CHAPTER 1
INTRODUCTION TO COMPUTER

1.1 INTRODUCTION

Nowadays, computers are an integral part of our lives. They are used for the reservation of tickets for airplanes and railways, payment of telephone and electricity bills, deposit and withdrawal of money from banks, processing of business data, forecasting of weather conditions, diagnosis of diseases, searching for information on the Internet, etc.
Computers are also used extensively in schools, universities, organizations, music industry, movie industry, scientific research, law firms, fashion industry, etc.

The term computer is derived from the word *compute*. The word *compute* means *to calculate*. A *computer* is an electronic machine that accepts data from the user, processes the data by performing calculations and operations on it, and generates the desired output results. Computer performs both simple and complex operations, with speed and accuracy.

This chapter discusses the history and evolution of computer, the concept of input-process-output and the characteristics of computer. This chapter also discusses the classification of digital computers based on their size and type, and the application of computer in different domain areas.

1.2 DIGITAL AND ANALOG COMPUTERS

A *digital computer* uses distinct values to represent the data internally. All information are represented using the digits 0s and 1s. The computers that we use at our homes and offices are digital computers.

Analog computer is another kind of a computer that represents data as variable across a continuous range of values. The earliest computers were analog computers. Analog computers are used for measuring of parameters that vary continuously in real time, such as temperature, pressure and voltage. Analog computers may be more flexible but generally less precise than digital computers. Slide rule is an example of an analog computer.

This book deals only with the *digital computer* and uses the term *computer* for them.

1.3 CHARACTERISTICS OF COMPUTER

Speed, accuracy, diligence, storage capability and versatility are some of the key characteristics of a computer. A brief overview of these characteristics are—

- ***Speed*** The computer can process data very fast, at the rate of millions of instructions per second. Some calculations that would have taken hours and days to complete otherwise, can be completed in a few seconds using the computer. For example, calculation and generation of salary slips of thousands of employees of an organization, weather forecasting that requires analysis of a large amount of data related to temperature, pressure and humidity of various places, etc.
- ***Accuracy*** Computer provides a high degree of accuracy. For example, the computer can

accurately give the result of division of any two numbers up to 10 decimal places.

- ***Diligence*** When used for a longer period of time, the computer does not get tired or fatigued. It can perform long and complex calculations with the same speed and accuracy from the start till the end.
- ***Storage Capability*** Large volumes of data and information can be stored in the computer and also retrieved whenever required. A limited amount of data can be stored, temporarily, in the primary memory. Secondary storage devices like floppy disk and compact disk can store a large amount of data permanently.
- ***Versatility*** Computer is versatile in nature. It can perform different types of tasks with the same ease. At one moment you can use the computer to prepare a letter document and in the next moment you may play music or print a document.

Computers have several limitations too. Computer can only perform tasks that it has been programmed to do. Computer cannot do any work without instructions from the user. It executes instructions as specified by the user and does not take its own decisions.

1.4 HISTORY OF COMPUTER

Until the development of the first generation computers based on vacuum tubes, there had been several developments in the computing technology related to the mechanical computing devices. The key developments that took place till the first computer was developed are as follows—

- ***Calculating Machines*** ABACUS was the first mechanical calculating device for counting of large numbers. The word ABACUS means calculating board. It consists of bars in horizontal positions on which sets of beads are inserted. The horizontal bars have 10 beads each, representing units, tens, hundreds, etc. An abacus is shown in Figure 1.1

Figure 1.1 Abacus

- ***Napier's Bones*** was a mechanical device built for the purpose of multiplication in 1517 ad. by an English mathematician John Napier.
- ***Slide Rule*** was developed by an English mathematician Edmund Gunter in the 15th century. Using the slide rule, one could perform operations like addition, subtraction, multiplication and division. It was used extensively till late 1970s. Figure 1.2 shows a slide rule.

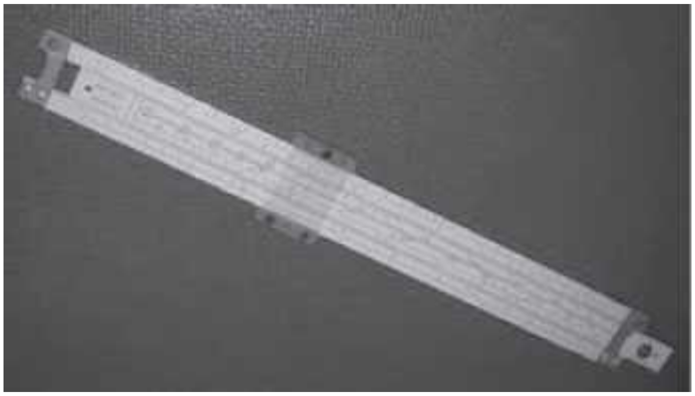

Figure 1.2 Slide rule

- ***Pascal's Adding and Subtraction Machine*** was developed by Blaise Pascal. It could add and subtract. The machine consisted of wheels, gears and cylinders.
- ***Leibniz's Multiplication and Dividing Machine*** was a mechanical device that could both multiply and divide. The German philosopher and mathematician Gottfried Leibniz built it around 1573.
- ***Punch Card System*** was developed by Jacquard to control the power loom in 1801. He invented the punched card reader that could recognize the presence of hole in the punched card as binary one and the absence of the hole as binary zero. The Os and 1s are the basis of the modern digital computer. A punched card is shown in Figure 1.3.

Figure 1.3 Punched card

- ***Babbage's Analytical Engine*** An English man Charles Babbage built a mechanical machine to do complex mathematical calculations, in the year 1823. The machine was called as difference engine. Later, Charles Babbage and Lady Ada Lovelace developed a general-purpose calculating machine, the analytical engine. Charles Babbage is also called the father of computer.
- ***Hollerith's Punched Card Tabulating Machine*** was invented by Herman Hollerith. The machine could read the information from a punched card and process it electronically.

The developments discussed above and several others not discussed here, resulted in the development of the first computer in the 1940s.

1.5 GENERATIONS OF COMPUTER

The computer has evolved from a large—sized simple calculating machine to a smaller but much more powerful machine. The evolution of computer to the current state is defined in terms of the generations of computer. Each generation of computer is designed based on a new technological development, resulting in better, cheaper and smaller computers that are more powerful, faster and efficient than their predecessors. Currently, there are five generations of computer. In the following subsections, we will discuss the generations of computer in terms of—

1. the technology used by them (hardware and software),
2. computing characteristics (speed, i.e., number of instructions executed per second),
3. physical appearance, and
4. their applications.

1.5.1 First Generation (1940 to 1955): Using Vacuum Tubes

- ***Hardware Technology*** The first generation of computers used vacuum tubes (Figure 1.4) for circuitry and magnetic drums for memory. The input to the computer was through punched cards and paper tapes. The output was displayed as printouts.

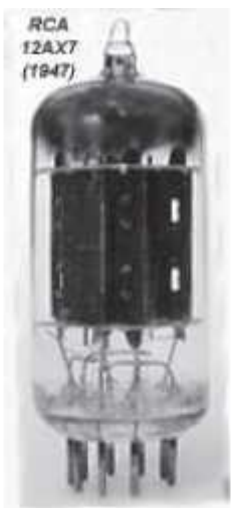

Figure 1.4 Vacuum tube

- ***Software Technology*** The instructions were written in machine language. Machine language uses 0s and 1s for coding of the instructions. The first generation computers could solve one problem at a time.
- ***Computing Characteristics*** The computation time was in milliseconds.
- ***Physical Appearance*** These computers were enormous in size and required a large room for installation.
- ***Application*** They were used for scientific applications as they were the fastest computing device of their time.
- ***Examples*** UNIVersal Automatic Computer (UNIVAC), Electronic Numerical Integrator And Calculator (ENIAC), and Electronic Discrete Variable Automatic Computer (EDVAC).

The first generation computers used a large number of vacuum tubes and thus generated a lot of heat. They consumed a great deal of electricity and were expensive to operate. The machines were prone to frequent malfunctioning and required constant maintenance. Since first generation computers used machine language, they were difficult to program.

1.5.2 Second Generation (1955 to 1953): Using Transistors

- ***Hardware Technology*** Transistors (Figure 1.5) replaced the vacuum tubes of the first generation of computers. Transistors allowed computers to become smaller, faster, cheaper, energy efficient and reliable. The second generation computers used *magnetic core technology* for primary memory. They used magnetic tapes and magnetic disks for secondary storage. The input was still through punched cards and the output using printouts. They used the concept of a stored program, where instructions were stored in the memory of computer.

Figure 1.5 Transistors

- ***Software Technology*** The instructions were written using the *assembly language*. Assembly language uses mnemonics like ADD for addition and SUB for subtraction for coding of the instructions. It is easier to write instructions in assembly language, as compared to writing instructions in machine language. High-level programming languages, such as early versions of COBOL and FORTRAN were also developed during this period.
- ***Computing Characteristics*** The computation time was in microseconds.
- ***Physical Appearance*** Transistors are smaller in size compared to vacuum tubes, thus, the size of the computer was also reduced.
- ***Application*** The cost of commercial production of these computers was very high, though less than the first generation computers. The transistors had to be assembled manually in second generation computers.
- ***Examples*** PDP-8, IBM 1401 and CDC 1504.

Second generation computers generated a lot of heat but much less than the first generation computers. They required less maintenance than the first generation computers.

1.5.3 Third Generation (1954 to 1971): Using Integrated Circuits

- ***Hardware Technology*** The third generation computers used the *Integrated Circuit (IC)* chips. Figure 1.5 shows IC chips. In an IC chip, multiple transistors are placed on a silicon chip. Silicon is a type of semiconductor. The use of IC chip increased the speed and the efficiency of computer, manifold. The keyboard and monitor were used to interact with the third generation computer, instead of the punched card and printouts.

Figure 1.5 IC chips

- ***Software Technology*** The keyboard and the monitor were interfaced through the *operating system*. Operating system allowed different applications to run at the same time. *High-level languages* were used extensively for programming, instead of machine language and assembly language.
- ***Computing Characteristics*** The computation time was in nanoseconds.
- ***Physical Appearance*** The size of these computers was quite small compared to the second generation computers.
- ***Application*** Computers became accessible to mass audience. Computers were produced commercially, and were smaller and cheaper than their predecessors.
- ***Examples*** IBM 370, PDP 11.

The third generation computers used less power and generated less heat than the second generation computers. The cost of the computer reduced significantly, as individual components of the computer were not required to be assembled manually. The maintenance cost of the computers was also less compared to their predecessors.

1.5.4 Fourth Generation (1971 to present): Using Microprocessors

- ***Hardware Technology*** They use the *Large Scale Integration (LSI)* and the *Very Large Scale Integration (VLSI)* technology. Thousands of transistors are integrated on a small silicon chip using LSI technology. VLSI allows hundreds of thousands of components to be integrated in a small chip. This era is marked by the development of microprocessor. *Microprocessor* is a chip containing millions of transistors and components, and,

designed using LSI and VLSI technology. A microprocessor chip is shown in Figure 1.7. This generation of computers gave rise to Personal Computer (PC). Semiconductor memory replaced the earlier magnetic core memory, resulting in fast random access to memory. Secondary storage device like magnetic disks became smaller in physical size and larger in capacity. The *linking of computers* is another key development of this era. The computers were linked to form networks that led to the emergence of the Internet.
This generation also saw the development of pointing devices like mouse, and handheld devices.

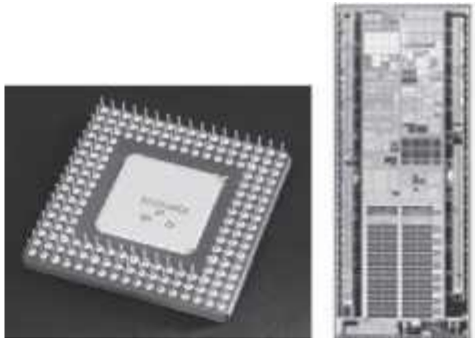

Figure 1.7 Microprocessors

- ***Software Technology*** Several new operating systems like the MS-DOS and MS- Windows developed during this time. This generation of computers supported *Graphical User Interface (GUI)*. GUI is a user-friendly interface that allows user to interact with the computer via menus and icons. High-level programming languages are used for the writing of programs.
- ***Computing Characteristics*** The computation time is in picoseconds.
- ***Physical Appearance*** They are smaller than the computers of the previous generation. Some can even fit into the palm of the hand.
- ***Application*** They became widely available for commercial purposes. Personal computers became available to the home user.
- ***Examples*** The Intel 4004 chip was the first microprocessor. The components of the computer like Central Processing Unit (CPU) and memory were located on a single chip. In 1981, IBM introduced the first computer for home use. In 1984, Apple introduced the Macintosh.

The microprocessor has resulted in the fourth generation computers being smaller and cheaper than their predecessors. The fourth generation computers are also portable and more reliable. They generate much lesser heat and require less maintenance compared to their predecessors. GUI and pointing devices facilitate easy use and learning on the computer. Networking has resulted in resource sharing and communication among different computers.

1.5.5 Fifth Generation (Present and Next): Using Artificial Intelligence

The goal of fifth generation computing is to develop computers that are capable of learning and self-organization. The fifth generation computers use *Super Large Scale Integrated (SLSI)* chips that are able to store millions of components on a single chip. These computers have large memory requirements.

This generation of computers uses *parallel processing* that allows several instructions to be executed in parallel, instead of serial execution. Parallel processing results in faster processing speed. The Intel dualcore microprocessor uses parallel processing.

The fifth generation computers are based on *Artificial Intelligence (AI)*. They try to simulate the human way of thinking and reasoning. Artificial Intelligence includes areas like Expert System (ES), Natural Language Processing (NLP), speech recognition, voice recognition, robotics, etc.

1.6 CLASSIFICATION OF COMPUTER

The digital computers that are available nowadays vary in their sizes and types. The computers are broadly classified into four categories (Figure 1.8) based on their size and type—(1) Microcomputers, (2) Minicomputers, (3) Mainframe computers, and (4) Supercomputer.

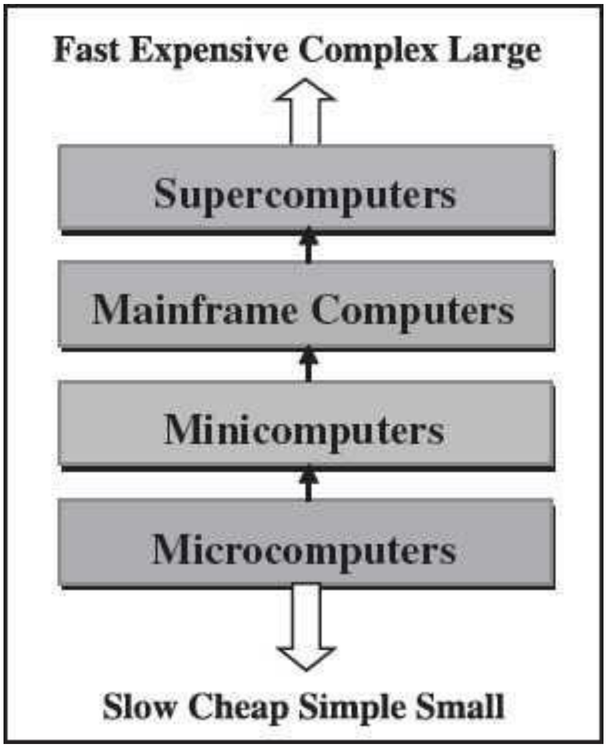

Figure 1.8 Classification of computers based on size and type

1.6.1 Microcomputers

Microcomputers are small, low-cost and single-user digital computer. They consist of CPU, input unit, output unit, storage unit and the software. Although microcomputers are stand-alone machines, they can be connected together to create a network of computers that can serve more than one user. IBM PC based on Pentium microprocessor and Apple Macintosh are some examples of microcomputers. Microcomputers include desktop computers, notebook computers or laptop, tablet computer, handheld computer, smart phones and netbook, as shown in Figure 1.9.

Figure 1.9 Microcomputers

- ***Desktop Computer or Personal Computer (PC)*** is the most common type of microcomputer. It is a stand-alone machine that can be placed on the desk. Externally, it consists of three units—keyboard, monitor, and a system unit containing the CPU, memory, hard disk drive, etc. It is not very expensive and is suited to the needs of a single user at home, small business units, and organizations. Apple, Microsoft, HP, Dell and Lenovo are some of the PC manufacturers.
- ***Notebook Computers or Laptop*** resemble a notebook. They are portable and have all the features of a desktop computer. The advantage of the laptop is that it is small in size (can be put inside a briefcase), can be carried anywhere, has a battery backup and has all the functionality of the desktop. Laptops can be placed on the lap while working (hence the name). Laptops are costlier than the desktop machines.
- ***Netbook*** These are smaller notebooks optimized for low weight and low cost, and are designed for accessing web-based applications. Starting with the earliest netbook in late 2007, they have gained significant popularity now. Netbooks deliver the performance needed to enjoy popular activities like streaming videos or music, emailing, Web surfing or instant messaging. The word *netbook* was created as a blend of Internet and note*book*.
- ***Tablet Computer*** has features of the notebook computer but it can accept input from a stylus or a pen instead of the keyboard or mouse. It is a portable computer. Tablet computer are the new kind of PCs.
- ***Handheld Computer or Personal Digital Assistant (PDA)*** is a small computer that can be held on the top of the palm. It is small in size. Instead of the keyboard, PDA uses a pen or a stylus for input. PDAs do not have a disk drive. They have a limited memory and are less powerful. PDAs can be connected to the Internet via a wireless connection. Casio and Apple are some of the manufacturers of PDA. Over the last few years, PDAs have merged into mobile phones to create smart phones.
- ***Smart Phones*** are cellular phones that function both as a phone and as a small PC. They may use a stylus or a pen, or may have a small keyboard. They can be connected to the Internet wirelessly. They are used to access the electronic-mail, download music, play

games, etc. Blackberry, Apple, HTC, Nokia and LG are some of the manufacturers of smart phones.

1.6.2 Minicomputers

Minicomputers (Figure 1.10) are digital computers, generally used in multi-user systems. They have high processing speed and high storage capacity than the microcomputers. Minicomputers can support 4–200 users simultaneously. The users can access the minicomputer through their PCs or terminal. They are used for real-time applications in industries, research centers, etc. PDP 11, IBM (8000 series) are some of the widely used minicomputers.

Figure 1.10 Minicomputer

1.6.3 Mainframe Computers

Mainframe computers (Figure 1.11) are multi-user, multi-programming and high performance computers. They operate at a very high speed, have very large storage capacity and can handle the workload of many users. Mainframe computers are large and powerful systems generally used in centralized databases. The user accesses the mainframe computer via a terminal that may be a dumb terminal, an intelligent terminal or a PC. A *dumb terminal* cannot store data or do processing of its own. It has the input and output device only. An *intelligent terminal* has the input and output device, can do processing, but, cannot store data of its own. The dumb and the intelligent terminal use the processing power and the storage facility of the mainframe computer. Mainframe computers are used in organizations like banks or companies, where many people

require frequent access to the same data. Some examples of mainframes are CDC 5500 and IBM ES000 series.

Figure 1.11 Mainframe computer

1.6.4 Supercomputers

Supercomputers (Figure 1.12) are the fastest and the most expensive machines. They have high processing speed compared to other computers. The speed of a supercomputer is generally measured in FLOPS (FLoating point Operations Per Second). Some of the faster supercomputers can perform trillions of calculations per second. Supercomputers are built by interconnecting thousands of processors that can work in parallel.

Supercomputers are used for highly calculation-intensive tasks, such as, weather forecasting, climate research (global warming), molecular research, biological research, nuclear research and aircraft design. They are also used in major universities, military agencies and scientific research laboratories. Some examples of supercomputers are IBM Roadrunner, IBM Blue gene and Intel ASCI red. PARAM is a series of supercomputer assembled in India by C-DAC (Center for Development of Advanced Computing), in Pune. PARAM Padma is the latest machine in this series. The peak computing power of PARAM Padma is 1 Tera FLOP (TFLOP).

Figure 1.12 Supercomputer

1.7 THE COMPUTER SYSTEM

Computer is an electronic device that accepts data as input, processes the input data by performing mathematical and logical operations on it, and gives the desired output. The computer system consists of four parts•(1) Hardware, (2) Software, (3) Data, and (4) Users. The parts of computer system are shown in Figure 1.13.

Hardware consists of the mechanical parts that make up the computer as a machine. The hardware consists of physical devices of the computer. The devices are required for input, output, storage and processing of the data. Keyboard, monitor, hard disk drive, floppy disk drive, printer, processor and motherboard are some of the hardware devices.

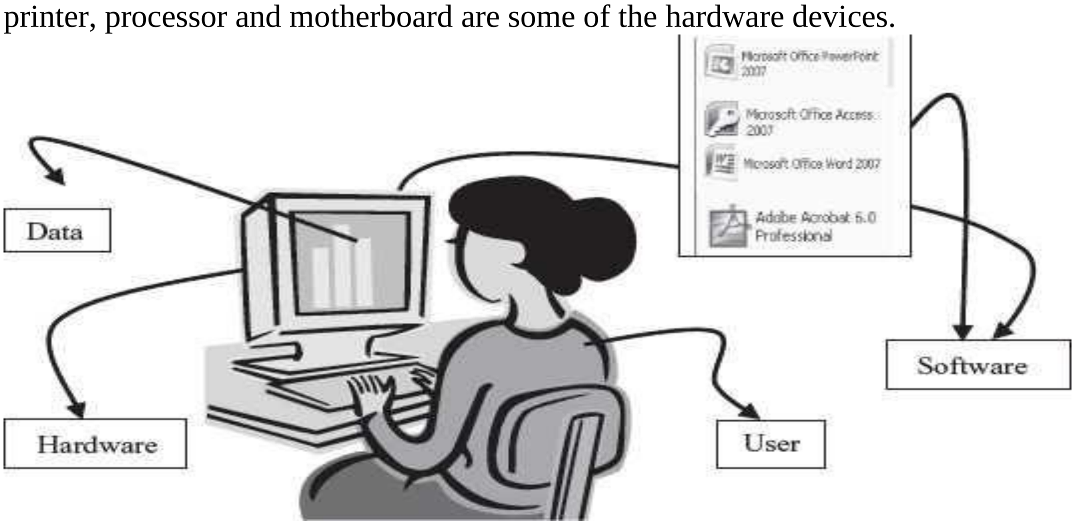

Figure 1.13 Parts of computer system

Software is a set of instructions that tells the computer about the tasks to be performed and how these tasks are to be performed. *Program* is a set of instructions, written in a language understood by the computer, to perform a specific task. A set of programs and documents are collectively called software. The hardware of the computer system cannot perform any task on its own. The hardware needs to be instructed about the task to be performed. Software instructs the computer about the task to be performed. The hardware carries out these tasks. Different software can be loaded on the same hardware to perform different kinds of tasks.

Data are isolated values or raw facts, which by themselves have no much significance. For example, the data like 29, *January*, and *1994* just represent values. The data is provided as input to the computer, which is processed to generate some meaningful information. For example, 29, January and 1994 are processed by the computer to give the date of birth of a person.

Users are people who write computer programs or interact with the computer. They are also known as *skinware, liveware, humanware or peopleware*. Programmers, data entry operators, system analyst and computer hardware engineers fall into this category.

1.7.1 The Input-Process-Output Concept

A computer is an electronic device that (1) accepts data, (2) processes data, (3) generates output,

and (4) stores data. The concept of generating output information from the input 4 data is also referred to as *input-process-output* concept.The input-process-output concept of the computer is explained as follows—

- ***Input*** The computer accepts input data from the user via an input device like keyboard. The input data can be characters, word, text, sound, images, document, etc.
- ***Process*** The computer processes the input data. For this, it performs some actions on the data by using the instructions or program given by the user of the data. The action could be an arithmetic or logic calculation, editing, modifying a document, etc. During processing, the data, instructions and the output are stored temporarily in the computer's main memory.
- ***Output*** The output is the result generated after the processing of data. The output may be in the form of text, sound, image, document, etc. The computer may display the output on a monitor, send output to the printer for printing, play the output, etc.
- ***Storage*** The input data, instructions and output are stored permanently in the secondary storage devices like disk or tape. The stored data can be retrieved later, whenever needed.

