary Number	010 (MSB)	111 (LSB)
Octal Number	2	7

The octal equivalent of $(010111)_2$ is $(27)_8$.

Example 2: Determine the octal equivalent of $(1010111110110010)_2$.

Binary Number	001 (MSB)	010	111	110	110	010(LSB)
Octal Number	1	2	7	6	6	2

The octal equivalent of $(1010111110110010)_2$ is $(127662)_8$.

Note: In the above example, we have added two 0's in the MSB so as to complete the required grouping of 3-bits.

1.6.3.6 Conversion of Octal to Binary

Since it is easier to read large numbers in octal form than in the binary form, the primary application of octal numbers is representing binary numbers. Besides, each octal digit can be represented by a three bit binary number; it is very easy to convert from octal to binary. The following steps are involved:

1. Convert the decimal number to its 3-bit binary equivalent.
2. Combine the 3-bit sections by removing the spaces to get the binary number.

Example 1: Determine the binary equivalent of $(231)_8$.

Octal Number	2	3	1
Binary Coded Value	010	011	001

Combining the 3-bits of the binary coded values, we have 010011001.
Thus, the binary equivalent of $(231)_8$ is $(010011001)_2$.

Example 2: Determine the binary equivalent of $(453267)_8$.

Octal Number	4	5	3	2	6	7
Binary Coded Value	100	101	011	010	110	111

Combining the 3-bits of the binary coded values, we have 100101011010110111.
Thus, the binary equivalent of $(453267)_8$ is $(100101011010110111)_2$.

1.6.3.7 Conversion of Decimal to Hexadecimal

To convert a decimal number into its hexadecimal equivalent, the same procedure is adopted as decimal to binary conversion but the decimal number is divided by 16 (the base of the hexadecimal number system).

Example 1: Determine the hexadecimal equivalent of $(5112)_{10}$.

		Remainder	
16	5112		
16	319	8 = 8	Least Significant Bit (LSB) ↑
16	19	15 = F	
16	1	3 = 3	
16	0	1 = 1	Most Significant Bit (MSB)

Taking remainders in the reverse order, we have 13F8. Thus, the hexadecimal equivalent of $(5112)_{10}$ is $(13F8)_{16}$.

Note: *Here in the hexadecimal conversion, the concept of LSB and MSB is similar to that of the binary and octal conversions.*

Example 2: Determine the hexadecimal equivalent of $(584666)_{10}$.

16	584666	**Remainder**	
16	36541	10 = A	**Least Significant Bit (LSB)**
16	2283	13 = D	↑
16	142	11 = B	
16	8	14 = E	
16	0	8 = 8	**Most Significant Bit (MSB)**

Thus, the hexadecimal equivalent of $(584666)_{10}$ is $(8EBDA)_{16}$.

1.6.3.8 Conversion of Hexadecimal to Decimal

In the hexadecimal to decimal conversion, each digit of the hexadecimal number is multiplied by its weighted position and each of the weighted values is added together to get the decimal number.

Example 1: Determine the decimal equivalent of $(B14)_{16}$.

Hexadecimal Number	B = 11	1	4
Weight of Each Bit	16^2	16^1	16^0
Weighted Value	256 x 11	16 x 1	1 x 4
Solved Multiplication	2816	16	4

Sum of weight of all bits = 2816 + 16 + 4
= 2836

Thus, the decimal equivalent of $(B14)_{16}$ is $(2836)_{10}$.

Example 2: Determine the decimal equivalent of $(8AFE2B)_{16}$.

Hexadecimal Number	8	A=10	F=15	E=14	2	B =11
Weight of Each Bit	16^5	16^4	16^3	16^2	16^1	16^0
Weighted Value	1048576 x 8	65536xl 0	4096 x 15	256 x 14	16x 2	1 x 11
Solved Multiplication	8388608	655360	61440	3584	32	11

Sum of weight of all bits = 8388608 + 655360 + 61440 + 3584 + 32 + 11
= 9109035

Thus, the decimal equivalent of $(8AFE2B)_{16}$ is $(9109035)_{10}$.

1.6.3.9 Conversion of Binary to Hexadecimal

The conversion of an integer binary number to hexadecimal is accomplished by the following steps:

1. Break the binary number into 4-bit sections starting from the LSB to the MSB.
2. Convert the 4-bit binary number to its hexadecimal equivalent.

For whole numbers, it may be necessary to add a zero as the MSB in order to complete a grouping of four bits.

Note: *By adding a zero, the MSB will not change the value of the binary number.*

Example 1: Determine the hexadecimal equivalent of $(11001011)_2$.

Binary Number	1100	1011
Decimal Number	12	11
Hexadecimal Number	C (MSB)	B (LSB)

The hexadecimal equivalent of $(11001011)_2$ is $(CB)_{16}$.

Example 2: Determine the hexadecimal equivalent of (1010111100110110001)2.

Binary Number	0010	1011	1100	1101	1001
Decimal Number	2	11	12	13	9
Hexadecimal Number	2 (MSB)	B	C	D	9 (LSB)

The hexadecimal equivalent of $(1010111100110110001)_2$ is $(2BCD9)_{16}$.

Note: *In the above example, we have added two 0s in the MSB so as to complete the required grouping of four bits.*

1.6.3.10 Conversion of Hexadecimal to Binary

Converting a hexadecimal (base 16) number to a binary (base 2) number is a precise process. Since a single digit in a hexadecimal number corresponds directly to a 4-digit binary number, so in order to convert the hexadecimal number into its binary equivalent, the following steps are involved:

1. Convert each hexadecimal digit to its 4-bit binary equivalent.
2. Combine the 4-bit sections by removing the spaces to get the binary number.

Example 1: Determine the binary equivalent of $(5AF)_{16}$.

Hexadecimal Number	5	A	F
Binary Coded Value	0101	1010	1111

Combining the 4-bits of the binary coded values, we have 010110101111.
Thus, the binary equivalent of $(5AF)_{16}$ is $(010110101111)_2$.

Example 2: Determine the binary equivalent of $(86DB45C)_{16}$.

Hexadecimal Number	8	6	D	B	4	5	C
Binary Coded Value	1000	0110	1101	1011	0100	0101	1100

Combining the 4-bits of the binary-coded values, we have 1000011011011011010001011100.

Thus, the binary equivalent of $(86DB45C)_{16}$ is $(1000011011011011010001011100)_2$.

1.6.3.11 Conversion of Octal to Hexadecimal

Octal and hexadecimal have certain relations with binary. The first digit in octal corresponds to the first three digits in its binary equivalent and so on. The same is true for hexadecimal and this time each digit represents four binary digits. This makes the conversion of octal to hexadecimal and vice versa quite easy. This conversion involves the following steps:

1. Convert each octal digit to 3-bit binary form.
2. Combine all the 3-bits binary numbers.
3. Divide the binary numbers into the 4-bit binary form by starting the first number from the right bit to the first number from the left bit.
4. Finally, convert these 4-bit blocks into their respective hexadecimal symbols.

Example 1: Determine the hexadecimal equivalent of $(2327)_8$.

Octal Number	2	3	2	7
Binary Coded Value	010	011	010	111

Combining the 3-bit binary blocks, we have 010011010111.

Dividing the group of binary numbers into the 4-bit binary blocks and by converting these blocks into their respective hexadecimal symbols, we have:

0100 1101 0111
4 D 7

Thus, the hexadecimal equivalent of $(2327)_8$ is $(4D7)_{16}$.

Example 2: Determine the hexadecimal equivalent of $(5473261)_8$.

Octal Number	5	4	7	3	2	6	1
Binary Coded Value	101	100	111	011	010	110	001

Combining the 3-bit binary blocks, we have 101100111011010110001.

Dividing the group of binary numbers into the 4-bit binary blocks and by converting these blocks into their respective hexadecimal symbols, we have:

0001 0110 0111 0110 1011 0001
1 6 7 6 B 1

Thus, the hexadecimal equivalent of $(5473261)_8$ is $(1676B1)_{16}$.

***Note**: We have added three 0's in the MSB in order to get the desired grouping of bits.*

1.6.3.12 Conversion of Hexadecimal to Octal

This conversion follows the same steps of octal to hexadecimal conversion except that each hexadecimal digit is converted into 4-bit binary form and then after grouping of all the 4-bit binary blocks, it is converted into the 3-bit binary form. Finally, these 3-bit binary forms are converted into octal symbols.

Example 1: Determine the octal equivalent of $(2B6)_{16}$.

Hexadecimal Number	2	B	6
Binary Coded Value	0010	1011	0110

Combining all the 4-bit binary blocks, we have 001010110110.

Dividing the group of binary numbers into the 3-bit binary blocks and by converting these blocks into their respective octal symbols, we have:

001 010 110 110
1 2 6 6

Thus, the octal equivalent of $(2B6)_{16}$ is $(1266)_8$.

Example 2: Determine the octal equivalent of $(5DE247)_{16}$.

Hexadecimal Number	5	D	E	2	4	7
Binary Coded Value	0101	1101	1110	0010	0100	0111

Combining all the 4-bit binary blocks, we have 010111011110001001000111.

Dividing the group of binary numbers into the 3-bit binary blocks and by converting these blocks into their respective octal symbols, we have:

010	111	011	110	001	001	000	111
2	7	3	6	1	1	0	7

Thus, the octal equivalent of $(5DE247)_{16}$ is $(27361107)_8$.

1.6.4 Summary

Number systems have been around for thousands of years. It defines a set of values used 10 represent the quantity and other special characters. Number systems basically are of two types: non-positional and positional number systems.

In a non-positional number system, special symbols or characters are used to indicate the value. It is very difficult to perform arithmetic with such a number system, as it has no symbol for zero.In a positional number system, the value of each digit in a number is defined by the symbols but also by the symbol's position. These symbols are called as *digits.*

The positional number system, which is being used nowadays is called as the *decimal number system.* Apart from this number system, there are some other positional number systems such as binary number system, octal number system, and hexadecimal number system.

The base or radix of the number system tells the number of symbols or digits used in the system The base of the decimal number system is 10, of binary number system is 2, of octal number system is 8 and of hexadecimal number system is 16. The primary number system used in our day-to-day life is the decimal number system. This number system includes ten digits (0, 1, 2, 3, 4, 5, 6, 7, 8 and 9).

The modem computer systems are operating by using the binary number system. This system is based on the number two and deals with only two numbers: 0 and 1.

In the hexadecimal number system, each hexadecimal number represents a power of 16. To represent the decimal numbers, this system uses 0 to 9 numbers and A to F characters to represent 10 to 15 numbers, respectively.

Every number system can be converted into another number system such as decimal to binary and vice versa, decimal to octal and vice versa, decimal to hexadecimal and vice versa, binary to octal and vice versa and so on. However, the method of each conversion is different from one another.

1.6.5 Self Check Exercise

Q.1 What is number system? Write the difference between a positional and a non-positional number system.

Q.2 Give the reasons as to why the binary number system is utilized for modern electronic digital computers.

Q.3 What is a radix or base of the system? With the help of this system, brief the various types of number systems.

Q.4 Explain how a decimal number is converted into binary, octal and hexadecimal number and vice versa. Give an example of each conversion.

