VENKATESHWARA
OPEN UNIVERSITY
www.vou.ac.in

INTRODUCTION TO COMPUTERS

BBA
[BBA-103]

INTRODUCTION TO COMPUTERS

BBA

[BBA-103]

VENKATESHWARA
OPEN UNIVERSITY
www.vou.ac.in

BOARD OF STUDIES

Prof Lalit Kumar Sagar
Vice Chancellor

Dr. S. Raman Iyer
Director
Directorate of Distance Education

SUBJECT EXPERT

Dr. Ashok Upadhyay	*Programme Director*
Mohit Kumar Anand	*Assistant Professor*
Ms. Likha Maryam	*Assistant Professor*
Miss Tako Yami	*Assistant Professor*

COURSE CO-ORDINATOR

Mr. Tauha Khan
Registrar

Authors:
Sanjay Saxena, Managing Director, Total Synergy Cosulting Pvt. Ltd.
(Units: 1.0-1.2, 1.3, 1.4.4, 1.6-1.10, 2) © Sanjay Saxena, 2019
Bharat Kumar, Assistant Librarian, Management Development Institute, Gurgaon
(Unit: 1.5) © Bharat Kumar, 2019
Richa Bhargava (Unit: 3) © Richa Bhargava, 2019
Vikas Publishing House (Unit: 1.3.1-1.3.2, 1.4-1.4.3, 1.4.5) © Reserved, 2019

All rights reserved. No part of this publication which is material protected by this copyright notice may be reproduced or transmitted or utilized or stored in any form or by any means now known or hereinafter invented, electronic, digital or mechanical, including photocopying, scanning, recording or by any information storage or retrieval system, without prior written permission from the Publisher.

Information contained in this book has been published by VIKAS® Publishing House Pvt. Ltd. and has been obtained by its Authors from sources believed to be reliable and are correct to the best of their knowledge. However, the Publisher and its Authors shall in no event be liable for any errors, omissions or damages arising out of use of this information and specifically disclaim any implied warranties or merchantability or fitness for any particular use.

VIKAS®

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT LTD
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Introduction to Computers

Syllabi	Mapping in Book
Unit I: Basics of Computers Computers and their Components - Elements of Data Communication - Storage Media. Need for Computerization of Library.	**Unit 1:** Basics of Computers **(Pages 3-81)**
Unit II: Software System and Application Software - Operating Systems - DOS, Windows, Unix & Linux.	**Unit 2:** Software **(Pages 83-115)**
Unit III: Computers and Society Information Society - National Information Infrastructure (NII) - Technological Convergence.	**Unit 3:** Computers and Society **(Pages 117-141)**

CONTENTS

INTRODUCTION

NOTES

Today, the fact that computers have made a big impact on many aspects of our lives can hardly be questioned. They have opened up an entire world of knowledge and information that is readily accessible. Thus, information about computers right from the first mechanical adding machine to the latest microprocessors has become imperative for students as well as anybody who has something to do with a computer system.

Today, we are using the fifth generation of computers. The term ‘generation’ is used to distinguish between varying hardware and software technologies. The hardware by itself cannot do any calculation or manipulation of data without being instructed what to do and how to do it. Thus, there is a need of software in a computer system.

The software used in a computer system is grouped into applications software, system software and utility software.

We have tried to familiarize students with the basic components of a computer system, which include the control unit, memory, processor, and input and output devices. Input devices are used to transfer information into the memory unit of a computer. Output devices are electromechanical devices that accept data from the computer and translate it into a form that can be understood by the outside world.

In computing, memory refers to the physical devices used to store programs (sequences of instructions) or data on a temporary or permanent basis for use in a computer or other digital electronic device. The term primary memory is used for the information in physical systems which function at high speed, i.e., RAM (Random Access Memory) as a distinction from secondary memory, which are physical devices for program and data storage which are slow to access but offer higher memory capacity. Primary memory stored on secondary memory is called virtual memory. The term storage is often used for traditional secondary memory, such as tape, magnetic disks and optical discs (CD-ROM and DVD-ROM). There are two main types of semiconductor memory: volatile and non-volatile. Examples of non-volatile memory are flash memory (sometimes used as secondary, sometimes primary computer memory) and ROM, PROM, EPROM and EEPROM memory. Examples of volatile memory are primary memory (typically Dynamic RAM or DRAM), and fast CPU cache memory (typically Static RAM or SRAM).

This book, *ICT Fundamentals*, has been written in the SIM format or the self-instructional mode wherein each unit begins with an ‘Introduction’ to the topic followed by an outline of the ‘Unit Objectives’. The detailed content is then presented in a simple and an organized manner, interspersed with ‘Check Your Progress’ questions to test the understanding of the students. A ‘Summary’ along with a list of ‘Key Terms’ and a set of ‘Questions and Exercises’ is also provided at the end of each unit for effective recapitulation.

UNIT 1 BASICS OF COMPUTERS

NOTES

Structure

1.0 INTRODUCTION

In this unit, you will learn about the definition, types, classification and uses of computers. A computer may be defined as a device that can operate upon information or data. The information is processed based on a set of instructions provided to generate the output. The increasing popularity of the computer has proved that it is a powerful and useful tool. The size, shape, cost and performance of computers have changed over the years, but the basic logical structure has not. You will learn about the history of computers and the various generations of its evolution.

You will also learn about digital computers and its various components. Input devices are used in order to transfer data to the memory of the computer. The Arithmetic Logic Unit (ALU) is responsible for calculations, to which this data from the memory is stored. You will learn about computer software and its need. A computer cannot operate without any instructions and is based on a logical sequence of instructions in order to perform a function. These instructions are known as a 'computer program' and constitute the computer software.

You will also learn how data is represented and processed in computer. Data processing can be defined as the process of converting raw data into suitable information using a series of operations like classifying, sorting, summarizing and tabulating it for easy storage and retrieval. There are two methods of processing data, i.e., batch processing and real-time processing. This unit has elaborated discussions about the how various components of the CPU work.

NOTES

You will also learn about various types of storage media, their capacity and their access time. Storage devices are mainly of two kinds: primary storage devices and secondary storage devices. Primary storage devices, also known as versatile memory, loses entire data once the power is switched off. Whereas, secondary storage devices, also known as non-versatile memory, can hold the data even after the power is switched off. In secondary storage; you can retrieve your data as soon as the power is switched on. However, access time for primary memory is very less as compared to the secondary memory. This unit also discusses various kinds of magnetic disks, such as CD-ROM, pen drive, external hard disk, and Blu-Ray disk.

1.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Describe various components of a computer
- List various advantages and disadvantages of a computer
- Outline the history of computers
- Recognize various elements of data communication
- Distinguish between various kinds of registers according to their properties
- List various kinds of storage media
- Name various magnetic storage devices
- Explain how data is organised on a magnetic disk
- Describe various types of communication transfer modes in USB
- Handle various troubleshooting if hard disk is configured during the time of troubleshooting process
- List various needs for computerization of library
- Describe various types of library automation systems
- Explain various steps in library automation

1.2 COMPUTERS AND THEIR COMPONENTS

Computers have undergone great transformation over the past decade; however, the basic logical structure remains the same. A computer primarily constitutes of three integral components, viz. input unit, Central Processing Unit (CPU) and output unit. The CPU constitutes of the main memory, the Arithmetic Logic Unit (ALU) and the control unit.

Apart from these three basic components, computers have secondary storage devices known as auxiliary storage or backing storage that store data and instructions on a long-term basis.

The following are the primary functions of a computer:

- **Inputting**: The process in which the user specify a set of commands to process data into the computer system.

NOTES

- **Storing**: The process of recording data and information so that it can be retrieved for use whenever required.
- **Processing**: This process implies performing arithmetic or logical operations on data to convert them into useful information. Arithmetic operations include addition, subtraction, multiplication and division, and logical operations include comparisons, such as equal to, less than and greater than.
- **Outputting**: This is the process of providing results to the user. These can be in the form of visual display and/or printed reports.
- **Controlling**: This refers to directing the sequence and the manner in which all the previous functions are carried out.

A detailed description of the components that perform these tasks is as follows:

1. Input Unit

Programs and data are required to be present in a computer system before any operation can be performed. A program denotes the set of instructions which the computer has to carry out and data is the information on which these instructions are to be operated. If the task is to rearrange a list of telephone subscribers in alphabetical order, the sequence of instructions that will guide the computer through this operation is the program, while the list of names to be sorted is the data.

The input unit is responsible for transferring data and instructions from the external environment into the computer system. Instructions and data enter the input unit through the particular input device (such as keyboard, scanner and card reader). These instructions and data are then converted into binary codes (computer acceptable form) and sent to the computer system for further processing.

2. Central Processing Unit

The central processing unit is known as the brain of the computer. It is an important part of the computer and includes the control unit, the ALU and the primary memory that are described as follows:

- **Main Memory (Primary Storage):** The main memory or the primary storage of the computer system is responsible for storing all the instructions and data temporarily. The data is then transferred to the arithmetic logical unit for processing. After this, the final output is again stored back in the primary storage, until it is further sent to the output device.

  The primary storage also temporarily stores any intermediate result generated by the ALU. So data and instructions move frequently between the ALU and the primary storage before the processing is complete. It should be noted that no processing occurs within the primary storage.
- **Arithmetic Logic Unit:** In addition to the basic four arithmetic operations, viz. addition, subtraction, multiplication and division, the ALU also performs logic comparison operations including equal to, less than and greater than.
- **Control Unit:** The function of the control unit is to ensure that according to the stored instructions, the right operation is done on the right data at the right time. The control unit receives instructions and commands from the programs in the primary memory, processes them and ensures that the commands are executed in the desired order by all the other units of the computer system. In fact, the control unit is comparable to the central nervous system of the human body.

NOTES

3. Output Unit

Computers understand, process data and return the output in a binary form. The basic function of the output unit is to convert these results into a human readable form before providing the output through various output devices, such as terminals and printers.

The storage capacity of the primary memory of the computer is limited. Often, it is necessary to store large amounts of data. So, additional memory, called secondary storage or auxiliary memory, is used in most computer systems.

Secondary storage is storage other than the primary storage. These are peripheral devices connected to and controlled by the computer to allow permanent storage of data and programs. Usually, hardware devices, such as magnetic tapes and magnetic disks, fall in this category.

1.2.1 Advantages and Disadvantages of Computers

The increasing popularity of the computer has proved that it is a powerful and useful tool. Its usefulness is due to its following features:

- **Speed**: Computers are very fast. They can process millions of instructions every second. The speed is related to the amount of data it processes and the time it takes to complete the processing task.
- **Storage**: Computers can store vast amounts of information in the form of files which can be recalled at any time. These files help in easy and speedy retrieval of information. This type of storage is known as electronic storage system.
- **Accuracy**: In addition to being fast, computers are also accurate. The degree of accuracy for a particular computer depends upon its design. Most errors in computers are not of a technical nature but are human. Usually, programmers are responsible for these errors.
- **Diligence**: Computers can perform any complicated task accurately without making any error. Computers do not suffer from carelessness, boredom or tiredness. Moreover, their efficiency does not decrease with age.
- **Versatility**: Computers perform various tasks depending upon the instructions given to them and their hardware characteristics. They are capable of performing any task, provided the task is reduced to a series of logical steps. A computer can be used to prepare a Word document and in between called to search for another document that is stored in its memory. It can perform both tasks simultaneously.

Though computers can do better than human beings in terms of accuracy, speed and memory, there are certain disadvantages of computer systems as they depend on human beings for their operations and functions. The following are some of the disadvantages of computers:

- They depend on human beings who program them for efficient, accurate and fast functioning.
- Computers do not have their own intelligence and, thus, cannot think intelligently or work independently like human beings. They can just follow instructions given by programs or by users.
- They can neither take decisions nor can correct wrong instructions. Programmers or users maintain and update them.
- As with many other modern appliances, computers also need electric power to run.

NOTES

1.2.2 History of Computers

The first mechanical adding machine was invented by a French mathematician, physicist, inventor, writer and chritian philosopher Blaise Pascal in 1642. Later, in 1671, a German philosopher, mathematician and political advisor Baron Gottfried Wilhelm von Leibniz invented the first calculator. Around this time, an American inventor Herman Hollerith developed the concept of punched cards, which were extensively used as an input medium in mechanical adding machines.

A mathematician, philosopher, inventor and mechanical engineer, Charles Babbage, a 19th century professor at Cambridge University, is considered the father of the modern digital computer. During this period, mathematical and statistical tables were prepared by a group of clerks. However, utmost care and precaution could not eliminate human errors.