1.7.2 Components of Computer Hardware

The computer system hardware comprises of three main components —

1. Input/Output (I/O) Unit,
2. Central Processing Unit (CPU), and
3. Memory Unit.

The I/O unit consists of the input unit and the output unit. CPU performs calculations and processing on the input data, to generate the output. The memory unit is used to store the data, the instructions and the output information. Figure 1.14 illustrates the typical interaction among the different components of the computer.

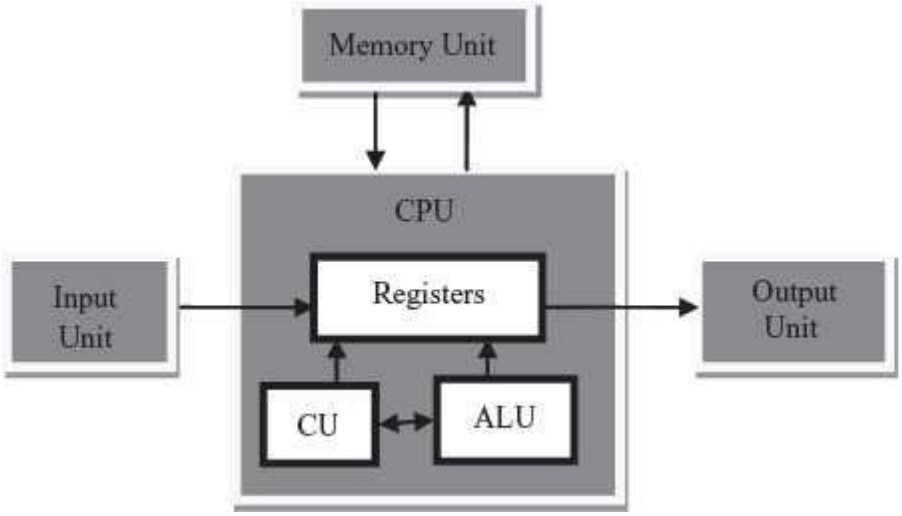

Figure 1.14 The computer system interaction

- ***Input/Output Unit*** The user interacts with the computer via the I/O unit. The Input unit accepts data from the user and the Output unit provides the processed data i.e. the information to the user. The Input unit converts the data that it accepts from the user, into a form that is understandable by the computer. Similarly, the Output unit provides the output in a form that is understandable by the user. The input is provided to the computer using input devices like keyboard, trackball and mouse. Some of the commonly used output devices are monitor and printer.
- ***Central Processing Unit*** CPU controls, coordinates and supervises the operations of the computer. It is responsible for processing of the input data. CPU consists of Arithmetic Logic Unit (ALU) and Control Unit (CU).
    - ALU performs all the arithmetic and logic operations on the input data.
    - CU controls the overall operations of the computer i.e. it checks the sequence of execution of instructions, and, controls and coordinates the overall functioning of the units of computer.

  Additionally, CPU also has a set of *registers* for temporary storage of data, instructions, addresses and intermediate results of calculation.

- ***Memory Unit*** Memory unit stores the data, instructions, intermediate results and output, *temporarily*, during the processing of data. This memory is also called the *main memory or primary memory* of the computer. The input data that is to be processed is brought into the main memory before processing. The instructions required for processing of data and any intermediate results are also stored in the main memory. The output is stored in memory before being transferred to the output device. CPU can work with the information stored in the main memory. Another kind of storage unit is also referred to as the *secondary memory* of the computer. The data, the programs and the output are stored *permanently* in the storage unit of the computer. Magnetic disks, optical disks and magnetic tapes are examples of secondary memory.

1.8 APPLICATION OF COMPUTERS

Computers have proliferated into various areas of our lives. For a user, computer is a tool that provides the desired information, whenever needed. You may use computer to get information about the reservation of tickets (railways, airplanes and cinema halls), books in a library, medical history of a person, a place in a map, or the dictionary meaning of a word. The information may be presented to you in the form of text, images, video clips, etc.

Figure 1.15 shows some of the applications of computer. Some of the application areas of the computer are listed below—

- ***Education*** Computers are extensively used, as a tool and as an aid, for imparting education. Educators use computers to prepare notes and presentations of their lectures. Computers are used to develop computer-based training packages, to provide distance education using the e-learning software, and to conduct online examinations. Researchers use computers to get easy access to conference and journal details and to get global access to the research material.
- ***Entertainment*** Computers have had a major impact on the entertainment industry. The user can download and view movies, play games, chat, book tickets for cinema halls, use multimedia for making movies, incorporate visual and sound effects using computers, etc. The users can also listen to music, download and share music, create music using computers, etc.
- ***Sports*** A computer can be used to watch a game, view the scores, improve the game, play games (like chess, etc.) and create games. They are also used for the purposes of training players.
- ***Advertising*** Computer is a powerful advertising media. Advertisement can be displayed on different websites, electronic-mails can be sent and reviews of a product by different customers can be posted. Computers are also used to create an advertisement using the visual and the sound effects. For the advertisers, computer is a medium via which the advertisements can be viewed globally. Web advertising has become a significant factor in the marketing plans of almost all companies. In fact, the business model of Google is mainly dependent on web advertising for generating revenues.

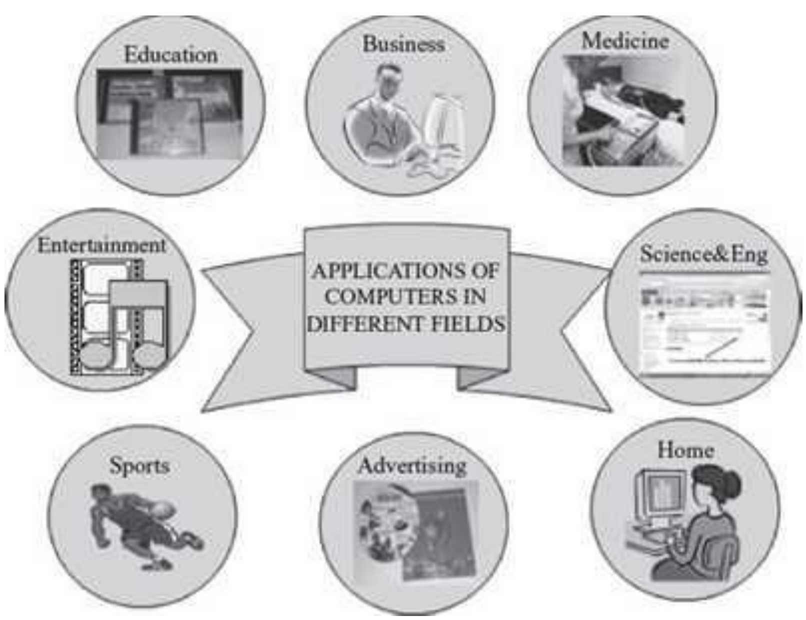

Figure 1.15 Applications of computer

- ***Medicine*** Medical researchers and practitioners use computers to access information about the advances in medical research or to take opinion of doctors globally. The medical history of patients is stored in the computers. Computers are also an integral part of various kinds of sophisticated medical equipments like ultrasound machine, CAT scan machine, MRI scan machine, etc. Computers also provide assistance to the medical surgeons during critical surgery operations like laparoscopic operations, etc.
- ***Science and Engineering*** Scientists and engineers use computers for performing complex scientific calculations, for designing and making drawings (CAD/CAM applications) and also for simulating and testing the designs. Computers are used for storing the complex data, performing complex calculations and for visualizing 3–dimensional objects. Complex scientific applications like the launch of the rockets, space exploration, etc., are not possible without the computers.
- ***Government*** The government uses computers to manage its own operations and also for e-governance. The websites of the different government departments provide information to the users. Computers are used for the filing of income tax return, paying taxes, online submission of water and electricity bills, for the access of land record details, etc. The police department uses computers to search for criminals using fingerprint matching, etc.
- ***Home*** Computers have now become an integral part of home equipment. At home, people use computers to play games, to maintain the home accounts, for communicating with friends and relatives via Internet, for paying bills, for education and learning, etc. Microprocessors are embedded in house hold utilities like, washing machines, TVs, food processors, home theatres, security devices, etc.

The list of applications of computers is so long that it is not possible to discuss all of them here. In addition to the applications of the computers discussed above, computers have also proliferated into areas like banks, investments, stock trading, accounting, ticket reservation, military operations, meteorological predictions, social networking, business organizations, police department, video conferencing, telepresence, book publishing, web newspapers, and information sharing.

CHAPTER 2
THE COMPUTER SYSTEM HARDWARE

2.1 INTRODUCTION

When we talk of computer hardware, the three related terms that require introduction are—computer architecture, computer organization and computer design. *Computer architecture* refers to the structure and behavior of the computer. It includes the specifications of the components, for example, instruction format, instruction set and techniques for addressing memory, and how they connect to the other components. Given the components, *computer organization* focuses on the organizational structure. It deals with how the hardware components operate and the way they are connected to form the computer. Given the system specifications, *computer design* focuses on the hardware to be used and the interconnection of parts. Different kinds of computer, such as a PC or a mainframe computer may have different organization; however, basic organization of the computer remains the same.

A computer consists of three main components—(1) Input/Output (I/O) Unit, (2) Central Processing Unit (CPU), and (3) Memory Unit. The computer user interacts with the computer via the I/O unit. The purpose of I/O unit is to provide data and instructions as input to the computer and to present relevant information as output from the computer. CPU controls the operations of the computer and processes the received input to generate the relevant output. The memory unit stores the instructions and the data during the input activity, to make instructions readily available to CPU during processing. It also stores the processed output. This chapter discusses the hardware components of the computer and the interaction between them.

2.2 CENTRAL PROCESSING UNIT

Central Processing Unit (CPU) or the processor is also often called the *brain of computer*. CPU (Figure 2.1) consists of Arithmetic Logic Unit (ALU) and Control Unit (CU). In addition, CPU also has a set of registers which are temporary storage areas for holding data, and instructions. *ALU* performs the arithmetic and logic operations on the data that is made available to it. *CU* is responsible for organizing the processing of data and instructions. *CU* controls and coordinates the activity of the other units of computer. CPU uses the registers to store the data, instructions during processing.

CPU executes *the stored program instructions*, i.e. instructions and data are stored in memory before execution. For processing, CPU gets data and instructions from the memory. It interprets the program instructions and performs the arithmetic and logic operations required for the processing of data. Then, it sends the processed data or result to the memory. CPU also acts as an administrator and is responsible for supervising operations of other parts of the computer.

The CPU is fabricated as a single Integrated Circuit (IC) chip, and is also known as the *microprocessor*. The microprocessor is plugged into the motherboard of the computer (*Motherboard* is a circuit board that has electronic circuit etched on it and connects the microprocessor with the other hardware components).

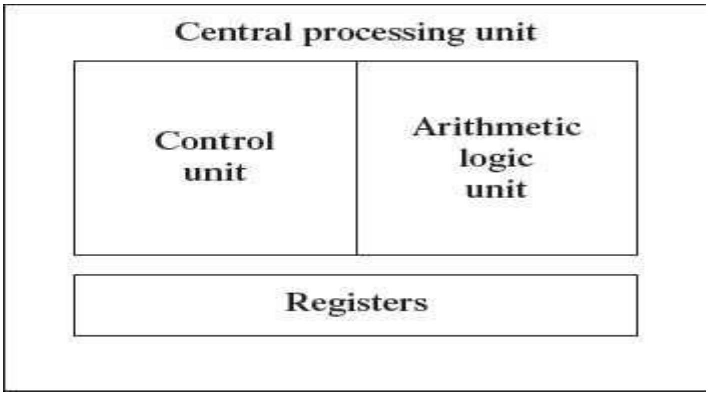

Figure 2.1 CPU

2.2.1 Arithmetic Logic Unit

- ALU consists of two units—arithmetic unit and logic unit.
- The arithmetic unit performs arithmetic operations on the data that is made available to it. Some of the arithmetic operations supported by the arithmetic unit are—addition, subtraction, multiplication and division.
- The logic unit of ALU is responsible for performing logic operations. Logic unit performs comparisons of numbers, letters and special characters. Logic operations include testing for greater than, less than or equal to condition.
- ALU performs arithmetic and logic operations, and uses *registers* to hold the data that is being processed.

2.2.2 Registers

- Registers are high-speed storage areas within the CPU, but have the least storage capacity. Registers are not referenced by their address, but are directly accessed and manipulated by the CPU during instruction execution.
- Registers store data, instructions, addresses and intermediate results of processing. Registers are often referred to as the CPU's *working memory*.
- The data and instructions that require processing must be brought in the registers of CPU before they can be processed. For example, if two numbers are to be added, both numbers are brought in the registers, added and the result is also placed in a register.
- Registers are used for different purposes, with each register serving a specific purpose. Some of the important registers in CPU (Figure 2.2) are as follows—
  - Accumulator (ACC) stores the result of arithmetic and logic operations.
  - Instruction Register (IR) contains the current instruction most recently fetched.
  - Program Counter (PC) contains the address of next instruction to be processed.
  - Memory Address Register (MAR) contains the address of next location in the memory to be accessed.
  - Memory Buffer Register (MBR) temporarily stores data from memory or the data to be sent to memory.
  - Data Register (DR) stores the operands and any other data.

Figure 2.2 CPU registers

- The number of registers and the size of each (number of bits) register in a CPU helps to determine the power and the speed of a CPU.
- The overall number of registers can vary from about ten to many hundreds, depending on the type and complexity of the processor.
- The size of register, also called *word size*, indicates the amount of data with which the computer can work at any given time. The bigger the size, the more quickly it can process data. The size of a register may be 8, 15, 32 or 54 bits. For example, a 32–bit CPU is one in which each register is 32 bits wide and its CPU can manipulate 32 bits of data at a time. Nowadays, PCs have 32–bit or 54–bit registers.
- 32-bit processor and 54-bit processor are the terms used to refer to the size of the registers. Other factors remaining the same, a 54-bit processor can process the data twice as fast as one with 32-bit processor.

2.2.3 Control Unit

- The control unit of a computer does not do any actual processing of data. It organizes the processing of data and instructions. It acts as a supervisor and, controls and coordinates the activity of the other units of computer.
- CU coordinates the input and output devices of a computer. It directs the computer to carry out stored program instructions by communicating with the ALU and the registers. CU uses the instructions in the Instruction Register (IR) to decide which circuit needs to be activated. It also instructs the ALU to perform the arithmetic or logic operations. When a program is run, the Program Counter (PC) register keeps track of the program instruction to be executed next.
- CU tells when to fetch the data and instructions, what to do, where to store the results, the sequencing of events during processing etc.
- CU also holds the CPU's Instruction Set, which is a list of all operations that the CPU can perform.

The function of a (CU) can be considered synonymous with that of a conductor of an orchestra. The conductor in an orchestra does not perform any work by itself but manages the orchestra and ensures that the members of orchestra work in proper coordination.

2.3 MEMORY UNIT

The memory unit consists of cache memory and primary memory. *Primary memory or main memory* of the computer is used to store the data and instructions during execution of the instructions. Random Access Memory (RAM) and Read Only Memory (ROM) are the primary memory. In addition to the main memory, there is another kind of storage device known as the secondary memory. Secondary memory is non-volatile and is used for permanent storage of data and programs. A program or data that has to be executed is brought into the RAM from the secondary memory.

2.3.1 Cache Memory

- The data and instructions that are required during the processing of data are brought from the secondary storage devices and stored in the RAM. For processing, it is required that the data and instructions are accessed from the RAM and stored in the registers. The time taken to move the data between RAM and CPU registers is large. This affects the speed of processing of computer, and results in decreasing the performance of CPU.
- Cache memory is a very high speed memory placed in between RAM and CPU. Cache memory increases the speed of processing.
- Cache memory is a storage buffer that stores the data that is used more often, temporarily, and makes them available to CPU at a fast rate. During processing, CPU first checks cache for the required data. If data is not found in cache, then it looks in the RAM for data.
- To access the cache memory, CPU does not have to use the motherboard's system bus for data transfer. (The data transfer speed slows to the motherboard's capability, when data is passed through system bus. CPU can process data at a much faster rate by avoiding the system bus.)

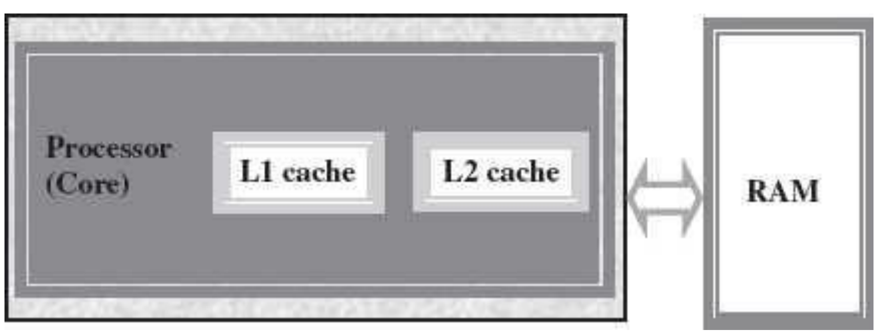

Figure 2.3 Illustration of cache memory

- Cache memory is built into the processor, and may also be located next to it on a separate chip between the CPU and RAM. Cache built into the CPU is faster than separate cache, running at the speed of the microprocessor itself. However, separate cache is roughly twice as fast as RAM.
- The CPU has a built-in *Level 1 (L1)* cache and *Level2 (L2)* cache, as shown in Figure 2.3. In addition to the built-in L1 and L2 cache, some CPUs have a separate cache chip on the

motherboard. This cache on the motherboard is called *Level 3 (L3)* cache. Nowadays, high-end processor comes with built-in L3 cache, like in Intel core i7. The L1, L2 and L3 cache store the most recently run instructions, the next ones and the possible ones, respectively. Typically, CPUs have cache size varying from 255KB (L1), 5 MB (L2), to 12MB (L3) cache.

- Cache memory is very expensive, so it is smaller in size. Generally, computers have cache memory of sizes 255 KB to 2 MB.

2.3.2 Primary Memory

- Primary memory is the main memory of computer. It is used to store data and instructions during the processing of data. Primary memory is semiconductor memory.
- Primary memory is of two kinds—Random Access Memory (RAM) and Read Only Memory (ROM).
- RAM is volatile. It stores data when the computer is on. The information stored in RAM gets erased when the computer is turned off. RAM provides *temporary storage* for data and instructions.
- ROM is non-volatile memory, but is a read only memory. The storage in ROM is permanent in nature, and is used for storing standard processing programs that permanently reside in the computer. ROM comes programmed by the manufacturer.
- RAM *stores data and instructions during the execution* of instructions. The data and instructions that require processing are brought into the RAM from the storage devices like hard disk. CPU accesses the data and the instructions from RAM, as it can access it at a *fast* speed than the storage devices connected to the input and output unit (Figure 2.4).
- The input data that is entered using the input unit is stored in RAM, to be made available during the processing of data. Similarly, the output data generated after processing is stored in RAM before being sent to the output device. Any intermediate results generated during the processing of program are stored in RAM.
- RAM provides a *limited storage capacity*, due to its *high cost*.

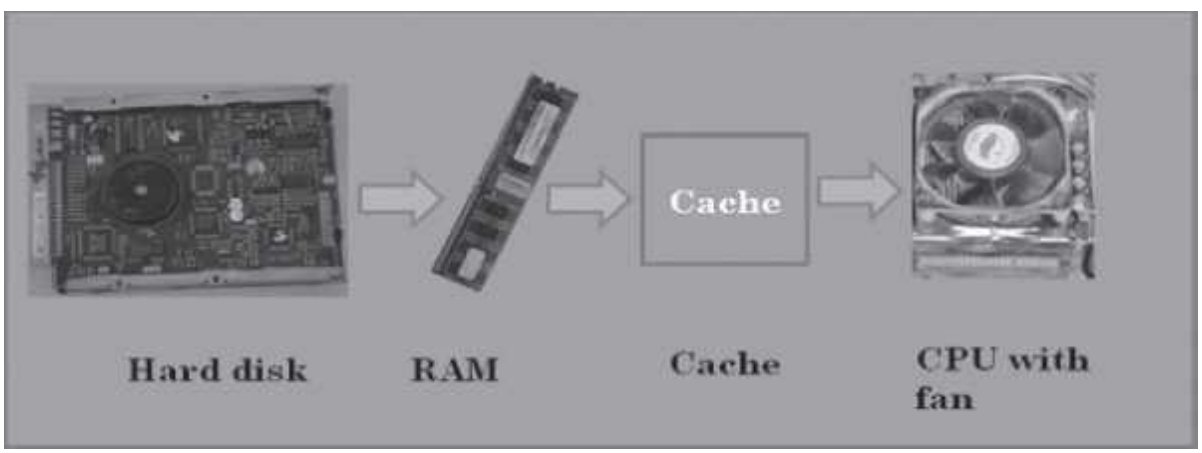

Figure 2.4 Interaction of CPU with memory

2.3.3 Secondary Memory

- The secondary memory stores data and instructions *permanently*. The information can be stored in secondary memory for a long time (years), and is generally permanent in nature unless erased by the user. It is a non-volatile memory.

- It provides *back-up storage* for data and instructions. Hard disk drive, floppy drive and optical disk drives are some examples of storage devices.
- The data and instructions that are currently not being used by CPU, but may be required later for processing, are stored in secondary memory.
- Secondary memory has a *high storage capacity* than the primary memory.
- Secondary memory is also *cheaper* than the primary memory.
- It takes *longer time to access* the data and instructions stored in secondary memory than in primary memory.

Magnetic tape drives, disk drives and optical disk drives are the different types of storage devices.

2.4 INSTRUCTION CYCLE

The primary responsibility of a computer processor is to execute a sequential set of instructions that constitute a program. CPU executes each instruction in a series of steps, called *instruction cycle* (Figure 2.8).

- A instruction cycle involves four steps (Figure 2.9)—
    - ***Fetching*** The processor fetches the instruction from the memory. The fetched instruction is placed in the *Instruction Register*. *Program Counter* holds the address of next instruction to be fetched and is incremented after each fetch.
    - ***Decoding*** The instruction that is fetched is broken down into parts or decoded. The instruction is translated into commands so that they correspond to those in the CPU's instruction set. The instruction set architecture of the CPU defines the way in which an instruction is decoded.
    - ***Executing*** The decoded instruction or the command is executed. CPU performs the operation implied by the program instruction. For example, if it is an ADD instruction, addition is performed.
    - ***Storing*** CPU writes back the results of execution, to the computer's memory.

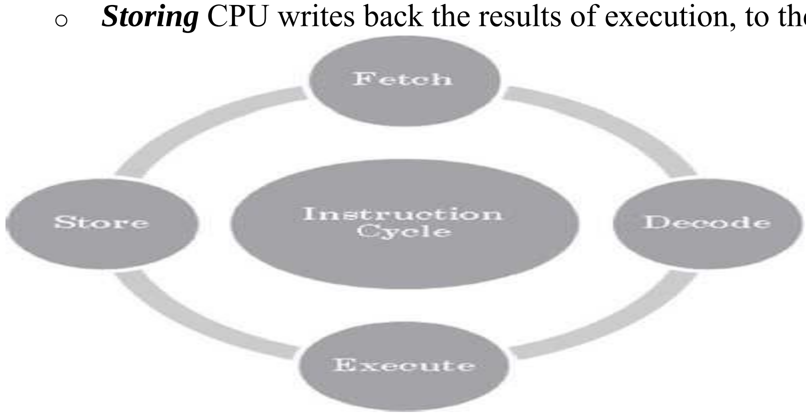

Figure 2.8 Instruction cycle

- Instructions are of different categories. Some categories of instructions are—
    - Memory access or transfer of data between registers.
    - Arithmetic operations like addition and subtraction.
    - Logic operations such as AND, OR and NOT.