Q.5 With an appropriate example, explain the conversion of:

a. Binary to octal and vice versa

b. Binary to hexadecimal and vice versa

c. Octal to hexadecimal and vice versa

1.6.6 Suggested Readings:

1. Computer Fundamentals By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Fundamentals of Information Technology By Shiv Kumar Anand and Harmohan Sharma (Kalyani Publishers)
3. Fundamentals of Information Technology by V.Rajaraman (PHI, New Delhi).
4. Digital Design by M. Morris Mano (Pearson Education)
5. Computer Fundamentals, Architecture & Organisation by B.Ram, New Age International.
6. The Number System by Hugh Thurston
7. Multiple Base Number System : Theory and Applications by Vissil Dimitrov, C & C Press

Lesson No. 1.7 **Author : Arun Sharma**

COMPUTER CODES

Chapter Outline:

1.7.0 Objectives

1. *Represent decimal numbers using the BCD*
2. *Understand the difference between BCD and straight binary*
3. *Represent decimal numbers using the excess 3 code*
4. *Understand the purpose of ASCII code and EBCDIC code*
5. *Understanding Gray code*

In today's technology, the binary number system is used by the computer system to represent the data in the computer understandable format. Numeric data (consists of only numbers 0, 1,2….. 9) is not the only form of data, which is handled by the computer. Alphanumeric data (it is a string of symbols of the letters A, B, C……. Z or the digits 0,1,2…….9) and some special characters such as =, -, +, *, /, (,), etc. are also required to be processed by the computer.

There are lot of ways to represent numeric, alphabetic and special characters in computer's internal storage area. In computers, the code is made up of fixed size groups of binary positions. Each binary position in a group is assigned a specific value; for example 8, 4, 2, or 1. In this way, every character can be represented by a combination of bits that is different from any other combination. Moreover, data can also be arranged in a way that's very simple and easy to decode or transmitted with varying degrees of redundancy for error detection and correction. The following are the most commonly used coding systems:

- Binary Coded Decimal (BCD)
- Excess-3 code
- American Standard Code for Information Interchange (ASCII)
- Extended Binary Coded Decimal Interchange Code (EBCDIC)
- Gray code

1.7.1 BCD

Binary Coded Decimal (BCD) is a method of using binary digits to represent the decimal digits 0 to 9. A decimal digit is represented by four binary digits. The BCD coding is the binary equivalent of the decimal digit. BCD system was developed by the IBM (International Business Machines) corporation. With BCD, each digit of a number is converted into its binary equivalent rather than converting the entire decimal number to its binary form. Similarly, letters and special characters can be coded in the binary form.

Let us determine the BCD value for the decimal number 5319. Since there are four digits in our decimal number, there are four bytes in our BCD number. They are:

Thousands-Hundreds	Tens-Units
53	19
0 1 0 1 0 0 1 1	0 0 0 1 1 0 0 1

Binary code decimal digits (0-9) are represented using 4-bits. The valid combinations of bits and their respective values are shown in Table 7.1

Decimal Code	BCD Digit
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

Table 7.1 Binary Coded Decimal

One of the advantages of the BCD system is that there is no limit to the size of a number. For adding another digit, we just have to add a new 4-bit sequence. In contrast, numbers represented in binary format are generally limited to the largest number, which can be represented by 8, 16, 32 or 64 bits. Moreover, this is a fast way to convert numbers from decimal to binary. However, this coding is not sufficient for business purposes as it can represent only 16, that is, 2^4 symbols. The later version of BCD used a 6-bit code. These BCD codes defined six-bit words, which allowed representing a maximum of 64, that is, 2^6 symbols. Computers using BCD codes could work only with upper case letters and 0 to 9 numbers and few characters. However, the modern computers do not use BCD numbers as they have to process names and other non-numeric data.

1.7.2 EXCESS-3 CODE

The Excess-3 is a digital code that is formed by adding 3 to each decimal digit and then converting the result to 4-bit binary. Since no definite weights can be assigned to the four digit positions, Excess-3 is an unweighted code.

For instance, to form the Excess-3 representation of 4, first 3 is added to 4 yielding 7, and equivalent binary is 0111.

The Excess-3 code for the decimal 7 is

$$\begin{array}{r} 7 \\ +3 \\ \hline 10 \\ \hline \end{array} \longrightarrow 1010$$

The Excess-3 code_for each decimal digit is found by the same procedure, and the entire code is shown in the following table 7.2

Decimal	BCD	Excess-3
0	0000	0011
1	0001	0100
2	0010	0101
3	0011	0110
4	0100	0111
5	0101	1000
6	0110	1001
7	0111	1010
8	1000	1011
9	1001	1100

Table 7.2 Excess-3 code

Notice that ten of a possible 16 code combinations are used in the Excess-3 code. The six invalid combinations are 0000, 000 1, 00 1 0, 110 1, 1110 and 1111.

Convert 136 to Excess-3 code.

First add 3 to each digit in the decimal number and then convert each resulting sum to its equivalent binary code

1	3	6	
+3	+3	+3	
4	6	9	
↓	↓	↓	
0100	0110	1001	Excess-3 for 136_{10}

1.7.3 ASCII

For the data representation, there is another 8-bit code known as the ***American Standard Code for Information Interchange*** **(ASCII).** This code was originally designed as a 7-bit code. Several computer manufacturers cooperated to develop this code for transmitting and processing data. Later on, IBM developed a new version of ASCII called as ASCII-8. They made use of all eight bits providing 256 symbols. Nevertheless, IBM had not changed the original set of 128 codes so that the original instructions and data could still work with the new character set. ASCII is commonly used in the transmission of data through data communication and is used almost exclusively to represent the data internally in the microcomputers. In ASCII, rather than breaking letters into three groups, upper case letters are assigned codes beginning with hexadecimal value 41 and continuing sequentially through hexadecimal value 5A. Similarly, lower case letters are assigned hexadecimal values of 61 through 7 A.

The decimal values 1 to 9 are assigned the zone code 0011 in ASCII. Table 7.3 of ASCII coding chart shows upper case and lower case alphabetic characters and numeric digits 0 to 9. The standard ASCII code defines 128 character codes (from 0 to 127), of which, the first 32 are control codes (nonprintable) and the other 96 are representable characters.

Table 7.3 ASCII Coding Chart

Value	Character	Value	Character	Value	Character	Value	Character
0		32		64	@	96	'
1	☺	33	!	65	A	97	A
2	☻	34	“	66	B	98	B
3	♥	35	#	67	C	99	C

Value	Character	Value	Character	Value	Character	Value	Character
4	♦	36	$	68	D	100	D
5	♣	37	%	69	E	101	E
6	♠	38	&	70	F	102	F
7	•	39	‘	71	G	103	G
8	◘	40	(	72	H	104	H
9	○	41	)	73	I	105	I
10	◙	42	*	74	J	106	I
11	♂	43	+	75	K	107	K
12	♀	44	,	76	L	108	I
13	♪	45	-	77	M	109	M
14	♫	46	.	78	N	110	N
15	☼	47	/	79	O	111	O
16	►	48	0	80	P	112	P
17	◄	49	1	81	Q	113	Q
18	↕	50	2	82	R	114	R
19	‼	51	3	83	S	115	S
20	¶	52	4	84	T	116	T
21	§	53	5	85	U	117	U
22	▬	54	6	86	V	118	V
23	↨	55	7	R7	W	119	W
24	↑	56	8	88	X	120	X
25	↓	57	9	89	Y	121	Y
26	→	58	:	90	Z	122	Z
27	←	59	;	91	[	123	{
28	∟	60	<	92	\	124	|

29	↔	61	=	93	]	125	}
30	▲	62	>	94	^	126	~
31	▼	63	?	95	_	127	⌂
Value	**Character**	**Value**	**Character**	**Value**	**Character**	**Value**	**Character**
128	Ç	160	á	192	└	224	A
129	ü	161	í	193	┴	225	ß
130	é	162	ó	194	┬	226	Γ
131	â	163	ú	195	├	227	Π
132	ä	164	ñ	196	─	228	Σ
133	à	165	Ñ	197	┼	229	Σ
134	å	166	ª	198	╞	230	µ
135	ç	167	º	199	╟	231	T
136	ê	168	¿	200	╚	232	Φ
137	ë	169	⌐	201	╔	233	Θ
138	è	170	¬	202	╩	234	Ω
139	ï	171	½	203	╦	235	Δ
140	î	172	¼	204	╠	236	∞
141	ì	173	¡	205	═	237	Φ
142	Ä	174	«	206	╬	238	E
143	Å	175	»	207	╧	239	∩
144	É	176	░	208	╨	240	≡
145	æ	177	▒	209	╤	241	±
146	Æ	178	▓	210	╥	242	≥
147	ô	179	│	211	╙	243	≤
148	ö	180	┤	212	╘	244	⌠
149	ò	181	╡	213	╒	245	⌡

150	û	182	╢	214	╓	246	÷
151	ù	183	╖	215	╫	247	≈
152	ÿ	184	╕	216	╪	248	°
153	Ö	185	╣	217	┘	249	·
154	Ü	186	║	218	┌	250	·
155	¢	187	╗	219	█	251	√
Value	**Character**	**Value**	**Character**	**Value**	**Character**	**Value**	**Character**
156	£	188	╝	220	▄	252	^{n}
157	¥	189	╜	221	▌	253	^{2}
158	Pts	190	╛	222	▐	254	■
159	*f*	191	┐	223	▀	255	

Example: Determine the binary coding of 'words' in the ASCII form.

0111 0111 0110 1111 0111 0010 01100110 0111 0011
(w o r d s)

The corresponding ASCII codes for 'words' are:

119 111 114 100 115
(w o r d s)

1.7.4 EBCDIC

EBCDIC or ***Extended Binary Coded Decimal Interchange Code*** uses 8 bits for each character, it is possible to represent 256 different characters or bit combinations. This provides a unique code for each decimal value 0 to 9 (for a total of 10), each upper case and lower case letter (for a total of 52) and for a variety of special characters. Since it is an 8-bit code, each group of the eight bits makes up one alphabetic, numeric or special character and is called a *byte.*

In EBCDIC, the bit pattern 1100 is the zone combination (zone and digit) used for the alphabetic characters A through 1, 1101 is used for the characters J through R, and 1110 is the zone combination used for characters S through Z. The bit pattern 1111 is the zone combination used when representing decimal digits. For example, the code 11000001 is equivalent to the letter A; the code 1111 0001 is equivalent to the decimal digit 1. Other zone combinations are used when forming special characters. The concepts and advantages of ASCII are identical to those of EBCDIC. The important difference between the two coding systems lies in the 8-bit combinations assigned to represent the various alphabetic, numeric and special characters. While using ASCII 8-bit code, we notice that the selection of bit patterns used in the positions differs from those used in EBCDIC. For example, let us look at the characters DP3 in both EBCDIC and ASCII to see how they compare.