In 1842, Babbage came up with a new idea of the Analytical Engine, which was meant to be completely automatic. This machine was capable of performing basic arithmetic functions. However, these machines were difficult to manufacture because the precision required to manufacture them was not available at that time.

The following is a brief description of the various generations of computers.

- **Mark I Computer (1937–44)**: This was the first fully automatic calculating machine designed by an American physicist and a pioneer in computing Howard H. Aiken, the design of which was based on the technique of punching card machinery. In this technique, both mechanical and electronic components were used.
- **Atanasoff-Berry Computer (1939–42):** This computer was developed by an American physicist and inventor Dr John Atanasoff to solve certain mathematical equations. It used forty-five vacuum tubes for internal logic and capacitors for storage.
- **The Electronic Numerical Integrator And Computer (ENIAC) It (1943–46):** was the first electronic computer developed for military requirements and was used for many years to solve ballistic problems.
- **EDVAC (1946–52)**: One of the drawbacks of ENIAC was that its programs were wired on boards, which made it difficult to change them. To overcome the drawbacks of ENIAC, the Electronic Discrete Variable Automatic Computer (EDVAC) was designed. The basic idea behind this concept was that sequences of instructions could be stored in the memory of the computer for automatically directing the flow of operations.
- **EDSAC (1947–49)**: A British computer scientist Professor Maurice Wilkes developed the Electronic Delay Storage Automatic Calculator (EDSAC), by which addition and multiplication operations could be accomplished.
- **The UNIVersal Automatic Computer (UNIVAC) (1951):** It was the first digital computer to be installed in the Census Bureau in 1951 and was used continuously for 10 years. In 1952, International Business Machines (IBM) introduced a commercial computer IBM 701. This computer could be used for scientific and business purposes.

Generations of Computer

The history of computer development can be divided into different phases which are often referred to as generations of computing devices. ‘Generation’ in computer

NOTES

terminology is a 'step' in technology. Each generation of computers is characterized by a major technological development that fundamentally changes the way computers operate, resulting in increasingly smaller, cheaper, more powerful, efficient and reliable devices which has decreased the energy consumption and the heat decipation.

Originally, the term 'generation' was used to distinguish between varying hardware technologies, but nowadays, it includes both hardware and software.

The following are the characteristics of each generation of computers:

1. First Generation (1940–1956): Vacuum Tubes

The first generation computers used vacuum tubes in their electronic circuits and magnetic drums for memory. A vacuum tube was a delicate glass device that used filaments as a source of electrons, and could control and amplify electronic signals. Figure 1.1 displays a vacuum tube.

Fig. 1.1 *A Vacuum Tube*

These computers could perform computations in milliseconds but were enormous in size, occupying almost an entire room. They were very expensive to operate and in addition to using a great deal of electricity, generated a lot of heat resulting in malfunctioning.

First generation computers relied on machine language (binary-coded programs) to perform operations and could solve only one problem at a time. Input was based on punch cards and paper tape, and output was displayed on printouts.

Early computers, such as ENIAC, EDVAC and UNIVAC I, all can be classified as first generation computers.

2. Second Generation (1956–1963): Transistors

Transistors, developed in 1947, replaced vacuum tubes in the second generation computers. The transistor was far superior compared to vacuum tube, making computers smaller, faster, cheaper, more energy efficient and more reliable than their first generation predecessors. Although transistors also generated a great deal of heat that could damage the computer, it was a great improvement over the vacuum tube. Second generation computers still relied on punched cards for input and printouts for output.

The cryptic binary machine language was followed by the symbolic or assembly language that allowed programmers to specify instructions in words. High-level programming languages, such as COBOL and FORTRAN, were also being developed at this time.

These were also the first computers that stored their instructions in the memory, which advanced the technology from magnetic drum to magnetic core. The first computers of this generation were specifically developed for the atomic energy industry.

NOTES

3. Third Generation (1964–1971): Integrated Circuits (ICs)

Transistors were clearly an improvement over the vacuum tube but still generated a lot of heat resulting in computer damage.

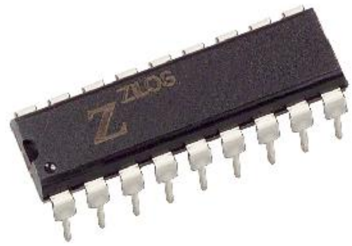


Fig. 1.2 *An IC Chip*

The development of integrated circuit (see Figure 1.2) by Jack Kilby in 1958, an electrical engineer with Texas Instruments, was the greatest achievement for initiation of the third generation of computers. The invention of IC chips made it possible to greatly reduce the size of computers with improvement in operation speed and reliability.

Instead of punched cards and printouts, users interacted with third generation computers through devices, such as keyboards and monitors. They also interfaced with an operating system that allowed the device to run many different applications at one time with a central program that monitored the memory.

Now, the computers became accessible to the masses because they were substantially smaller and cheaper than their predecessors.

4. Fourth Generation (1971–Present): Microprocessors

Large Scale Integration (LSI) were developed which could fit hundreds of components onto a single chip. By 1980s, Very Large Scale Integration (VLSI) squeezed thousands of components onto a single chip. Ultra Large Scale Integration (ULSI) increased that number to millions.

The ability to fit so much processing capability in an area so small, helped to reduce the size and price of the computers. It also increased its power, efficiency and reliability.

Initially, the IC technology was used only for constructing the processor, but it was soon discovered that the same technology could also be used for the construction of memory. The first memory chip was constructed in 1970 and could hold 256 bits. Figure 1.2 displays an IC chip.

As more and more components were fabricated on a single chip, fewer and fewer chips were needed to construct the processor. The Intel 4004 chip, developed in 1971, located all the components of the computer — from Central Processing Unit and Memory to Input/Output controls—on a single chip. This was the first microprocessor. Figure 1.3 displays the Intel pentium microprocessor chip.

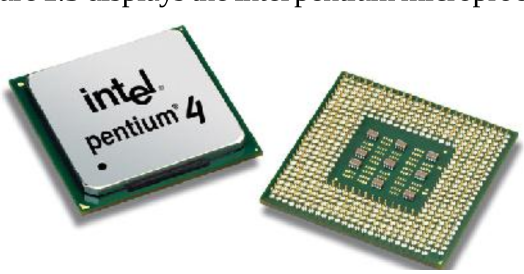


Fig. 1.3 *The Intel Pentium Microprocessor Chip*

NOTES

IBM introduced its first personal computer in 1981 and in 1984 Apple also introduced a personal computer, the Macintosh. Microprocessors also advanced from the realm of desktop computers to advanced technologies, and many areas of life as more and more everyday devices began to use microprocessors.

As computers increased in computing power, it was possible to connect them together to form networks, which eventually led to the development of the Internet. Fourth-generation computers also marked the development of Graphical User Interfaces (GUIs), the mouse and various handheld devices.

5. Fifth Generation (Present and Beyond): Artificial Intelligence

The fifth generation computers are being developed using the technology of artificial intelligence; for instance, voice recognition systems. Parallel processing and supercomputers have lead to the further development of artificial intelligence. In the future, quantum computation and molecular technology will tremendously transform computers. The fifth generation aims at creating devices that respond to input in natural language and are capable of learning and self organization.

Table 1.1 provides a list of various computer generations.

Table 1.1 *Generation of Computers*

Generation	Time	Hardware	Features	Examples
I	**1942-1955**	**Vacuum Tubes**	High-speed electronic switching device; memory type was electromagnetic; bulky in size; generated a large amount of heat; frequent technical faults; required constant maintenance; used for scientific purposes; air-conditioning required	ENIAC, EDVAC, EDSAC, UNIVAC I
II	**1955-1964**	**Transistors**	Better electronic switching devices than vacuum tubes; made of germanium semiconductors; memory type was magnetic cores; powerful and more reliable; easy to handle; much smaller than vacuum tubes; generated less heat as compared to vacuum tubes; used for business and industries for commercial data processing; air-conditioning required	Livermore Atomic Research Computer (LARC), IBM
III	**1964-1975**	**Integrated Circuits (ICs)** made up of transistors, resistors and capacitors fixed on single silicon chip	ICs were smaller than transistors; consumed less power; dissipated less heat as compared to transistors; more reliable and faster than earlier generations; capable of performing about 1 million instructions per second; large storage capacity; used for both scientific and commercial purposes; air-conditioning required	Mainframe, Minicomputers
IV	**1975-1989**	**Microprocessor** made up of Large Scale Integration Circuits (LSI) and Very Large Scale Integration Circuits (VLSI)	Microprocessor had control on logical instructions and memory; semiconductor memories; personal computers were assembled; used in LAN and WAN to connect multiple computers at a time; used graphical user interface; smaller, more reliable and cheaper than third-generation computers; larger primary and secondary storage memories; had Computer Supported Cooperative Working (CSCW); air-conditioning not required	Personal Computers (PCs), LAN, WAN, CSCW
V	**1989-Present**	Ultra Scale Large Integration (USLI), Optical Disks	PCs were assembled – portable and non-portable, powerful desktop PCs and workstations; less prone to hardware failure; user-friendly features – Internet, e-mailing; air-conditioning not required	Portable PCs, Palmtop Computers, Laptop

NOTES

1.2.3 More Parts of a Computer System

In order to transfer data to the memory of the computer, input devices are used. The ALU is responsible for calculations. Once the calculations are done, the data is transferred back to the memory. The memory is responsible for storing data, according to which different functions are carried out. This memory is also known as the main memory or the Immediate Access Store (IAS).

The control unit is responsible for controlling various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer. The main function of the control unit is to make sure that the instructions are correctly followed and all operations are done exactly according to the instructions at the correct time. This process leads to outcomes that are stored in memory. Figure 1.4 displays a computer system.

Fig. 1.4 A Computer System

Motherboard

The main PCB (Printed Circuit Board) is sometimes alternatively known as a logical board or a main board of a personal computer. In fact, any complex electronic system is known as a motherboard. It includes a flat fibreglass platform which hosts the CPU, the main electronic components, device controller chips, main memory slots, slots for attaching the storage devices and other subsystems. Figure 1.5 displays a motherboard.

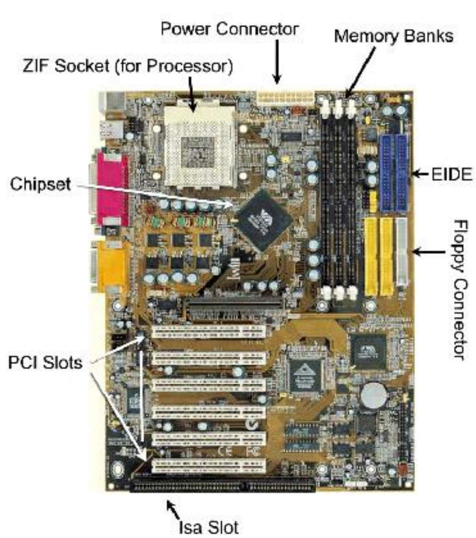


Fig. 1.5 A Motherboard

NOTES

Sockets and Ports

- **Main Power Socket:** The top part of the rear of the computer locates the main power cable socket, which supplies power from the electric mains to the computer system. This socket is the part of the main power supply unit of the computer.
- **Monitor Power Socket:** The socket that supplies the power from the computer system to the computer monitor and is located below the main power cable socket. However, you might not find this socket in all computers and you can plug in the monitor directly in main power supply.
- **PS/2 Mouse Port:** Next you will find a small, round, green colored port with seven holes and a small logo of the mouse printed next to it. This is where your PS/2 mouse will be plugged in.
- **PS/2 Keyboard Port:** Right next to mouse port you will find another similar purple colored port with the keyboard logo printed next to it. This is where your PS/2 keyboard will be plugged in.
- **Fan Housings:** You will notice two fan housings at the back of your computer. One fan housing is a part of the power supply unit and the other will be somewhere below it to cool off the heat generated by the CPU.
- **Serial Ports:** It is a 9-pin connector normally used to attach the old serial port mouse, hand-held scanners, modems, joysticks, game pads and other such devices.
- **Parallel Port:** It is a 25-pin connector used to attach parallel port printers, modems, external hard disk drives and other such devices.
- **Audio Jacks:** There are three audio jacks in your computer system. One jack is used for connecting your speakers or headphones, the second is used to connect the microphone and the third to connect to another audio device, such as a music system.
- **LAN Port:** The LAN port (where LAN is Local Area Network) is where the RJ45 connector of your LAN cable is plugged-in to connect your computer to other computers or the Internet.
- **USB Ports:** The USB port (where USB is Universal Serial Bus) is designed to connect multiple peripheral devices in a single standardized interface, and has a plug and play option that allows devices to be connected and disconnected without having to restart the computer. It has replaced many serial and parallel ports for devices, such as mouse, printers, modems, joysticks, game pads, scanners, digital cameras and other such devices.
- **VGA Port:** This is a 15-pin connector that connects the signal cable of the monitor to the computer. Here VGA means Video Graphics Array.