- Control the sequence, conditional connections, etc.

A CPU performance is measured by the number of instructions it executes in a second, i.e., *MIPS* (million instructions per second), or *BIPS* (billion instructions per second).

2.5 MICROPROCESSOR

A processor's instruction set is a determining factor in its architecture. On the basis of the instruction set, microprocessors are classified as—Reduced Instruction Set Computer (RISC), and Complex Instruction Set Computer (CISC). The *x85* instruction set of the original Intel 8085 processor is of the CISC type. The PCs are based on the *x85* instruction set.

- ***CISC*** architecture hardwires the processor with complex instructions, which are difficult to create otherwise using basic instructions. CISC combines the different instructions into one single CPU.
  - CISC has a large instruction set that includes simple and fast instructions for performing basic tasks, as well as complex instructions that correspond to statements in the high level language.
  - An increased number of instructions (200 to 300) results in a much more complex processor, requiring millions of transistors.
  - Instructions are of variable lengths, using 8, 15 or 32 bits for storage. This results in the processor's time being spent in calculating where each instruction begins and ends.
  - With large number of application software programs being written for the processor, a new processor has to be backwards compatible to the older version of processors.
  - AMD and Cyrix are based on CISC.
- ***RISC*** has simple, single-cycle instructions, which performs only basic instructions. RISC architecture does not have hardwired advanced functions. All high-level language support is done in the software.
  - RISC has fewer instructions and requires fewer transistors, which results in the reduced manufacturing cost of processor.
  - The instruction size is fixed (32 bits). The processor need not spend time in finding out where each instruction begins and ends.
  - RISC architecture has a reduced production cost compared to CISC processors.
  - The instructions, simple in nature, are executed in just one clock cycle, which speeds up the program execution when compared to CISC processors.
  - RISC processors can handle multiple instructions simultaneously by processing them in parallel.
  - Apple Mac G3 and PowerPC are based on RISC.

Processors like Athlon XP and Pentium IV use a hybrid of both technologies.

2.6 INTERCONNECTING THE UNITS OF A COMPUTER

CPU sends data, instructions and information to the components inside the computer as well as to the peripherals and devices attached to it. ***Bus*** is a set of electronic signal pathways that allows information and signals to travel between components inside or outside of a computer. The different components of computer, i.e., CPU, I/O unit, and memory unit are connected with each other by a bus. The data, instructions and the signals are carried between the different components via a bus. The features and functionality of a bus are as follows—

- A bus is a set of wires used for interconnection, where each wire can carry one bit of data.
- A bus width is defined by the number of wires in the bus.
- A computer bus can be divided into two types—Internal Bus and External Bus.
- The Internal Bus connects components inside the motherboard like, CPU and system memory. It is also called the *System Bus*. Figure 2.10 shows interaction between processor and memory.

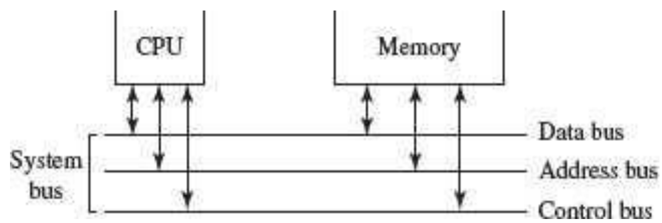

Figure 2.10 Interaction between CPU and memory

- The External Bus connects the different external devices, peripherals, expansion slots, I/O ports and drive connections to the rest of computer. The external bus allows various devices to be attached to the computer. It allows for the expansion of computer's capabilities. It is generally slower than the system bus. It is also referred to as the *Expansion* Bus.
- A system bus or expansion bus comprise of three kinds of buses — data bus, address bus and control bus.
- The interaction of CPU with memory and I/O devices involves all the three buses.
  - The command to access the memory or the I/O device is carried by the ***control bus***.
  - The address of I/O device or memory is carried by the ***address bus***.
  - The data to be transferred is carried by the ***data bus***.

Figure 2.11 shows interaction between processor, memory and the peripheral devices.

2.6.1 System Bus

The functions of data bus, address bus and control bus, in the system bus, are as follows—

- ***Data Bus*** transfers data between the CPU and memory. The bus width of a data bus affects the *speed of computer*. The size of data bus defines the size of the processor. A processor can be 8, 15, 32 or 54-bit processor. An 8–bit processor has 8 wire data bus to carry 1 byte of data. In a 15–bit processor, 15–wire bus can carry 15 bits of data, i.e., transfer 2 bytes, etc.

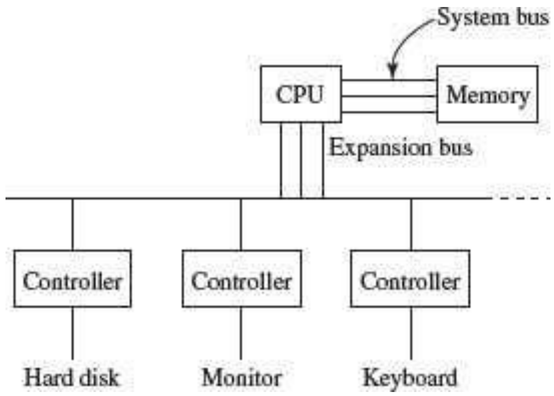

Figure 2.11 Interaction between CPU, memory and peripheral devices

- ***Address Bus*** connects CPU and RAM with set of wires similar to data bus. The width of address bus determines the *maximum number of memory locations* the computer can address. Currently, Pentium Pro, II, III, IV have 35–bit address bus that can address 2^{35} bytes or 54 GB of memory.
- ***Control*** Bus specifies whether data is to be read or written to the memory, etc.

2.6.2 Expansion Bus

The functions of data bus, address bus and control bus, in the expansion bus, are as follows—

- The expansion bus connects external devices to the rest of computer. The external devices like monitor, keyboard and printer connect to ports on the back of computer. These ports are actually a part of the small circuit board or *expansion card* that fits into an *expansion slot* on the motherboard. Expansion slots are easy to recognize on the motherboard.
- Expansion slots make up a row of long plastic connectors at the back of the computer with tiny copper ‘finger slots’ in a narrow channel that grab the connectors on the expansion cards. The slots are attached to tiny copper pathways on the motherboard (the expansion bus), which allows the device to communicate with the rest of computer.
- ***Data Bus*** is used to transfer data between I/O devices and CPU. The exchange of data between CPU and I/O devices is according to the industry standard data buses. The most commonly used standard is Extended Industry Standard Architecture (EISA) which is a 32-bit bus architecture. Some of the common bus technologies are—
    - Peripheral Component Interconnect (PCI) bus for hard disks, sound cards, network cards and graphics cards,
    - Accelerated Graphics Port (AGP) bus for 3–D and full motion video,
    - Universal Serial Bus (USB) to connect and disconnect different devices.
- ***Address Bus*** carries the addresses of different I/O devices to be accessed like the hard disk, CD ROM, etc.
- ***Control Bus*** is used to carry read/write commands, status of I/O devices, etc.

2.6.3 External Ports

The peripheral devices interact with the CPU of the computer via the bus. The connections to the bus from the peripheral devices are made via the ports and sockets provided at the sides of the computer. The different ports and sockets facilitate the connection of different devices to the computer. Some of the standard port connections available on the outer sides of the computer are— port for mouse, keyboard, monitor, network, modem, and, audio port, serial port, parallel port and USB port. The different ports are physically identifiable by their different shapes, size of contact pins and number of pins. Figure 2.12 shows the interaction of serial and parallel port interfaces with the devices.

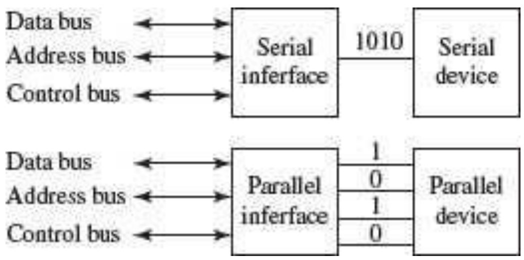

Figure 2.12 Interaction of serial and parallel port interfaces

2.9 PERFORMANCE OF A COMPUTER

There are a number of factors involved that are related to the CPU and have an effect on the overall speed and performance of the computer. Some of the factors that affect the performance of the computer include—

- ***Registers*** The size of the register (word size) indicates the amount of data with which the computer can work at any given time. The bigger the size, the more quickly it can process data. A 32–bit CPU is one in which each register is 32 bits wide.
- ***RAM*** It is used to store data and instructions during execution of the instructions. Anything you do on your computer requires RAM. When the computer is switched on, the operating system, device drivers, the active files and running programs are loaded into RAM. If RAM is less, then the CPU waits each time the new information is swapped into memory from the slower devices. Larger the RAM size, the better it is. PCs nowadays usually have 1 GB to 4 GB of RAM.
- ***System Clock*** The clock speed of a CPU is defined as the frequency with which a processor executes instructions or the data is processed. Higher clock frequencies mean more clock ticks per second. The computer's operating speed is linked to the speed of the system clock. The clock frequency is measured in millions of cycles per second or megahertz (MHz) or gigahertz (GHz) which is billions of cycles per second. A CPU's performance is measured by the number of instructions it executes in a second, i.e., *MIPS* or *BIPS*. PCs nowadays come with a clock speed of more than 1 GHz. *In Windows OS, you can select the System Properties dialog box to see the processor name and clock frequency*.
- ***Bus*** *Data bus* is used for transfering data between CPU and memory. The data bus width affects the speed of computer. In a 15–bit processor, 15–bit wire bus can carry 15 bits of data. The bus speed is measured in MHz. Higher the bus speed the better it is. *Address bus* connects CPU and RAM with a set of wires similar to data bus. The address bus width determines the *maximum number of memory locations* the computer can address. Pentium Pro, II, III, IV have 35–bit address bus that can address 235 bytes or 54 GB of memory. PCs nowadays have a bus speed varying from 100 MHz to 400 MHz.

- ***Cache Memory*** Two of the main factors that affect a cache's performance are its size (amount of cache memory) and level L1, L2 and L3. Larger the size of cache, the better it is. PCs nowadays have a L1 cache of 255KB and L2 cache of 1MB.

2.10 INSIDE A COMPUTER CABINET

The computer cabinet encloses the components that are required for the running of the computer. The components inside a computer cabinet include the power supply, motherboard, memory chips, expansion slots, ports and interface, processor, cables and storage devices.

2.10.1 Motherboard

The computer is built up around a *motherboard*. The motherboard is the most important component in the PC. It is a large Printed Circuit Board (PCB), having many chips, connectors and other electronics mounted on it. The motherboard is the hub, which is used to connect all the essential components of a computer. The RAM, hard drive, disk drives and optical drives are all plugged into interfaces on the motherboard. The motherboard contains the processor, memory chips, interfaces and sockets, etc.

The motherboard may be characterized by the form factor, chipset and type of processor socket used. *Form factor* refers to the motherboard's geometry, dimensions, arrangement and electrical requirements. Different standards have been developed to build motherboards, which can be used in different brands of cases. Advanced Technology Extended (ATX) is the most common design of motherboard for desktop computers. *Chipset* is a circuit, which controls the majority of resources (including the bus interface with the processor, cache memory and RAM, expansion cards, etc.) Chipset's job is to coordinate data transfers between the various components of the computer (including the processor and memory). As the chipset is integrated into the motherboard, it is important to choose a motherboard, which includes a recent chipset, in order to maximize the computer's upgradeability. The *processor socket* may be a rectangular connector into which the processor is mounted vertically (slot), or a square-shaped connector with many small connectors into which the processor is directly inserted (socket). The Basic Input Output System (BIOS) and Complementary Metal-Oxide Semiconductor (CMOS) are present on the motherboard.

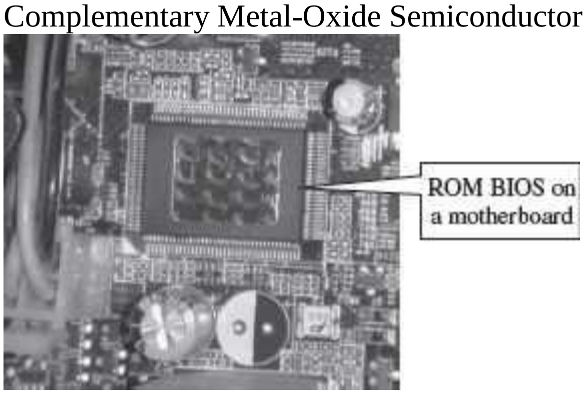

Figure 2.14 ROM BIOS

- ***BIOS*** It is the basic program used as an interface between the operating system and the motherboard. The BIOS (Figure 2.14) is stored in the ROM and cannot be rewritten. When the computer is switched on, it needs instructions to start. BIOS contain the instructions for the starting up of the computer. The BIOS runs when the computer is switched on. It performs a Power On Self Test (POST) that checks that the hardware is functioning properly and the hardware devices are present. It checks whether the operating system is present on the hard drive. BIOS invokes the bootstrap loader to load the operating system into memory. BIOS can be configured using an interface named BIOS setup, which can be accessed when the computer is booting up (*by pressing the DEL key*).

- ***CMOS Chip*** BIOS ROMs are accompanied by a smaller CMOS (CMOS is a type of memory technology) memory chip. When the computer is turned off, the power supply stops providing electricity to the motherboard. When the computer is turned on again, the system still displays the correct clock time. This is because the CMOS chip saves some system information, such as time, system date and essential system settings. CMOS is kept powered by a button battery located on the motherboard (Figure 2.15). The CMOS chip is working even when the computer power is switched off. Information of the hardware installed in the computer (such as the number of tracks or sectors on each hard drive) is stored in the CMOS chip.

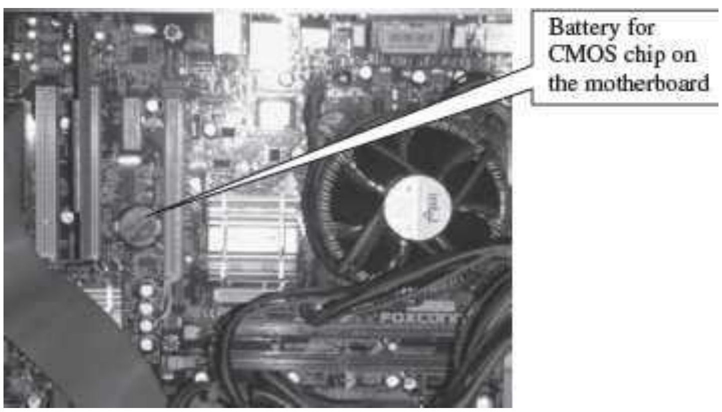

Figure 2.15 Battery for CMOS chip

2.10.2 Ports and Interfaces

Motherboard has a certain number of I/O sockets that are connected to the ports and interfaces found on the rear side of a computer (Figure 2.15). You can connect external devices to the ports and interfaces, which get connected to the computer's motherboard.

- Serial Port— to connect old peripherals.
- Parallel Port— to connect old printers.

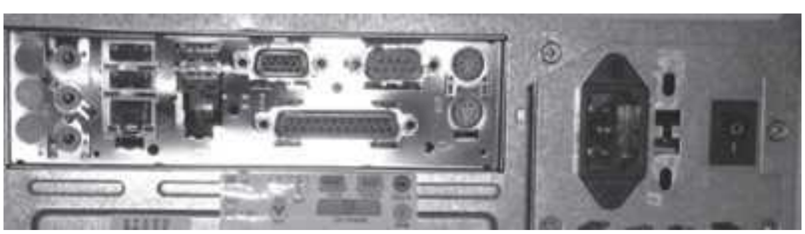

Figure 2.15 Ports on the rear side of a PC

- USB Ports—to connect newer peripherals like cameras, scanners and printers to the computer. It uses a thin wire to connect to the devices, and many devices can share that wire simultaneously.

- Firewire is another bus, used today mostly for video cameras and external hard drives.
- RJ45 connector (called LAN or Ethernet port) is used to connect the computer to a network. It corresponds to a network card integrated into the motherboard.
- VGA connector for connecting a monitor. This connector interfaces with the built-in graphics card.
- Audio plugs (line-in, line-out and microphone), for connecting sound speakers and the microphone. This connector interfaces with the built-in sound card.
- PS/2 port to connect mouse and keyboard into PC.
- SCSI port for connecting the hard disk drives and network connectors.

2.10.3 Expansion Slots

The expansion slots (Figure 2.17) are located on the motherboard. The expansion cards are inserted in the expansion slots. These cards give the computer new features or increased performance. There are several types of slots:

- ISA (Industry Standard Architecture) slot—To connect modem and input devices.
- PCI (Peripheral Component InterConnect) slot—To connect audio, video and graphics. They are much faster than ISA cards.
- AGP (Accelerated Graphic Port) slot—A fast port for a graphics card.
- PCI (Peripheral Component InterConnect) Express slot—Faster bus architecture than AGP and PCI buses.
- PC Card—It is used in laptop computers. It includes Wi-Fi card, network card and external modem.

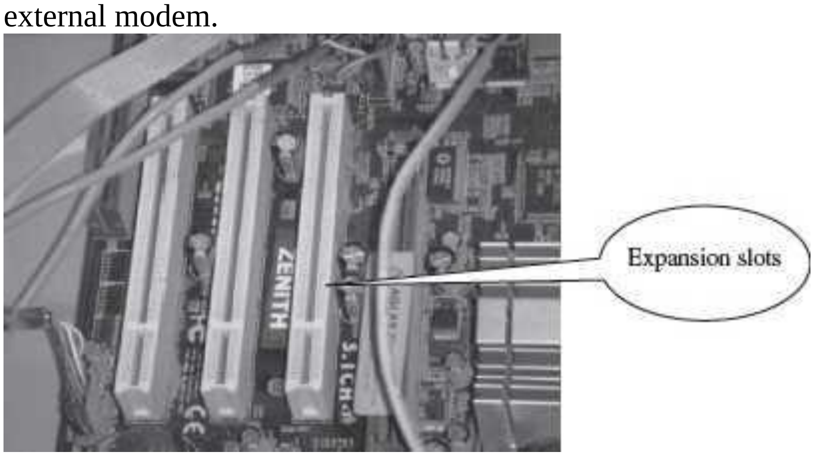

Figure 2.17 Expansion slots

2.10.4 Ribbon Cables

Ribbon cables (Figure 2.18) are flat, insulated and consist of several tiny wires moulded together that carry data to different components on the motherboard. There is a wire for each bit of the word or byte and additional wires to coordinate the activity of moving information. They also connect the floppy drives, disk drives and CD-ROM drives to the connectors in the motherboard. Nowadays, Serial Advanced Technology Attachment (SATA) cables have replaced the ribbon cables to connect the drives to the motherboard.

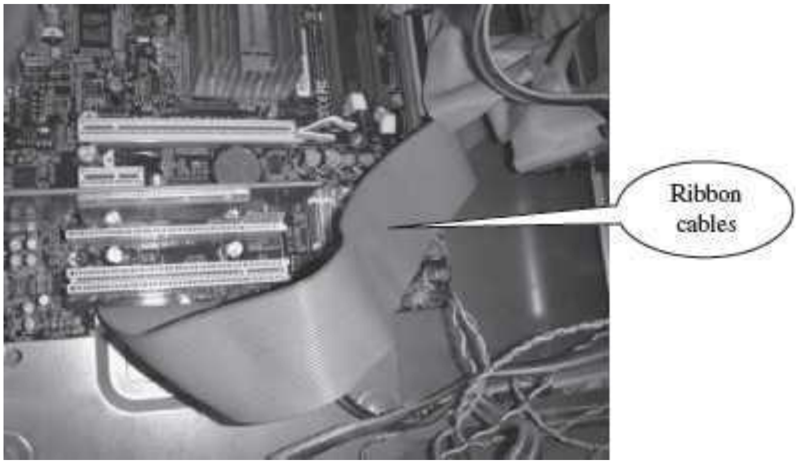

Figure 2.18 Ribbon cables inside a PC

2.10.5 Memory Chips

The RAM consists of chips on a small circuit board (Figure 2.19). Two types of memory chips—Single In-line Memory Module (SIMM) and Dual In-line Memory Module (DIMM) are used in desktop computers. The CPU can retrieve information from DIMM chip at 54 bits compared to 32 bits or 15 bits transfer with SIMM chips. DIMM chips are used in Pentium 4 onwards to increase the access speed.

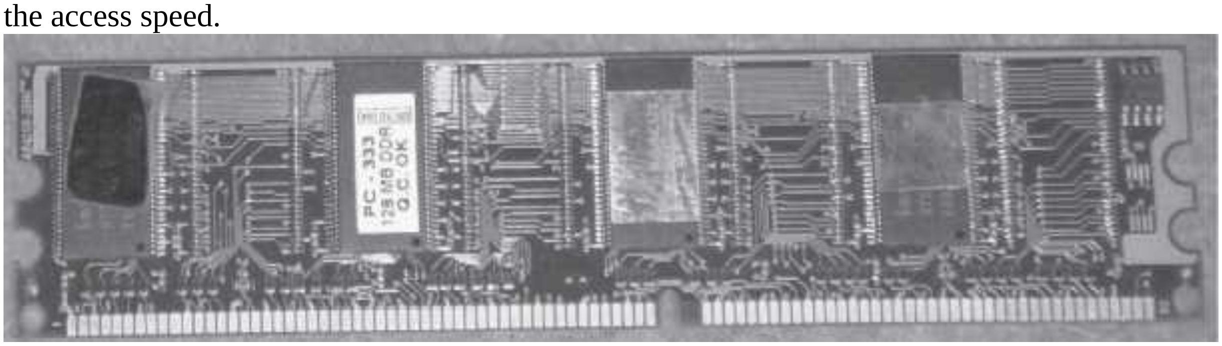

Figure 2.19 RAM memory chip

2.10.6 Storage Devices

The disk drives are present inside the machine. The common disk drives in a machine are hard disk drive, floppy drive (Figure 2.20 (i & ii)) and CD drive or DVD drive. High-storage devices like hard disk, floppy disk and CDs (Figure 2.20 (iii) & (iv)) are inserted into the hard disk drive, floppy drive and CD drive, respectively. These storage devices can store large amounts of data, permanently.

2.10.7 Processor

The processor or the CPU is the main component of the computer. Select a processor based on factors like its speed, performance, reliability and motherboard support. Pentium Pro, Pentium 2 and Pentium 4 are some of the processors.

Figure 2.20 Storage devices (i) Hard disk drive, (ii) DVD drive, (iii) Floppy disk, (iv) CD

CHAPTER 3
COMPUTER MEMORY

3.1 INTRODUCTION

The computer's memory stores data, instructions required during the processing of data, and output results. Storage may be required for a limited period of time, instantly, or, for an extended period of time. Different types of memories, each having its own unique features, are available for use in a computer. The cache memory, registers, and RAM are fast memories and store the data and instructions temporarily during the processing of data and instructions. The secondary memory like magnetic disks and optical disks have large storage capacities and store the data and instructions permanently, but are slow memory devices. The memories are organized in the computer in a manner to achieve high levels of performance at the minimum cost.

In this chapter, we discuss different types of memories, their characteristics and their use in the computer.

3.2 MEMORY REPRESENTATION

The computer memory stores different kinds of data like input data, output data, intermediate results, etc., and the instructions. ***Binary digit*** or ***bit*** is the basic unit of memory. A *bit* is a single binary digit, i.e., 0 or 1. A bit is the smallest unit of representation of data in a computer. However, the data is handled by the computer as a combination of bits. A group of 8 bits form a **byte**. One byte is the smallest unit of data that is handled by the computer. One byte can store 2^8, i.e., 255 different combinations of bits, and thus can be used to represent 255 different symbols. In a byte, the different combinations of bits fall in the range 00000000 to 11111111. A group of bytes can be further combined to form a **word**. A word can be a group of 2, 4 or 8 bytes.