Character	**D**	**P**	**3**
EBCDIC	1100 0100	1101 0111	1111 0011
ASCII	0100 0100	0101 0000	0011 0011

Table 7.4 EBCDIC Codes

ALPHABETIC CHARACTERS							
UPPER CASE				**LOWER CASE**			
Prints as	**EBCDIC**			**Prints as**	**EBCDIC**		
	In Binary		**In Hexadecimal**		**In Binary**		**In Hexadecimal**
	Zone	**Digit**			**Zone**	**Digit**	
A	1100	0001	C1	a	1000	0001	81
B	1100	0010	C2	b	1000	0010	82
C	1100	0011	C3	c	1000	0011	83
D	1100	0100	C4	d	1000	0100	84
E	1100	0101	C5	e	1000	0101	85
F	1100	0110	C6	f	1000	0110	86
G	1100	0111	C7	g	1000	0111	87
H	1100	1000	C8	h	1000	1000	88
I	1100	1001	C9	i	1000	1001	89
J	1101	0001	D1	J	1001	0001	91
K	1101	0010	D2	k	1001	0010	92
L	1101	0011	D3	1	1001	0011	93
M	1101	0100	D4	m	1001	0100	94
N	1101	0101	D5	n	1001	0101	95

0	1101	0110	D6	0	1001	0110	96
P	1101	0111	D7	p	1001	0111	97
Q	1101	1000	D8	q	1001	1000	98
R	1101	1001	D9	r	1001	1001	99
S	1110	0010	E2	s	1010	0010	A2
T	1110	0011	D3	t	1010	0011	A3
U	1110	0100	E4	u	1010	0100	A4
V	1110	0101	E5	v	1010	0101	A5
W	1110	0110	E6	w	1010	0110	A6
X	1110	0111	E7	x	1010	0111	A7
Y	1110	100 0	E8	Y	1010	1000	A8
Z	1110	1001	E9	z	1010	1001	A9
NUMERIC CHARACTERS							
0	1111	0000	FO	5	1111	0101	F5
1	1111	0001	F1	6	1111	0110	F6
2	1111	0010	F2	7	1111	0111	F7
3	1111	0011	F3	8	1111	1000	F8
4	1111	0100	F4	9	1111	1001	F9

1.7.5 GRAY CODE

Gray code is an unweighted code, meaning that the bit positions in the code groups do not have any specific weight assigned to them. The Gray code exhibits only one bit in the code group change when going from one step to the next. Gray code is not suited for arithmetic operations.

Table 7.5 shows the Gray code representation for the decimal numbers 0 through 15, together with the straight binary code. If we examine the Gray code groups for each decimal number, it can be seen that in going from one decimal number to the next only one bit of the Gray code changes.

Decimal	Binary	Gray Code
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

Table 7.5 Gray code

For example, in going from 3 to 4, the Gray code changes from 0010 to 0110, with only the second bit from the left changing, while the binary code changes from 0011 to 0100, a change of three bits. This is a principal characteristic of the Gray code.

7.5.1 Binary-to-Gray Conversion

To convert from a binary number to a Gray code number, apply the following steps:

1. The most significant digit (left-most) in the Gray code is the same as the corresponding digit in the binary number.
2. Going from left to right, add each adjacent pair of binary digits to get the next Gray code digit. Disregard carries.

For example, let us assume the binary number 11010 to convert to Gary code

Step-1 The left-most Gray digit is the same as the left most binary digit

1	1	0	1	0	Binary
↓					
1					Gray

Step-II Add the left-most binary digit to the adjacent one and discard carry

[1 +	1]	0	1	0	Binary
	↓				
1	0				Gray

Step-III Add the next adjacent pair

1	[1 +	0]	1	0	Binary
		↓			
1	0	1			Gray

Step-IV Add the next adjacent pair

1	1	[0 +	1]	0	Binary
			↓		
1	0	1	1		Gray

Step- V Add the last adjacent pair

1	1	0	[1 +	0]	Binary
				↓	
1	0	1	1	1	Gray

The conversion is now complete and the Gray code equivalent to binary 11010 is 10111.

7.5.2 Gray-to-Binary Conversion

To convert from Gray code to binary, a similar method is used with little difference. The following steps apply:

1. The left-most significant digit in the binary code is same as the corresponding digit in the Gray code.
2. Add each binary digit generated to the Gray digit in the next adjacent position. Disregard carries.

For example, the conversion of the Gray code number 11011 to binary is as follows:

Step-I The left most Gray digit is the same.

1	1	0	1	1	Gray
↓					
1					Binary

Step-II Add the last binary digit just generated to the Gray digit in the next position. Discard carry.

1	1	0	1	1	Gray
+	↓				
1	0				Binary

Step-III Add the last binary digit generated to the next Gray code.

1	1	0	1	1	Gray
	+	↓			
1	0	0			Binary

Step-IV Add the last binary digit generated to the next Gray digit

1	1	0	1	1	Gray
		+	↓		
1	0	0	1		Binary

Step- V Add the last adjacent pair

1	1	0	1	1	Gray
			+	↓	
1	0	0	1	0	Binary

The conversion is now complete and the Binary equivalent to Gray 11011 is 10010.

1.7.6 Summary

The binary coding schemes are used to represent the internal storage area of the computers. In binary coding, every character is represented by a combination of bits. The most commonly used computer coding systems are BCD, ASCII and EBCDIC.

BCD (Binary Coded Decimal) is a method that represents the decimal digits with the help of binary digits. It is a 6-bit code, which can represent a maximum of 64 different characters.

The Excess-3 is a digital code that is formed by adding 3 to each decimal digit. Excess-3 is an unweighted code.

ASCII is a 8-bit code and is exclusively used to represent the data internally in the microcomputers. It can represent 128 different characters.

EBCDIC or Extended Binary Coded Decimal Interchange Code uses 8 bits for each character and can represent 256 different characters. It provides a unique code for each decimal value 0 through 9 and for a variety of special characters.

Gray code is an unweighted code. Gray code is not suited for arithmetic operations.

1.7.7 *Self Check Exercise*

Q.1 What is the purpose of the binary coding system? Briefly explain the terms: BCD, ASCII and EBCDIC.

Q.2 What do you mean by BCD code?

Q.3 Encode these binary number in BCD

- 45
- 247
- 1029

Q.4 What is Gray code?

1.7.8 *Suggested Readings:*

1. Computer Fundamentals By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Fundamentals of Information Technology By Shiv Kumar Anand and Harmohan Sharma (Kalyani Publishers)
3. Fundamentals of Information Technology by V.Rajaraman (PHI, New Delhi).
4. Digital Design by M. Morris Mano (Pearson Education)
5. Computer Fundamentals, Architecture & Organisation by B.Ram, New Age International.
6. Code : the sudden language of computer hardware and software by Charles Petzold

COMPUTER ARITHMETIC

Chapter Outline:

1.8.0 Objectives

- *Addition and subtraction of binary, octal and hexadecimal number systems*
- *Know about signed and unsigned numbers*
- *1's and 2's complement*

Arithmetic is a branch of mathematics that involves combining numbers by addition, subtraction, multiplication and division. During school days, arithmetic was restricted only to decimal number system. However, in computer, we require arithmetic on other number systems such as binary, octal and hexadecimal. In the following few sections, we will discuss how to perform basic arithmetic on these number systems.

1.8.1 Binary Arithmetic

Everything that is stored in or manipulated by the computer is a number. The computer only understands the numbers 1 and 0. Therefore, every number has to be converted to binary (0s and 1s) digits. The basic arithmetic operations of the binary number system are:

- Addition
- Subtraction

1.8.1.1 Binary Addition

Binary addition is carried out in the same way as the decimal addition is performed. In decimal addition, the unit column is added first, then the tens column, the hundreds, and so on. If the sum is greater than or equal to ten, the least significant digit is written as a partial sum and a carry of 1 is added to the sum of the next column. This process is repeated for each larger significant digit. These steps are also followed in the binary addition. The addition table of the binary arithmetic is very simple because this system has only two digits. As a result, there are only four outcomes or rules of the binary addition. These are listed below:

Table 1.8.1 Addition of Binary Numbers

INPUT		OUTPUT	
X	**Y**	**SUM(S)**	**CARRY(C)**
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

In the table above, the results of the four addition operations between the two binary digits are divided between the 'sum' and the 'carry' part. The first three outcomes are the simple arithmetic operations but in the fourth operation, a *'carry-over'* condition occurs. This has been performed in the same manner as in decimal arithmetic according to which 1 is carried to the next higher column. However, since 1 is the largest possible digit in the binary system, any value which will be greater than 1 requires the digit to be carried over.

For instance, 10 plus 10 in the binary system requires addition of two 1s in the second position. Here, 1 + 1 = 0 plus a carry of 1. Hence, in the binary addition the sum of 10 + 10 is 100.

Example 1: Add the binary numbers 1111 and 1010 and check the answer with the help of decimal addition.

Binary

	1	1		
	+	+		
	1	1	1	1
+	1	0	1	0
1	1	0	0	1

Decimal

	1	5
+	1	0
	2	5

According to the last step of the above binary addition, 1 + 1 + 1 = 10 + 1= 11 = 1 + carry of 1 into higher column.

Example 2: Calculate the sum of 110011, 10010, 1100 and 101 and check the answer with the help of decimal addition.

Binary

		1	1	1	1	1	
		+	+	+	+	+	
		1	1	0	0	1	1
			1	0	0	1	0
				1	1	0	0
+					1	0	1
	1	0	1	0	1	1	0

Decimal

	1	
	+	
	5	1
	1	8
	1	2
+		5
	8	6

Example 3: Add the binary fractional numbers 11.10 and 10.10 and check the result with the help of decimal addition.

Binary

		1	1	
		+	+	
		1	1. 1	0
+		1	0. 1	0
	1	1	0. 0	0

Decimal

	1
	+
	3 . 5
	+ 2 .
5	
	6 . 0

Example 4: Calculate the sum *of* 11010.0100, 1001.01,001.11, and 10.1010 and check the answer with the help *of* decimal addition.

Binary

		1		1	1	1	1			
		+		+	+	+	+			
		1	1	0	1	0 . 0		1	0	0
			1	0	0	1 . 0		1		
				0	0	1 . 1		1		
+					1	0 . 1		0	1	0
	1	0	0	1	1	1 . 1		1	1	0

Decimal

	1	1	1		
	+	+	+		
	2	6	. 2	5	
		9	. 2	5	
		1	. 7	5	
+		2	. 6	2	5
	3	9	. 8	7	5

1.8.1.2 Binary Subtraction

Subtraction is generally simple in comparison to addition since only two numbers are involved and the upper value representation is greater than the lower value representation. In binary subtraction, the problem *of* 'borrow' is similar to that in decimal. If the subtrahend bit is equal to or smaller than the minuend bit, then perform the subtraction, otherwise borrow one from its left most neighbour. If its neighbour is zero, then proceed to the left until a borrow can be performed. For the left most bit, a borrow is made from the outside.