Figure 1.6 displays power cables and sockets of monitor and CPU, respectively.

NOTES

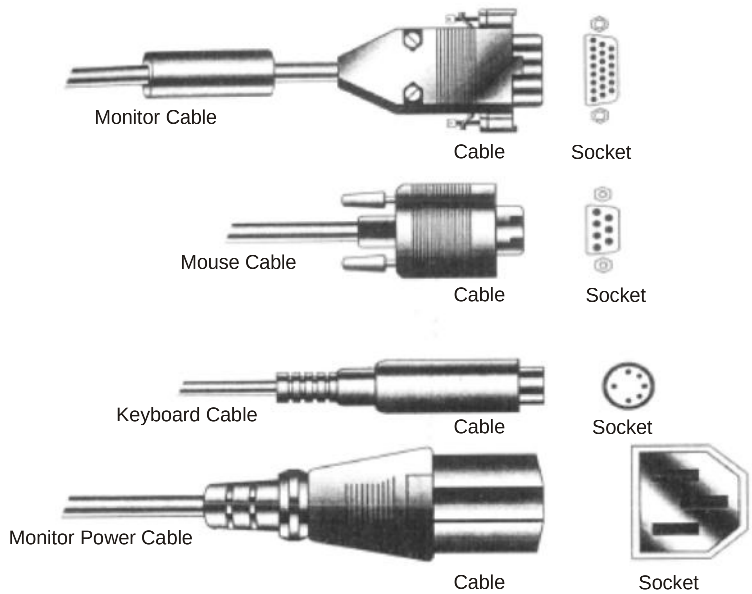
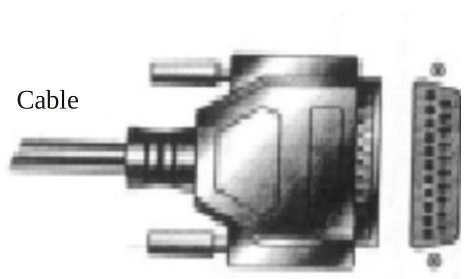


Fig. 1.6 *Monitor and CPU: Power Cable and Sockets*

Figure 1.7 displays a LAN cable and a printer cable with its socket.

LAN Cable



Printer Cable and Socket

Fig. 1.7 *LAN Cable and Printer Cable with Socket*

Memory

Storage and retrieval of instructions and data in a computer system is the responsibility of the memory. In order to store data and instructions, the CPU constitutes many registers, though these are capable of storing very few bytes. All computers need storage space for temporarily storing instruction and data during the execution of the program as the CPU can process data at a higher speed than the speed at which data can be transferred from disks to registers. This could lead to the idle CPU which remains free most of the time. The primary or the main memory is the temporary storage located in the computer hardware. Secondary storage or auxiliary memory

NOTES

constitutes devices that can give backup storage, such as magnetic tapes and disks. The memory is classified as follows:

(i) **Internal Processor Memory:** A small set of high-speed registers placed inside a processor and used for storing temporary data while processing.

(ii) **Primary Storage Memory:** The main memory of the computer which communicates directly with the processor. This memory is large in size and fast, but not as fast as the internal memory of the processor. It comprises a couple of integrated chips mounted on a printed circuit board plugged directly on the motherboard. Random-Access Memory (RAM) is an example of primary storage memory.

(iii) **Secondary Storage Memory:** This stores all the system software and application programs, and is basically used for data backups. It is much larger in size and slower than primary storage memory. Hard disk drives, floppy disk drives and flash drives are a few examples of secondary storage memory.

Memory Capacity: Capacity, in computers, refers to the number of bytes that it can store in its main memory. This is usually stated in terms of Kilobytes (KB) which is 1024 bytes or Megabytes (MB) which is equal to 1024 KB (10,48,576 bytes). The rapidly increasing memory capacity of computer systems has resulted in defining the capacity in terms of Gigabytes (GB) which is 1024 MB (1,07,37,41,824 bytes). Thus, a computer system having a memory of 256 MB is capable of storing ($256 \times 1024 \times 1024$) 26,84,35,456 bytes or characters.

Processors Used in PCs

The most significant part of the computer is the CPU. The CPU is mostly a microprocessor-based chip located on a single or sometimes a multiple printed circuit boards and is an internal component of the system. It is directly connected to the motherboard; however, the compatibility of the motherboard and the CPU depends on the specific series of the latter. Due to the tremendous amount of heat generated by the CPU, it contains a heat sink and a cooling fan.

Popular microprocessors include Intel and AMD, which are compatible with IBM CPUs.

The brands of CPUs are not the only differentiating factors between different processors, there are various technical aspects to these processors which allow us to differentiate between CPUs of different power, speed and processing capability. Accordingly, each of these manufacturers sells numerous product lines offering CPUs of different architecture, speed, price range and so on. The following are the most common aspects of modern CPUs that enable us to judge their quality or performance:

- **32 or 64-bit Architecture:** A bit is the smallest unit of data that a computer processes. 32 or 64-bit architecture refers to the number of bits that the CPU can process at a time.
- **Clock Rate:** The speed at which the CPU performs basic operations, measured in Hertz (Hz) or in modern computers Megahertz MHz or Gigahertz GHz.
- **Number of Cores:** CPUs with more than one core are essentially multiple CPUs running in parallel to enable more than one operation to be performed simultaneously. Current ranges of CPUs offer up to eight cores. Currently,

the Dual-core (i.e., two cores) CPU is most commonly used for standard desktops and laptops, and Quad-core (i.e., four cores) is popular for entry-level servers.

- **Additional Technology or Instruction Sets:** These refer to unique features that a particular CPU or range of CPUs offer to provide additional processing power or reduced running temperature. These range from Intel's MMX, SSE3, and HT to AMD's 3DNOW, and Cool n Quiet.

NOTES

These technical factors are the basic way to judge how a CPU will perform. It is important to consider multiple factors when looking at a CPU rather than just the clock speed or any one specification on its own. It is easy for a single-core processor to run music videos, Internet applications or games individually, but when multiple applications are run together, it starts to slow down. A system running on a dual-core processor would be able to multitask better then a single-core processor, while it is very easy for an 8-core processor to run all these applications plus a lot more without showing any signs of slowing down. However, Intel's 4-core processors are actually two dual-core processors combined in a single processor, whereas AMD's 4-core processors are actually four processors built in a single chip.

A combination of the above mentioned specifications, along with the operating systems that the processor supports and the specific purpose for which the computer is to be used, are the factors to be considered when deciding which CPU is the most suitable for your needs.

1.3 ELEMENTS OF DATA COMMUNICATION

Information is handled in the computer by electrical components, such as transistors, integrated circuits, semiconductors and wires, all of which can indicate only two states or conditions. Transistors are either conducting or non-conducting; magnetic materials are either magnetised or non-magnetised in a direction; a pulse or voltage is either present or not present. All information can, therefore, be represented within the computer by the presence (ON) or absence (OFF) of these various signals.

Thus, all data to be stored and processed in computers are transformed or coded as strings of two symbols, one symbol to represent each state. The two symbols normally used are 0 and 1. These are known as BITS, an abbreviation for BInary digiTS.

Let us now understand some commonly used terms:

Value	***Meaning***
0	Off
1	On

BITS: The smallest element that a computer uses. It can hold one of two possible values.

A bit which is ON is seen as SET or TRUE; and a bit which is OFF is seen as NOT SET or FALSE.

Since, only two values can be stored in a single bit, there can only be four unique combinations:

00 01 10 11

Check Your Progress

1. Name three basic components of computer.
2. State at least two advantages and disadvantages of computers.
3. List all the generations of the computers and also name the hardware used in them.
4. Define primary and secondary memory.
5. List the three qualities of CPUs on which their performance can be judged.

NOTES

Bits, therefore, have to be combined together into larger units for holding greater ranges of values.

NIBBLE: A set of four bits. There can only be a maximum of sixteen different values possible:

$$2^4 = 16 \text{ (2 to the power of the number of bits)}$$

BYTES: A set of two nibbles or eight bits. Characters are often stored in bytes along with numeric values.

$$2^8 = 256 \text{ (2 to the power of the number of bits)}$$

WORD: Just like we express information in words, so do computers. A computer 'word' is a group of bits, the length of which varies from machine to machine, but is normally predetermined for each machine. The word may be as long as 64 bits or as short as 8 bits.

Character Representation

Binary data is not the only data handled by the computer. We also need to process alphanumeric data, like alphabets (upper and lower case), digits (0 to 9) and special characters, such as +, –, *, /, (,), space and blank. These also must be internally represented as bits.

Table 1.2 shows the representation of data in bits.

Table. 1.2 *Data Representation in Bits*

Decimal Number	Binary Equivalent
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

BCD (Binary Coded Decimal)

It is one of the early memory codes. It is based on the concept of converting each digit of a decimal number into its binary equivalent rather than converting the entire decimal value into a pure binary form. It further uses 4 digits to represent each of the digits.

Converting 42_{10} into its BCD equivalent would result in:

$$42_{10} = \underset{4}{0100} \quad \underset{2}{0010} \quad \text{or} \quad 01000010 \text{ in BCD}$$

As seen, 4-bit BCD code can be used to represent decimal numbers only. Since, 4 bits are insufficient to represent the various other characters used by the computer, instead of using only 4 bits (giving 16 possible combinations), computer designers commonly use 6 bits to represent characters in BCD code. In this, the 4 BCD numeric place positions are retained, but two additional zone positions are added.

With 6 bits, it is possible to represent 2^6 or 64 different characters. This is, therefore, sufficient to represent the decimal digits (10), alphabetic characters (26) and special characters (28).

EBCDIC

A major drawback of the BCD code is that only 64 distinct characters can be represented by it. Since, this proved insufficient for representing many special characters (28 plus), upper case letters (26), lower case letters (26) and decimal numbers (10), the BCD code was extended from a 6-bit to an 8-bit code. The additional 2 bits function as zone bits, thus, expanding the zone bits to 4. The code that results is known as the Extended Binary Coded Decimal Interchange Code (EBCDIC). It was developed by IBM and is used in many computers, particularly, the majority of IBM models. It allows 2^8 or 256 characters to be represented. This takes care of the character requirement along with a large quantity of non-printable (such as movement of the cursor) and printable control characters (such as vertical spacing on printer).

Since, EBCDIC is an 8-bit code, it is possible to divide it into two groups of 4-bits each. One hexadecimal digit represents each group. Thus, computers using the EBCDIC code for internal representation of characters can use the hexadecimal number system as a notation for memory dump.

ASCII

A computer code that is very widely used for data interchange is called the ‘American Standard Code for Information Interchange’ or ASCII. Several computer manufacturers have adopted it as their computers’ internal code. This code uses 7 digits to represent 128 characters. Now an advanced ASCII is used having 8-bit character representation code allowing 256 different characters. This representation is being used in microcomputers.

We will now study the encoding method. Table 1.3 represents the bit combinations that are required for each character.

Table 1.3 *Bit Combinations for Each Character*

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	TAB	LF	VT	FF	CR	SO	SI
10	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20		!	"	#	$	%	&	‘	(	)	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
60	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

Thus, the text string ‘Hello.’ is coded in ASCII using hexadecimal digits as follows:

H	e	l	l	o	.
48	65	6C	6C	6F	2E

The byte sequence 48 65 6C 6C 6F 2E represents the string ‘Hello.’

1.3.1 Processing of Data

NOTES

Data processing can be defined as the process of converting raw data into suitable information using a series of operations, like classifying, sorting, summarizing and tabulating, it for easy storage and retrieval. Processed data is called information.

Data, especially large volumes of it, unless processed properly, are not of much use in the current information driven world. Relevant information can give a definite edge to a business to stay ahead of its competition and plan for the future. In fact, the speed at which information can be extracted from data (a process called data processing) is just as crucial as the information itself. Information usually loses its value if it's not up to date. Automatic Data Processing (ADP) applications are gaining wide popularity in the market to solve this very problem. They not only save time but also reduce the cost of data processing. An ADP application, once configured, is ideal for converting similarly structured data into specific sets of information using predefined rules of selection, processing and presentation. Data processing can also include the conversion of one type of information into another for legacy systems transfer.