1 bit = 0 or 1

1 Byte (B) = 8 bits

1 Kilobyte (KB) = 2^{10} = 1024 bytes

1 Megabyte (MB) = 2^{20} = 1024KB

1 Gigabyte (GB) = 2^{30} = 1024 MB = 1024 *1024 KB

1 Terabyte (TB) = 2^{40}= 1024 GB = 1024 * 1024 *1024 KB

3.3 MEMORY HIERARCHY

The memory is characterized on the basis of two key factors—capacity and access time. *Capacity* is the amount of information (in bits) that a memory can store. *Access time* is the time interval between the read/ write request and the availability of data. The lesser the access time, the faster is the *speed of memory*. Ideally, we want the memory with *fastest speed and largest capacity*.

However, the cost of fast memory is very high. The computer uses a hierarchy of
memory that is organized in a manner to enable the fastest speed and largest capacity of memory. The hierarchy of the different memory types is shown in Figure 3.2.

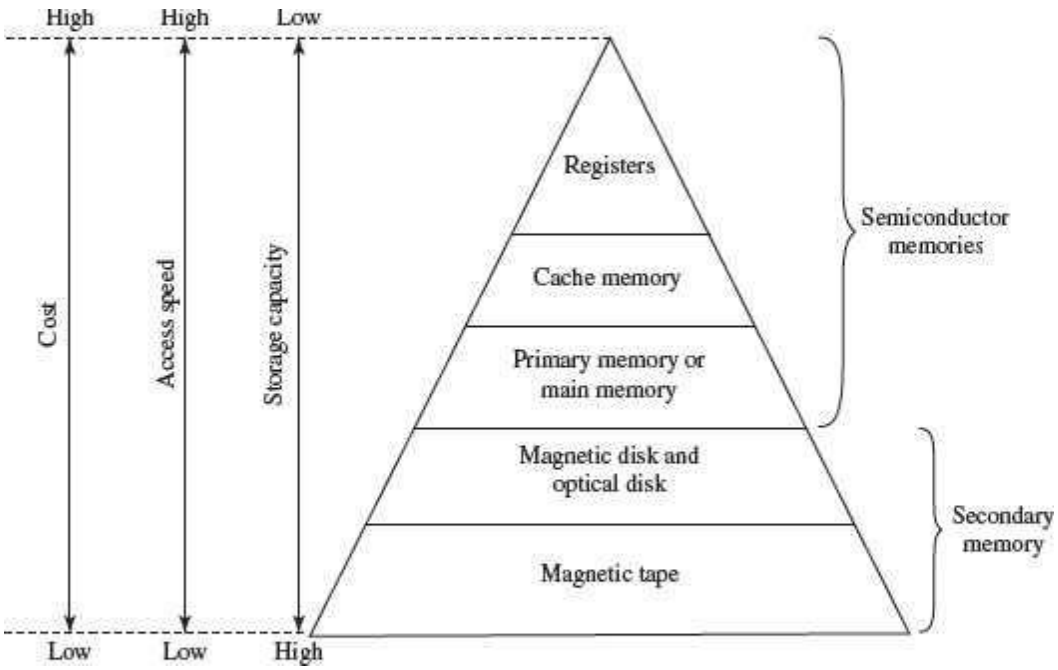

Figure 3.2 Memory hierarchy

The internal memory and external memory are the two broad categories of memory used in the computer. The internal memory consists of the CPU registers, cache memory and primary memory. The internal memory is used by the CPU to perform the computing tasks. The external memory is also called the secondary memory. The secondary memory is used to store the large amount of data and the software .

In general, referring to the computer memory usually means the internal memory.

- ***Internal Memory***—The key features of internal memory are—(1) limited storage capacity, (2) temporary storage, (3) fast access, and (4) high cost. Registers, cache memory, and primary memory constitute the internal memory. The primary memory is further of two kinds—RAM and ROM. Registers are the fastest and the most expensive among all the memory types. The registers are located inside the CPU, and are directly accessible by the CPU. The speed of registers is between 1—2 ns (nanosecond). The sum of the size of registers is about 200B. Cache memory is next in the hierarchy and is placed between the CPU and the main memory. The speed of cache is between 2—10 ns. The cache size varies between 32 KB to 4MB. Any program or data that has to be executed must be brought into RAM from the secondary memory. Primary memory is relatively slower than the cache memory. The speed of RAM is around 50ns. The RAM size varies from 512KB to 3GB.

- ***Secondary Memory***—The key features of secondary memory storage devices are—(1) very high storage capacity, (2) permanent storage (non-volatile), unless erased by user, (3) relatively slower access, (4) stores data and instructions that are not currently being used by CPU but may be required later for processing, and (5) cheapest among all memory. The storage devices consist of two parts—drive and device. For example, magnetic tape drive and magnetic tape, magnetic disk drive and disk, and, optical disk drive and disk. The speed of magnetic disk is around 50ms. The capacity of a hard disk ranges from 150 GB to 1,500 GB (1.5 Tera Bytes). Figure 3.3 shows the interaction between CPU and memory.

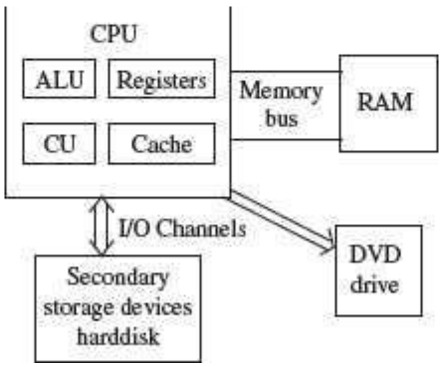

Figure 3.3 CPU and the memory

To get the fastest speed of memory with largest capacity and least cost, the fast memory is located close to the processor. The secondary memory, which is not as fast, is used to store information permanently, and is placed farthest from the processor. With respect to CPU, the memory is organized as follows—

- Registers are placed inside the CPU (small capacity, high cost, very high speed)
- Cache memory is placed next in the hierarchy (inside and outside the CPU)
- Primary memory is placed next in the hierarchy
- Secondary memory is the farthest from CPU (large capacity, low cost, low speed)

The speed of memories is dependent on the kind of technology used for the memory. The registers, cache memory and primary memory are semiconductor memories. They do not have any moving parts and are fast memories. The secondary memory is magnetic or optical memory, has moving parts and has slow speed.

3.4 CPU REGISTERS

- Registers are very high-speed storage areas located inside the CPU. After CPU gets the data and instructions from the cache or RAM, the data and instructions are moved to the registers for processing. Registers are manipulated directly by the control unit of CPU during instruction execution. That is why registers are often referred to as the CPU's *working memory*. Since CPU uses registers for the processing of data, the number of registers in a CPU and the size of each register affect the power and speed of a CPU. The more the number of registers (ten to hundreds) and bigger the size of each register (8 bits to 54 bits), the better it is.

3.5 CACHE MEMORY

- Cache memory is placed in between the CPU and the RAM. Cache memory is a fast memory, faster than the RAM. When the CPU needs an instruction or data during processing, it first looks in the cache. If the information is present in the cache, it is called a *cache hit*, and the data or instruction is retrieved from the cache. If the information is not present in cache, then it is called a *cache miss* and the information is then retrieved from RAM. The content of cache is decided by the cache controller (a circuit on the motherboard). The most recently accessed information or instructions help the controller to guess the RAM locations that may be accessed next. To get good system performance, the number of hits must far outnumber the misses. The two main factors that affect the performance of cache are its size and level (L1, L2 and L3).

The CPU registers and the cache memory have been discussed in detail in the previous chapter.

3.6 PRIMARY MEMORY

Primary memory is the main memory of computer. It is a chip mounted on the motherboard of computer. Primary memory is categorized into two main types-

- Random Access Memory (RAM), and
- Read Only Memory (ROM)

RAM is used for the temporary storage of input data, output data and intermediate results. The input data entered into the computer using the input device, is stored in RAM for processing. After processing, the output data is stored in RAM before being sent to the output device. Any intermediate results generated during the processing of program are also stored in RAM. Unlike RAM, the data once stored in ROM either cannot be changed or can only be changed using some special operations. Therefore, ROM is used to store the data that does not require a change.
Flash memory is another form of rewritable read-only memory that is compact, portable, and requires little energy.

3.6.1 Random Access Memory

- RAM is used to *store data and instructions during the operation of computer*.
  - The data and instructions that need to be operated upon by CPU are first brought to RAM from the secondary storage devices like the hard disk.
  - CPU interacts with RAM to get the data and instructions for processing.
- RAM loses information when the computer is powered off. It is a *volatile memory*. When the power is turned on, again, all files that are required by the CPU are loaded from the hard disk to RAM. Since RAM is a volatile memory, any information that needs to be saved for a longer duration of time must not be stored in RAM.

- RAM provides *random access* to the stored bytes, words, or larger data units. This means that it requires same amount of time to access information from RAM, irrespective of where it is located in it.
- RAM can be *read from and written to* with the same speed.
- The *size ofRAM is limited due to its high cost*. The size of RAM is measured in MB or GB.
- The performance of RAM is affected by—
    - Access speed (how *quickly* information can be retrieved). The speed of RAM is expressed in nanoseconds.
    - Data transfer unit size (how *much* information can be retrieved in one request).
- RAM affects the speed and power of a computer. More the RAM, the better it is. Nowadays, computers generally have 512 MB to 4 GB of RAM.
- RAM is a microchip implemented using semiconductors.
- There are two categories of RAM, depending on the technology used to construct a RAM— (1) Dynamic RAM (DRAM), and (2) Static RAM (SRAM).
- ***DRAM*** is the most common type of memory chip. DRAM is mostly used as main memory since it is small and cheap.
    - It uses transistors and capacitors. The transistors are arranged in a matrix of rows and columns. The capacitor holds the bit of information 0 and 1. The transistor and capacitor are paired to make a *memory cell*. The transistor acts as a switch that lets the control circuitry on the memory chip read the capacitor or change its state.
    - DRAM must be refreshed continually to store information. For this, a memory controller is used. The memory controller recharges all the capacitors holding a 1 before they discharge. To do this, the memory controller reads the memory and then writes it right back.
    - DRAM gets its name from the refresh operation that it requires to store the information; otherwise it will lose what it is holding. The refresh operation occurs automatically thousands of times per second. DRAM is slow because the refreshing takes time.
    - Access speed of DRAM ranges from 50 to 150 ns.
- ***SRAM*** chip is usually used in *cache memory* due to its high speed.
    - SRAM uses multiple transistors (four to six), for each memory cell. It does not have a capacitor in each cell.
    - A SRAM memory cell has more parts so it takes more space on a chip than DRAM cell.
    - It does not need constant refreshing and therefore is faster than DRAM.
    - SRAM is more expensive than DRAM, and it takes up more space.
    - It stores information as long as it is supplied with power.
    - SRAM are easier to use and very fast. The access speed of SRAM ranges from 2–10 nanosecond.
- The memory chips (Figure 3.4) are available on a separate Printed Circuit Board (PCB) that is plugged into a special connector on the motherboard. Memory chips are generally available as part of a card called a *memory module*. There are generally two types of RAM modules—Single Inline Memory Module (SIMM) and Dual Inline Memory Module (DIMM).

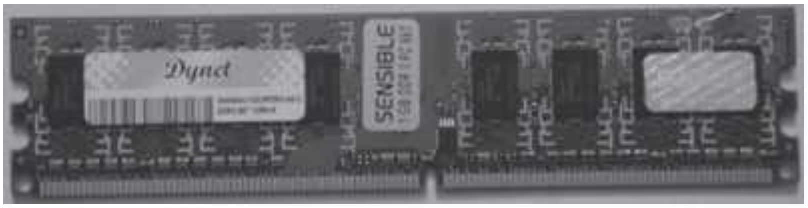

Figure 3.4 PCB containing RAM chip of 1 GB

- SIMM modules have memory chip on one side of the PCB. SIMM modules can store 8 bits to 32 bits of data simultaneously.
- DIMM modules have memory chips on both sides of the PCB. DIMM format are 54–bit memories. Smaller modules known as Small Outline DIMM (SO DIMM) are designed for portable computers. SO DIMM modules have 32–bit memory.

3.6.2 Read Only Memory

ROM is a *non-volatile* primary memory. It does not lose its content when the power is switched off. The features of ROM are described as follows—

- ROM, as the name implies, has only read capability and no write capability. After the information is stored in ROM, it is permanent and cannot be corrected.
- ROM comes programmed by the manufacturer. It stores standard processing programs that permanently reside in the computer. ROM stores the data needed for the start up of the computer. The instructions that are required for initializing the devices attached to a computer are stored in ROM.
- The ROM memory chip (Figure 3.5) stores the *Basic Input Output System (BIOS)*. BIOS provides the processor with the information required to boot the system. It provides the system with the settings and resources that are available on the system. BIOS is a permanent part of the computer. It does not load from disk but instead is stored in a ROM memory chip. The program code in the BIOS differs from ordinary software since it acts as an integral part of the computer. When the computer is turned on, the BIOS does the following things—

Figure 3.5 ROM BIOS and CMOS battery on a motherboard

- Power On Self Test (POST) is a program that runs automatically when the system is booted. BIOS performs the power-on self-test. It checks that the major hardware components are working properly.
- BIOS setup program, which is a built-in utility in BIOS, lets the user set the many functions that control how the computer works. BIOS displays the system settings and finds the bootable devices. It loads the interrupt handlers and device drivers. It also initializes the registers.
- *Bootstrap Loader* is a program whose purpose is to start the computer software for operation when the power is turned on. It loads the operating system into RAM and launches it. It generally seeks the operating system on the hard disk. The bootstrap loader resides in the ROM. The BIOS initiates the bootstrap sequence.

- ROMs are of different kinds. They have evolved from the fixed read only memory to the ones that can be programmed and re-programmed. They vary in the number of re-writes and the method used for the re-writing. Programmable ROM (PROM), Erasable Programmable ROM (EPROM) and Electrically Erasable Programmable ROM (EEPROM) are some of the ROMs. All the different kinds of ROM retain their content when the power is turned off.
  - ***PROM*** can be programmed with a special tool, but after it has been programmed the contents cannot be changed. PROM memories have thousands of fuses (or diodes). High voltage (12 V) is applied to the fuses to be burnt. The burnt fuses correspond to 0 and the others to 1.
  - ***EPROM*** can be programmed in a similar way as PROM, but it can be erased by exposing it to ultra violet light and re-programmed. EPROM chips have to be removed from the computer for re-writing.
  - ***EEPROM*** memories can be erased by electric charge and re-programmed. EEPROM chips do not have to be removed from the computer for re-writing.
- ***Flash Memory*** is a kind of semiconductor-based non-volatile, rewritable computer memory that can be electrically erased and reprogrammed (Figure 3.5). It is a specific type of EEPROM.

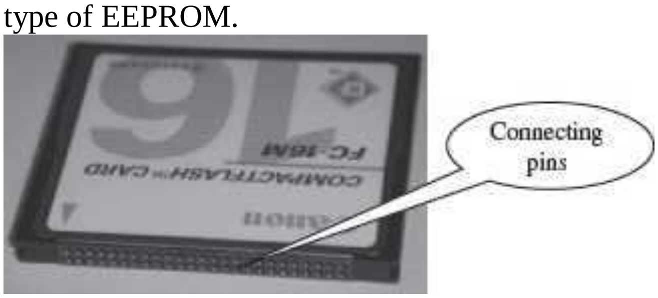

Figure 3.5 Flash memory

- It combines the features of RAM and ROM. It is a random access memory and its content can be stored in it at any time. However, like ROM, the data is not lost when the machine is turned off or the electric power is cut. Flash memory stores bits of data in memory cells.
- Flash memories are high-speed memories, durable, and have low-energy consumption. Since flash memory has no moving part, it is very shock-resistant. Due to these features, flash memory is used in devices such as digital camera,

mobile phone, printer, laptop computer, and record and play back sound devices, such as MP3 players.

3.7 SECONDARY MEMORY

In the previous section, we saw that RAM is expensive and has a limited storage capacity. Since it is a volatile memory, it cannot retain information after the computer is powered off. Thus, in addition to primary memory, an auxiliary or secondary memory is required by a computer. The secondary memory is also called the storage device of computer. *In this chapter, the terms secondary memory and storage device are used interchangeably*. In comparison to the primary memory, the secondary memory stores much larger amounts of data and information (for example, an entire software program) for extended periods of time. The data and instructions stored in secondary memory must be fetched into RAM before processing is done by CPU.

Magnetic tape drives, magnetic disk drives, optical disk drives and magneto-optical disk drives are the different types of storage devices.

3.8 ACCESS TYPES OF STORAGE DEVICES

The information stored in storage devices can be accessed in two ways—

1. Sequential access
2. Direct access

3.8.1 Sequential Access Devices

Sequential access means that computer must run through the data in sequence, starting from the beginning, in order to locate a particular piece of data. Magnetic tape is an example of sequential access device. Let us suppose that magnetic tape consists of 80 records. To access the 25th record, the computer starts from first record, then reaches second, third etc. until it reaches the 25th record. Sequential access devices are generally slow devices.

3.8.2 Direct Access Devices

Direct access devices are the ones in which any piece of data can be retrieved in a non-sequential manner by locating it using the data's address. It accesses the data directly, from a desired location. Magnetic disks and optical disks are examples of direct access devices. There is no predefined order in which one can read and write data from a direct access device. In a magnetic disk consisting of 80 records, to access the 25th record, the computer can directly access the 25th record, without going past the first 24 records. Based on access, magnetic tapes are sequential access devices, and, magnetic disks, optical disk and magneto-optical disks are direct access devices.

3.9 MAGNETIC TAPE

Magnetic tape is a plastic tape with magnetic coating (Figure 3.7). It is a storage medium on a large open reel or in a smaller cartridge or cassette (like a music cassette). Magnetic tapes are cheaper storage media. They are durable, can be written, erased, and re-written. Magnetic tapes are sequential access devices, which mean that the tape needs to rewind or move forward to the location

where the requested data is positioned in the magnetic tape. Due to their sequential nature, magnetic tapes are not suitable for data files that need to be revised or updated often.
They are generally used to store back-up data that is not frequently used or to transfer data from one system to other.

Figure 3.7 A 10.5-inch reel of 9-track tape

The **working of magnetic tape** is explained as follows—

- Magnetic tape is divided horizontally into tracks (7 or 9) and vertically into frames (Figure 3.8). A frame stores one byte of data, and a track in a frame stores one bit. Data is stored in successive frames as a string with one data (byte) per frame.

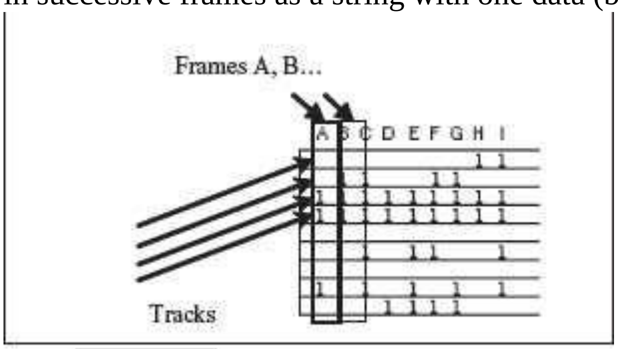

Figure 3.8 A portion of magnetic tape

- Data is recorded on tape in the form of blocks, where a block consists of a group of data also called as records. Each block is read continually. There is an *Inter-Record Gap (IRG)* between two blocks that provides time for the tape to be stopped and started between records (Figure 3.9).

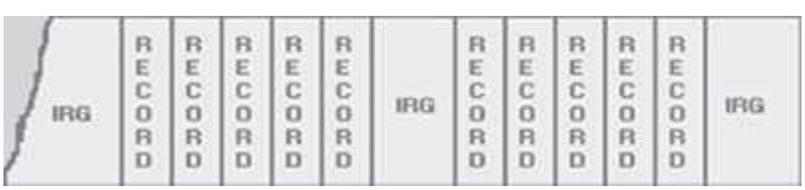

Figure 3.9 Blocking of data in a magnetic tape

- Magnetic tape is mounted on a magnetic tape drive for access. The basic ***magnetic tape drive*** mechanism consists of the supply reel, take-up reel, and the read/write head assembly. The magnetic tape moves on tape drive from the supply reel to take-up reel, with its magnetic coated side passing over the read/write head.
- Tapes are categorized based on their width - ¼ inch, ½ inch, etc.
- The storage capacity of the tape varies greatly. A 10–inch diameter reel of tape which is 2400 feet long can store up to 180 million characters.

The features of magnetic tape are—

- Inexpensive storage device
- Can store a large amount of data
- Easy to carry or transport
- Not suitable for random access data
- Slow access device
- Needs dust prevention, as dust can harm the tape
- Suitable for back-up storage or archiving

3.10 MAGNETIC DISK

Magnetic disk is a direct access secondary storage device. It is a thin plastic or metallic circular plate coated with magnetic oxide and encased in a protective cover. Data is stored on magnetic disks as magnetized spots. The presence of a magnetic spot represents the bit 1 and its absence represents the bit 0.

The working of magnetic disk is explained as follows—

- The surface of disk is divided into concentric circles known as ***tracks***. The outermost track is numbered 0 and the innermost track is the last track. Tracks are further divided into ***sectors***. A sector is a pie slice that cuts across all tracks. The data on disk is stored in sector. Sector is the smallest unit that can be read or written on a disk. A disk has eight or more sectors per track (Figure 3.10).

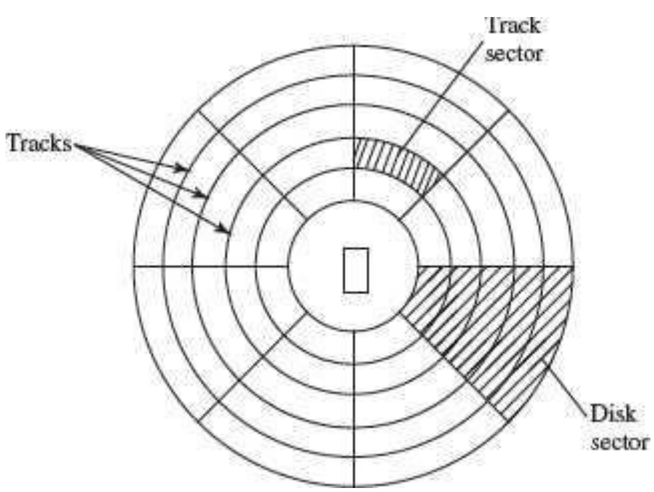

Figure 3.10 Tracks and sectors of a disk

- Magnetic disk is inserted into a magnetic disk drive for access. The drive consists of a read/write head that is attached to a disk arm, which moves the head. The disk arm can move inward and outward on the disk.
- During reading or writing to disk, the motor of disk drive moves the disk at high speed (50–150 times/sec.)
- The storage capacity of disk drive is measured in gigabytes (GB).
- Large disk storage is created by stacking together multiple disks. A set of same tracks on all disks forms a ***cylinder***. Each disk has its own read/write head which work in coordination.
- A disk can also have tracks and sectors on both sides. Such a disk is called ***double-sided disk***.

The features of magnetic disk are—

- Cheap storage device
- Can store a large amount of data
- Easy to carry or transport
- Suitable for frequently read/write data
- Fast access device
- More reliable storage device
- To be prevented from dust, as the read/write head flies over the disk. Any dust particle in between can corrupt the disk.

Finding data on a magnetic disk is as follows—

- In order to use a disk, it has to be formatted. Formatting includes assigning addresses to various locations on disk, assigning location of root directory and checking for defects on the surface of disk.
- During formatting, the tracks and sectors of a disk are labeled, which provides an address to each location of the disk.
- There are different methods to format a disk. File Allocation Table (FAT) is the commonly used logical format for disk formatting performed by Windows.
- Four areas are created when a disk is formatted using FAT—
    - ***Boot Sector*** It contains the program that runs when the computer is started. The program checks if the disk has files required to run the operating system. It then transfers control to an operating system program which continues the startup process. Boot sector also contains information about the disk, like number of bytes per sector and number of sectors per track. This information is required by the operating system to access the data on the disk.
    - ***File Allocation Table*** It records the location of each file and status of each sector. While reading or writing to disk, operating system checks the FAT to find free area or locate where data is stored on disk, respectively.
    - ***Root Directory*** This is the main folder of disk. It contains other folders in it, creating a hierarchical system of folders. The root directory contains information about all folders on the disk.
    - ***Data Area*** The remaining area of the disk (after boot sector, FAT, root directory) is the data area. It stores the program files and data files that are stored on the disk.
- The Windows XP and the Windows 2000 operating system use the New Technology File

System (NTFS) 5 file system. The NTFS 5 file system offers better security and increased performance. It allows using of filenames that are more than eight characters long.