We can construct a subtraction table (as shown in Table 2.6 below) that has two parts - the three cases *of* subtracting without borrow, and the one case *of* the involvement *of* a borrow digit, no matter how far to the left is the next available binary digit. Like the binary addition, binary subtraction also follows four rules for the operation. These rules are discussed below:

Table 1.8.2 Subtraction *of* Binary Numbers

INPUT		OUTPUT	
X	**Y**	**Difference(D)**	**Borrow(B)**
0	0	0	0
0	1	1	1
1	0	1	0
1	1	0	0

The rules, given in the above table, can be explained by the following example *of* subtraction:

	1				
	~~10~~	10		10	
~~1~~	~~0~~	~~0~~	~~1~~	~~0~~	1
- 0	0	1	0	1	1
0	1	1	0	1	0

The following steps are involved:

a. First, for the least significant bit (the right most bit), 1 - 1 is 0.

b. For the next bit, 0 - 1 cannot be computed since the subtrahend is smaller than the minuend. Borrow 1 from the third bit to form the binary number 10 (decimal 2) and do the subtraction. The operation is 10 - 1 = 1 which in the decimal number system is 2 - 1 = 1.

c. For the third bit, since we borrowed 1 for the second bit, we have 0 - 0 that is 0.

d. For the fourth bit again, we cannot perform the subtraction. However, the fifth bit in the minuend is zero, so we must borrow from the sixth bit. This makes the fifth bit 10 (decimal 2). Borrowing 1 from the fifth bit makes it 1 and the fourth bit becomes 10 (decimal 2). Now the subtraction in binary is 10 - 1 = 1 which is the result of the fourth bit.

e. For the fifth bit, we now have 1 - 0 = 1.

f. Since we borrowed 1 from the sixth bit for the fourth bit, so for the sixth bit, the subtraction is 0 - 0 = 0.

Example 1: Find the binary difference of (1101-10110) and check the answer with the help of decimal subtraction.

Binary

			10	
	1	~~1~~	~~0~~	1
-	1	0	1	1
	0	0	1	0

.:

Decimal

	1	3
-	1	1
	0	2

Note: *Here, we borrowed* 1 *from 3rd-column because of the difference 0-1 in the 2nd-column.*

Example 2: Calculate the binary difference of (11100011-10101000) and check the answer with the help of binary subtraction.

Binary

			1					
			0	1				
			~~0~~	~~10~~	10			
	1	~~1~~	~~1~~	~~0~~	~~0~~	0	1	1
-	1	0	1	0	1	0	0	0
	0	0	1	1	1	0	1	1

Decimal

	2	2	7
-	1	6	8
		5	9

1.8.2 Octal Arithmetic

In the computer, everything that is stored in or manipulated is in a form of binary number. Nevertheless, the octal number system is also a common system, which has been used with the computers. The essential arithmetic operations of the octal number system are:

- Addition
- Subtraction

1.8.2.1 Octal Addition

Addition of the octal number is carried out in the same way as the decimal addition is performed. The steps are given below:

1. First, add the two digits of the unit column of the octal number in decimal.
2. This process is repeated for each larger significant digit of the octal number.

3. During the process of addition, if the sum is less than or equal to 7, then it can be directly written as a octal digit.
4. If the sum is greater than 7, then subtract 8 from that particular digit and carry 1 to the next digit position.

Note: *In this addition, we should remember that the largest octal digit is* 7 *instead of 9.*

Example 1: Add the octal numbers 26 and 17.

	1(Carry)	
2	6	
1	7	
4	13	
	8	(modification)
4	5	

Thus, the resultant octal sum is 45.

Example 2: Add the octal numbers 5647 and 1425.

5	6	4	7	
+ 1	4	2	5	
7	10	7	12	
	-8		-8	(modification)
7	2	7	4	

Thus, the resultant octal sum is 7274.

1.8.2.2 Octal Subtraction

In the octal subtraction, the method, which we have adopted, is similar to that of binary subtraction method. The only difference lies in the carry part. During octal subtraction, instead of 1, we will borrow 8 and the rest of the steps are similar to that of binary subtraction.

Example 1: Subtract $(677)_8$ from $(770)_{1.8.}$

8+6=14

6	~~6~~	8	(Borrow)
~~7~~	~~7~~	~~0~~	
- 6	7	7	
0	7	1	

Thus, the difference is $(71)_8$.

Note: *Here, we borrowed 8 from the 2nd column for the difference 0-7 and 8 from the* 3^{rd} *column for the difference* 6-7.

Example 2: Subtract $(2761)_8$ from $(6357)_8$.

5	8+2=10	8+5=13	
~~6~~	~~3~~	~~5~~	7
2	7	6	1
3	3	7	6

Thus, the difference is $(3376)_8$.

1.8.3 Hexadecimal Arithmetic

The hexadecimal number system is extensively used in the memories of the computer system and in the computer instructions. The basic arithmetic operations that are to be performed are listed below:

- Addition
- Subtraction

1.8.3.1 Hexadecimal Addition

The addition operation performed with the hexadecimal numbers is analogous to the decimal addition except with a few differences that are discussed in the following steps:

1. First add the unit column of the hexadecimal digits in decimal.
2. This process is repeated for each larger significant digit of the hexadecimal number.
3. During the process of addition, observe if the sum is 15 or less, then it can be directly expressed as a hexadecimal digit.
4. If the sum is greater than 15, then subtract 16 from that particular digit and carry 1 to the next digit position.

Example 1: Add the hexadecimal numbers $(76)_{16}$ and $(45)_{16}$.

	7	6	
+	4	5	
	11	11	
	-	-	(modification)
	B	B	

The hexadecimal sum is $(BB)_{16}$.

Note: *In the above example, since the decimal sums are less than* 15 *so they are expressed directly in the hexadecimal form.*

Example 2: Add the hexadecimal numbers $(A27E9)_{16}$ and $(6FB43)_{16}$.

	1	1	1	1		
	A	2	7	E	9	
+	6	F	B	4	3	
1	17	18	19	18	12	
	-16	-16	-16	-16	-	(modification)
1	1	2	3	2	12	
1	1	2	3	2	C	(Hex Form)

The hexadecimal sum is $(11232C)_{16}$.

1.8.3.2 Hexadecimal Subtraction

The hexadecimal subtraction is based on the same principles as that of binary subtraction. In this subtraction, 16 will be used as borrow instead of 1. The rest of the steps are similar to the binary subtraction.

Example 1: Subtract $(75)_{16}$ from $(527)_{16}$.

	4 16+2=18		(Borrow)	
	5	2	7	
+		7	5	
	4	11	2	
	4	B	2	(Hex Form)

The hexadecimal difference is $(4B2)_{16}$.

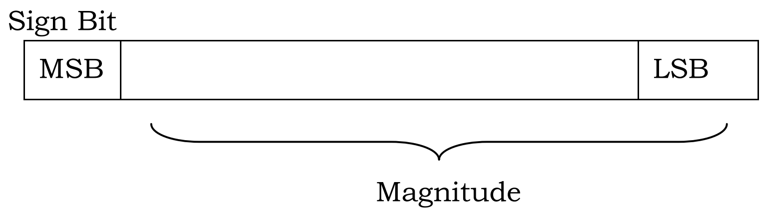
Example 2: Subtract $(1F65)_{16}$ from $(7E2CA)_{16}$.

		13		16+2=18 (Borrow)		
	7	E	2	C	A	
-		1	F	6	5	
	7	12	3	6	5	
	7	C	3	6	5	(Hex Form)

The hexadecimal difference is $(7C365)_{16}$.

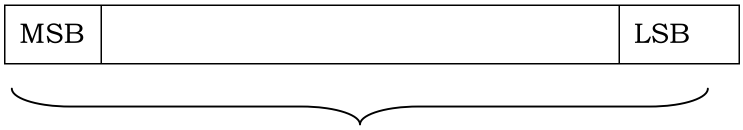
1.8.4 Signed and Unsigned Numbers

An n-bit signed binary number consists of two parts. A part denoting the sign of the number and a part denoting the magnitude of the number, The MSB is always a sign bit which denotes the sign of the number and the convention is that 0 and 1 denote '+' and '-', respectively. The remaining (n-l) bit denotes the magnitude of the number. The 8-bit sign-magnitude system ranges from -127 to +127.



In some applications, all data is either positive or negative, for example, smallest 8-bit number is 00000000 and largest is 11111111. Therefore, it ranges from 0 to 255. Here, positive and negative sign are not included with these numbers. All the bits in the binary number are used to represent magnitude of the corresponding decimal number.

There are certain restrictions in unsigned binary number. With 8-bit unsigned arithmetic, all magnitude must be between 0 and 255. Therefore, each number being added or subtracted must be between 0 and 255 and answer must fall in the range of 0 to 255. If the magnitude is greater than 255, one should use 16-bit arithmetic.



Magnitude

So far, we have considered all binary numbers as unsigned numeric values. However, we can also use signed binary numbers. Whether a number is a signed number or an unsigned number depends solely on how we treat the number in our operation. We assign a bit, the MSB, as a sign bit that helps us to place a minus sign in a binary position.

The rules for signed and unsigned binary numbers are simple:

- In an unsigned number, the MSB is a weighted position bit.
- In a signed number, the MSB (the sign bit) is 0 for a positive number.
- In a signed number, the MSB (the sign bit) is 1 for a negative number.

1.8.4.1 Complements

The complement of a number is the number which when added to the original will make it equal to a multiple of the base number system.

The complement of a number can be used as a representation of that number as a *negative* and as a *positive* number that represents a negative. It is a method, which can be used to make the subtraction easier for machines. Consequently, complements are used in the digital computers for simplifying the subtraction operation and for the logical operation.

For every base 'r' system, there are two types of complements: rs complement and (r-l)s complement. For decimal r = 10, we have 9s and l0s complement.

For binary r = 2, we have 1s and 2s complement.

For octal r = 8, we have 7s and 8s complement.

For hexadecimal r = 16, we have 15s and 16s complement.

1.8.4.2 Negative Binary Numbers - the 1s Complement

Positive numbers are same in both sequences, but we need to define the negative numbers in the system. All the negative numbers have the binary MSB = 1, which is helpful in identifying the sign of the number. Indeed, the binary MSB is commonly known as the *sign bit.* This bit is useful in differentiating between positive and negative numbers. In addition, the sign bit allows us to divide the counting sequence evenly between positive and negative numbers.

To form the negative of any number, first complement all the bits of the number. This result is known as the *one's complement* of the original number. This requires us to change every logic 1 bit in a number to logic 0 and every logic 0 bit to a logic 1. For instance, let us find the 1 s complement of 0011 0110 in binary.

Number Format	D7	D6	D5	D4	D3	D2	D1	D0
Unsigned Number	0	0	1	1	0	1	1	0
1s Complement	1	1	0	0	1	0	0	1

1.8.4.3 Negative Binary Numbers - the 2s Complement

We do not just place 1 in the MSB of a binary number to make it negative. We must take the 2s complement of the number. Taking the 2s complement of the number will cause the MSB to become 1.

To obtain the 2s complement of a number, there is a two-step process:

1. Take the 1 s complement of the number by changing every logic 1 bit in the number to logic 0 bit and change every logic 0 bit to logic 1 bit.
2. Add 1 to the 1's complement of the binary number. Now, we have the 2s complement of the original number. Here, we can notice that the MSB has become 1.

1s complement and 2s complement of 0011 0110 in binary is shown in the following table:

Number Format	D7	D6	D5	D4	D3	D2	D1	D0
Unsigned Number	0	0	1	1	0	1	1	0
1s Complement	1	1	0	0	1	0	0	1
2s Compliment	1	1	0	0	1	0	1	0

If we are using signed binary numbers and the MSB is already logic 1, it means the value is the 2s complement of the number.