Typically, a data processing cycle can be broadly divided into five stages which areas follows:

- Data Collection
- Data Preparation
- Data Input
- Data Processing
- Information Output

Just as there are different types of data (classified either by usage, attributes or content), there are different methods of processing data. These are as follows:

1. **Real-Time Processing:** In this mode, data is processed almost immediately and in a continuous flow. This is of particular advantage when the lifespan of information is small and core business activities are involved. The advantages of real-time processing are that the derived information is up to date and, hence, it is more relevant for decision making. For example, in a bank or in an ATM (Automatic Teller Machine), as soon as you deposit money in your account, your account status (balance standing to your credit) is updated, instantaneously. This enables you as well as the bank to know the exact status of funds in real-time mode, or in other words, as of now. Similarly, in a railway reservation system, a train ticket booked from anywhere in the world must update the central database in real time to ensure that the seat once booked is not sold again. Real-time processing also requires relatively little storage as compared to batch processing.
2. **Batch Processing:** Real-time processing requires high-speed broadband connections, so that the data inputted from different computers or locations can be used to update a centralized server and database. Setting up such networks is expensive and not always feasible because sometimes the data need not to be processed immediately. For example, in a BPO (Business Processing Outsourcing) outfit, hundreds of operators may be inputting data, which can be made available to the client only after it is checked and verified by a supervisor(s).

NOTES

Such situations call for batch mode processing, which is used when the conversion of data into information is not required immediately and, therefore, this data processing is done in lots or batches. The advantages of batch processing are that it is cheaper and processing can be done offline.

It should be noted that data processing and data conversion are technically quite different; while data conversion only means converting data from one form to another, whereas data processing means conversion of data into information or sometimes vice versa.

1.3.2 CPU

Execution of programs is the main function of the computer. The program to be executed is a set of instructions that is stored in the computer's memory. Tasks are completed when the instructions of the program are executed by the CPU. Also, all the major calculations and comparisons are carried out inside the CPU. Additionally, the CPU is responsible for activating and controlling the operations of various units of the computer system. It activates the peripherals to perform input or output.

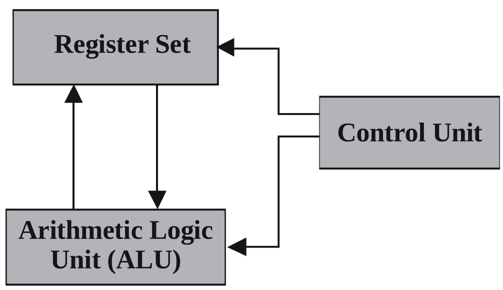


Fig. 1.8 *Major Components of a CPU*

The CPU is made up of three major components (see Figure 1.8): the register set (associated with the main memory) that stores the intermediate data during the execution of instructions, the Arithmetic Logic Unit (ALU) that performs the required micro-operations for executing the instructions, and the control unit that supervises the transfer of information among the registers and instructs the ALU as to which operation to perform.

Control Unit

The control unit is necessary if the CPU is to function efficiently and information/data is to be transferred between the CPU and other devices. It does not perform the actual processing of the data, but manages and coordinates the entire computer system, including the input and output devices. It retrieves and interprets the instructions from the program stored in the main memory, and issues signals that cause the other units of the system to execute them.

This is done through some special purpose registers and a decoder. The special purpose register, called the **instruction register**, holds the current instruction to be executed, and the **program control register** holds the next instruction to be executed. The decoder interprets the meaning of each instruction supported by the CPU. Each instruction is also accompanied by a **microcode,** i.e., the basic directions to tell the CPU how to execute the instruction.

NOTES

Arithmetic Logic Unit (ALU)

It performs arithmetic and logic operations. This means that when the control unit encounters an instruction that involves an arithmetic operation (add, subtract, multiply or divide) or a logic operation (equal to, less than or greater than), it passes the control to ALU. The ALU has the necessary circuitry to carry out these arithmetic and logic operations.

As an example, a comparison of two numbers (a logical operation) may require the control unit to load the two numbers in the requisite registers and then pass on the execution of the 'compare' function to the ALU.

Figure 1.9 represents the basic structure of a CPU.

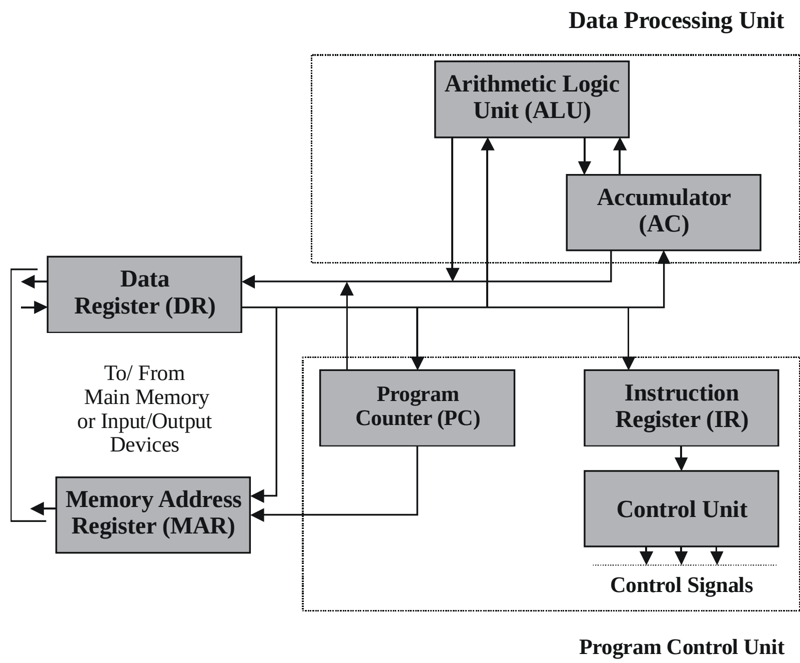


Fig. 1.9 *Basic Structure of a CPU*

Instruction Set

The primary function of the processing unit in the computer is to interpret the instructions given in a program and carry out the instructions. Processors are designed to interpret a specified number of instruction codes. Each instruction code is a string of binary digits. All processors have input/output instructions, arithmetic instructions, logic instructions, branch instructions and instructions to manipulate characters. The number and type of instructions differ from processor to processor. The list of specific instructions supported by the CPU is termed as its **instruction set.**

An instruction in the computer should specify the following:

- The task or operation to be carried out by the processor. This is termed as the **opcode.**
- The address(s) in memory of the operand(s) on which the data processing is to be performed. This is termed as **operand address**.
- The address in the memory that may store the results of the data processing operation performed by the instruction.

- The address in the memory for the next instruction to be fetched and executed.

Note that the next instruction, which is to be executed, is normally the next instruction following the current instruction in the memory. Therefore, normally no explicit reference to the next instruction is provided.

Instruction Representation

An instruction is divided into a number of fields and is represented as a sequence of bits. Each of the fields constitutes an element of the instruction. A layout of an instruction is termed as the **instruction format** (see Figure 1.10).

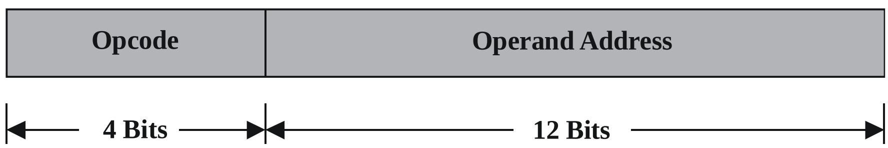


Fig. 1.10 *A Sample Instruction Format*

In most instruction sets, many instruction formats are used. An instruction is first read into an Instruction Register (IR), then it is decoded by the CPU which extracts the required operands on the basis of references made on the instruction fields, and processes it. Since, the binary representation of the instruction is difficult to comprehend, it is rarely used for representation. Instead, a symbolic representation is used. (see Table 1.4)

Table 1.4 *Examples of Typical Instructions*

Instruction	Interpretation	Number of Addresses
ADD A,B,C	Operation A = B + C is executed	3
ADD A,B	A = A + B. In this case the original content of operand location is lost	2
ADD A	AC = AC + A. Here A is added to the accumulator	1

Typically, CPUs manufactured by different manufacturers have different instruction sets. This is why, machine language programs developed for a particular CPU do not run on a computer with a different CPU (having a different instruction set).

Registers

The primary task that the CPU performs is the execution of instructions. It executes every instruction by means of a number of small operations known as **micro-operations**. Thus, it can be seen that:

- The CPU needs an extremely large main memory.
- The speed of the CPU must be as fast as possible.

To understand further, let us define two relevant terms:

Memory Cycle Time: Time taken by the CPU to access the memory.

Cycle Time of the CPU: The time that the CPU takes for executing the shortest well-defined micro-operation.

NOTES

It has been observed that the time taken by the CPU to access the memory is about 10 times higher than the time that the CPU takes for executing the shortest well-defined micro-operation. Therefore, CPU registers serve as the temporary storage areas within the CPU. They are termed as fast memory and can be accessed almost instantaneously.

Further, the number of bits a register can store at a time is called the length of the register. Most CPUs sold today have 32-bit or 64-bit registers. The size of the register is also called the word size, and indicates the amount of data that a CPU can process at a time. Thus, the bigger the word size, the faster the computer can process data.

The number of registers varies among computers, but typical registers found in most computers include:

Memory Buffer Register (MBR): Data is received from the memory (in the case of read operations), and it is held in the memory (in the case of write operations) by MBR.

Memory Address Register (MAR): The memory location's address where data is to be stored (in case of write operations), and the location from where data is to be accessed (in case of read operations) is specified by MAR.

Accumulator (AC): Interactions with the ALU are carried out by the accumulator, in which the output and input operands are stored. This register, therefore, holds the initial data to be operated upon, the intermediate results and the final results of processing operations.

Program Counter (PC): The address next instruction to be executed subsequent to the current instruction is kept by the program counter.

Instruction Register (IR): Instructions are loaded in the instruction register prior to being executed, i.e., the instruction register holds the current instruction that is being executed.

Instruction processing, in its simplest form, can be defined as a two-step process:

1. Codes or instructions are read (fetched) from the CPU one by one.
2. The operation indicated by that particular instruction is performed or executed.

The program counter fetches the instruction and tracks which instruction to fetch next. Since, the execution of a program is in a sequential manner, the program counter usually fetches the next instruction in the sequence. This instruction appears in the binary code form and is loaded into an IR. It is then interpreted by the CPU and the desired action is carried out. Generally, the following categories can be identified for these actions:

- ***Data Transfer:*** Transferring of data from I/O (Input/Output Unit) to CPU, from CPU to I/O, from memory to CPU or from CPU to memory.
- ***Data Processing:*** A logic or an arithmetic operation may be carried out on the data by the CPU.
- ***Sequence Control:*** This action is typically required for altering the sequence in which the instructions are executed. For instance, if an instruction from location 50 specifies that the next instruction to be fetched should be from

location 100, then the program counter will need to be modified to contain the location 100 (which otherwise would have contained 51).

Instructions can be executed involving many combinations of these actions.

NOTES

Processor Speed

The term processor has replaced the term Central Processing Unit (CPU). The processor in a personal computer that is embedded in small devices is often called a **microprocessor**.

The speed at which the processor executes commands is called the processor speed or **clock speed**. Every computer contains an internal clock (known as the system clock) that regulates the rate at which the instructions are executed, and synchronises the various computer components. The processor requires a fixed number of clock cycles (electric pulses per second) to execute each instruction. Clock cycles are required to fetch, decode and execute a single program instruction. Thus, the shorter the clock cycle, the faster the processor.

In a computer, clock speed, therefore, refers to the number of pulses per second generated by an oscillator that sets the tempo for the processor. It is usually measured in MHz (**Megahertz** - Million pulses per second) or GHz (**Gigahertz** - Billions pulses per second).

Computer clock speed has been, roughly, doubling every year. The Intel 8088, common in computers around the year 1990, ran at 4.77 MHz. Today's personal computers run at clock speeds of 100 to 1000 MHz and some even exceed one gigahertz.

Although the processing speed in personal computers is measured in terms of megahertz, the processing speed of minicomputers and mainframe systems is measured in terms of MIPS (Millions of Instructions Per Second) or BIPS (Billions of Instructions Per Second). This is because personal computers generally employ a single microprocessor chip as their CPU, while other classes of computers employ multiple processors to speed up their overall performance. Thus, a minicomputer having a speed of 500 MIPS is capable of executing 500 million instructions per second.

Clock speed is a measure of computer 'power', but it is not always directly proportional to the performance level. If you double the speed of the clock leaving all other hardware unchanged, you will not necessarily double the processing speed. The type of microprocessor, the bus architecture and the nature of the instruction set, all make a difference.