- Floppy disk, hard disk and zip disk are the different types of magnetic disks.

3.10.1 Floppy Disk

- Floppy disk (FD) is a flat, round, single disk made of Mylar plastic and enclosed in square plastic jacket (Figure 3.11).
- Floppy Disk Drive (FDD) is the disk drive for floppy disk.
- The floppy disk is inserted into the floppy disk drive to read or write data to it.
- Floppy disk has a write-protect slide tab that prevents a user from writing to it.
- A floppy disk may be single-sided or double-sided disk, i.e., data can be read and written on one and both sides of floppy disk, respectively.

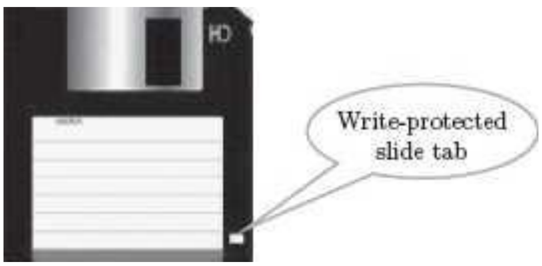

Figure 3.11 Floppy disk

- They are portable. They can be removed from the disk drive, carried or stored separately.
- They are small and inexpensive.
- Floppy disks are slower to access than hard disk. They have less storage capacity and are less expensive than hard disk.
- They come in two basic sizes—5-¼ inch and 3-½ inch.
- The 5-¼ inch disk came around 1987. It can store 350 KB to 1.2 MB of data.
- The 3-½ inch disk has capacity of 400 KB to 1.44 MB. It usually contains 40 tracks and 18 sectors per track and can store 512 bytes per sector.

3.10.2 Hard Disk

- A hard disk (HD) consists of one or more platters divided into concentric tracks and sectors. It is mounted on a central spindle, like a stack. It can be read by a read/write head that pivots across the rotating disks. The data is stored on the platters covered with magnetic coating (Figure 3.12).

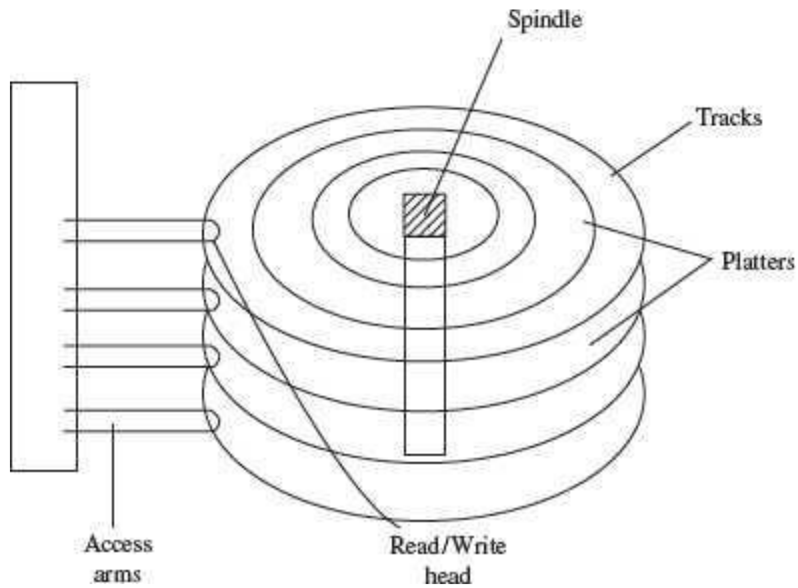

Figure 3.12 Parts of hard disk

- Hard disk is a fixed disk. The disk is not removable from the drive, unlike floppy disk.
- The hard disk and Hard Disk Drive (HDD) is a single unit.
- Hard disk can store much more data than floppy disk. The data in hard disk are packed more closely (because fast spinning uses smaller magnetic charges) and they have multiple platters, with data being stored on both sides of each platter. Large capacity hard disks may have 12 or more platters.
- Unlike floppy disk, the read/write head of hard disk does not touch the disk during accessing.
- Hard disk can spin at the speed of up to 10,000 revolutions per minute and have an access time of 9—14 ms. It stores 512 bytes per sector but the number of sectors are more per track (54 or more) than floppy disk.
- Nowadays, hard disks are available that can store up to 500 GB of data. Generally, PCs come with 150 GB hard disk.
- Hard disk is the key secondary storage device of computer. The operating system is stored on the hard disk. The performance of computer like speed of computer boot up, loading of programs to primary memory, loading of large files like images, video, audio etc., is also dependent on the hard disk.
- Nowadays, *portable external hard disk drive* is available which can be attached to the USB drive of the computer. They come in the storage capacities of 80 GB to 500 GB.

3.10.3 Zip Disk

- They are high-capacity removable disk and drive.
- They have the speed and capacity of hard disk and portability of floppy disk.
- Zip disk are of the same size as floppy disk, i.e., 3–½ inch but have a much higher capacity than the floppy disk (Figure 3.13).

- Zip disk and drive were made by Iomega Corp. It comes as a complete unit—disk, drive, connection cable, power cord and operating system. It can be connected to the computer system externally using a parallel chord or SCSI cable.
- Their capacity ranges from 100 MB to 750 MB. They can be used to store large files, audio and video data.

Figure 3.13 Zip disk

3.11 OPTICAL DISK

Optical disk (Figure 3.14) is a flat and circular disk which is coated with reflective plastic material that can be altered by laser light. Optical disk does not use magnetism. The bits 1 and 0 are stored as spots that are relatively bright and light, respectively.

- An optical disk consists of a single spiral track that starts from the edge to the centre of disk. Due to its spiral shape, it can access large amount of data sequentially, for example music and video. The random access on optical disk is slower than that of magnetic disk, due to its spiral shape.
- The tracks on optical disk are further divided into sectors which are of same length. Thus, the sectors near the centre of disk wrap around the disk longer than the sectors on the edges of disk. Reading the disk thus requires spinning the disk faster when reading near the centre and slower when reading near the edge of disk. Optical disks are generally slower than hard disks. Figure 3.15 shows the tracks and sectors in a magnetic disk and optical disk.

Figure 3.14 Optical disk

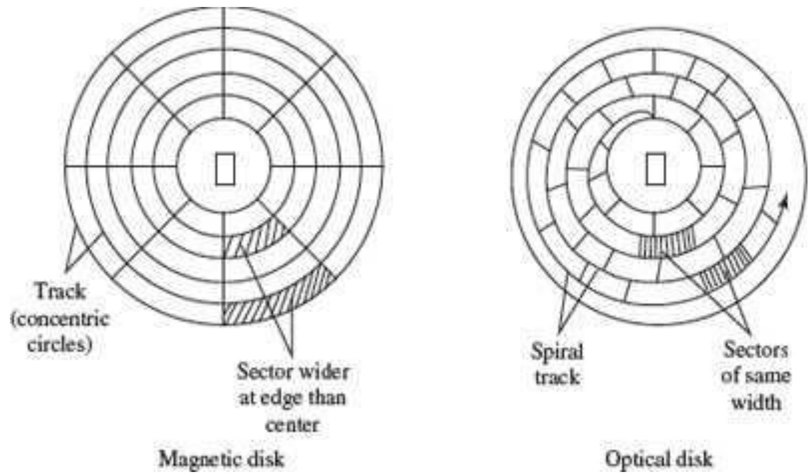

Figure 3.15 Sectors and track in magnetic disk and optical disk

- Optical disks can store large amount of data, up to 5 GB, in a small space. Commonly used optical disks store 500–700 MB of data.
- The access time for an optical disk ranges from 100 to 200 ms.
- There are two most common categories of optical disks—read-only optical disks and recordable optical disks.

3.11.1 CD-ROM

- Originally, Compact Disk (CD) was a popular medium for storing music. Now, it is used in computers to store data and is called Compact Disk-Read Only Memory (CD-ROM).
- As the name suggests, CD-ROM (Figure 3.15) is an optical disk that can only be read and not written on. CD-ROM is written on by the manufacturer of the CD-ROM using the laser light.
- A CD-ROM drive reads data from the compact disk. Data is stored as pits (depressions) and lands (flat area) on CD-ROM disk. When the laser light is focused on the disk, the pits scatter the light (interpreted as 0) and the lands reflect the light to a sensor (interpreted as 1).
- As CD-ROM is read only, no changes can be made into the data contained in it.
- Since there is no head touching the disk, but a laser light, CD-ROM does not get worn out easily.
- The storage density of CD-ROM is very high and cost is low as compared to floppy disk and hard disk.
- Access time of CD-ROM is less. CD-ROM drives can read data at 150Kbps. They come in multiples of this speed like—2x, 4x, 52x, 75x, etc.
- It is a commonly used medium for distributing software and large data.

3.11.2 DVD-ROM

- Digital Video Disk-Read Only Memory (DVD-ROM) is an optical storage device used to store digital video or computer data (Figure 3.17).
- DVDs look like CDs, in shape and physical size.

Figure 3.15 CD-ROM

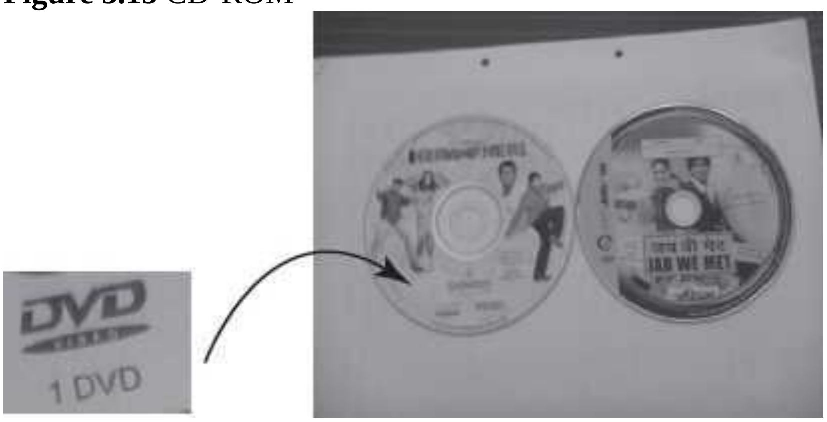

Figure 3.17 DVDs

- It improves on CD technology.
- It is a high-density medium with increased track and bit density.
- DVD-ROM uses both sides of the disk and special data compression technologies. The tracks for storing data are extremely small.
- A full-length movie can be stored on a single disk.
- Each side of DVD-ROM can store 4.7 GB of data, so a single DVD can store 9.4 GB of data.
- New DVD-ROMs use layers of data track, to double its capacity. Such dual layer disks can store 17 GB of data.

3.11.3 Recordable Optical Disk

In addition to the read only CDs and DVDs, recordable optical disks (Figure 3.18) are also available. Users can record music, video, audio and data on it. The recordable optical disks are—

- ***Compact Disk-Recordable (CD-R)*** is a Write Once-Read Many (WORM) disk. A CD-R disk allows the user to write data permanently on to the disk. Once the data is written, it cannot be erased. CD-R disk uses a laser that burns pits into the disk surface. It looks like a CD disk externally. To write to a CD-R disk, a device named CD-Writer or CD–burner is required. A CD-R disk can store 700 MB of data that can run for 80 minutes. CD-R is used to create music CDs in home computers, back up data from other storage devices, archives of large data, etc.

- ***Compact Disk-ReWritable (CD-RW)*** allows data to be written, erased and re-written on. The capacity of CD-RW is same as a CD. They generally do not play on all CD-ROM drives.
- ***Digital Video Disk-Recordable (DVD-R)*** allows recording of data on a DVD. A DVD writer device is required to write the data to DVD. The data once written on a DVD cannot be erased or changed.

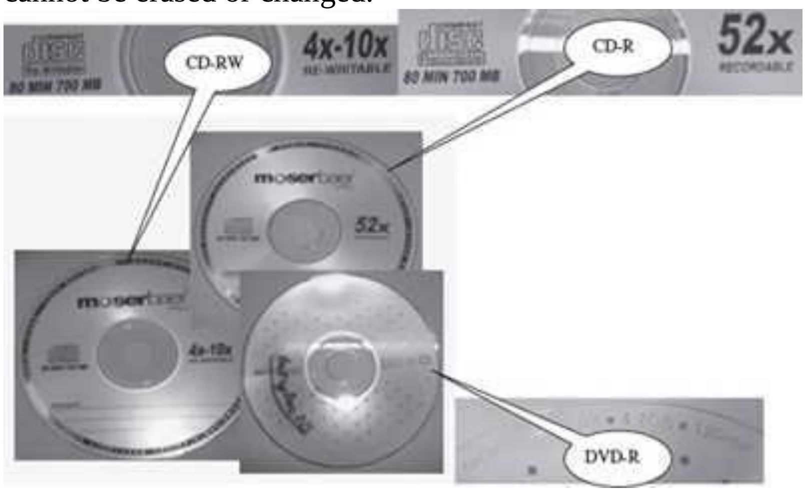

Figure 3.18 CD-R, CD-RW and DVD-R

3.12 MAGNETO-OPTICAL DISK

- Magneto-optical disks use laser beam to read data and magnetic field to write data to disk. • These are optical disks where data can be written, erased and re-written.
- They are expensive and outdated. They were used during the mid 1990s. They have now been replaced by CD-RW and DVD-R.

3.13 USING THE COMPUTER MEMORY

The computer starts using the memory from the moment the computer is switched on, till the time it is switched off. The list of steps that the computer performs from the time it is switched on are—

- Turn the computer on.
- The computer loads data from ROM. It makes sure that all the major components of the computer are functioning properly.
- The computer loads the BIOS from ROM. The BIOS provides the most basic information about storage devices, boot sequence, security, plug and play capability and other items.
- The computer loads the OS from the hard drive into the system's RAM. CPU has immediate access to the OS as the critical parts of the OS are maintained in RAM as long as the computer is on. This enhances the performance and functionality of the overall system.
- Now the system is ready for use.
- When you load or open an application it is loaded in the RAM. Since the CPU looks for information in the RAM, any data and instructions that are required for processing (read,

write or update) is brought into RAM. To conserve RAM usage, many applications load only the essential parts of the program initially and then load other pieces as needed. Any files that are opened for use in that application are also loaded into RAM.

- The CPU requests the data it needs from RAM, processes it and writes new data back to RAM in a continuous cycle. The shuffling of data between the CPU and RAM happens millions of times every second.
- When you save a file and close the application, the file is written to the secondary memory as specified by you. The application and any accompanying files usually get deleted from RAM to make space for new data.
- If the files are not saved to a storage device before being closed, they are lost.

Sometimes, when you write a program and the power goes off, your program is lost if you have not saved it. This is because your program was in the RAM and was not saved on the secondary memory; the content of the RAM gets erased when the power is switched off.

CHAPTER 4
INPUT AND OUTPUT DEVICES

4.1 INTRODUCTION

A computer interacts with the external environment via the input-output (I/O) devices attached to it. Input device is used for providing data and instructions to the computer. After processing the input data, computer provides output to the user via the output device. The I/O devices that are attached, externally, to the computer machine are also called *peripheral devices*. Different kinds of input and output devices are used for different kinds of input and output requirements. In this chapter, we shall discuss different kinds of input devices and output devices.

4.2 INPUT-OUTPUT UNIT

An I/O unit is a component of computer. The I/O unit is composed of two parts—input unit and output unit. The input unit is responsible for providing input to the computer and the output unit is for receiving output from the computer.

4.2.1 Input Unit

- The input unit gets the data and programs from various input devices and makes them available for processing to other units of the computer.
- The input data is provided through input devices, such as—keyboard, mouse, trackball and joystick. Input data can also be provided by scanning images, voice recording, video recording, etc.
- Irrespective of the kind of input data provided to a computer, all input devices must translate the input data into a form that is understandable by the computer, i.e., in machine readable form. The transformation of the input data to machine readable form is done by the input interface of input device.

In brief, the input unit accepts input data from the user via input device, transforms the input data in computer acceptable form using input interface for the input device and provides the transformed input data for processing.

4.2.2 Output Unit

- The output unit gets the processed data from the computer and sends it to output devices to make them available to the user of computer.
- The output data is provided through output devices like display screen, printer, plotter and speaker.
- The processed data sent to the output device is in a machine understandable form. This processed data is converted to human readable form by the *output interface* of output device.

In brief, the output unit accepts output data from computer via output device, transforms the output information to human readable form using the output interface of output device and provides the transformed output to user.

In addition to input devices and output devices, some devices function as both input and output devices. The I/O devices provide the input to computer as well as get output from computer. The I/O devices are used by both the input unit and the output unit. Hard disk drive, floppy disk drive, optical disk drives are examples of I/O devices. Table 4.1 lists the different I/O devices.

Input Devices	Keyboard, Mouse, Digitizing Tablet, Track Ball, Joystick, TouchScreen, Light Pen, Speech Recognition System, Digital camera, Scanner, Magnetic Ink Character Recognition (MICR), Optical Character Recognition (OCR), Optical Mark Recognition (OMR), Barcode Reader
Output Devices	Monitor, Visual Display Terminal, Printer, Plotter, Computer Output on Microfi lm (COM), Video Output System, Audio Response System
Input-Output Devices	Hard disk drive, Floppy disk drive, USB drive, CD drive, DVD drive

Table 4.1 I/O devices

4.3 INPUT DEVICES

Input devices allow users and other applications to input data into the computer, for processing. The data input to a computer can be in the form of text, audio, video, etc. Figure 4.1 shows some users working in an office. The data is entered manually by the user or with minimal user intervention. Input devices are classified as follows—

- Human data entry devices
  - Keyboard
  - Pointing devices—mouse, trackball, joystick, digitizing tablet
  - Pick devices—light pen, touch screen
- Source data entry devices
  - Audio input—speech recognition
  - Video input—digital camera
  - Scanner—hand-held scanner, flat-bed scanner
  - Optical Scanner—OCR, OMR, MICR, barcode reader

The input is provided to the computer using an input device, and must be translated to a form that the computer can understand. The translation is done by the input interface of the input device.

In addition to the above devices, the input to a computer can also be provided from a storage device on the computer, another computer, or another piece of equipment, such as a musical instrument, thermometer or sensors.

4.4 HUMAN DATA ENTRY DEVICES

Input devices that require data to be entered manually to the computer are identified as human data entry devices. The data may be entered by typing or keying in, or by pointing a device to a particular location.

Keyboard

Features Keyboard is a common input device. It is provided along with the computer, and is easy to use. It is used for entering the text data. For inputting the data, the user types the data using the keyboard. When the data is being typed, the display monitor displays the typed data.
Cursor is a vertical line, an underscore, blinking line, etc. Cursor moves with each typed character. The position of cursor indicates the location on monitor where the typed-in character will be displayed. A keyboard is shown in Figure 4.2.

Description The design of a keyboard is similar to a standard typewriter. The modern keyboards are QWERTY keyboard (Q, W, E, R, T, Y are the sequence of keys in top row of letters).
Standard keyboard contains 101 keys which are arranged in the same order as a typewriter. The keyboard has five sections (1) Typing keys (1, 2, 3…, A, B, C…), (2) Numeric keypad (numeric keys on right side), (3) Function keys (F1, F2…. on top side), (4) Control keys (cursor keys, ctrl, alt.…), and (5) Special-purpose keys (Enter, shift, spacebar…). Some keyboards have 110 keys, where the extra keys are designed to work with the Windows operating system.

Working When a key is pressed, keyboard interacts with a keyboard controller and keyboard buffer. The keyboard controller stores the code of pressed key in keyboard buffer and informs the computer software that an action has happened on the keyboard. The computer software checks and reads the keyboard buffer and passes the code of pressed character to the system software. Due to a time gap between pressing of a key on keyboard and reading by the system software, keyboard buffer is designed to store many keystrokes together.

Figure 4.2 Keyboard

4.4.1 Pointing Devices

Pointing devices are used for providing the input to computer by moving the device to point to a location on computer monitor. The input data is not typed; instead, the data is entered by moving the pointing device. The cursor on the computer monitor moves with the moving pointing device. Operations like move, click and drag can be performed using the pointing devices. Mouse, trackball, joystick and digitizing tablet are some of the common pointing devices.

Mouse

Features It is the most common pointing input device. The data is entered by pointing the mouse to a location on the computer screen. The mouse may also be used to position the cursor on screen, move an object by dragging, or select an object by clicking. The key benefit of using a mouse is that the cursor moves with the mouse. So, the cursor can be positioned at any location on the screen by simply moving the mouse. Moreover, it provides an easy way to select and choose commands from menus, dialog boxes, icons, etc. Mouse is used extensively, while working with graphics elements such as line, curve, shapes, etc.

Description Mouse is a small hand-held device having two or three buttons on its upper side. In addition to the buttons, mouse also has a small wheel between the buttons. Figure 4.3 (i) shows a mouse. The wheel of the mouse is used for the up and down movement, for example, scrolling a long document. A mouse is classified as physical mouse or optical mouse.

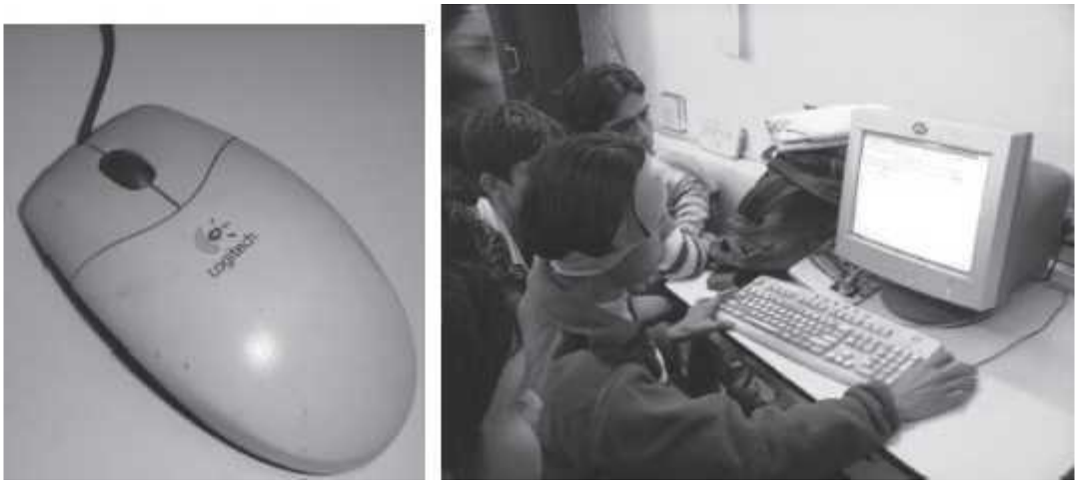

Figure 4.3 (i) Mouse, (ii) A user working with a mouse

Physical Mouse has a rubber ball on the bottom side that protrudes when the mouse is moved. It requires a smooth, dust free surface, such as a mouse pad, on which it is rolled.

Optical Mouse uses a Light Emitting Diode (LED) and a sensor to detect the movement of mouse. Optical mouse requires an opaque flat surface underneath it. Optical mouse was introduced by Microsoft in 1999. Optical mouse is better than physical mouse as there is no moving part that can cause wear and tear, and dirt cannot get inside it. A user is working with an optical mouse in Figure 4.3 (ii).

Working In a ***physical mouse***, rollers and sensors are used to sense the direction and rate of movement of mouse. When the ball of mouse moves, the rollers sense the horizontal and vertical movement and sensors sense the speed of movement. This information is passed to computer via the mouse chord. When an *optical mouse* is moved, a beam of light is reflected from its underside. These pulses of light determine the direction and rate of movement. This information is sent to computer via the mouse chord.