1.8.4.4 Representation of signed numbers using 2s complement

We have discussed *that* the signed numbers can be represented by taking out the 2s complement of the original number. However, this representation varies between positive and negative numbers.

If the number is positive, the magnitude remains in its binary form and a sign bit of 0 is placed in front of the MSB.

Example 1: Represent + $(12)_{10}$ in 2s complement form.

Binary Number		1	1	0	0
1s Complement		0	0	1	1
2s Compliment		0	1	0	0
With Sign Bit	**0**	0	1	0	0

If the number is negative, the magnitude is represented in its 2s complement form and a sign bit 1 is placed in front of the MSB.

Example 2: Represent - $(14)_{10}$ in 2s complement form.

Binary Number		1	1	1	0
1s Complement		0	0	0	1
2s Compliment		0	0	1	0
With Sign Bit	**1**	0	0	1	0

1.8.4.5 Addition-subtraction of signed numbers using 2s complement addition The addition of signed binary numbers represented in the radix complement form is similar to the unsigned case. However, when the 2s complement of a number is added to any other binary number, it will be equivalent to its subtraction from that number. As a result, subtraction of the signed numbers by 2s complement method is performed by using the following steps:

1. Convert both the numbers into the binary equivalent form.
2. Find the 2s complement form of the number, which is subtracting, that is, subtrahend.
3. Add this 2s complement number to the minuend.
4. If there is carry of **1,** ignore it from the result to obtain the correct result.
5. If there is no carry, recomplement the result and attach the negative sign to the obtained result.

Example 1: Add $(27)_{10}$ and $(-11)_{10}$ using complementary representation for the negative value. Binary form of $(27)_{10} = (011011)_2$ and of $(11)_{10} = (001011)_2$

Hence, the result is $(010000)_2$ or $(16)_{10}$.

Note: *Here, carry* is 1, *so ignore it and the result* is *(010000)2.*

Example 2: Subtract $(25)_{10}$ from $(42)_{10}$.

Binary form of $(25)_{10} = (011001)_2$ and of $(42)_{10} = (101010)_2$

Get the 2s complement of the $(011001)_2$

Here, ignore the carry 1 and the result is $(010001)_2$ or $(17)_{10}$.

1.8.4 Summary

All the computers perform the arithmetic operations in the binary mode. The basic arithmetic operations that have been performed by all the number systems are addition and subtraction.

The rules of binary addition are as follows:

0 + 0 = 0
0 + 1 = 1
1 + 0 = 1
1 + 1 = 0 plus a carry of 1 to next higher column

The rules of binary subtraction are as follows:

0 - 0 = 0
1 - 0 = 1
1 - 1 = 0
0 - 1 = 1 with a borrow from the next column

The *complement* of a number is the number which when added to the original will make it equal to a multiple of the base number system. The complement of a number can be used to represent a number as a *negative* and a *positive* number. The addition and subtraction of the signed numbers is dependent on the 2s complement of the numbers and whenever the 2s complement of a number is added to any other binary number, it will be equivalent to its subtraction from that number.

1.8.5 *Self Check Exercise:*

Q.1 Why computers have deigns to use the binary number system?

Q.2 Perform the binary addition

- 1010 + 1101
- 111011 + 101011
- 1010110 + 1011010

Q.3 Add the binary numbers 1011 and 101 in both decimal and binary forms.

Q.4 Subtract 0110111_2 from 1101110_2

Q.5 Subtract 011011_2 from 110111_2

1.8.6 *Suggested Readings:*

1. Computer Fundamentals By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Fundamentals of Information Technology By Shiv Kumar Anand and Harmohan Sharma (Kalyani Publishers)
3. Fundamentals of Information Technology by V.Rajaraman (PHI, New Delhi).
4. Digital Design by M. Morris Mano (Pearson Education)
5. Computer Fundamentals, Architecture & Organisation by B.Ram, New Age International.

PGDCA | **Paper : PGDCA-I**
Introduction to Information Technology

Lesson No. 1.9 | **Author : Arun Sharma**

BOOLEAN ALGEBRA

Chapter Outline:

1.9.0 Objectives

- *Learn basic concepts of Boolean algebra*
- *Describe laws of Boolean algebra*
- *Explain the theorems of Boolean Algebra*
- *Defining Boolean function*

In the mid-eighteenth century, Boolean algebra was developed by the English mathematician, George Boole. It became known as Boolean algebra, after its developer's name. *Boolean algebra* is a mathematical system, which is used for formulation of the logical statements with symbols so that problems can be solved in a definite manner of ordinary algebra. In short, Boolean algebra is the mathematics of digital systems.Since an ordinary algebraic expression may be simplified by means of the basic theorems, the expressions that describe a given digital network can also be reduced or simplified by using the Boolean algebra. Nowadays Boolean algebra is extensively used in designing the circuitry that is used in computers.

1.9.1 Fundamental Concepts of Boolean Algebra Boolean algebra is based on the fundamental concepts described below:

1.9.1.1 Logical Addition and Logical Multiplication In the normal algebraic expression, a variable can take any numerical value. For example, in expression 4B + 6 A = C, the value of A, B and C may range through the entire field of real numbers.Since Boolean algebra deals with the binary number system, the variables used in the Boolean equations have only two possible values (0 or 1). Thus, for performing the logical algebraic operations, that is, 'addition' and 'multiplication', Boolean algebra follows certain rules. These rules are listed below in Table 1.9.1

Table 1.9.1 Boolean Addition and Multiplication

Addition Rules	Multiplication Rules
0 + 0 = 0	0 . 0 = 0
0 + 1 = 1	0 . 1 = 0
1 + 0 = 1	1 . 0 = 0
1 + 1 = 1	1 . 1 = 1

The symbol '+' is used for logical addition operator. It is also known as 'OR' operator. We can define the + symbol (OR operator) by listing all possible combinations of A and B, and the resulting value of C, in the equation A + B = C. Since, the variables A and B can have only two possible values (0 or 1), only four (2^2) combinations of inputs are possible, as shown in table 1.9.2. The resulting output values for each of the four input combinations are given in the table. Such a table is known as a *truth table.* Hence, table 1.9.2 is the truth table for the logical OR operator.

Table 1.9.2 Truth table for logical OR (+) operator

INPUT		OUTPUT
A	**B**	**C=A+B**
0	0	0
0	1	1
1	0	1
1	1	1

Observe that, the result is 0, only when both the input variables are 0; it is 1, when any of the input variables is I; and it is also I, when both the input variables are I. This is the reason why the + symbol does not have the "normal" meaning, but is a logical addition operator. This concept of logical addition may be extended to any number of variables. For example, in the equation A + B + C + D = E, even if A, B, C, and D, all had the value of 1, the sum of the values (the result E) would be 1 only. The equation A + B = C is normally read as "A or B equals C".

The symbol '.' is used for logical multiplication operator. It is also known as 'AND' operator. We can again define the symbol (AND operator) by listing all possible combinations of A and B and the resulting value of C, in the equation A . B = C. The truth table for logical AND operator is shown in table 1.9.3 Observe from the truth table that, the result C is equal to 1, only when both the input variables A and B are 1, otherwise it is O. The equation A . B = C is normally read as "A and B equals C".

Table 1.9.3 Truth table for logical AND (.) operator.

INPUT		**OUTPUT**
A	**B**	**C=A.B**
0	0	0
0	1	0
1	0	0
1	1	1

1.9.1.2 Complementation

The OR and AND operations are binary operations, because they define an operation on two variables. The complementation operation is a unary operation, which is defined on a single variable.

The symbol '-' is normally used for complementation operator. It is also known as 'NOT' operator. Hence, we write **$\bar{A}$**, meaning "complement of A", or (A + B), meaning "complement of A + B." The complementation of a variable is the reverse of its value. Hence, if A = 0, then **$\bar{A}$** = 1, and if A = 1, then **$\bar{A}$** = 0. The truth table for logical NOT (-) operator is shown in Table 1.9.4. **$\bar{A}$** is read as "complement of A" or "not *of* A",

Table 1.9.4 Truth table for logical NOT (-) operator

INPUT	**OUTPUT**
A	**X = $\bar{A}$**
0	1
1	0

1.9.1.3 Operator Precedence

Does A + B · C mean (A + B) · C or A + (B · C)? The two generate different values for A = 1, B = 0, and C = 0, because we have (1 + 0) · 0 = 0 and 1 + (0 · 0) = 1, which differ. Hence, it is necessary to define operator precedence, *to* correctly evaluate Boolean expressions. The precedence of Boolean operators is as follows:

1. The expression is scanned from left *to* right.
2. Expressions enclosed within parentheses are evaluated first.
3. All complement (NOT) operations are performed next.
4. All '·' (AND) operations are performed after that.
5. Finally, all '+' (OR) operations are performed in the end.

According *to* this precedence rule, A + B · C means A + (B · C). Similarly, *for* the expression $\bar{A} \cdot B$, the complement *of* A and B are both evaluated first, and the results are then ANDed. Again, *for* the expression (A + B)', the expression inside the parenthesis (A + B) is evaluated first, and the result is then complemented.

1.9.2 Axioms of Boolean Algebra

Premise: There is a set B and two operators: + and * that satisfy the following axioms:

Axiom #1: Closure

If a and b are elements of B, (a + b) and (a * b) are in B.

Axiom #2: Cardinality

There are at least two elements a and b in B such that a != b.

Axiom #3: Identity Element

(a) A + 0 = A

(b) A . 1 = A

Axiom #4: Commutative Law

The commutative law of addition for two variables is algebraically expressed as:

(a) A + B = B + A

The commutative law of multiplication for two variables is expressed as:

(b) A . B = B . A

In summary, the orders in which the variables are ORed or ANDed make no difference.

Axiom #5: Associative Law

The associative law of addition of three variables is expressed as:

(a) A + (B + C) = (A + B) + C

The associative law of multiplication of three variables is expressed as:

(b) A (BC) = (AB) C

In summary, ORing or ANDing a group of variables produces the same result regardless of the group of the variables.

Axiom #6: Distributive Law

The distributive law of three variables is expressed as follows:

(a) A (B + C) = AB + AC

(b) A + (BC) = (A + B) (A + C)

This law states that ORing several variables and ANDing the result is equivalent of ANDing the single variable with each of the variables in the grouping, then ORing the result.

Axiom #7: Complement Element

(a) A + A' = 1

(b) A . A' = 0

1.9.3 Duality Principle

It states that every algebraic expression deducible from the postulates of Boolean algebra remains valid if the operator and identity elements are interchanged. In Boolean algebra there is precise duality between the oprators . (AND) and + (OR) and the digits 0 and 1. For example in table second row is obtained from the first row and vice-versa, simply by interchanging '+' with '.' And '0' with '1'.

Row-1	1 + 1 = 1	1 + 0 = 0 + 1 = 1	0 + 0 = 0
Row-2	0 . 0 = 0	0 . 1 = 1 . 0 = 0	1 . 1 = 1

If the dual of an algebraic expression is desired, we simply interchange OR and AND operators and replace 1's by 0's and 0's by 1's.

1.9.4 Theorems of Boolean Algebra

From the axioms above we can derive the following theorems.