Check Your Progress

6. Define bits.
7. What is the full form of ASCII?
8. List the three major components of CPU.
9. State the basic functionality of registers.
10. What is clock speed or processor speed?

1.4 STORAGE MEDIA

The most common properties used for characterizing and evaluating the storage unit of the computer system are expressed as follows:

1. **Storage Capacity:** Represents the size of the memory. It is the amount of data that can be stored in the storage unit. Primary storage units have less storage capacity as compared to the secondary storage units. The capacity of internal memory and main memory can be expressed in terms of the number of words or bytes.The capacity of external or secondary storage, on the other hand, is measured in terms of megabytes and gigabytes.

NOTES

2. **Storage Cost:** Cost is another key factor that is of prime concern in a memory system . It is usually expressed per bit. It is obvious that lower costs are desirable. It is worth noting that with the increase in the access time for memories , the cost decreases.
3. **Access Time:** It is the time required to locate and retrieve the data from the storage unit. It is dependant on the physical characteristics and access mode used for that device.

   Primary storage units have faster access time as compared to the secondary storage units.
4. **Access Mode:** Memory comprises various locations. Access mode is the mode in which information is accessed from the memory. The user can access memory devices in any of the following ways:
   (a) ***Random-Access Memory (RAM):*** This refers to the mode in which any memory location can be accessed in any order in the same amount of time. Ferrite and semiconductor memories, which usually constitute the primary storage or main memory, are of this nature.
   (b) ***Sequential Access:*** Memories that can be accessed only in a predefined sequence are sequential access memories. Here, since sequencing through other locations precedes the arrival at a desired location, the access time varies according to the location. Information on a sequential device can be retrieved in the same sequence in which it was stored. Songs stored on a cassette, that can be accessed only one by one, is an example of sequential access. Typically, magnetic tapes are sequential access memory.
   (c) ***Direct Access:*** Sometimes, the information is neither accessed randomly nor in sequence, but something in between. In this type of access, a separate read/write head, exists for each track, and on a track, you can access the information serially. This semi-random mode of access exists in magnetic disks.
5. **Permanence of Storage:** If the storage unit can retain the data even after the power is turned off or interrupted, it is termed as **non-volatile storage**. However, if the data is lost once the power is turned off or interrupted, it is called **volatile storage**. It is obvious from these properties that the primary storage units of the computer systems are volatile, while the secondary storage units are non-volatile. A non-volatile storage is definitely more desirable and feasible for storage of large volumes of data.

1.4.1 Static and Dynamic RAM

The main memory is the central storage unit in a computer system. It is a relatively large and fast memory. It is used to store programs and data during computer operations. The principal technology used for the main memory is based on semiconductor integrated circuits. There are two possible modes in which the integrated circuit RAM chips are available. These modes are static and dynamic.

The Static RAM (SRAM) stores binary information using clocked sequential circuits. The stored information remains valid only as long as power is applied to the unit. On the other hand, Dynamic RAM (DRAM) stores binary information in the

NOTES

form of electric charges that are applied to capacitors inside the chip. The stored charge on the capacitors tends to discharge with time, and so, must be periodically recharged by refreshing the dynamic memory. The dynamic RAM offers larger storage capacity and reduced power consumption. Therefore, large memories use dynamic RAM, while static RAM is mainly used for specialized applications.

The different types of memory discussed above are both of the read/write type. What about a memory where only one of the operations is possible, e.g., if we allow only reading from the memory (cannot change the information in the memory)? The memory might have some major importance, like an important bit of the computer's operating system, which normally does not change can be stored in this kind of memory. Such a memory is called **ROM (Read-Only Memory)**.

1.4.2 Read-Only Memory (ROM)

Most of the memory in a general-purpose computer is made of RAM integrated circuit chips, but a portion of the memory may be constructed using ROM chips. Originally, RAM was used to refer to random-access memory, but now we use the term read/write memory to distinguish it from read-only memory (since ROM is also random access). RAM is used for storing the bulk of programs and data that are subject to change, while ROM is used to store programs that are permanently resident in the computer and do not change.

Among other things, the ROM portion of the main memory is used for storing an initial program called the **bootstrap loader**, whose function is to get the computer software operate when the power is turned on. Since RAM is volatile, its contents are destroyed when the power is turned off. The contents of ROM remain unchanged even after the power is turned off and on again.

Read-only memory can be manufacturer-programmed or user-programmed. When the data is burnt into the circuitry of the computer by the manufacturer, it is called **manufacturer-programmed ROM**. For example, a personal computer manufacturer may store the boot program permanently in the ROM chip of the computers manufactured by it. Such chips are supplied by the manufacturer and are not modifiable by users. This is an inflexible process and requires mass production. Thus, a new kind of ROM , known as Programmable Read-Only Memory (PROM), was designed. This is also non-volatile in nature. It can be written only once using some special equipment. The supplier or the customer can electrically perform the writing process in PROM.

In both ROM and PROM, you can perform write operations only once and you cannot change whatever you have written. However, what about the cases where you mostly read but also write a few times? Another type of memory chip called EPROM (Erasable Programmable Read-Only Memory) was developed to take care of such situations. EPROMs are typically used by R&D personnel who experiment by changing microprograms on the computer system to test their efficiency.

Further, EPROM chips are of two types: EEPROMs (Electrically EPROM) in which high voltage electric pulses are used to erase stored information, and UVEPROM (UltraViolet EPROM) in which stored information is erased by exposing the chip for a while to ultraviolet light.

NOTES

Figure 1.11 summarizes the various types of Random-Access Memory.

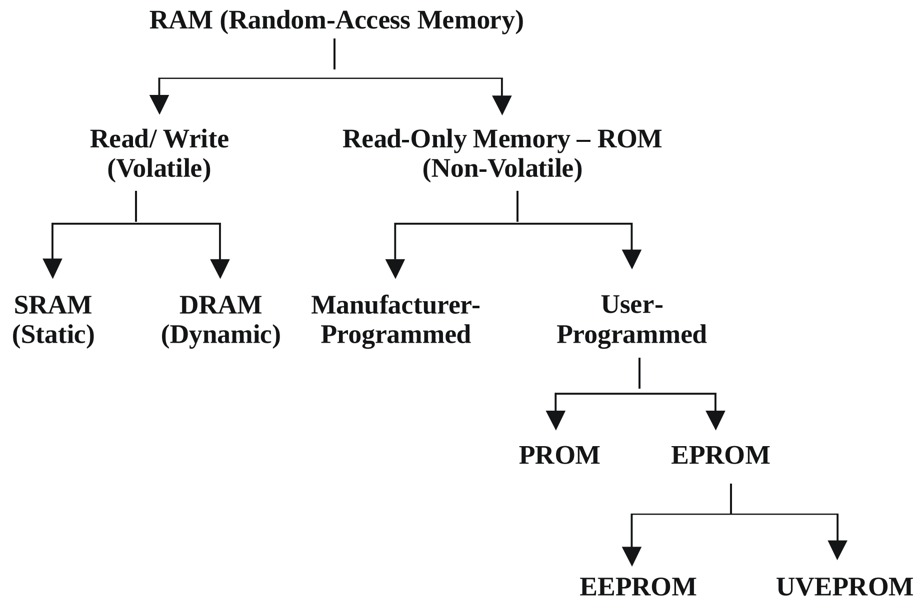


Fig. 1.11 *Types of Random-Access Memory*

1.4.3 Cache Memory

Cache memories are small, fast memories placed between the CPU and the main memory. They are faster than the main memory with access times closer to the speed of the CPU. Caches are fast, but very expensive. So, they are used only in small quantities. As an example, caches of size 64 KB, 128 KB are normally used in PC-386 and PC-486, which can have 1 to 8 MB of RAM or even more. Cache memories provide fast speed memory retrieval without compromising the size of the memory.

If the memory is so small, would it be advantageous to increase the overall speed of memory? This can be answered with the help of a phenomenon known as locality of reference. Let us examine what this means.

Locality of Reference: If we analyse a large number of typical programs,we would find that memory references at any given interval of time tend to be confined to a few localized areas in the memory. This phenomenon is called the property of locality of reference. This is true because most of the programs typically contain iterative loops (like 'for' or 'while' loops). During the execution of such programs, the same set of instructions (within the loop) are executed many times. The CPU repeatedly refers to the set of instructions in the memory that constitute the loop. Every time a specific subroutine is called, its set of instructions is fetched from the memory. Thus, loops and subroutines tend to localize the references to memory for fetching instructions.

Figure 1.12 explains the functioning of the cache memory.

Based on the locality of reference, we understand that the cache has a copy of certain portions of main memory. First, the memory read or write operation is checked with the cache, and in case of availability of desired data in the cache, it is used directly by the CPU. Otherwise, a block of words is read from main memory to cache and the word is then used by the CPU from cache.

NOTES

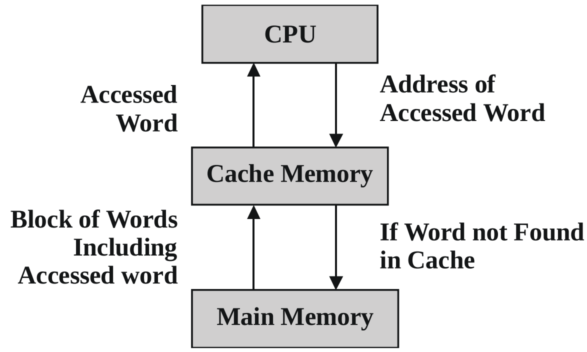


Fig. 1.12 *Functioning of the Cache Memory*

1.4.4 Secondary Storage Devices

As discussed earlier, RAM is a volatile memory with limited storage capacity. The cost of RAM is also relatively higher as compared to the secondary memory. Logic dictated that a relatively cheaper media, showing some sort of permanence of storage, be used. As a result, additional memory called external or auxiliary memory or secondary storage is used in most computers.

The magnetic medium was found to be long lasting and fairly inexpensive, and therefore, became an ideal choice for large storage requirements. The use of magnetic tapes and disks as storage media is very common nowadays. As optical technology is advancing, optical disks are turning out to be one of the major secondary storage devices.

Magnetic Storage Devices

Magnetic tapes are used for storing files of data that are sequentially accessed or not used very often and are stored off line. They are typically used as backup storage for archiving of data.

In case of magnetic tapes, a tape (plastic ribbon usually 1/2 inch or 1/4 inch wide and 50 to 2400 feet long) is wound on a spool and its other end is threaded manually on a take-up spool. The Beginning Of Tape (BOT) is indicated by a metal foil called a **marker**. When a write command is given, a block of data (records are usually grouped in blocks of two or more) is written on the tape. The next block is then written after a gap (called Inter Block Gap or IBG). A series of blocks are written in this manner. The End Of Tape (EOT) is indicated by an end-of-tape marker which is a metal foil stuck in the tape. After the data is written, the tape is rewound and kept ready for reading. Figure 1.13(a) shows the data organization on a magnetic tape. Figure 1.13(b) and 1.13(c) show magnetic tape reels and magnetic tape cartridges.

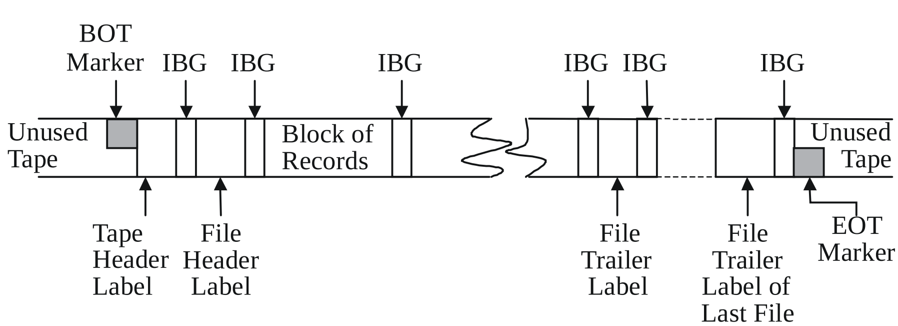


Fig. 1.13(a) *Data Organization on a Magnetic Tape*

NOTES

Fig. 1.13(b) *Magnetic Tape Reel*

Fig. 1.13(c) *Magnetic Tape Cartridge*

The tape is read sequentially, i.e., data can be read in the same order in which the data has been written. This implies that if the desired record is at the end of the tape, all the earlier records have to be read before it is reached. A typical example of a tape can be seen in a music cassette, where to listen to the fifth song one must listen to or traverse the earlier four songs. The access time of information stored on tape is, therefore, very high as compared to that stored on a magnetic disk.