Using the mouse The mouse can be used in five different ways, as follows—

- ***Pointing*** points to a location or object on the computer screen. Moving the mouse by hand moves the cursor on computer screen. The cursor moves in the direction in which the mouse moves.
- ***Left Click or Click*** means pressing the left button of mouse and releasing it. Clicking is used to select a button, command or icon on the screen.
- ***Right Click*** involves pressing the right button on mouse and releasing it. Right click displays a menu that contains options like cut, copy, paste, font, paragraph, etc. for the item on which the mouse is pointing.
- ***Double Click*** means pressing the left button of mouse twice successively, without moving the mouse, and then releasing the mouse. It is used to start a program or open a folder.
- ***Drag and Drop*** drags an object and drops it at another location. Drag means pointing mouse to an object on screen, pressing the left button of mouse, keeping it pressed and moving the mouse to point to a new location. The object gets dragged to the new location along with the mouse. When the left button of mouse is released, the object gets dropped at the new location. Drag and drop is used for moving folders, files and icons to new locations on the screen.

TrackBall

Features Trackball is a device that is a variant of the mouse but has the functionality of mouse. It is easy to use and takes less space than a mouse. Trackball is generally built in laptops since there is no space for the mouse to move on the lap. Trackballs come in various sizes—small and big.

Description Trackball looks like an upside-down mouse. Instead of moving the whole device to move the cursor on computer screen, trackball requires the ball to be rotated manually with a finger. The trackball device remains stationary. The cursor on the computer screen moves in the direction in which the ball is moved. The buttons on trackball are used in the same way as mouse buttons. A trackball is shown in Figure 4.4.

Working Trackball works in the same way as a physical mouse.

Figure 4.4 Trackball

Joystick

Features Joystick (Figure 4.5) is a device which is commonly used for playing video games. Joystick is mainly used to control the speed of the cursor and is thus popular in games involving speed like racing and flying games. The direction of push of the stick and the amount of deflection determines the change in position and the change in speed, respectively.

Description It is a stick with its base attached to a flexible rubber sheath inside a plastic cover. The plastic cover contains the circuit that detects the movement of stick and sends the information to computer. The position of the stick movement is given by the x and y coordinates of the stick.

.

Figure 4.5 Joystick

Digitizing Tablet

Features It is an input device used primarily to input drawings, sketches, etc. Digitizing tablet is used for Computer Aided Design (CAD) for the design of buildings, automotive designs, and designing of maps, etc. Figure 4.5 shows a digitizing tablet.

Description Digitizing tablet consists of two parts—electronic tablet and pen. The electronic tablet is a flat bed tablet. The pen looks like a ball pen but has an electronic head. The pen in moved on the tablet. Each position on the tablet corresponds to a fixed position on the screen. Drawings can be made on the tablet using a pen, and is provided as input to computer, where, a location on the tablet corresponds to a specific location on the screen.

Working The tablet contains circuit that can detect the movement of pen on the tablet, convert the movements into digital signals and send the digital signal to the computer.

Figure 4.5 Digitizing tablet

Pick Devices

Pick devices are used for providing input to the computer by pointing to a location on the computer monitor. The input data is not typed; the data is entered by pointing the pick device directly on the computer screen. Light pen and touch screen are some common pick devices.

Light Pen

Features It is a light sensitive pen-like input device and is used to select objects directly on the computer screen. It is used for making drawing, graphics and for menu selection. Figures and drawings can be made by moving the pen on computer screen.

Description and Working The pen contains a photocell in a small tube. When the pen is moved on the screen, light from the screen at the location of pen causes the photocell to respond. The electric response is transmitted to the computer that can identify the position on screen at which the light pen is pointing. Figure 4.7 shows a user using a light pen on the screen.

Figure 4.7 Using a light pen

Touch Screen

Features It is an input device that accepts input when the user places a fingertip on the computer screen. The computer selects the option from the menu of screen to which the finger points.

Touch screen are generally used in applications like Automated Teller Machine (ATM), public information computers like hospitals, airline reservation, railway reservation, supermarkets, etc.

Figure 8.4 Touch screen of an ATM

Description Touch screen consists of a clear glass panel that is placed over the view area of computer screen. In addition to the glass panel with sensors, it has a device driver, and a controller that translates the information captured by the glass panel sensors to a form that the computer can understand.

Working Touch screens have an infrared beam that criss-cross the surface of screen. When a fingertip is touched on the screen, the beam is broken, and the location is recorded. Some touch screens have ultrasonic acoustic waves that cross the surface of screen. When a fingertip is touched on the screen, the wave is interrupted, and the location is recorded. The recorded location is sent to the computer via the controller of touch screen, in a form that the computer can understand.

4.5 SOURCE DATA ENTRY DEVICES

Source data entry devices are used for audio input, video input and to enter the source document directly to the computer. Source data entry devices do not require data to be typed-in, keyed-in or pointed to a particular location.

4.5.1 Audio Input Device

Audio input can be provided to the computer using human voice or speech. Audio input to the computer can be used for different purposes. It can be used for making telephone calls, for audio and video conferencing over Internet, to record voice, to create audio files and embed these files to be sent over e-mail, or, to translate spoken words into text, etc.

Audio input devices like a *microphone* is used to input a person's voice into the computer. *A sound card* (Figure 4.9 (i)) translates analog audio signals from microphone into digital codes that the computer can store and process. Sound card also translates back the digital sound into analog signals that can be sent to the speakers. Translating spoken words into text is also known as *speech recognition or voice recognition*. The audio input along with the software for voice recognition forms the speech recognition system or voice recognition system.

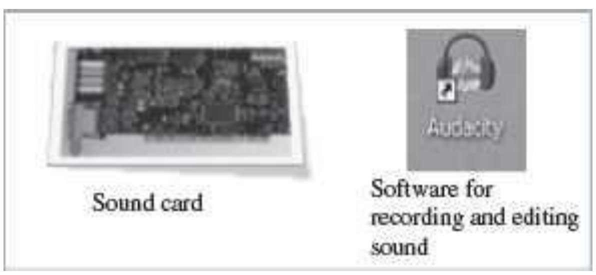

Figure 4.9 (i) Sound card, (ii) Audacity software

The computer can be operated using voice commands. The user can dictate the commands to the computer, instead of typing them. The computer has to be trained to recognize the voice of user using the speech patterns and pronunciation of words. The system thus adapts to the voice of user. Speech recognition systems are costly and difficult to develop. They are generally used by people who have difficulty in typing, people with disabilities or by corporate world for dictation. Audio input can be recorded on an mp3 recorder and provided as an input to computer. Open source software like Audacity is used for recording and editing of audio files (Figure 4.9 (ii)).

4.5.2 Video Input Device

Video input is provided to the computer using *video camera and digital camera* (Figure 4.10). Video camera can capture full motion video images. The images are digitized and can be compressed and stored in the computer disk. Webcam is a common video camera device. It is placed on the computer above the screen to capture the images of the user who is working on the computer. A video capture card allows the user to connect video devices like camcorders to the computer.

Figure 4.10 Video input devices

Digital camera works like video camera but can capture still images. The digital camera digitizes images, compresses them and stores them on a memory card like flash memory. The information from the digital camera can be brought into the computer and stored. The video files can be edited using software like VLC media player. Computer vision is an area of computer science that deals with images. Computer vision has applications in areas like robotics and industrial processing.

4.5.3 Optical Input Devices

Optical input devices allow computers to use light as a source of input. Scanner is an example of optical input device. Other common optical input devices are magnetic ink character reader used for Magnetic Ink Character Recognition (MICR), optical mark reader used for Optical Mark Recognition (OMR), optical character reader for Optical Character Recognition (OCR) and Barcode Reader.

4.5.3.1 Scanner

Scanner is an input device that accepts paper document as an input. Scanner is used to input data directly into the computer from the source document without copying and typing the data. The input data to be scanned can be a picture, a text or a mark on a paper. It is an optical input device and uses light as an input source to convert an image into an electronic form that can be stored on the computer. Scanner accepts the source paper document, scans the document and translates it into a bitmap image to be stored on the computer. The denser the bitmap, the higher is the resolution of the image. The quality of scan increases with the increase in resolution. Scanners come with utility software that allow the stored scanned documents to be edited, manipulated and printed. Hand-held scanner and flat-bed scanner are the two common types of scanners.

- ***Hand-held Scanners*** are portable and are placed over the document to be scanned. They consist of light emitting diodes. The scanned documents are converted and stored as an image in the computer memory. Hand-held scanners have to be moved at a constant speed over the document to be scanned, to get good quality scans. They are preferably used for low volume of documents, small pictures or photos. They are difficult to use if there is a need to scan a full page document. Some of the documents that are primarily scanned using hand-held scanners are price tags, label and ISBN number on books.

- ***Flat-bed Scanners*** provide high quality scan in a single pass. It is a box shaped machine similar to a photocopy machine and has a glass top and a lid that covers the glass (Figure 4.11). The document to be scanned is placed on the glass top, which activates the light beam beneath the glass top and starts the scan from left to right. They are largely used to scan full page documents.

Figure 4.11 Flat bed scanner

4.5.3.2 Optical Character Recognition (OCR)

OCR is a technique for the scanning of a printed page, translating it, and then using the OCR software to recognize the image as ASCII text that is editable. OCR uses optical character reader for recognition. The optical character reader stores the scanned image as bitmap image which is a grid of dots. Thus, you cannot edit the text that has been scanned. To edit the scanned text, you need OCR software. The OCR software translates the array of dots into text that the computer can interpret as words and letters. To recognize the words and letters of text, the OCR software compares the pattern on the scanned image with the patterns stored inside the computer. The text files generated via OCR can be stored in different formats. Figure 4.12 shows the working of the OCR system.

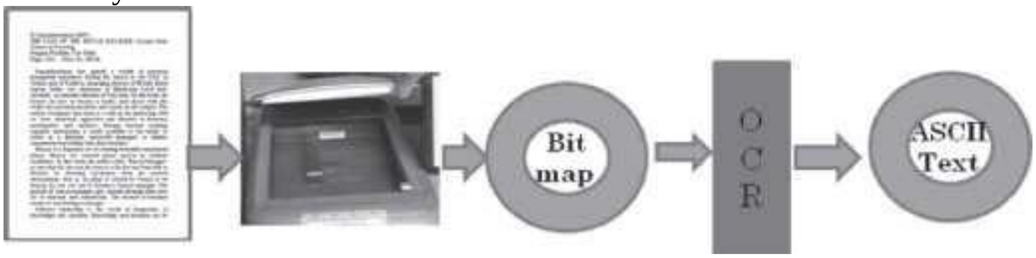

Figure 4.12 OCR system

4.5.3.3 Magnetic Ink Character Recognition (MICR)

MICR is used in banks to process large volumes of cheques (Figure 4.13). It is used for recognizing the magnetic encoding numbers printed at the bottom of a cheque. The numbers on the cheque are human readable, and are printed using an ink which contains iron particles. These numbers are magnetized. MICR uses magnetic ink character reader for character recognition. When a cheque is passed through Magnetic Ink Character Reader, the magnetic field causes the read head to recognize the characters or numbers of cheque. The readers are generally used in banks to process the cheques. The numbers in the bottom of the cheque include the bank number, branch number and cheque number. The reading speed of MICR is faster than OCR.

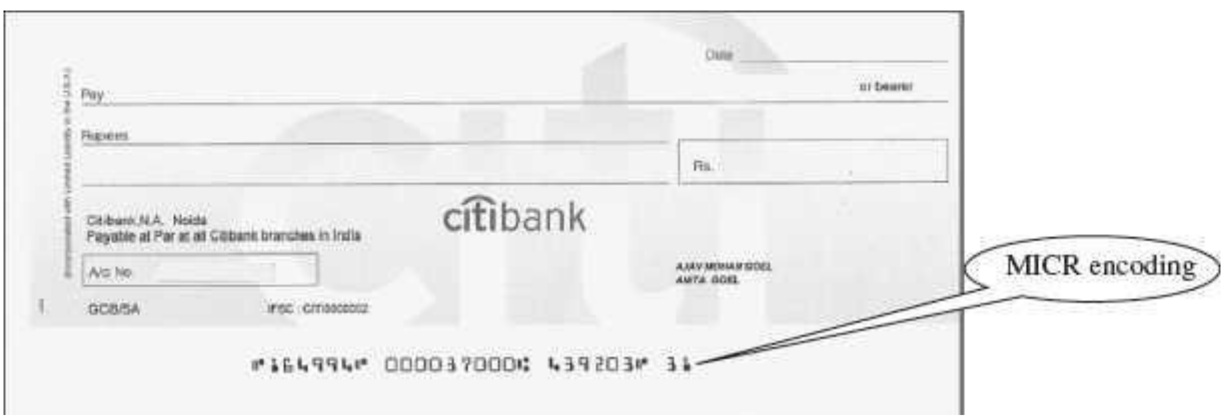

Figure 4.13 MICR encoded cheque

4.5.3.4 Optical Mark Recognition (OMR)

OMR is used to detect marks on a paper. The marks are recognized by their darkness. OMR uses an optical mark reader to read the marks. The OMR reader scans the forms, detects the mark that is positioned correctly on the paper and is darker than the surrounding paper, and passes this information to the computer for processing by application software. For this, it uses a beam of light that is reflected on the paper with marks, to capture presence and absence of marks. The optical mark reader detects the presence of mark by measuring the reflected light. The pattern of marks is interpreted and stored in the computer.

OMR is widely used to read answers of objective type tests, where the student marks an answer by darkening a particular circle using a pencil. OMR is also used to read forms, questionnaires, order forms, etc. Figure 4.14 shows a marked OMR answer sheet.

4.5.3.5 Barcode Reader

Barcodes are adjacent vertical lines of different width that are machine readable. Goods available at supermarkets, books, etc. use barcode for identification. Barcodes are read using reflective light by barcode readers. This information is input to the computer which interprets the code using the spacing and thickness of bars. Hand-held barcode readers are generally used in departmental stores to read the labels, and in libraries to read labels on books. Figure 4.15 (i) shows a barcode printed at the back of a book.

Barcode readers (Figure 4.15 (ii)) are fast and accurate. They enable faster service to the customer and are also used to determine the items being sold, number of each item sold or to retrieve the price of item.

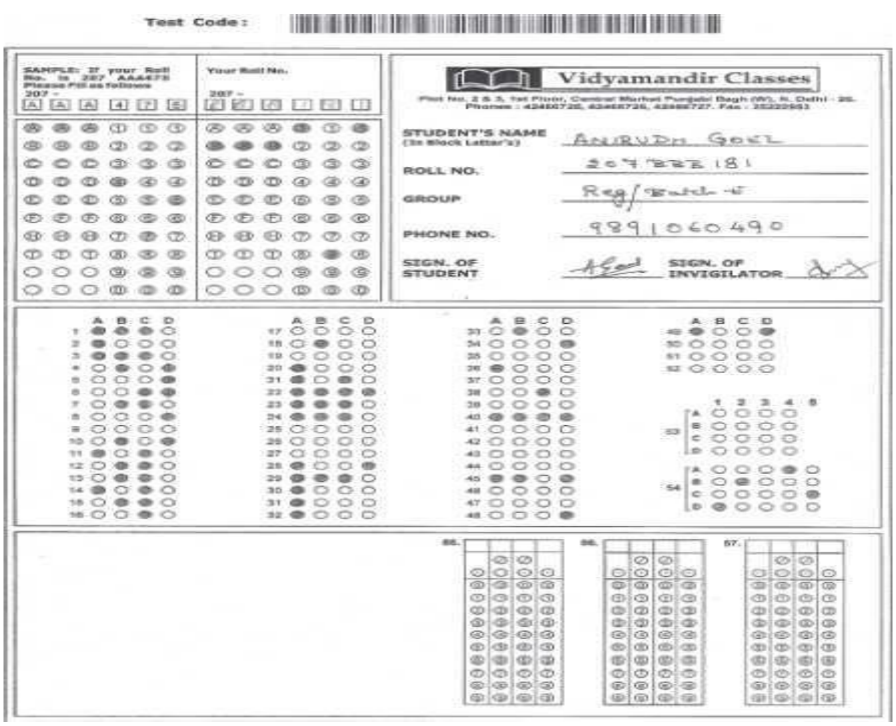

Figure 4.14 OMR answer sheet

Figure 4.15 (i) Barcode of a book, (ii) Barcode reader

4.6 OUTPUT DEVICES

Output devices provide output to the user, which is generated after processing the input data. The processed data, presented to the user via the output devices could be text, graphics, audio or video. The output could be on a paper or on a film in a tangible form, or, in an intangible form as audio, video and electronic form. Output devices are classified as follows—

- Hard Copy Devices
  - Printer
  - Plotter
  - Computer Output on Microfilm (microfiche)
- Soft Copy Devices
  - Monitor
  - Visual Display Terminal
  - Video Output
  - Audio Response

Figure 4.15 (i) shows a soft copy output on an LCD monitor and Figure 4.15 (ii) is a hard copy output on paper. The output device receives information from computer in a machine readable form. The received output is translated to a human understandable form. The translation is done using the output interface of output device.

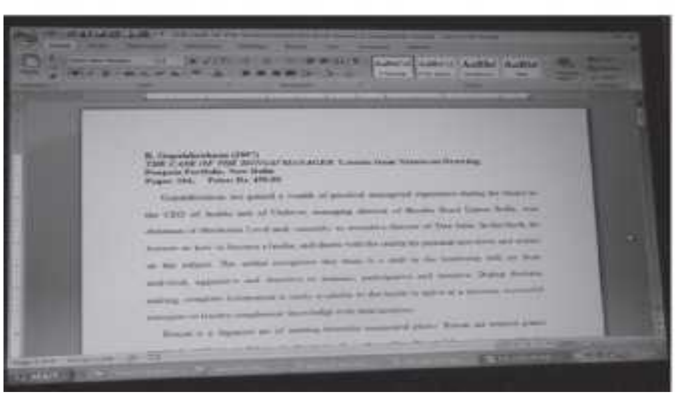


Figure 4.15 (i) Soft copy output, (ii) Hard copy output

4.6.1 Hard Copy Devices

The output obtained in a tangible form on a paper or any surface is called hard copy output. The hard copy can be stored permanently and is portable. The hard copy output can be read or used without a computer. The devices that generate hard copy output are called hard copy devices. Printer, plotter and microfiche are common hard copy output devices.

4.6.1.1 Printer

A printer prints the output information from the computer onto a paper. Printers are generally used to print textual information, but nowadays printers also print graphical information. The print quality (sharpness and clarity of print) of the printer is determined by the resolution of the printer. Resolution is measured in dots per inch (dpi). Printers with a high resolution (more dpi) provide better quality output. Different kinds of printers are available for different types of applications. Printers are classified into two categories—impact printer and non-impact printer.

Impact printers use the typewriter approach of physically striking a typeface against the paper and inked ribbon. Impact printers can print a character or an entire line at a time. Impact printers are low-cost printers useful for bulk printing. Dot matrix printers, daisy wheel printers and drum printers are examples of impact printers.

- ***Dot Matrix Printers*** (Figure 4.17) print one i character at a time. The speed of dot matrix printer lies between 200 and 500 characters per second (cps) and their resolution ranges from 72 to 350 dpi. Dot matrix printers normally come in two sizes—80 column printer and 132 column printer. Dot matrix printers can print alphanumeric characters, special characters, charts and graphs. They can print only in black and white. Some dot matrix printers can print in both directions - left to right and right to left. Dot matrix printers are commonly used for printing in applications like payroll and accounting.

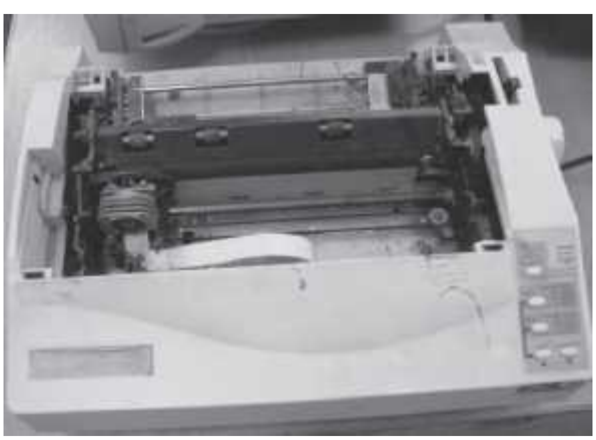

Figure 4.17 Dot matrix printer

- ***Daisy Wheel Printers*** (Figure 4.18) print one character at a time. They produce letter quality document which is better than a document printed by a dot matrix printer. The speed of daisy wheel printers is about 100 cps. The print head of the printer is like a daisy flower, hence the name. These printers are slow, can only print text (not graphics), and are costly in comparison to dot matrix printers. Daisy wheel printers are used where high quality printing is needed and no graphics is needed.

Figure 4.18 Daisy wheel for printers

- ***Drum Printers are line printers*****.** They are expensive and faster than character printers but produce a low quality output. They can print 200–2500 lines per minute. Drum printers are generally used for voluminous print outputs.

Non-Impact Printers do not hit or impact a ribbon to print. They use electro-static chemicals and ink-jet technologies. Non-impact printers are faster and quieter than impact printers. They produce high quality output and can be used for printing text and graphics both in black and white, and color. Ink-jet printers and laser printers are non-impact printers.

- ***Ink-jet Printers*** spray ink drops directly on the paper like a jet (Figure 4.19 (i)). Their resolution is more than 500 dpi. They produce high quality graphics and text. Ink-jet printers are commonly found in homes and offices.
- ***Laser Printers*** (Figure 4.19 (ii)) provide highest quality of text and graphics printing. Laser printers process and store the entire page before printing and are also known as *page printers*. The laser printer can print 5–24 pages of text per minute and their resolution ranges from 400 to 1200 dpi. They are faster and expensive than impact printers. Laser printers are used in applications requiring high quality voluminous printing.

Figure 4.19 (i) Inkjet printer, (ii) Laser printer

4.6.1.2 Plotter

A plotter (Figure 4.20) is used for vector graphics output to draw graphs, maps, blueprints of ships, buildings, etc. Plotters use pens of different colors (cyan, magenta, yellow and black) for drawing. Plotters draw continuous and accurate lines, in contrast to printers where a line is drawn as closely spaced dots. Plotter is a slow output device and is expensive. Plotters are of two kinds—drum plotter and flatbed plotter. In a *drum plotter*, pens mounted on the carriage are stationary and move only horizontally; for vertical movement, the drum on which the paper is fixed moves clockwise and anti-clockwise. In a *flatbed plotter*, the paper is fixed on a flat bed.
The paper is stationary and the pens mounted on the carriage move horizontally and vertically to draw lines. Plotters are mainly used for drawings in AUTOCAD (computer assisted drafting), Computer Aided Design (CAD) and Computer Aided Manufacturing (CAM) applications

Figure 4.20 Plotter

4.6.1.3 Computer Output on Microfilm

A microfilm (Figure 4.21) is in a fiche or roll format, and is used to record computer output directly from the computer tape or cartridge. Computer Output on Microfilm (COM) is a high speed and low cost process. It can produce data in microfilm form at a much faster speed from that of a paper printer. The standard roll film is 15 mm wide with a film image that is 1/24 of the original document. The copy of the image on microfilm retains its original clarity. Microfilm can be indexed to facilitate retrieving information from it. For reading images stored on microfilm, a microfilm reader is used. A screen is used for viewing the enlarged images. COM is suited for storing large amounts of data for manuals and archive records for long periods of time that have to be referenced occasionally. COM is used for storing output in banking and insurance applications, medical X rays, etc.

Figure 4.21 Microfilm

4.6.2 Soft Copy Devices

The output obtained in an intangible form on a visual display, audio unit or video unit is called soft copy output. The soft copy allows corrections to be made, can be stored, and, can be sent via E– to other users. The soft copy output requires a computer to be read or used. The devices that generate soft copy output are called soft copy devices. Visual output devices like computer monitor, visual display terminal, video system and audio response system are common soft copy output devices.