Theorem #1: Idempotent

i.) A + A = A

ii.) A * A = A

Proof: (i)

= L.H.S.

= A + A

= (A + A) * 1 (Identity)

= (A + A) * (A + A') (Complement)

= A + AA' (Distributive)

= A + 0 (Complement)

= A (Identity)

Proof: (ii)

= L.H.S.

= A * A

= AA + 0 (Identity)

= AA + AA' (Complement)

= A (A + A') (Distributive)

= A * 1 (Complement)

= A (Identity)

Theorem #2: Operations with 0 and 1

i.) A + 1 = 1

ii.) A * 0 = 0 (dual)

Proof: (i)

= L.H.S.

= A + 1

= (A + 1) * 1 (Identity)
= (A + 1) * (A +A') (Complement)
= A + 1 * A' (Distributive)
= A + A' * 1 (Commutative)
= A + A' (Identity)
= 1 (Complement)

Proof of (ii) holds by duality.

Theorem #3: Absorption

i.) A + A * B = A

ii.) A * (A + B) = A (dual)

Proof: (i)

= L.H.S.
= A + A * B
= A * 1 + A * B (Identity)
= A (1 + B) (Distributive)
= A (B + 1) (Commutative)
= A * 1 (Identity)
= A (Identity)

Proof of (ii) holds by duality.

Theorem #4: De-Morgan's Law

i.) (A + B)' = A' * B'

ii.) (A * B)' = A' + B' (dual)

De Morgan's first theorem (i) states that the complement of a logical sum equals the logical product of the compliments.

De Morgan's second theorem (ii) states that compliment of a logical product equal the logical sum of the compliments

Proof:

The proof for De-Morgan's law using the axioms of Boolean algebra is long. Another method (that also works for the other theorems we just discussed) is to show equality with a truth table:

A B	A + B	(A+B)'	A'	B'	A'*B'
0 0	0	1	1	1	1
0 1	1	0	1	0	0
1 0	1	0	0	1	0
1 1	1	0	0	0	0

By combining the laws of Boolean algebra and our knowledge of logic gates, we form several useful rules that may be used in manipulating and simplifying Boolean algebra expressions.

Table 1.9.5 Summary of basic Boolean identities

Sr. No.	Boolean Expression	Duality
1	A + 0 = A	A * A = A
2	A + 1 = 1	A * 0 = 0
3	A + A = A	A* A = A
4	A + A' = 1	A * A' = 0
5	A+ B = B + A	A * B = B* A
6	(A+B) + C = A + (B+C)	(A*B)* C = A*(B*C)
7	A * (B+C) = A*B + A*C	A + (B*C) = (A+B) * (A+C)
8	A + A*B = A	A * (A+B) = A
9	A + A'*B = A + B	A* (A' + B) = A*B
10	(A + B)' = A' + B'	(A * B)' = A' + B'

1.9.5 Boolean Functions

A Boolean function is an expression formed with binary variables and logical operators (OR, AND, NOT, and equal sign). In essence, a truth table is a list, which defines a Boolean function. For example, X = A · B + A · C. Let us consider the truth table of the given value as shown in Table 1.9.6. Note that the Function (X) is equal to 1 at three conditions, that is, A = 1, B = 0, C = 1; A = 1, B = 1, C = 0 and A = 1, B = 1, C = 1 otherwise X = 0. The algebraic expression representing this function is therefore expressed as:

X=ABC+ABC+ABC

Table 1.9.6 Truth Table for Boolean Functions

Input			Output
A	**B**	**C**	**X**
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Accordingly, the logic circuit of the given example is shown in Figure 1.9.1.

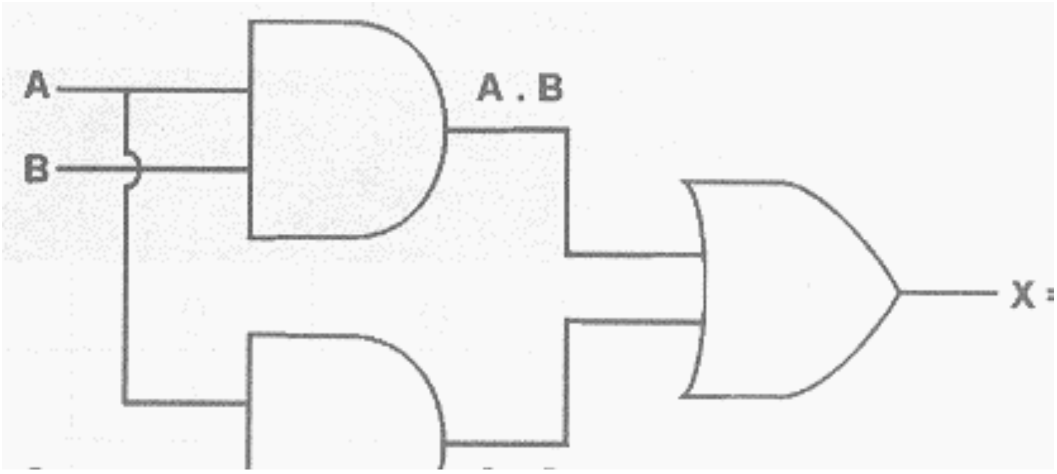


Figure 1.9.1 Logic Circuit

1.9.5.1 Minimization of Boolean Functions

When a Boolean function is implemented with logic gates (discussed in next chapter), each literal in the function designates an input to a gate, and each term is implemented with a gate. Hence, for a given Boolean function, the minimization of the number of literals, and the number of terms, will result in a circuit with less components. To find simpler circuits, one must know how to manipulate Boolean functions, to obtain equal and simpler expressions. However, we will give consideration only to the criterion of component minimization, which is achieved by literal minimization.

There are several methods used for minimizing the number of literals in a Boolean function. Here we will consider only the method of algebraic manipulations. Unfortunately, in, this method, there are no specific rules or guidelines to be followed, which will guarantee the final answer. The following examples illustrate this procedure.

Example: Simplify the following Boolean functions to a minimum number of literals.

(a) A + A' * B

(b) A * (A' + B)

(c) A' * B' * C + A' * B * C + A *B'

Solution:

a) A + A' * B

= (A + A') * (A + B)	(Distributive)
= 1 * (A + B)	(Complement)
= (A + B) * 1	(Commutative)
= A + B	(Identity)

(b) A * (A' + B)

=A * A' + A * B (Distributive)

= 0 + A * B (Complement)

=A * B + 0 (Commutative)

=A * B (Identity)

(c) A' * B' * C + A' * B * C + A *B'

= A' * C (B' + B) + A * B' (Distributive)

= A' * C.(B + B') + A * B' (Commutative)

= A' * C * 1 + A * B' (Complement)

= A'* C + A * B' (Identity)

1.9.6 Summary

Boolean Algebra is a mathematical system, which is used for the formulation' of the logical statements with symbols so that problems can be solved in a definite manner to ordinary algebra. In short, Boolean algebra is the mathematics of digital systems. Boolean algebra deals with the binary number system, the variables used in the Boolean equations have only two possible values (0 or 1). Boolean algebra is extensively used in designing the circuitry that is used in computers.

For manipulating and simplifying Boolean algebra expressions, we follow certain rules and laws of Boolean algebra.

A Boolean function is an expression formed with binary variables and logical operators (OR, AND, NOT, and equal sign). In essence, a truth table is a list, which defines a Boolean function.

1.9.7 Self Check Exercise:

- Discuss the various laws of the Boolean algebra.
- Explain the term Boolean function with a suitable example.
- Explain the principle of duality in Boolean algebra. How it is useful?
- State and prove the two basic De Morgan's theorems.

1.9.8 Suggested Readings:

1. Computer Fundamentals By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Fundamentals of Information Technology By Shiv Kumar Anand and Harmohan Sharma (Kalyani Publishers)
3. Fundamentals of Information Technology by V.Rajaraman (PHI, New Delhi).
4. Digital Design by M. Morris Mano (Pearson Education)
5. Computer Fundamentals, Architecture & Organisation by B.Ram, New Age International.

LOGIC GATES

Chapter Outline:

1.10.0 Objectives

- *What are logic gates?*
- *Their types and operation*
- *Combination of logic gates*

The term *gate* is used to describe the set of the basic electronic components which when combined with each other, are able to perform complex logical and arithmetic operations. As discussed earlier, everything in the digital world is based on the binary number system. Numerically, this involves only two symbols - 0 and 1. According to the need, we can use these symbols or we can equate them with others. Thus, when dealing with digital logic, we can specify that:

0 = False = No
1 = True = Yes

Using this two-valued logic system, either every statement or condition must be 'true' or 'false'. It cannot be partly true and partly false.

Thus, a *logic gate* is an elementary building block of a digital circuit. A simple logic gate has two inputs and one output. At any given moment, every terminal is in one of the two binary conditions, low (0) or high (1), represented by different voltage levels. In most logic gates, the low state is approximately zero volts (0 V), while the high state is approximately five volts positive (+5 V). In other words, a logic gate is an electronic circuit, which operates on one or more input signals to produce the desired output signals. Besides, the properties of a logic gate or of a combination of gates, may be defined and presented in the form of a diagram called a truth table, which lists the output that will be triggered by each of the possible combinations of input signals.

1.10.1 Basic Logic Gates

Information is fed to a gate in the form of binary coded input signals (logic value 0 stands for 'OFF' or 'low-voltage pulse' and logic 1 for 'ON' or 'high-voltage'). By using different combination of the input signals and the output signals, different types of logic gates are made and are used to perform the necessary operation of the computer circuits. Essentially, there are three basic logic gates:

1. AND gate
2. OR gate
3. NOT gate

1.10.1.1 AND Gate

The AND gate is composed of two or more inputs and a single output and performs logical multiplication. The standard symbol for the AND gate is shown in Figure 1.10.1 and its truth table is listed in Table 1.10.1. The logical operation of the AND gate is such that the output is HIGH (1) when all the inputs are HIGH, otherwise it is LOW (0). A dot (**.**) is used to show the AND operation.

Three Inputs AND Gate

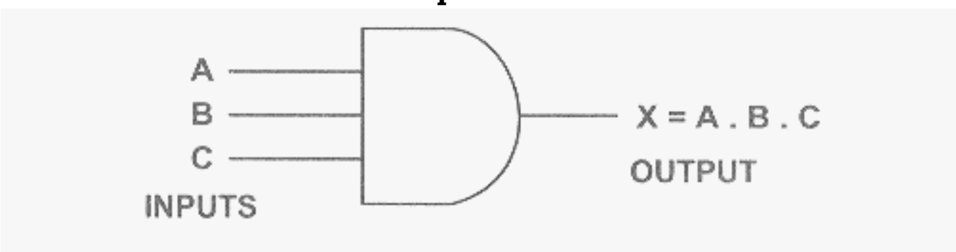


Figure 1.10.1 Logic Symbol of AND Gate

A	B	C	X=A.B.C
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Table 1.10.1 Truth table of AND Gate

The Venn diagram shown in Figure 1.10.2 provides an insight into the AND function. The highlighted area represents the function X = ABC.