The storage capacity of the tape depends on its data recording density and the length of the tape. The data recording density is the amount of data that can be stored or the number of bytes that can be stored per linear inch of tape. The data recording density is measured in BPI (Bytes Per Inch).

Thus,

Storage capacity of a tape = Data recording density × Length of tape

It is worth noting that the actual storage capacity for storing user data, is much less owing to the file header labels, file trailer labels, BOT and EOT markers, and the use of IBGs.

Some commonly used magnetic tapes are as follows:

- 1/2 inch tape reel.
- 1/2 inch tape cartridge.
- 1/4 inch streamer tape.
- 4 mm DAT (Digital Audio Tape) – typical capacity of 4 GB to 14 GB.

Magnetic Disks

Magnetic disks are direct-access medium, and so, are the most popular online secondary storage devices. Direct-access devices are also called **random-access devices** because information is literally available at random or in any order. There is direct access to any location on the device. Thus, approximately equal access time is required for each location. An example of this is a music CD where if you wish to listen to the fifth song, you can directly select the fifth track without having to fast forward the previous four.

A magnetic disk refers to a circular plate made of metal or plastic, and coated with magnetized material as shown in Figure 1.14. Often both sides of the disk are used. Data is recorded on the disk in the form of magnetized and non-magnetized spots (not visible to the naked eye) representing 1s and 0s.

NOTES

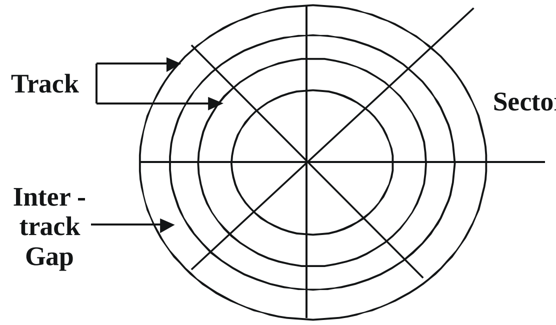


Fig. 1.14 Logical Layout of a Magnetic Disk

Organization of Data on a Magnetic Disk

Data is stored in concentric rings or tracks. Inter-track gaps are used to separate the adjacent tracks, so that, the interference of magnetic fields is minimized.Tracks are commonly divided into sections called **sectors**. In most systems, the minimum quantity of information that can be transferred is a sector. Usually, 8 or more sectors per track are found.

A track in a given sector near the circumference is longer than the track near the centre of the disk. If bits are recorded with equal density, some tracks would contain more bits than other tracks. To ensure that each sector can store equal amounts of data, some disks use variable recording density with higher density on the tracks near centre than on tracks near the circumference.

Multiple disks are usually stacked and used together to create disk storage systems having large capacities. In this case, multiple disks are fixed on a central shaft, one below the other, to form a disk pack. This is then mounted on a disk drive that has a motor to rotate this disk pack about its axis. The disk drive also has an access arm assembly with a separate read/write head for each surface of the disk pack. The access arms for all the disk surfaces move together. A disk system, is thus, addressed by the disk number, the disk surface, the sector number and the track within the sector.

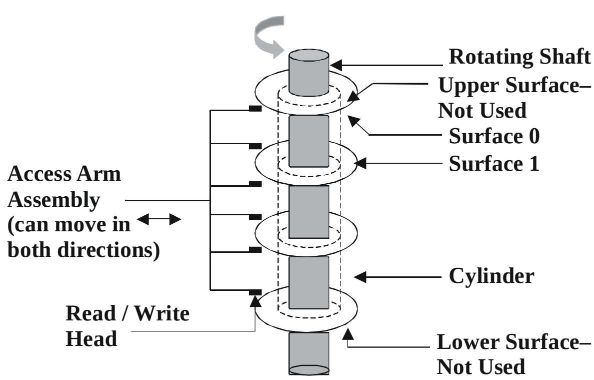


Fig. 1.15 A Disk Pack with Four Disks

Usually, the upper surface of the topmost disk and the lower surface of the bottommost disk are not used, since, these are prone to get scratched easily. For faster access of data from disk packs, a concept called cylinders is used. As can be seen in Figure 1.15, a set of corresponding tracks on all recording surfaces of the disk pack together form a cylinder. Thus, if there are 100 tracks on each disk surface, there are 100 cylinders in the disk pack.

NOTES

Cylinder-based organization provides faster data access. The related records of the file can be stored on the same cylinder (on multiple disks of a disk pack), and subsequently with one movement of the access arm, all records on, say, cylinder 5 (fifth track of every recording surface) can be simultaneously read.

The storage capacity of a disk system can be determined as follows:

Storage capacity = Number of recording surfaces × Number of tracks per surface × Number of sectors per track × Number of bytes per sector

Example: If a disk pack consists of 4 plates each having 2655 tracks having 125 sectors per track. Also, each sector can store 512 bytes then,

Storage capacity = $6 \times 2655 \times 125 \times 512$ = 1,01,95,20,000 bytes = 1×10^9 bytes approximately, or 1 GB or 1 Gigabyte

Note: We have 6 recording surfaces, since, there are 4 disk plates.

Access Time on Disks

As detailed earlier, the disk address is specified in terms of the surface number, track or cylinder number, and the sector number. The read/write heads need to be first positioned on the track on which the data are to be recorded or from which data needs to be read. Information is always written from the beginning of a sector and can be read only from the beginning of the desired track. Thus, the disk access time depends on the following factors:

- ***Seek Time:*** It is the time taken for positioning the head on a specific track. The seek time would vary depending on the position of the access arms at the time the read/write command was received, i.e., if the access arm was positioned on the outermost track and the current read operation required it to be positioned on the fifth track, then the time taken to position the access arm on track 5 is the seek time. It is obvious from this example that moving from the outermost to the innermost track or vice versa would result in the maximum seek time. The average seek time in most systems is 10–100 milliseconds.
- ***Latency Time:*** It is the time required by the desired sector to be positioned under the read/write head, i.e., the time required to spin the desired sector under the head. Latency is also known as rotational delay and varies depending on the distance of the desired sector from the initial position of the head on the specified track. The rotational speed of a disk is measured in rotations per minute (rpm) and can be anywhere between 300 to 7,200 rpm. On an average, latency equals half the time required for a rotation by the disk.

Besides these two times, the time taken to read a block of words (transfer rate) can also be taken into consideration. However, this is usually very small as compared to seek and latency times, and disk access time is generally considered to be a sum of seek time and latency times. Further, since access times to disk are large, a sizeable portion of the data is read in a single go. This is the reason, why disks are referenced in blocks.

Finding Data on a Disk

NOTES

The following are the ways for finding data on a disk:

Searching for Files and Folders: Windows Search, Wild Cards, Google Search

Over the last two decades, processors have typically doubled their capabilities in every eighteen months (Moore's Law). Processors have evolved continuously to become faster and capable of handling more complex resource intensive operations, and handling ever increasing volumes of data. This, speed and processing capability, has led to a stage where even simply entry-level desktops are capable of sophisticated applications, such as multimedia (covering graphical design, playing and editing movies and music, and creating presentations), database management and others. With the cost of storage media coming down constantly, this means users can today afford to store and share far more data than before. By some simple estimates, today's typical user handles at least fifty to hundred times the amount of data than just a decade ago.

In such a situation, where large amounts of data are stored on every computer, the need to be able to quickly find the piece that you are looking for becomes extremely critical. The users need a search facility which allows them to easily and quickly locate any file or folder, based upon any simple or complex search criteria. The search should be easy to understand and its options easily modifiable.

You can search files and folders in Windows on the local hard drive in several ways. The main parameters taken into consideration while searching for data are time and accuracy.

Windows Search

Microsoft Windows XP provides a default search function. Follow the instructions below for a basic overview of the search procedure.

1. Click **Start** and then click **Search**.

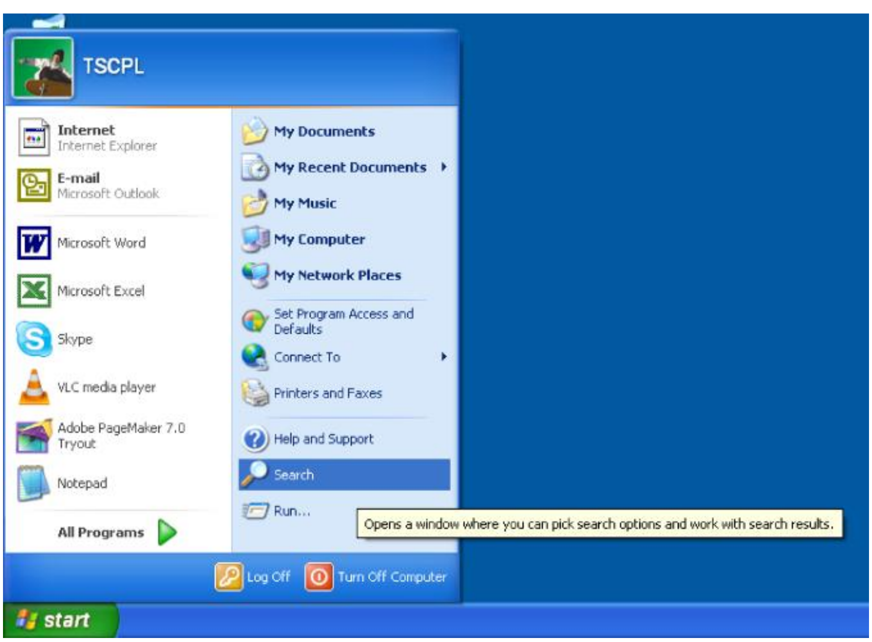


***Fig. 1.16** Search Option in the Start Menu*

NOTES

2. In the **Search Companion** dialog box, click **All files and folders**.

Fig. 1.17 *Search Companion Dialog Box*

3. Type part or the entire name of the file or folder, or type a word or phrase that is in the file.
4. In the **Look in** box, click the drive or drives, folder or network location that you want to search.

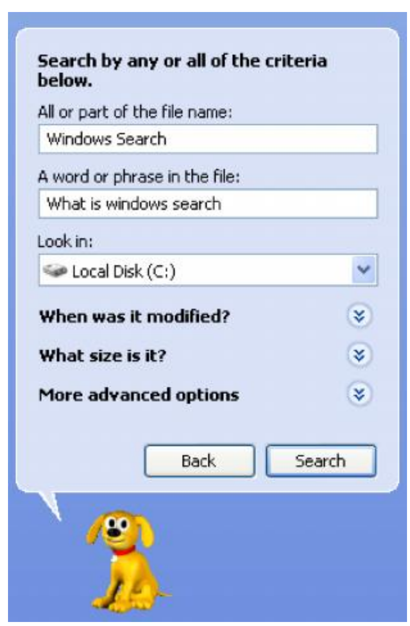


Fig. 1.18 *Selecting the Search Criteria*

NOTES

5. Choose one of the following options (see Figure 1.19):
   (a) Click **When was it modified** to look for files that were created or modified on or between specific dates.
   (b) Click **What size is it** to look for files that are a specific size.
   (c) Click **More advanced options** to specify additional search criteria, such as search in hidden folders as it's possible that it might have been saved in one.
6. Click **Search**.

Fig. 1.19 *Selecting the Search Criteria*

Wild Cards

Wildcards can act as placeholders for missing characters in filenames. Common wildcards are the asterisk (*) and the question mark (?). These can be valuable when searching for filenames in the operating system. They can often be used in search engines as well.

Let us assume that we have a file called 'Goodluck.doc' locally on my computer and we only remember the word 'good' as part of its filename. Now to search for this particular file we could type:

- 'good*.doc' to search for files that are of document type and their name starts with 'good'.
- '*luck.doc' to search for files that are of document type and their name ends with 'luck'.
- 'Good????.doc' to search for document files starting with 'good' and precisely four letters after that.
- 'Good*.*' to search for all files start with the word 'Good' and with any number of character or numbers after that in its filename.

NOTES

In addition to Windows' inbuilt search, you have the option of using one of the several search engines available. Search engines are applications designed to search for files or folders on your computer, network or over the entire Internet. They provide better file and folder indexing, quicker search results, and implement fuzzy logic (returning results that are not just exact matches to the text entered, but also, that may be related to the search subject).

Windows Desktop Search

This product, like its competitors, cannot only index files and office documents, but also offers real-time completion of names and good integration with the operating system. It even allows you to share search indexes across the entire network for finding files that are available on the hard drive of another computer. This is the same version that has been built into Windows Vista. Its biggest strength lies in searching files associated with Microsoft, such as Word, Excel, PowerPoint and HTML. Further features can be added to it by installing third party plug-ins available over the Internet. This helps in enhancing its search capabilities while keeping the system resources free. Windows desktop search can also share indexed information from other computers over the network and give search accessibility spanning over the entire network.