4.6.2.1 Monitor

Monitor is a common output device. The monitor is provided along with the computer, to view the displayed output. A monitor is of two kinds - monochrome display monitor and color display monitor. A monochrome display monitor uses only one color to display text and color display monitor can display 255 colors at one time. The number of colors displayed by a color monitor varies with the kind of color adapter attached to it—CGA, EGA, VGA, XGA and SVGA. The CGA monitor supports four colors and SVGA supports around 15,000,000 colors. Monitors are available in various sizes like 14, 15, 17, 19 and 21 inches.

An image on the monitor is created by a configuration of dots, also known as pixels. The clarity of image on the computer screen depends on three factors—

1. ***Resolution of Screen***—the number of pixels in horizontal and vertical direction. More the number of pixels, the sharper is the image. The common resolution of computer screen is 800x500 and 1024x758,

2. ***Dot Pitch***—the diagonal distance between two colored pixels on a display screen, and
3. ***Refresh Rate***—the number of times per second the pixels are recharged so that their glow remains bright.

Monitors may be *Cathode Ray Tube* (CRT) monitors (Figure 4.22) that look like a television or *Liquid Crystal Display* (LCD) monitors (Figure 4.23, Figure 4.24) that have a high resolution, flat screen, flat panel display. Nowadays, LCD monitors are generally used.

4.6.2.2 Visual Display Terminal

A monitor and keyboard together are known as *Visual Display Terminal* (VDT). A keyboard is used to input data and monitor is used to display the output from the computer. The monitor is connected to the computer by a cable. Terminals are categorized as dumb, smart and intelligent terminals. The dumb terminals do not have processing and programming capabilities. Smart terminals have built-in processing capability but do not have its own storage capacity. Intelligent terminals have both built-in processing and storage capacity.

Figure 4.22 CRT monitor

4.6.2.3 Video Output

Screen image projector or data projector (Figure 4.25 (i)) is an output device that displays information from the computer onto a large white screen. The projector is mainly used to display visual output to a large gathering of people required for the purposes of teaching, training, meetings, conference presentations, etc. (Figure 4.25 (ii)).

4.6.2.4 Audio Response

A complete sound system consists of sound card, microphone, speaker and the appropriate software. In addition to recording and playing the sound, the software allows editing of sound, like cutting, copy, amplification and creation of vibrant sound effects.

Audio response provides audio output from the computer. Audio output device like *speakers, headset or headphone* (Figure 4.25) is used for audio output sound from computer. The signals are sent to the speakers via the sound card that translates the digital sound back into analog signals. The audio response from the computer may be generated by synthesizing the input human speech to give audio output, or may be a result of a set of rules that are used to create artificial speech.

Audio output is commonly used for customer service in airlines, banks, etc. It is also used in video conferences, surveys, etc. Audio response is used by visually impaired to read information from the screen. For speech impaired people, audio response helps them to communicate with other people.

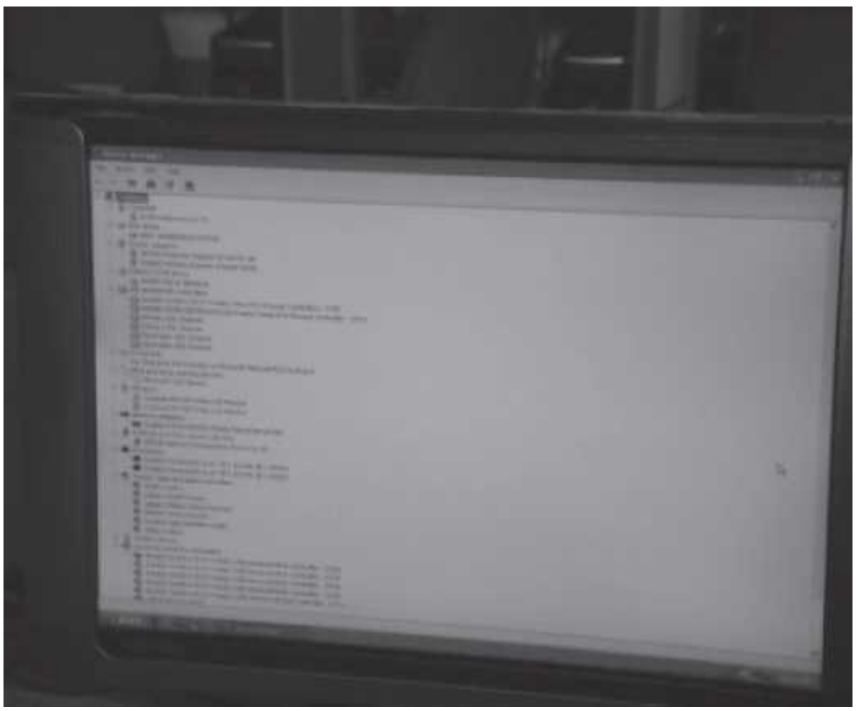

Figure 4.23 LCD monitor

Figure 4.24 A user viewing the output on a LCD monitor

Figure 4.25 (i) LCD projector, (ii) A presentation in progress using LCD projector

Figure 4.25 Headphone and speakers

4.7 I/O PORT

The peripheral devices can be connected to computer in several ways. Devices such as network adapters and sound cards are connected to expansion slots inside the computer. Printers and scanners are connected to ports on the backside of the computer. Also in a portable computer, the PC Card connects to the PC Card slot on it.

The I/O ports are the external interfaces that are used to connect input and output devices like printer, modem and joystick to the computer. The I/O devices are connected to the computer via the serial and parallel ports, Universal Serial Bus (USB) port, Firewire port, etc. (Figure 4.27).

- ***Parallel Port*** A parallel port is an interface for connecting eight or more data wires. The data flows through the eight wires simultaneously. They can transmit eight bits of data in parallel. As a result, parallel ports provide high speed data transmission. Parallel port is used to connect printer to the computer.
- ***Serial Port*** A serial port transmits one bit of data through a single wire. Since data is transmitted serially as single bits, serial ports provide slow speed data transmission.

Serial port is used to connect external modems, plotters, barcode reader , etc.

- **USB Port** Nowadays, USB is a common and popular external port available with computers. Normally, two to four USB ports are provided on a PC. USB allows different devices to be connected to the computer without requiring re-boot of the computer. USB also has the plug and play feature which allows devices ready to be run simply by plugging them to the USB port. A single USB port can support connection of up to 127 devices.
- ***Firewire (IEEE 1394)*** It is used to connect audio and video multimedia devices like video camera. It is an expensive technology and is used for large data movement. Hard disk drive and new DVD drives connect through firewire. It has data transfer rate of up to 400 MB/sec.

In addition to the above ports, other ports also exist like Musical Instrument Digital Interface (MIDI) port to connect musical instruments like synthesizers and drum machines, PC expansion boards, and PC card and many more.

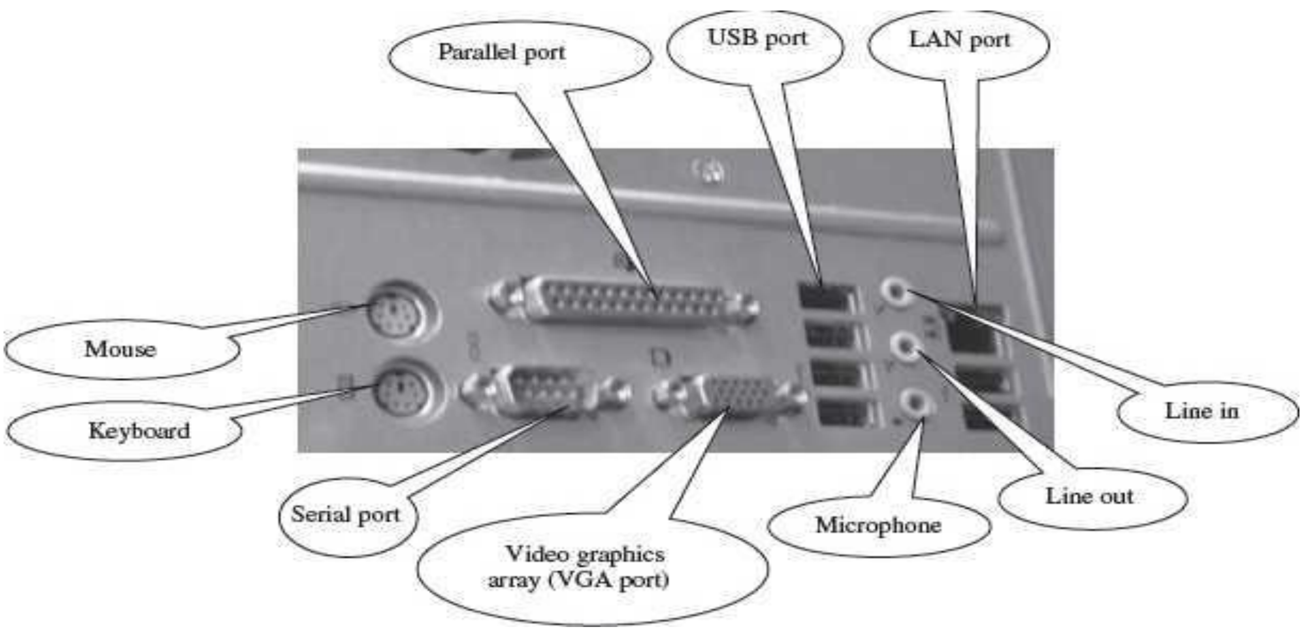

Figure 4.27 Backside of computer cabinet with different ports

4.8 WORKING OF I/O SYSTEM

The working of I/O system combines I/O hardware and I/O software. The I/O hardware includes ports, buses and device controllers for different devices, and I/O devices. The I/O software is the device driver software that may be embedded with operating system or comes with each device. The working of I/O system is described as follows—

- ***I/O Devices*** are attached to computer via the *ports* of computer. There are many standard ports available on the backside of the computer case like serial port and parallel port. If one or more devices use a common set of wires, it is called a bus. For example, PCI bus, PCI Express bus, etc.
- ***Device Controller*** operates on a bus, a port or a device. It controls the signals on the wires of port or bus. The controllers have one or more registers for data and control signals. Controller may be simple like a serial port controller for a serial port, or, complex like a SCSI controller. Some devices have their own built-in controllers.
- ***Device Driver*** is software via which the operating system communicates with the device

controllers. Each device has its own device driver, and a device controller which is specific to the device. The device drivers hide the differences among the different device controller and present a uniform interface to the operating system.

- Application programs use an I/O device by issuing commands and exchanging data with the device driver. The device driver provides correct commands to the controller,

  interprets the controller register, and transfers data to and from device controller registers as required for the correct device operation.

Operating system ------ Device Drivers -------- Device Controllers -------- D

CHAPTER 5
COMPUTER SOFTWARE

5.1 INTRODUCTION

A computer system consists of hardware and software. The computer hardware cannot perform any task on its own. It needs to be instructed about the tasks to be performed. Software is a set of programs that instructs the computer about the tasks to be performed. Software tells the computer how the tasks are to be performed; hardware carries out these tasks. Different sets of software can be loaded on the same hardware to perform different kinds of tasks

user can use the same computer hardware for writing a report or for running a payroll program. The components like monitor, keyboard, processor, and mouse, constitute the hardware (Figure 5.1). In this chapter, we will discuss the different categories of computer software.

Figure 5.1 Making diagrams using hardware and software

6.1 TYPES OF SOFTWARE

Software can be broadly classified in two categories:

1. System Software, and
2. Application Software.

System software provides the basic functions that are performed by the computer. It is necessary for the functioning of a computer. Application software is used by the users to perform specific tasks. The user may choose the appropriate application software, for performing a specific task, which provides the desired functionality. The system software interacts with hardware at one end and with application software at the other end. The application software interacts with the system software and the users of the computer. Figure 5.2 shows the hierarchy of software, hardware and users.

Figure 5.2 Software hierarchy

5.2 SYSTEM SOFTWARE

System software provides basic functionality to the computer. System software is required for the working of computer itself. The user of computer does not need to be aware about the functioning of system software, while using the computer. For example, when you buy a computer, the system software would also include different device drivers. When you request for using any of the devices, the corresponding device driver software interacts with the hardware device to perform the specified request. If the appropriate device driver for any device, say a particular model of a printer, is installed on the computer, the user does not need to know about the device driver, while printing on this printer.

The *purposes of the system software* are:

- To provide basic functionality to computer,
- To control computer hardware, and
- To act as an interface between *user, application software* and *computer hardware*.

On the basis of their functionality, system software may be broadly divided into two categories (Figure 5.3) as follows—

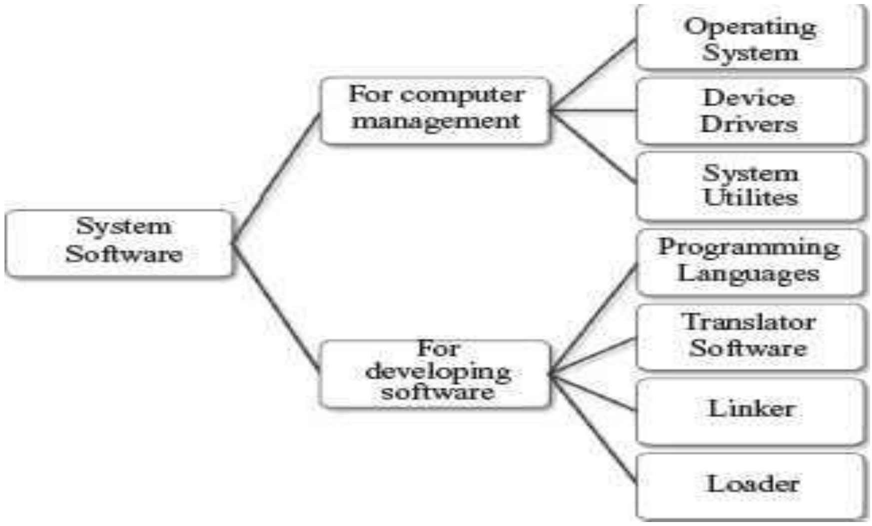

Figure 5.3 System software

- *System software for the management and functionality of computer relates* to the functioning of different components of the computer, like, processor, input and output devices etc. System software is required for managing the operations performed by the components of computer and the devices attached to the computer. It provides support for various services, as requested by the application software. Operating system, device drivers, and system utilities constitute the system software for management of computer and its resources.
- System software for the *development of application software* provides services required for the development and execution of application software. System software provides the software tools required for the development of application software. The programming language software, translator software, loader, and linker are also categorized as system software, and are required for the application software development.

5.3.1 Operating System

Operating System (OS) is an important part of a computer. OS intermediates between the user of a computer and the computer hardware. Different kinds of application software use specific hardware resources of a computer like CPU, I/O devices and memory, as needed by the application software. OS controls and coordinates the use of hardware among the different application software and the users. It provides an interface that is convenient for the user to use, and facilitates efficient operations of the computer system resources. The key functions of OS are—

- It provides an environment in which users and application software can do work.
- It manages different resources of the computer like the CPU time, memory space, file storage, I/O devices etc. During the use of computer by other programs or users, operating system manages various resources and allocates them whenever required, efficiently.

- It controls the execution of different programs to prevent occurrence of error.
- It provides a convenient interface to the user in the form of commands and graphical

interface, which facilitates the use of computer.

Some available operating systems are Microsoft Disk Operating System (MS-DOS), Windows 7, Windows XP, Linux, UNIX, and Mac OS X Snow Leopard.

5.3.2 Device Driver

A device driver acts as a translator between the hardware and the software that uses the devices. In other words, it intermediates between the device and the software, in order to use the device.

Some devices that are commonly connected to the computer are—keyboard, mouse, hard disk, printer, speakers, microphone, joystick, webcam, scanner, digital camera, and monitor. For proper working of a device, its corresponding device driver must be installed on the computer. For example, when we give a command to read data from the hard disk, the command is sent to the hard disk driver and is translated to a form that the hard disk can understand. The device driver software is typically supplied by the respective device manufacturers.

Programmers can write the higher-level application code independently of whatever specific hardware devices it will ultimately use, because code and device can interface in a standard way, regardless of the software superstructure, or of the underlying hardware. Each version of a device, such as a printer, requires its own hardware-specific specialized commands. In contrast, most applications instruct devices (such as a file to a printer) by means of high level generic commands for the device, such as PRINTLN (print a line). The device-driver accepts these generic high-level commands and breaks them into a series of low-level, device-specific commands, as required by the device being driven.

Nowadays, the operating system comes preloaded with some commonly used device drivers, like the device driver for mouse, webcam, and keyboard. The device drivers of these devices are pre-installed on the computer, such that the operating system can automatically detect the device when it is connected to the computer. Such devices are called *plug and play devices*. In case the computer does not find the device driver, it prompts the user to insert the media (like a CD which contains the corresponding device driver) provided along with the device. Most device manufacturers, host the device drivers for their devices on their companies' websites; users can download the relevant driver and install it on their computer.

Each device has its own device driver (Figure 5.4).
Whenever a new device is connected to a computer, its device driver has to be loaded in the computer·s memory, to enable use of the device. When you buy a new printer, you get the device driver CD with it. You must install the device driver on your computer, to use the new printer. Each printer comes with its own device driver. If you replace your old printer with a new model, you need to install the device driver for the new printer.
Device drivers can be *character* or *block device* drivers. Character device drivers are for character based devices like keyboard, which transfer data character by character. Block device driver are for devices that transfer data as a block, like in hard disk.

Figure 5.4 Device driver (i) CD of a printer (ii) Sony audio recorder

5.3.3 System Utilities

System utility software is required for the maintenance of computer. System utilities are used for supporting and enhancing the programs and the data in computer. Some system utilities may come embedded with OS and others may be added later on. Some examples of system utilities are:

- *Anti-virus* utility to scan computer for viruses (Figure 5.5).
- *Data Compression* utility to compress the files.

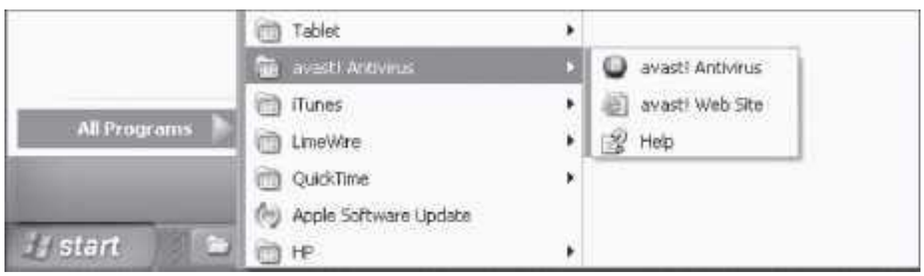

Figure 5.5 Antivirus software on a computer

- *Cryptographic* utility to encrypt and decrypt files.
- *Disk Compression* utility to compress contents of a disk for increasing the capacity of a disk.
- *Disk Partitioning* to divide a single drive into multiple logical drives. Each drive is then treated as an individual drive and has its own file system. Figure 5.5 shows a hard disk with three partitions.
- *Disk Cleaners* to find files that have not been used for a long time. It helps the user to decide what to delete when the hard disk is full.

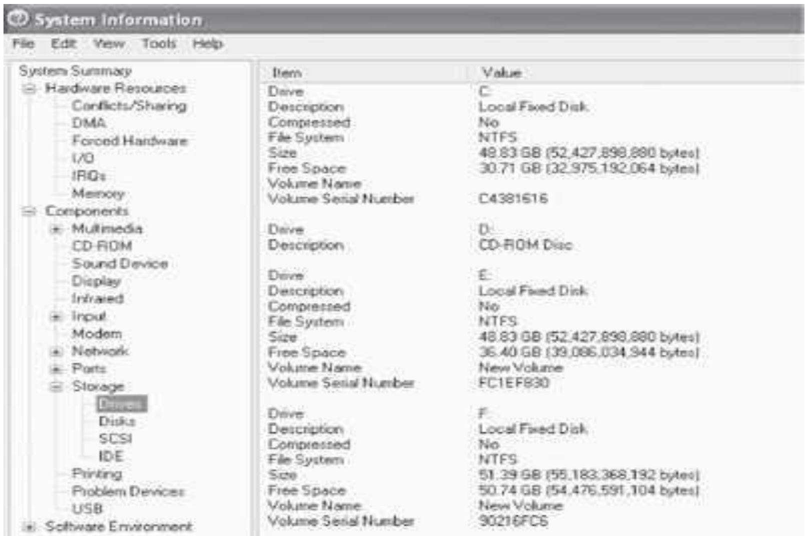

Figure 5.5 A hard disk having three partitions—C, E, and F

- *Backup Utility* to make a copy of all information stored on the disk. It also restores the backed up contents in case of disk failure.
- *System Profiling Utility* provides detailed information about the software installed on the computer and the hardware attached to it.
- *Network Managers* to check the computer network and to log events.

The system utilities on a computer working on Windows XP OS can be viewed by clicking <Start><All Programs><Accessories><System Tools>. Figure 5.7 shows system tools in Windows XP.

Figure 5.7 Some system tools in Windows XP

5.3.4 Programming Languages

A Programming Language consists of a set of vocabulary and grammatical rules, to express the computations and tasks that the computer has to perform. Programming languages are used to write a program, which controls the behavior of computer, codify the algorithms precisely, or enables the human-computer interface. Each language has a unique set of keywords (words that it understands) and a special syntax for organizing program instructions. The programming language should be understood, both by the programmer (who is writing the program) and the computer. A computer understands the language of 0·s and 1·s, while the programmer is more comfortable with English-like language. Programming Language usually refers to high-level languages like COBOL, BASIC, FORTRAN, C, C++, Java etc. Programming languages fall into three categories (Figure 5.8):

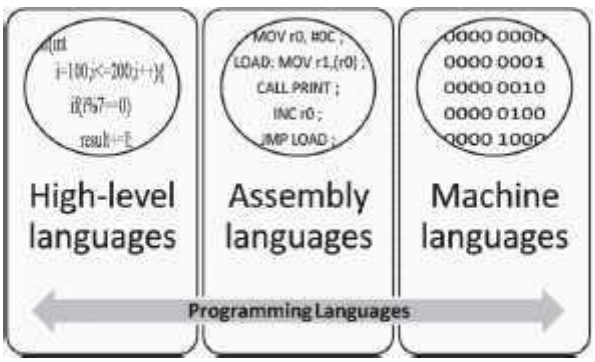

Figure 5.8 Programming languages

- *Machine Language* is what the computer can understand but it is difficult for the programmer to understand. Machine languages consist of numbers only. Each kind of CPU has its own unique machine language.
- *Assembly Language* falls in between machine language and high-level language. They are similar to machine language, but easier to program in, because they allow the programmer to substitute names for numbers.
- High-level Language is easier to understand and use for the programmer but difficult for the computer.

Regardless of the programming language used, the program needs to be converted into machine language so that the computer can understand it. In order to do this a program is either compiled or interpreted.

Figure 5.9 shows the hierarchy of programming languages. The choice of programming language for writing a program depends on the functionality required from the program and the kind of program to be written. Machine languages and assembly languages are also called *low-level languages*, and are generally used to write the system software. Application software is usually written in *high-level* languages. The program written in a programming language is also called the *source code*.