Note: *An AND gate will yield a logic 1 output only if it receives a logic 1 signal through all its inputs.*

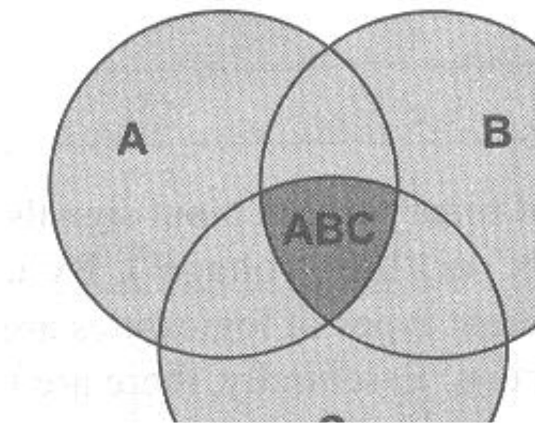
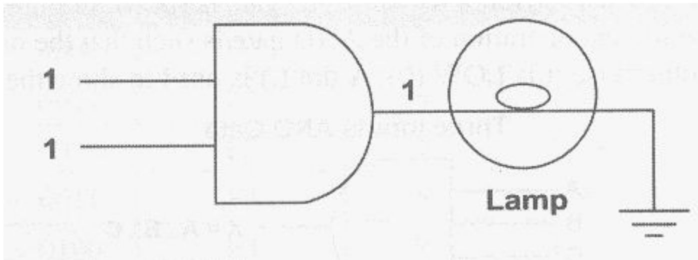


Figure 1.10.2 Venn diagram illustrating AND Gate

We can explain the function of the AND gate with the help of an example.



According to the above figure, a value of '1' is needed at all AND gate inputs to produce an output value of '1' from the AND gate.

1.10.1.2 OR Gate

The OR gate is composed of two or more inputs and a single output and performs logical addition. The standard symbol for the OR gate is shown in Figure 1.10.2 and its truth table is listed in Table 1.10.2. The logical operation of the OR gate is such that the output is HIGH (1) when any of the inputs are HIGH; otherwise it is LOW (0). A plus (+) is used to show the OR operation. In other words, the OR gate is an electronic circuit that gives a high output if one or more of its inputs are high.

Three Inputs OR Gate

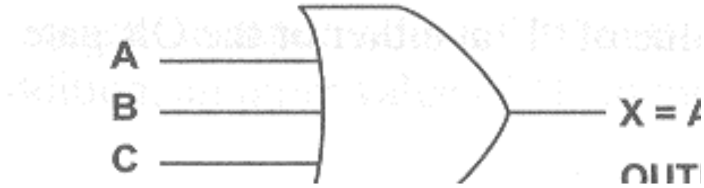


Figure 1.10.3 Logic Symbol of OR Gate

Table 1.10.2 Truth Table of OR Gate

A	**B**	**C**	**X=A+B+C**
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

The venn diagram shown in Figure 1.10.4 provides function of the OR gate. The highlighted area represents the function X = A + B + C.

Note: *An OR gate will give a logic* 1 *output if one or more of its inputs receives a logic* 1 *signal.*

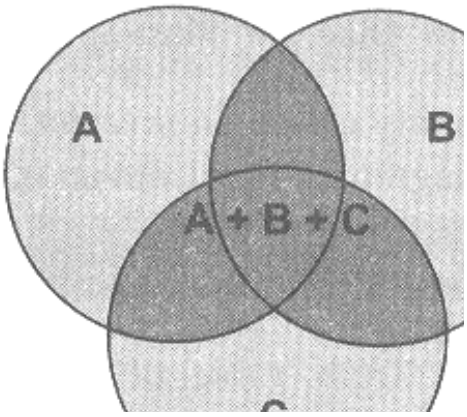
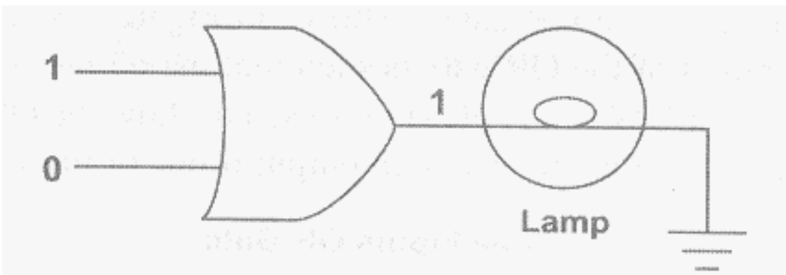


Figure 1.10.4 Venn Diagram illustrating OR Gate

We can explain the function of the OR gate with the help of an example that is discussed below:



According to this example, an input value of '1' at either of the OR gate inputs will result in an output value of '1' from the OR gate.

1.10.1.3 NOT gate

The NOT gate performs a basic logic function called *inversion* or *complementation.* The purpose of this gate is to change one logic level (HIGH *I* LOW) to the opposite logic level. In terms of bits, it changes' 1' to '0' and vice versa. The standard logic symbol for the NOT gate and a Venn diagram illustrating the relationship between the variables and the logic gate operation are shown in Figure 1.10.5 and Figure 1.10.6, respectively.

NOT Gate (an Inverter)

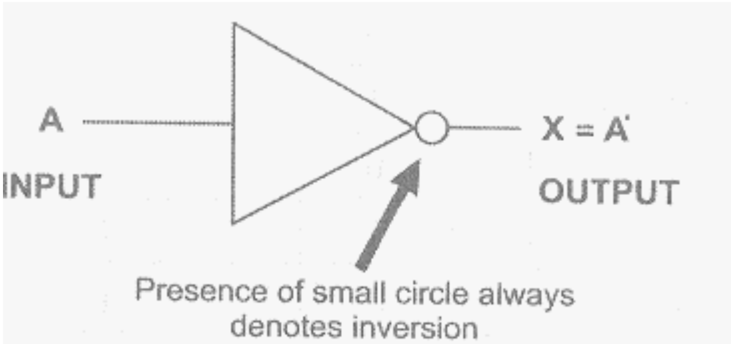


Figure 1.10.5 Logic Symbol of NOT Gate

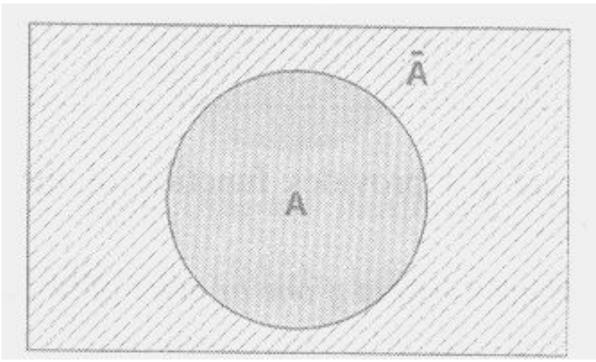


Figure 1.10.6 Venn Diagram Illustrating NOT Gate

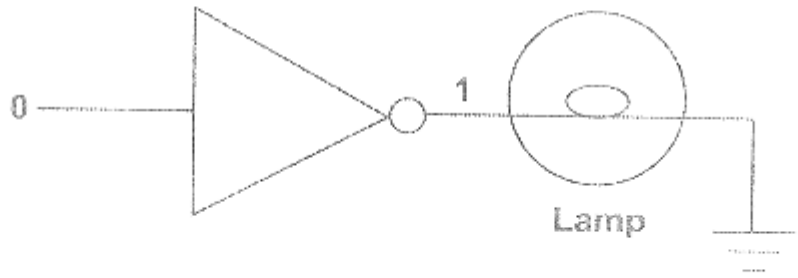
In electronics a NOT gate is more commonly called an *inverter.* The circle on the symbol is called a *bubble* and is generally used in circuit diagrams to indicate an inverted input or output.

The truth table for the NOT gate is shown in Table 1.10.3.

Table 1.10.3 Truth Table of NOT Gate

A	**X=A'**
0	1
1	0

We can explain the function of the NOT gate with the help of an example that is discussed below:



With reference to the above figure, an input value of '0' at the NOT gate produces an output value of '1' from the NOT gate. In addition, an input value of '1' at the NOT gate produces an output value of '0' from the NOT gate.

1.10.2 Logic Operations

Generally, information is fed to a gate in the form of binary coded input signals, that is, either '0' or '1' and after performing certain logical operations; each combination of input signals yields a specific output (logic 0 or 1). Probably, these logical operations are based on making decision between two logics. There are basically three logical operations:

1. AND Operation
2. OR Operation
3. NOT Operation

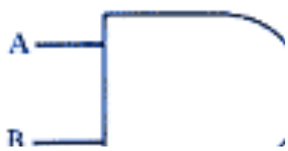
1.10.2.1 AND Operation

In the AND operation, a result of 1 occurs when all of the input variables are 1. Furthermore, the output of this operation is 0 for any case where one or more inputs are 0. The multiplication sign stands for the AND operation, same for ordinary multiplication of 1s and 0s.

The expression X = A . B reads as 'X equals A AND B'. An example of two inputs AND gate and its truth table is as follows:

Two Inputs AND Gate

INPUTS



X = A · B
OUTPUT

Figure 1.10.7 AND Gate

Table 1.10.4 Truth Table

A	B	X=A.B
0	0	0
0	1	0
1	0	0
1	1	1

With the AND operation, the outputs 1 x 1 = 1, 1 x 1 x 1 = 1, and so on can be obtained.

1.10.2.2 OR Operation

According to the OR operation, a result of 1 is obtained when any of the input variable is 1. **In** addition, the OR operation produces a result of 0 only when all the input variables are 0 The + sign stands for the OR operation, not for ordinary addition.

The expression X = A + B reads as 'X equals A OR B'.

Two Inputs OR Gate

Figure 1.10.8 OR Gate

Table 1.10.5 Truth Table

A	B	X=A+B
0	0	0
0	1	1
1	0	1
1	1	1

With the OR operation, the outputs 1 + 1 = 1, 1 + 1 + 1 = 1, and so on can be obtained.

1.10.2.3 NOT Operation

The NOT operation is unlike the OR and AND operations. This operation can be performed on a single input variable. For example, if the variable A is subjected to the NOT operation, the result x can be expressed as:

x = A' or A
where, the prime (1) represents the NOT operation.
This expression is read as:
x equals NOT A
x equals the inverse of A
x equals the complement of A.

Figure 1.10.9 NOT Gate

Each of these is in common usage and all indicate that the logic value of x = A is opposite to the logic value of A. The NOT operation is also referred to as inversion or complementation and these terms are used interchangeably.

The truth table of the NOT operation is as follows:

Table 1.10.6 Truth Table

A	**X=A'**
0	1
1	0

With reference to the above table,
1' = 0 because NOT 1 is 0
0' = 1 because NOT 0 is 1.

1.10.3 Combination of Logic Gates

When we look at logic circuit diagrams for digital equipment, we are not going to see just a single gate, but many combinations of gates. Using combinations of logic gates, complex operations can be performed. Although there is no limit to the number of gates that can be arrayed together in a single device, nevertheless, in practice, there is a limit to the number of gates that can be packed into a given physical space. In this section, we will analyse some basic combinations of gates. These gates can be listed as:

- NAND gate
- NOR gate
- Exclusive-OR (XOR) and Exclusive-NOR (XNOR) gate

1.10.3.1 NAND Gate

The NAND, which is composed of two or more inputs and a single output, is a very popular logic element because it may be used as a *universal function.* That is, it may be employed to construct an inverter, an AND gate, an OR gate, or any combination of these functions. The term NAND is formed by the combination of NOT-AND and implies an AND function with an inverted output. The standard symbol for the NAND gate is shown in Figure 1.10.10 and its truth table is listed in Table 1.10.7. The logical operation of the NAND gate is such that the output is LOW (0) only when all the inputs are HIGH (1).