Alternate to file and file content searching, Windows desktop search can also search for emails in Outlook and Outlook Express, and their attachments. This is an optional support and, therefore, plug-ins are required to run this feature. It can also search multimedia files from metadata and display results in a detailed format.

Other Popular Search Engines

Google Desktop Search: One of the most efficient desktop search applications, Google desktop search can be used for indexing documents of any and all kinds of files, such as Word, Excel, PowerPoint and Acrobat and can also index Outlook emails along with Web surfing history. Common files, such as music, videos and pictures, are indexed according to their metadata. Google desktop search predetermines the selection of a video, audio or picture file according to its database of recognized codes. Files located on machines over the network may also be searched by using third-party plug-ins, if they are also running copies of Google desktop.

The biggest advantage of Google desktop search is that it doesn't take too much hard drive and memory space in its default state. Google desktop search's standard version supports widely used email applications, like Outlook, Netscape Mail, and Thunderbird, and by adding plug-ins it can also include emails in Gmail, Yahoo mail and many others.

Google desktop search's user interface (see Figure 1.20) is similar to the online Google search and it is integrated into the various Web browsers, such as the Internet Explorer, Netscape, Firefox and Mozilla, as a toolbar.

NOTES

Fig. 1.20 *Google Desktop Search*

Yahoo Desktop Search: It is powered by X1 Technologies which is a well-known application that can index over 400 different file formats that can be previewed in their native file format. Due to this feature, Yahoo desktop search has found wide acceptance among advanced users. It is a very fast search application because the search results are displayed as you type the keyword. It can index the entire enterprise network for file search, moreover, multiple ways of sorting search results are also available.

Search results can be copied on an Excel or Word file, and immediate actions, such as delete, copy, move and add to zip, can be performed on the results. Security access can be restricted to permissions given based on users or groups. The user interface of Yahoo desktop search is shown in Figure 1.21.

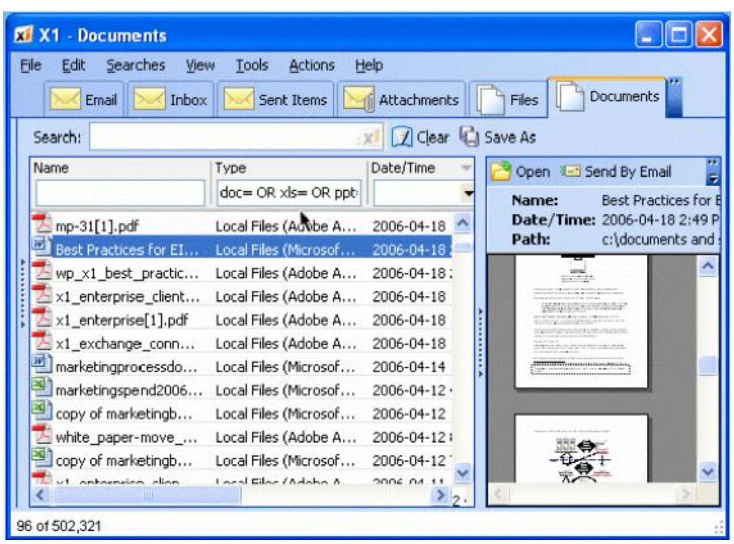


Fig. 1.21 *Yahoo Desktop Search*

ISYS Desktop Search: ISYS (not an acronym) was originally developed for Microsoft Disk Operating System (MS DOS) in the year 1988. Due to its long history of development and a well established reputation in the market, a lot of organizations and small businesses have adopted it. ISYS, unlike other desktop search applications (see Figure 1.22), is not freely available because it was originally developed for corporate users for power search. It has the capability of not only indexing the local hard drive but also the entire network and all the shared storages on it. ISYS can also search through 125 different types of file formats for keywords inside them. The other distinct

NOTES

options include advanced search queries and multilanguage support. Fuzzy Logic, which is a very important technique used in most search engines, is also available in the ISYS desktop search application. Fuzzy logic is primarily used to automatically adjust for typographical errors that occur when you scan in documents for searching.

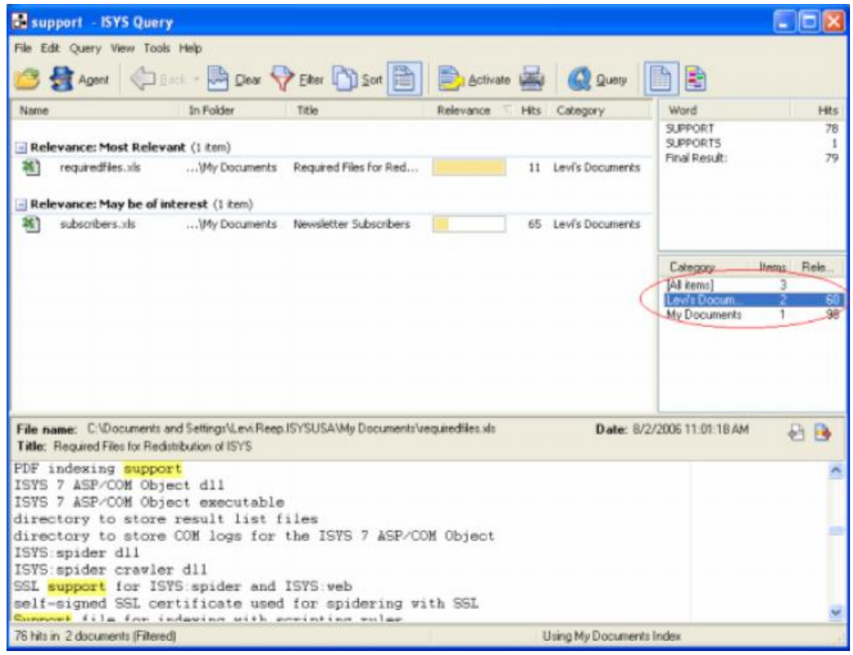


Fig. 1.22 *ISYS Desktop Search*

Some of the common features of all the four applications covered are that they index data on the hard drive and provide plug-ins for a feature rich experience. They are able to perform at various levels depending upon the user's requirements and the search criteria. This level of performance, for all the applications discussed above, is better in respect of speed and accuracy than the default search functionality available in most operating systems. However, a major concern of desktop search applications is that they do not encrypt indexed data for security. This makes the entire layout and content of the hard drive available without any authentication.

Disk Devices

A **disk drive** is a peripheral device used to store and collect information. It can be removable or fixed, high capacity or low capacity, fast or slow speed, and magnetic or optical.

Structurally, a drive is the object inside which a disk(s) is either permanently or temporarily stored. While a disk contains the media on which the data is stored, a drive contains the machinery and circuitry required for implementing the read/write operations on the disk.

The disk looks literally like a flat circular disk. The computer writes information onto the disk, where it is stored in the same form as it is stored on a cassette tape. Disks, as such, are just magnetically coated rolls or circular disks which are divided into sectors and tracks. The data is accordingly stored and numbered with respect to the track and sector number on the disk. Only, the structure of the media is different from one to another. Examples of removable disk drives are DVD, CD-ROM and floppy disk drive. The examples of non-removable disk drives include hard disk.

Check Your Progress

11. Define seek time.
12. Name three popular desktop search engines.

The method of accessing data could be sequential access (magnetic tape drives) or random access (HDD, DVD), where the read/write head can directly go to any location on the disk and perform action.

NOTES

Diskettes

Based on the size and packaging of the disks, they can be classified into two types: Floppy disks and hard disks. Further, disks that are permanently attached to the unit assembly and cannot be removed by the occasional user are called hard disks. A drive using removable disks is called a floppy disk drive.

Floppy Disks

The disks used with a floppy disk drive are small removable disks made of plastic and coated with magnetic recording material. There are two sizes which are used, with diameters of 5.25 and 3.5 inches.

- The 5.25 inch disk is a 5.25 inch diameter floppy disk. Earlier, such disks recorded data only on one side and were called Single Sided (SS) disks. Later both the surfaces were used for recording. They are called Double Sided (DS) disks. These are available in two capacities: Double Density (DD), and High Density (HD), where density refers to the number of bits that can be stored per square inch area.

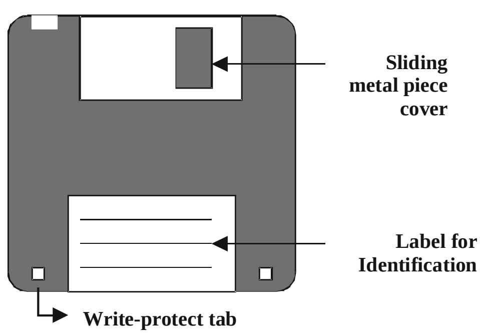


Fig. 1.23 *A 3.5 Inch Floppy Disk*

- The 3.5 inch floppy disk (see Figure 1.23) is a disk of 3½ inch diameter. These record data on both sides and are, therefore, double sided disks. These disks come in three different capacities: Double density, high density and very high density. These are smaller and can store more data than the 5.25 inch disks.

  The storage capacity for any disk can be calculated as:

  Storage capacity = Number of recording surfaces × Number of tracks per surface × Number of sectors per track × Number of bytes per sector

Thus, for a 3.5 inch high density disk which has 80 tracks, 18 sectors/track, and 512 bytes/sector, the disk storage capacity can be calculated as follows:

$$2 \times 80 \times 18 \times 512 = 14{,}74{,}560 \text{ bytes or } 1.4 \text{ MB (approximately)}$$

Table 1.5 provides the necessary details and associated storage capacities of various kinds of floppy disks.

NOTES

Floppy disks were extensively used in personal computers as a medium for distributing software to computer users. Nowadays, CDs or DVDs are used for that purpose. Now, floppy disks are obsolete.

Table 1.5 *Storage Capacities of Floppy Disks*

Size (diameter in inches)	No. of Recording Surfaces	No. of Tracks	No. of Sectors/ Tracks	No. of Bytes/ Sector	Storage Capacity (approx)
5.25	2	40	9	512	3,68,640 bytes or 360kB
5.25	2	80	15	512	12,28,800 bytes or 1.2 MB
3.5	2	40	18	512	7,37,280 bytes or 720 kB
3.5	2	80	18	512	14,74,560 bytes or 1.4 MB
3.5	2	80	36	512	29,49,120 or 2.8 MB

Hard Disks

Unlike floppy disks, hard disks are made up of rigid metal. The sizes for the disk platters range between 1 to 14 inches in diameter. Depending on the way they are packaged, hard disks can be categorized as disk packs or Winchester disks.

- ***Disk Packs*** consist of two or more hard disks mounted on a single central shaft as shown in Figure 1.24. Due to this, all disks in a disk pack rotate at the same speed. It consists of separate read/write heads for each surface (excluding the upper surface of the topmost disk platter and the lower surface of the bottommost disk platter, as mentioned earlier). Disk packs are removable in the sense that they can be removed and kept offline when not in use (typically stored away in plastic cases). They have to be mounted on the disk drive before they can be used. Thus, different disk packs can be mounted on the same disk drive at different instances, thereby, providing virtually unlimited (modular) storage capacity.

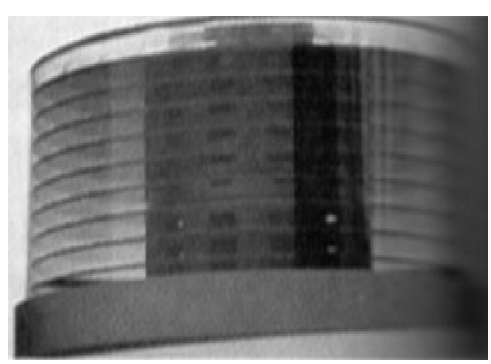


Fig. 1.24 *A Disk Pack*

- ***Winchester Disks*** also consist of two or more hard disk platters mounted on a single central shaft, but are of the fixed type as shown in Figure 1.25. The disk platters are sealed in a contamination-free container. Due to this fact, all the disk platters, including the upper surface of the topmost disk platter and the lower surface of the bottommost platter, are used for storing data. So, even though Winchester disks have limited storage capacity as opposed to disk packs, they can store larger amount of data as compared to the same number of disk platters.