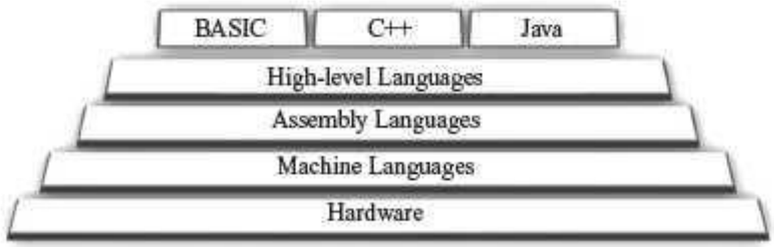

Figure 5.9 A program in machine language

5.4.3.1 Machine Language

A program written in machine language is a collection of binary digits or bits that the computer reads and interprets. It is a system of instructions and data executed directly by a computer·s CPU. It is also referred to as machine code or object code. It is written as strings of 0’s and 1·s, as shown in Figure 5.10. Some of the features of a program written in machine language are as follows:

```
00000000101000010000000000011000
00000000100011100001100000100001
10001100011000100000000000000000
10001100111100100000000000000100
10101100111100100000000000000000
10101100011000100000000000000100
00000011111000000000000000001000
```

Figure 5.10 Machine language code

- The computer can understand the programs written in machine language directly. No translation of the program is needed.
- Program written in machine language can be executed very fast (Since no translation is required).
- Machine language is defined by the hardware of a computer. It depends on the type of the processor or processor family that the computer uses, and is thus machine-dependent. A machine- level program written on one computer may not work on another computer with a different processor.
- Computers may also differ in other details, such as memory arrangement, operating systems, and peripheral devices; because a program normally relies on such factors, different computer may not run the same machine language program, even when the same type of processor is used.
- Most machine-level instructions have one or more opcode fields which specify the basic instruction type (such as arithmetic, logical, jump, etc), the actual operation (such as add or compare), and some other fields.
- It is difficult to write a program in machine language as it has to be written in binary code. For e.g., 00010001 11001001. Such programs are also difficult to modify.
- Since writing programs in machine language is very difficult, programs are hardly written in machine language.

5.3.4.2 Assembly Language

A program written in assembly language uses symbolic representation of machine codes needed to program a particular processor (CPU) or processor family. This representation is usually defined by the CPU manufacturer, and is based on abbreviations (called mnemonics) that help the programmer remember individual instructions, registers, etc. Small, English-like representation is used to write the program in assembly language, as shown in Figure 5.11. Some of the features of a program written in assembly language are as follows:

```
MOV   B, A
MVI   C, 06H
LXI   H, XX50H
ADD   M
JNC   NXTITM
INR   B
INX   H
DCR   C
JNZ   NXTBIT
```

Figure 5.11 Assembly language code

Assembly language programs are easier to write than the machine language programs, since assembly language programs use short, English-like representation of machine code. For e.g.:

ADD 2, 3

LOAD A

SUB A, B

The program written in assembly language is the source code, which has to be converted into machine code, also called object code, using translator software, namely, assembler.
Each line of the assembly language program is converted into one or more lines of machine code. Hence assembly language programs are also machine-dependent.
Although assembly language programs use symbolic representation, they are still difficult to write.
Assembly language programs are generally written where the efficiency and the speed of program are the critical issues, i.e. programs requiring high speed and efficiency.

5.3.4.3 High-level Language

A program in a high-level language is written in English-like language. Such languages hide the details of CPU operations and are easily portable across computers. A high-level language isolates the execution semantics of computer architecture from the specification of the program, making the process of developing a program simpler and more understandable with respect to assembly and machine level languages. Some of the features of a program written in high-level language are as follows:

Programs are easier to write, read or understand in high-level languages than in machine language or assembly language. For example, a program written in C++ is easier to understand than a machine language program (Figure 5.12).

Programs written in high-level languages is the source code which is converted into the object code (machine code) using translator software like interpreter or compiler.

A line of code in high-level program may correspond to more than one line of machine code.

Programs written in high-level languages are easily portable from one computer to another.

Different Generations of Programming Languages

In addition to the categorization of programming languages into machine language, assembly language, and high-level language, programming languages are also classified in terms of generations in which they have evolved. Table 5.1 shows the classification of programming languages based on generations.

Translator Software

Translator software is used to convert a program written in high-level language and assembly language to a form that the computer can understand. Translator software converts a program written in assembly language, and high-level language to a machine-level language program (Figure 5.13). The translated program is called the *object code*. There are three different kind of translator software:

First Generation	Machine language
Second Generation	Assembly language
Third Generation	C, COBOL, Fortran, Pascal, C++, Java, ActiveX (Microsoft) etc.
Fourth Generation	.NET (VB.NET, C#.NET etc.) Scripting language (Javascript, Microsoft Frontpage etc.)
Fifth Generation	LISP, Prolog

Table 5.1 Generations of programming languages

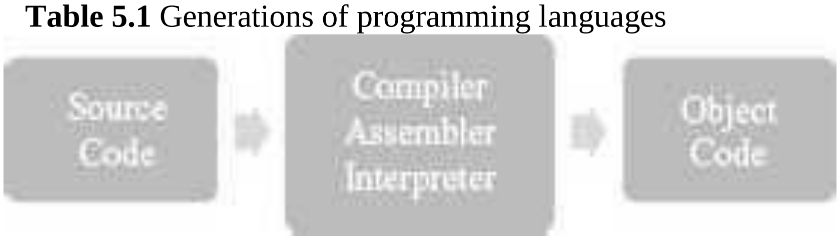

Figure 5.13 Translator software

- Assembler,
- Compiler, and
- Interpreter.

Assembler converts a program written in assembly language to machine language. Compiler and interpreter convert a program written in high-level language to machine language. Let's now discuss, briefly, the different kinds of translator software.

5.3.4.4 Assembler

Assembly language is also referred to as a symbolic representation of the machine code. Assembler is a software that converts a program written in assembly language into machine code (Figure 5.14). There is usually a one-to-one correspondence between simple assembly statements and machine language instructions. The machine language is dependent on the processor architecture, though computers are generally able to carry out the same functionality in different ways. Thus the corresponding assembly language programs also differ for different computer architectures.

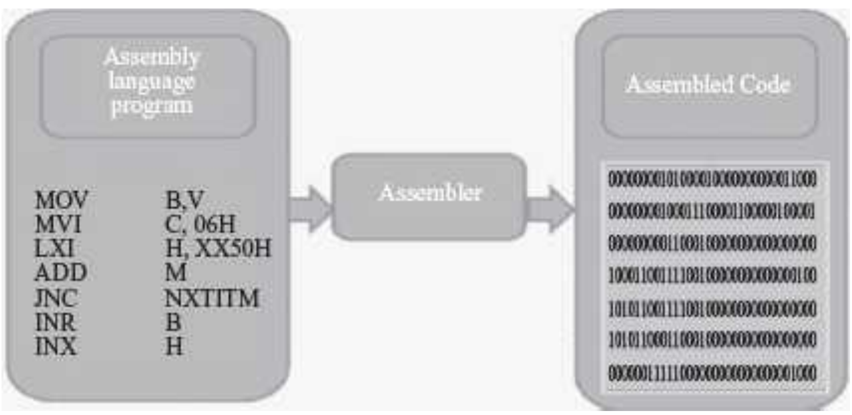

Figure 5.14 Assembler

5.3.4.5 Compiler

A program written in a high-level language has to be converted to a language that the computer can understand, i.e. *binary form*. Compiler is the software that translates the program written in a high-level language to machine language. The program written in high-level language is referred to as the *source code* and compiled program is referred as the *object code*. The object code is the *executable code*, which can run as a stand-alone code. It does not require the compiler to be present during execution. Each programming language has its own compiler. Some languages that use a compiler are C++, COBOL, Pascal, and FORTRAN. In some languages, compilation using the compiler and linking using the linker are required for creating the executable object code.

The compilation process generally involves two parts—breaking down the source code into small pieces and creating an intermediate representation, and, constructing the object code for the intermediate representation. The compiler also reports syntax errors, if any, in the source code.

5.3.4.6 Interpreter

The purpose of interpreter is similar to that of a compiler. The interpreter is used to convert the high-level language program into computer-understandable form. However, the interpreter functions in a different way than a compiler. Interpreter performs line-by-line execution of the source code during program execution. Interpreter reads the source code line-by-line, converts it into machine understandable form, executes the line, and then proceeds to the next line. Some languages that use an interpreter are BASIC and Python.

Difference Between a Compiler and An Interpreter: Compiler and Interpreter are used to convert a program written in high-level language to machine language; however, they work differently. The key differences between a compiler and an interpreter are as follows:

- Interpreter looks at a source code line-by-line. Compiler looks at the entire source code.
- Interpreter converts a line into machine executable form, executes the line, and proceeds with the next line. Compiler converts the entire source code into object-code and creates the object code. The object code is then executed by the user.
- For a given source code, once it is compiled, the object code is created. This object code can be executed multiple number of times by the user. However, interpreter executes line-by-line, so executing the program using an interpreter means that during each execution, the source code is first interpreted and then executed.
- During execution of an object code, the compiler is not required. However, for interpretation, both interpreter and the source code is required during execution (because source code is interpreted during execution).
- Since interpreter interprets line-by-line, the interpreted code runs slower than the compiled code.

5.4 APPLICATION SOFTWARE

The software that a user uses for accomplishing a specific task is the *application software*. Application software may be a single program or a set of programs. A set of programs that are written for a specific purpose and provide the required functionality is called software package. Application software is written for different kinds of applications—graphics, word processors, media players, database applications, telecommunication, accounting purposes etc.

Some examples of application software packages (Figure 5.17) are as follows:

- *Word Processing Software:* For writing letter, reports, documents etc. (e.g. MS-WORD).
- *Image Processing Software:* For assisting in drawing and manipulating graphics (e.g. Adobe Photoshop).

- *Accounting Software:* For assisting in accounting information, salary, tax returns (Tally software).

Figure 5.17 Some application software

- *Spreadsheet Software:* Used for creating budget, tables etc. (e.g. MS-Excel).
- *Presentation Software:* To make presentations, slide shows (e.g. MS-PowerPoint)
- *Suite of Software having Word Processor, Spreadsheet and Presentation Software:* Some examples are MS-Office, Google Docs, Sun Openoffice, Apple iWork.
- CAD/CAM *Software:* To assist in architectural design. (e.g. AutoCAD, Autodesk)
- *Geographic Information Systems:* It captures, stores, analyzes, manages, and presents data, images and maps that are linked to different locations. (e.g. ArcGIS)
- *Web Browser Software:* To access the World Wide Web to search documents, sounds, images etc. (e.g. Internet Explorer, Netscape Communicator, Chrome).

5.5 SOFTWARE ACQUISITION

Different kinds of software are made available for use to users in different ways. The user may have to purchase the software, can download for free from the Internet, or can get it bundled along with the hardware. Nowadays with the advent of Cloud computing, many application software are also available on the cloud for use through the Internet, e.g. Google Docs. The different ways in which the software are made available to users are:

- **Retail Software** is off-the-shelf software sold in retail stores. It comes with printed manuals and installation instructions. For example, Microsoft Windows operating system.
- **OEM Software stands** for "Original Equipment Manufacturer" software. It refers to software which is sold, and bundled with hardware. Microsoft sells its operating system as OEM software to hardware dealers. OEM software is sold at reduced price, without the manuals, packaging and installation instructions. For example, Dell computers are sold with the "Windows 7" OS pre-loaded on them.
- **Demo Software** is designed to demonstrate what a purchased version of the software is capable of doing and provides a restricted set of features. To use the software, the user

must buy a fully- functional version.

- **Shareware** is a program that the user is allowed to try for free, for a specified period of time, as defined in the license. It is downloadable from the Internet. When the trial period ends, the software must be purchased or uninstalled.
- **Freeware** is software that is free for personal use. It is downloadable from the Internet. The commercial use of this software may require a paid license. The author of the freeware software is the owner of the software, though others may use it for free. The users abide by the license terms, where the user cannot make changes to it, or sell it to someone else.
- **Public Domain Software** is free software. Unlike freeware, public domain software does not have a copyright owner or license restrictions. The source code is publicly available for anyone to use. Public domain software can be modified by the user.
- **Open-Source Software** is software whose source code is available and can be customized and altered within the specified guidelines laid down by the creator. Unlike public domain software, open-source software has restrictions on their use and modification, redistribution limitations, and copyrights. Linux, Apache, Firefox, OpenOffice are some examples of open-source software.

CHAPTER 6
OPERATING SYSTEM

Any computer system you use has an operating system. The user interacts with the machine via the operating system. A software on the machine interacts with the hardware via the operating system. Operating system intermediates between the hardware and the user. The purpose of this chapter is to introduce you to the Operating System.

6.1 INTRODUCTION

The computer system comprises of a functional set of hardware, software, user and data. Hardware consists of the components of computer like memory, processor, storage devices, and Input/Output devices. The software may be of different kinds—application software and system softwa re. A computer system may be a single stand-alone system or may consist of several interconnected systems. The user uses the application software to perform various tasks, for example, the user uses word processing software for document preparation. While using the application software, the user uses the storage of a computer—to store a document on the hard disk, to execute a command on the CPU, to retrieve a document from a peripheral device or to print document on printer. For using the hardware, there is a need for software that interacts with both the hardware and the application software. Operating system (OS) is the software that provides an interface between the computer hardware, and the application programs or users (Figure 6.1).

In this chapter, we discuss about the components of operating system, the different types of operating system and the functions of operating system. A brief description of some operating systems is also given.

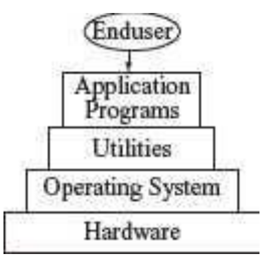

Figure 6.1 View of components of computer system

7.1 OBJECTIVES OF OPERATING SYSTEM

Operating system is system software that controls and coordinates the use of hardware among the different application software and users. OS intermediates between the user of computer and the computer hardware. The user gives a command and the OS translates the command into a form that the machine can understand and execute. OS has two main objectives—(1) to make the computer system convenient and easy to use, for the user, and—(2) to use the computer hardware in an efficient way, by handling the details of the operations of the hardware.

- OS hides the working of the hardware from the user and makes it convenient for the user to use the machine. The application program used by the user requires the use of the hardware during processing. Some examples are—display of application's user interface, loading a program into memory, using I/O devices, allocating CPU to different processes during execution, and store or load data from hard disk. When using the machine, the user gives the command to perform the required actions to the OS and the OS handles all the operational steps. The user is not bothered about how these actions will be performed. This is the job of OS. OS provides an interface to the application programs to interact with the hardware. The user need not get into the details of this interaction.

- At the other end, the different resources of computer hardware have to be managed and controlled. This includes managing the communication between different devices, controlling the sequence and execution of processes, allocating space on hard disk, providing error handling procedures etc. OS supervises and manages the hardware of the computer.

Some of the commonly used operating systems are Microsoft Disk Operating System (MS-DOS), Windows 6, Windows XP, Linux, UNIX, and Mac OS X Snow Leopard.

6.2 TYPES OF OPERATING SYSTEMS

OS are classified into different types depending on their capability of processing—(1) Single user, (2) Multiuser, (3) Multitasking, (4) Multiprocessing, (5) Real time, and (6) Embedded.

- **Single User and Single Task OS** is for use by a single user for a standalone single computer for performing a single task (Figure 6.2). Operating system for Personal Computers (PC) are singleuser OS. For example, if the user is editing a document, then a document cannot be printed on the printer simultaneously. Single user OS are simple operating system designed to manage one task at a time. MS-DOS is an example of single user OS.

Figure 6.2 A single user performing a single task

- **Single User and Multitasking OS** allows execution of more than one task or process concurrently. For this, the processor time is divided amongst different tasks. This division of time is also called *time sharing*. The processor switches rapidly between processes. For example, the user can listen to music on the computer while writing an article using a word processor software. The user can switch between the applications and also transfer data between them (Figure 6.3). Windows 95 and all later versions of Windows are examples of multitasking OS.

Figure 6.3 A single user performing multitasking (issuing print command and making drawings)

- **Multiuser OS** is used in computer networks that allow same data and applications to be accessed by multiple users at the same time (Figure 6.4). The users can also communicate with each other. Linux, UNIX, and Windows 6 are examples of multiuser OS.
- **Multiprocessing OS** have two or more processors for a single running process. Processing takes place in parallel and is also called *parallel processing*. Each processor

works on different parts of the same task, or, on two or more different tasks. Since execution takes place in parallel, they are used for high speed execution, and to increase the power of computer. Linux, UNIX and Windows 6 are examples of multiprocessing OS.

Figure 6.4 Multiple users working on connected computers

Real Time OS are designed to respond to an event within a predetermined time. These operating systems are used to control processes. Processing is done within a time constraint. OS monitors the events that affect the execution of process and respond accordingly. They are used to respond to queries in areas like medical imaging system, industrial control systems etc. LynxOS is an example of real time OS.

Embedded OS is embedded in a device in the ROM. They are specific to a device and are less resource intensive. They are used in appliances like microwaves, washing machines, traffic control systems etc.

6.3 FUNCTIONS OF OPERSATING SYSTEM

Operating system is a large and complex software consisting of several components. Each component of the operating system has its own set of defined inputs and outputs. Different components of OS perform specific tasks to provide the overall functionality of the operating system (Figure 6.5). Main functions of the operating system are as follows:

Figure 6.5 Functions of OS

- **Process Management**—The process management activities handled by the OS are—(1) control access to shared resources like file, memory, I/O and CPU, (2) control execution of applications, (3) create, execute and delete a process (system process or user process),

- cancel or resume a process (5) schedule a process, and (6) synchronization, communication and deadlock handling for processes.

- **Memory Management**—The activities of memory management handled by OS are—(1) allocate memory, (2) free memory, (3) re-allocate memory to a program when a used block is freed, and (4) keep track of memory usage.
- **File Management**—The file management tasks include—(1) create and delete both files and directories, (2) provide access to files, (3) allocate space for files, (4) keep back-up of files, and (5) secure files.
- **Device Management**—The device management tasks handled by OS are—(1) open, close and write device drivers, and (2) communicate, control and monitor the device driver.
- **Protection and Security**—OS protects the resources of system. User authentication, file attributes like read, write, encryption, and back-up of data are used by OS to provide basic protection.
- **User Interface or Command Interpreter**—Operating system provides an interface between the computer user and the computer hardware. The user interface is a set of commands or a graphical user interface via which the user interacts with the applications and the hardware.

6.4 EXAMPLES OF OPERATING SYSTEMS

MS-DOS, Windows family of operating systems, Unix OS, Linux OS, and Mac OS X are some of examples of commonly used OSs. Each operating system has specific characteristics. Here, we will discuss the features of the MS-DOS, Windows family of operating systems and Linux operating system.

MS-DOS

- MS-DOS was the first widely-installed operating system for PCs in 1980s.
- MS-DOS is easy to load and install. It neither requires much memory for the operating system, nor a very powerful computer to run on.
- MS-DOS is a command line user interface operating system. This means that the user has to type single line commands through the command interface. So, user has to remember the different commands and their syntax.
- It is a single-user and single-tasking operating system for the PC. Only one user can use it and only one task can be executed, at a given point of time. Also, it does not have a built-in support for networking.
- MS-DOS is a 16-bit OS, meaning thereby that it can send or receive 16 bits of data at a time and can process 16 bits of data. It is not able to take the advantage of 32-bit processors.

- To use MS-DOS, user must know where the programs and data are stored and how to interact with it. In the MS-DOS command mode, *command.com* routine interprets the typed in command from the keyboard.

To get the window of the command prompt 

in the Windows environment

- <Start> <Run> Type "*cmd*" <Enter>, or
- <Start> <All programs> <Accessories> <Command Prompt>

cmd.exe or command prompt is the command line interpreter on the current Windows-based OS. It is similar to command.com in MS-DOS. cmd.exe is a Windows program that acts as a DOS-like command line interpreter.

Windows Family of OS

- Windows is a personal computer operating system from Microsoft.
- The Windows family of OS which is currently in use includes the Windows 9x family (Windows 95, Windows 98 and Windows 2000), Windows XP, Windows Vista, and Windows 6 operating systems.
- Windows family of OS is GUI-based operating system. Since GUI interfaces are easy to use and are user-friendly, these have become very popular.
- Windows support multi-tasking. It means Windows OS allows simultaneous execution of multiple tasks (Figure 6.19).
- Windows contains built-in networking, which allows users to share files and applications with each other, if their PCs are connected to a network.
- Windows 6 comes in six different editions, Starter, Home Basic, Home Premium, Professional, Enterprise and Ultimate.
- With each new version of the Windows OS, the user interface undergoes some changes and the user has to learn to use the new interface. This becomes troublesome for the user.

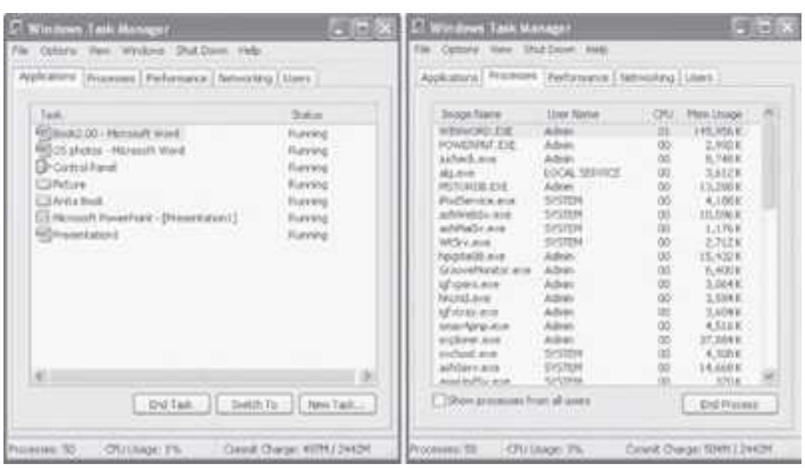

Figure 6.19 Windows task manager

Brief History of Windows OS

The Windows OS has evolved from the Windows 1.0 in the 1985 to the Windows 6 in 2009. In this span of 24 years, several versions of the Windows OS have been released. Table 6.1 gives an overview of the history of the Windows OS, along with their significance.

Year	Windows OS Released	Comments
1985	Windows 1.0	It was not successful
1990	Windows 3.0	The first commercially successful version of Windows. It is an upgrade to the interface over Windows 1 and 2
1993	Windows NT 3.1	The first Microsoft OS not based in DOS. Separate versions of NT with their DOS counterparts are released
1995	Windows 95	The first native 32 bit OS. Microsoft plans to merge the NT and DOS platforms but are unsuccessful due to backward compatibility issues and lack of hardware support of NT.
1998	Windows 98	Microsoft integrates its web browser in the GUI and file manager. Hackers can use the Internet to infiltrate a computer or network.
2000	Windows 2000	As with Windows 95, Microsoft planned Windows 2000 to merge the NT and DOS based OS's but was unsuccessful
2001	Windows XP	Windows XP successfully merges the compatibility found in Windows 98 with the stability found in Windows NT/2000. It provides enhanced stability over Windows 98.
2005	Windows XP Professional x64 Edition	OS was slow to take off due to the dearth of 64-bit software and drivers
2008	Windows Vista	First 3D operating system
2009	Windows 6	Some of the new features included in Windows 6 are advancements in touch, speech, and handwriting recognition, support for virtual hard disks, support for additional file formats, improved performance on multi-core processors, improved boot performance, and kernel improvements.

Table 6.1 Windows OS overview

Linux OS

- Linux is a Unix-like OS. Unix OS has a user interface called *shell*. The kernel provides interface for the programs to interact with the hardware, and provides services like process management and memory management. The shell interacts with the kernel through the system calls.
- Linux was developed by *Linus Torvalds* in 1992. Linux is copyright under the GNU Public License. Linux is a "free" operating system that is easily available. Since Linux follows the open development model, it is being constantly upgraded by programmers across the globe.
- Some organizations offer Linux with add-on features and capabilities. Red Hat, Mandrake, Debian and Novell are the popular vendors of Linux OS.
- Tux, the Linux penguin is the official mascot of Linux.
- Linux is a command line user interface OS. Linux has GUI interfaces called desktop environments like GNOME and K Desktop Environment (KDE). The GUI interface is convenient for the user to use.
- Linux is a 32-bit, multi-tasking OS. It supports multiple users and multiple processors.
- Linux is a reliable and secure OS, and is available almost for free. So, Linux is fast becoming very popular and powerful OS.
- Linux OS is easily available, such as Redhat Linux ver. 9, and, Debian·s—Ubuntu, Kubuntu, and Edubuntu