Two Input NAND Gate

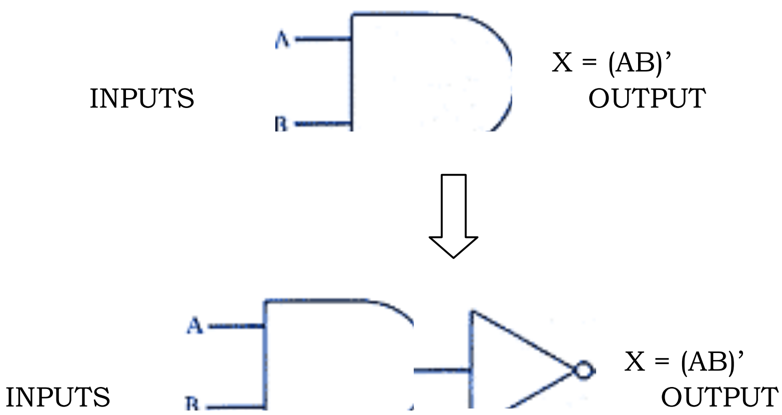


Figure 1.10.10 Standard Logic Symbol of NAND Gate

INPUT		OUTPUT
A	**B**	**X**
0	0	1
0	1	1
1	0	1
1	1	0

Table 1.10.7 Truth Table for NAND Gate

The NAND gate is called a universal gate because combinations of it can be used to accomplish all the basic functions.

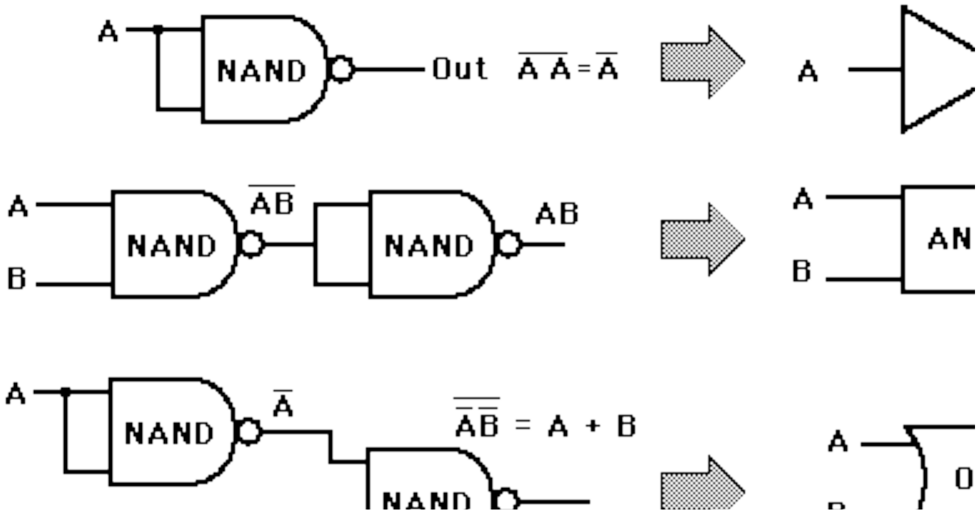


Figure 1.10.11 NAND Gate Operations

1.10.3.2 NOR Gate

The NOR gate, which is composed of two or more inputs and a single output, also has a universal property. The term NOR is formed by the concatenation NOT -OR and implies an OR function with an inverted output. The standard symbol for the NOR gate is shown in Figure 1.10.12 and its truth table is listed in Table 1.10.8 The logical operation of the NOR gate is such that the output is HIGH (1) only when all the inputs are LOW.

Two Input NOR Gate

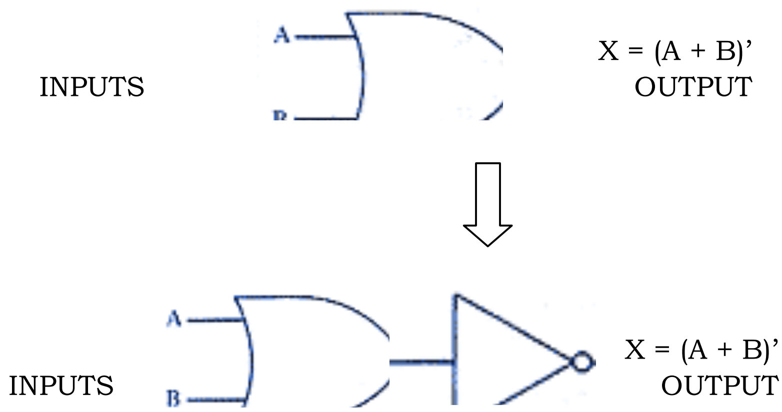


Figure 1.10.12 Standard Logic Symbol of NOR Gate

INPUT		OR	NOR
A	**B**	**X=A+B**	**X=A'.B'**
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

Table 1.10.8 Truth Table for NOR Gate

The NOR gate is called a universal gate because combinations of it can be used to accomplish all the basic functions.

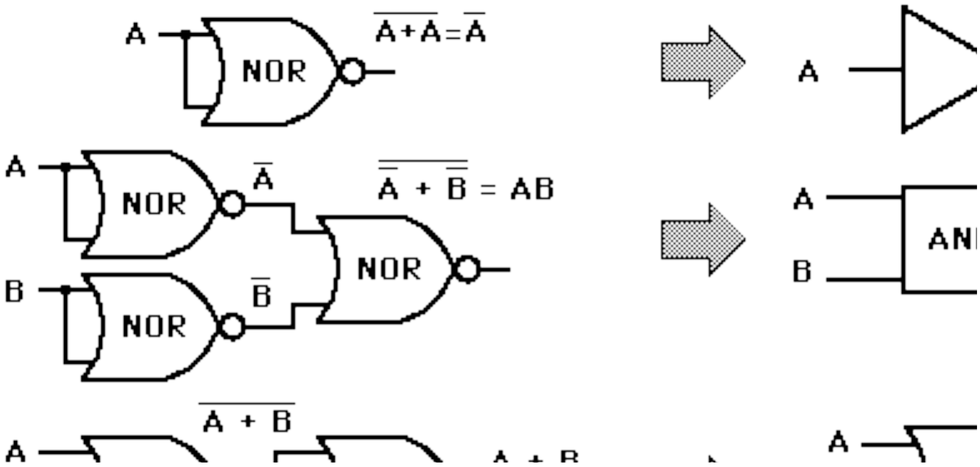


Figure 1.10.13 NOR Gate Operations

1.10.3.3 Exclusive-OR (XOR) and exclusive-NOR (XNOR) Gate

These gates are usually formed from the combination of the other logic gates already discussed. However, because of their functional importance, these gates are treated as basic gates with their own unique symbols. The truth tables for the XOR and XNOR gates, shown in Figure 1.10.14 are listed in Table 1.10.9. The exclusive-OR is an 'inequality' function and the output is HIGH (1), when the inputs are not equal to each other. Conversely, the exclusive-NOR is an 'equality' function and the output is HIGH (1) when the inputs are equal to each other.

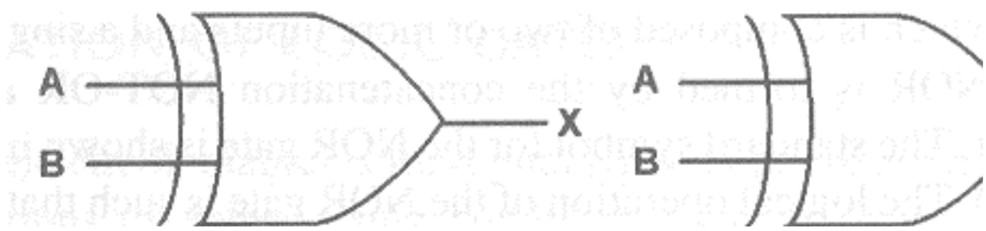


Figure 1.10.14 Standard Logic Symbols for (a) XOR (b) XNOR

INPUT		XOR	XNOR
A	**B**	**X**	**X**
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	1

Table 1.10.9 Truth Table for XOR and XNOR Logic Gates

The exclusive-OR gate and exclusive-NOR gate are denoted by the $\oplus$ and $\odot$, respectively.

In addition, these gates perform the following Boolean functions:

$A \oplus B = A \cdot B' + A' \cdot B$

$A \odot B = A \cdot B + A' \cdot B'$

1.10.4 Summary

A logic gate is a building block of a digital circuit that operates on one or more input signals to obtain the standard output signals.

An AND gate is an electronic circuit, which performs the logical multiplication operation. This circuit generates an output signal of 1, only if all input signals are also 1. An OR gate is an electronic circuit, which performs the logical addition operation. This gate generates an output signal of 1, if any of the input signals is 1.

The NOT gate performs a basic logic function called *inversion* or *complementation***.** In electronics, a NOT gate is more commonly called an inverter. The purpose of this gate is to change one logic level (HIGH/ LOW) to the opposite logic level. In terms of bits, it changes a '1' to a '0' and vice versa.

Information is fed to a gate in the form of binary coded input signals, that is either'0' or '1' and after performing certain logical operations; each combination of input signals yields a specific' output (logic 0 or 1). There are basically, three logical operations: AND operation, OR operation, NOT operation.

The combinational logic gates are the interconnection of the logic gates for forming other logic gates or networks. Using combinations of logic gates, complex operations can be performed. Some of these combinational logic gates are: NAND gate, NOR gate, and Exclusive-OR (XOR) and Exclusive-NOR (XNOR) gate.

The NAND gate is formed by the combination of NOT-AND and implies an AND function with an inverted output. It is a very popular logic element because it may be used as a *universal function.*

The NOR gate is formed by the combination of NOT-OR and implies an OR function with an inverted output. This gate is also accomplished with the universal property.

The exclusive-OR and exclusive-NOR gates are usually formed from the combination of the other logic gates. The exclusive-OR gate and exclusive-NOR gate are denoted by the $\oplus$ and $\odot$, respectively.

1.10.5Self Check Exercise :

Q.1 What are the different types of logic gates? Explain it with the help of truth tables and give an example of each gate

Q.2 Why are NAND and NOR gates called universal gates?

Q.3 Shoe the implementation of the logical operations AND, OR and NOT with only NAND gates and with only NOR gates.

1.10.6Suggested Readings:

1. Computer Fundamentals By Pradeep K. Sinha and Priti Sinha (BPB Publications)
2. Fundamentals of Information Technology By Shiv Kumar Anand and Harmohan Sharma (Kalyani Publishers)
3. Fundamentals of Information Technology by V.Rajaraman (PHI, New Delhi).
4. Digital Design by M. Morris Mano (Pearson Education)
5. Computer Fundamentals, Architecture & Organisation by B.Ram, New Age International.