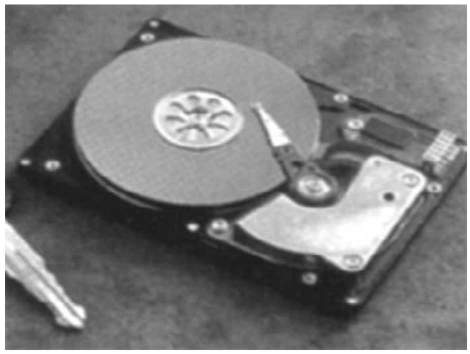


Fig. 1.25 *A Winchester Disk*

Zip disks consist of a single hard disk platter encased in a plastic cartridge (see Figure 1.26). Such a disk typically has a capacity of about 100 MB. The zip drive can further be fixed or portable. The fixed zip drive is permanently connected to the computer system while the portable ones can be carried around and connected to any computer system for the duration of its use. In both cases, however, the zip cartridge (the actual storage medium) is portable just like a floppy, though with a nearly 100 times larger storage capacity.

Fig. 1.26 *Zip Disks and Zip Drive*

Tape Drives

A **tape drive** is a data storage device that reads and writes data on a magnetic tape. Magnetic tape data storage is typically used for offline, archival data storage. A tape drive provides sequential access storage, unlike a disk drive, which provides random access storage. A tape drive physically winds tape between reels to read any particular piece of data. As a result, tape drives have very slow average seek times. For sequential access once the tape is positioned, however, tape drives can stream data very fast. Typically, it is a device, like a tape recorder, that reads data from and writes it onto a tape. Tape drives can range in capacity from a few megabytes to hundreds of gigabytes of uncompressed data.

Tape drives can be connected to a computer with SCSI (Small Computer System Interface), Fibre Channel, SATA (Serial Advanced Technology Attachment), USB

(Universal/Serial Bus), FireWire or other interfaces. Tape drives are used with autoloaders and tape libraries, which automatically load, unload and store multiple tapes. Hence, increasing the volume of data which can be stored without manual intervention.

NOTES

Computer Output Microfilm (COM)

Computer Output Microfilm or COM is a process for copying and printing data onto microfilm from electronic media found on personal, mini or mainframe computers. Characteristically, COM is a unique tape drive that is mostly used by organizations to store payroll, accounting, insurance, inventory or employee data. COM provides a sophisticated electronic records management tool that ensures proper retention of archival records, by decreasing any unauthorized destruction of records and increasing public access. It comprises a high-speed recorder, which transfers digital data onto a microfilm applying laser technology, and a processor which develops the microfilm once exposed to the light source.

A computer output microfilm device translates information normally held on magnetic tape into miniature images on a microfilm (also called microfiche–'fiche' pronounced as 'fish'). The device displays information as characters on a CRT screen and then using photographic methods, records the display onto the film. Drawings and images can also be displayed along with narrative text.

A special reader/printer can be subsequently used to view the processed film. The reader operates on a 'back projection' principle displaying one frame at a time on a translucent screen, typically about A4 size. The printer can then be used selectively to produce a hard copy of what is presented on screen.

Figure 1.27 identifies the various steps in COM production.

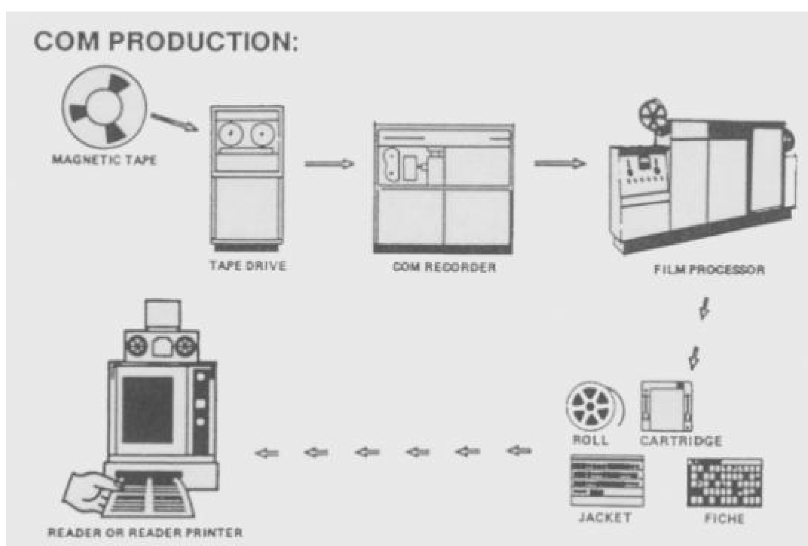


Fig. 1.27 *Steps in a COM Production*

A COM system provides an easy and compact way of recording and storing information, and subsequently retrieving desired pieces of information. It offers various advantages like reduction of paper, reduction in cost (since, it is cheaper than most electronic media), improved quality (COM technology provides superior image quality), and electronic record retention/archiving.

NOTES

COM is best suited for data requiring long-term storage. This is because microfilm is less volatile than magnetic media like disks and tapes. COM stores data in a very compact format. It is to be noted that up to 270 pages can be contained in a single 4 × 6 inch fiche.

Conversion of magnetic tapes to microfilms is cost-effective for closed files. However, in case of highly active data or data requiring regular updating, using microfilms may not be as efficient as retaining the information online. It is, therefore, useful for data that must be archived for long periods of time and referenced only occasionally, e.g., information that must be archived to comply with legal regulations, information maintained by insurance companies, banks, government agencies and various other organizations of this type.

1.4.5 Measuring Drive Performance

The performance of a disk is measured in terms of how fast it can read or write data. Over the years, there have been changes in disk drive interface, rotation speeds, number of heads and cylinders, and storage format, all of which have led to a decrease in data access time.

The various types of disks currently available in the market are as follows:

1. IDE – Integrated Drive Electronics
2. EIDE – Enhanced Integrated Drive Electronics
3. SCSI – Small Computer System Interface
4. SATA I & II – Serial Advanced Technology Attachment

There are two standard methods for accessing and writing data on a disk: Sequential access and random access.

Sequential access is when you read or write to the disk blocks in sequential or continuous order, that is, one block after another. Examples of where sequential access is used in computing or data storage would be the backing up of data onto tape drives or the process of writing data onto CDs and DVDs. Any storage medium based on magnetic tape, VHS, audio cassette and so on, are read and written by sequential access.

Random access, as the name suggests, is performed when the hard drive head needs to read/write data from/at various locations on the disk. In this case, the disk heads move rapidly from one place to another and the seek time to access data is increased because it involves mechanical operations. Most of the disk operations performed during routine computer work are random access. This is also the reason why random access time is more important while measuring disk performance than sequential access time. While data is written onto optical media sequentially, data on CDs and DVDs can be read randomly.

For random access, the average seek time and average latency time are added to come up with the total time it takes for the disk to read and write data on it. The average seek time is the time it takes to move the head arm from one position to another, and average latency time is the time it takes for the required data block to come under the head for the read/write operations. The average latency time depends on the Rotations Per Minute (RPM) of the disk, which is the speed at which the magnetic or optical disk rotates.

NOTES

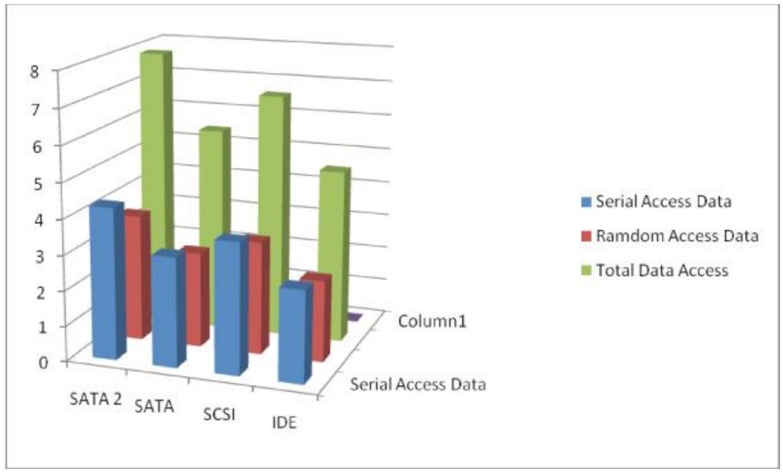


Fig. 1.28 Data accessed by Different Disk Drives

The bar chart in Figure 1.28 shows the relative amount of data accessed by each type of Disk Drive.

Optical and Solid State Storage Devices

The following sub sections discuss about optical and solid state storage devices:

Compact Disk (CD)

The Compact Disk (CD) was invented by an American inventor James Russell. A CD is a small, portable and easy to use device made of molded polymer. It is used to record, store, playback audio, video, text, graphics, and so on, in a digital form. It comes in the shape of circle. No other shape for CD is available in the market. The main feature of CD is that it is portable, and hence, it can be used in any type of portable and CD player devices. The other types of CD are popularly used as CD-ROM (Compact Disk-Read-Only Memory), CD-I (Compact Disk Interactive), CD-RW (Compact Disk-ReWritable), CD-RW XA (Compact Disk-ReWritable Extended Architecture), photo CD, video CD and so on. Figure 1.29 shows compact disk, which is frequently used to store data.

Fig. 1.29 Compact Disk

CD is an optical storage medium that reads the recorded data by the arrangement of optical beams. For example, the process of storing audio data (having large amount

NOTES

of data) in digital formats is known as audio encoding because one second of audio information can store one million bits of data. Therefore, CD is quite capable to store one millions of data and it takes area as tiny as pinhead. A CD has 4.75" diameter and made up of polycarbonate plastic disk. It can contain approximately 74 minutes of audio information. The information is encoded into the CD in terms of lands and pits, and is represented by binary highs and lows that are read as laser 'reads'. The future of this disk is that it would be common medium of exchanging the data for audio, video and other applications. Figure 1.30 shows the structure of compact disk, which is generally connected with circuit board.

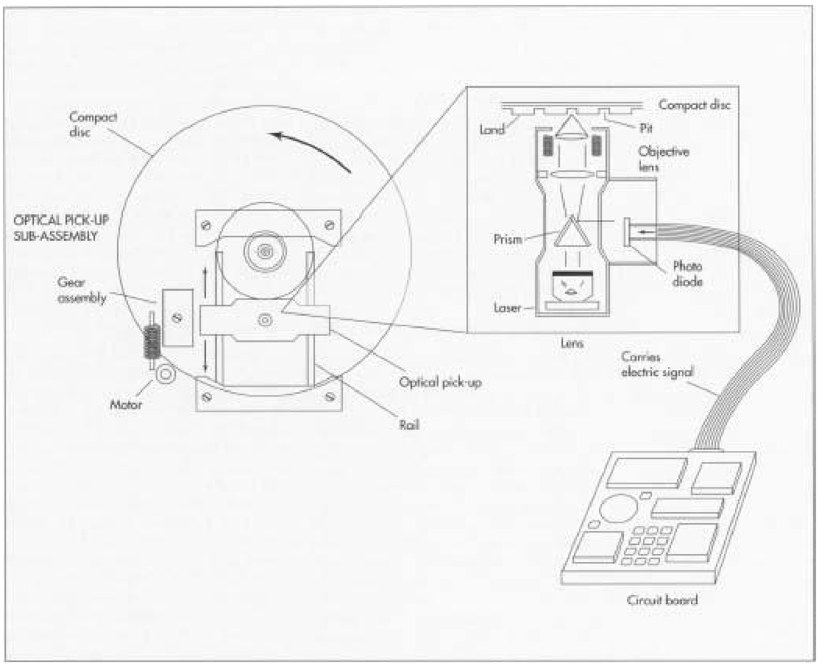


Fig. 1.30 Structure of Compact Disk

Compact Disk Interactive (CD-I)

CD-I represents the name of interactive multimedia CD player, which was developed by Royal Philips Electronics N.V. It is mainly useful for creating movies, game and videos. A CD-I movie disk is also known as video CD, holds approximately 70 minutes Video Home System (VHS) quality video. In 1990, Sony company launched a portable CD-I device with the 4" LCD monitor. Figure 1.31 shows compact disk interactive, which was designed for real-time animation, video and sound.

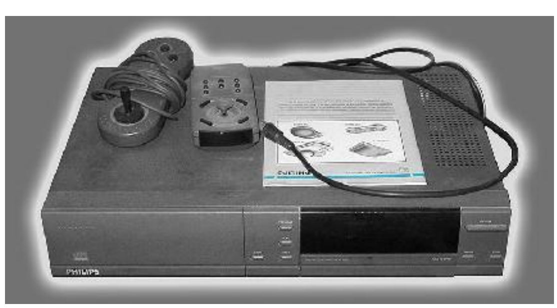


Fig. 1.31 Compact Disk Interactive