COMPUTER FUNDAMENTALS
AND
PROGRAMMING IN C

Second Edition

Pradip Dey
RCC Institute of Information Technology

Manas Ghosh
RCC Institute of Information Technology

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries.

Published in India by
Oxford University Press
YMCA Library Building, 1 Jai Singh Road, New Delhi 110001, India

First Edition published in 2006
Second Edition published in 2013

ISBN-13: 978-0-19-808456-3
ISBN-10: 0-19-808456-0

Typeset in Times New Roman
by Pee-Gee Graphics, New Delhi
Printed in India by Rajkamal Electric Press, Kundli, Haryana

Brief Contents

Contents

5. BASIC CONCEPTS OF OPERATING SYSTEMS 70

6. THE INTERNET 104

7. INTRODUCTION TO ALGORITHMS AND PROGRAMMING CONCEPTS 114

8. BASICS OF C 136

9. INPUT AND OUTPUT 183

10. CONTROL STATEMENTS 202

11. ARRAYS AND STRINGS 246

12. FUNCTIONS 282

13. POINTERS IN C 324

14. USER-DEFINED DATA TYPES AND VARIABLES 394

15. FILES IN C 420

16. ADVANCED C 445

CHAPTER 1

Computer Fundamentals

LEARNING OBJECTIVES

After studying this chapter, the readers will be able to

- trace the evolution of computers—generations and classification of computers
- explain the basic units of a computer system
- explain the hardware and software of a personal computer
- load an operating system (OS) in a personal computer

1.1 INTRODUCTION—WHAT IS A COMPUTER?

The *Oxford Dictionary* defines a computer as 'an automatic electronic apparatus for making calculations or controlling operations that are expressible in numerical or logical terms'.

The definition clearly categorizes the computer as an electronic apparatus although the first computers were mechanical and electro-mechanical apparatuses. The definition also points towards the two major areas of computer application: data processing and computer-assisted controls or operations. Another important conclusion of the definition is the fact that the computer can perform only those operations or calculations that can be expressed in logical or numerical terms.

A computer is a data processor. It can accept input, which may be either data or instructions or both. The computer remembers the input by storing it in memory cells. It then processes the stored input by performing calculations or by making logical comparisons or both. It gives out the result of the arithmetic or logical computations as output information. The computer accepts input and outputs data in an alphanumeric form. Internally it converts the input data to meaningful binary digits, performs the instructed operations on the binary data, and transforms the data from binary digit form to understandable alphanumeric form.

1.2 EVOLUTION OF COMPUTERS—A BRIEF HISTORY

Computing in the mechanical era

The concept of calculating machines evolved long before the invention of electrical and electronic devices. The first mechanical calculating apparatus was the **abacus**, which was invented in 500 BC in Babylon. It was used extensively without any improvement until 1642 when Blaise Pascal designed a calculator that employed gears and wheels. But it was not until the early 1800s that a practical, geared, mechanical computing calculator became available. This machine could calculate facts but was not able to use a program to compute numerical facts.

In 1823, Charles Babbage, aided by Augusta Ada Byron, the Countess of Lovelace, started an ambitious project of producing a programmable calculating machine for the Royal Navy of Great Britain. Input to this mechanical machine, named the **Analytical Engine**, was given through punched cards. This engine stored 1,000, 20-digit decimal numbers and a modifiable program, which could vary the operation of the machine so that it could execute different computing jobs. But even after several years of effort, the machine that had more than 50,000 mechanical parts could not operate reliably because the parts could not be machined to precision.

Computing in the electrical era

With the availability of electric motors in 1800, a host of motor-operated calculating machines based on Pascal's calculator was developed. A mechanical machine, driven by a single electric motor, was developed in 1889 by Herman Hollerith to count, sort, and collate data stored on punched cards. Hollerith formed the Tabulating Machine Company in 1896. This company soon merged into International Business Machines (IBM) and the mechanical computing machine business thrived.

In 1941, Konrad Zuse developed the first electronic calculating computer, **Z3.** It was used by the Germans in World War II. However, Alan Turing is credited with developing the first electronic computer in 1943. This computer system, named the **Colossus**, was a fixed-program computer; it was not programmable.

J.W. Mauchly and S.P. Eckert of the University of Pennsylvania completed the first general-purpose electronic digital computer in 1946. It was called the **ENIAC**, Electronic Numerical Integrator and Calculator. It used 17,000 vacuum tubes, over 500 miles of wires, weighed 30 tons, and performed around 100,000 operations per second. The **IAS** computer system, under development till 1952 by John von Neumann and others at the Princeton Institute, laid the foundation of the general structure of subsequent general-purpose computers. In the early 1950s, Sperry-Rand Corporation launched the UNIVAC I, UNIVAC II, UNIVAC 1103 series while IBM brought out Mark I and 701 series. All these machines used vacuum tubes.

The transistor was invented at Bell Labs in 1948. In 1958, IBM, International Computers Limited (ICL), Digital Equipment Corporation (DEC), and others brought out general-purpose computers using transistors that were faster, smaller in size, weighed less, needed less power, and were more reliable.

Meanwhile, at Texas Instruments, Jack Kilby invented the integrated circuit in 1958 that led to the development of digital integrated circuits in the 1960s. This led to the development of IBM 360/370, PDP 8/1, and HP 9810 in 1966. These computers used medium- and small-scale integrated circuits (MSI and SSI).

Thereafter, in 1971, Intel Corporation announced the development of the single-chip microprocessor 4004, a very large-scale integrated circuit. In 1972, the 8008 8-bit microprocessor was introduced. Subsequently, the 8080 and MC 6800 appeared in 1973, which were improved 8-bit microprocessors. The last of the 8-bit microprocessor family from Intel, 8085, was introduced as a general-purpose processor in 1974. In 1978, the 8086, and in 1979, the 8088 microprocessors were released.

Though desktop computers were available from 1975 onwards, none could gain as much popularity as the IBM PC. In 1981, IBM used the 8088 microprocessor in the personal computer. The 80286 16-bit microprocessor came in 1983 as an updated version of 8086. The 32-bit microprocessor 80386 arrived in 1986 and the 80486 arrived in 1989. With the introduction of the Pentium in 1993, a highly improved personal computer was available at an affordable price.

With the development of the desktop computers, in the form of personal computers, and networking, the whole scenario of computing has undergone a sea change. Now, portable computers such as the laptop and palmtop are available, which can execute programs, store data, and output information at speeds higher than that possible with all the earlier computers. Efforts are now being made to integrate a palmtop computer with a mobile phone unit.

Along with the development of computer hardware, programming languages were devised and perfected. In the 1950s, Assembly language was developed for UNIVAC computers. In 1957, IBM developed FORTRAN language. Then in the years that followed came programming languages such as ALGOL, COBOL, BASIC, PASCAL, C/C++, ADA, and JAVA.

Further, with the creation of the operating system (OS), a supervisor program for managing computer resources and controlling the CPU to perform various jobs, the computer's operational capability touched a new dimension. There are a variety of operating systems today. Some which gained popularity are UNIX for large and mini-computers and MSDOS and MS-WINDOWS for personal computers. However, with the availability of LINUX, a trend to change over to this operating system is on.

1.3 GENERATIONS OF COMPUTERS

What generation a computer belongs to is determined by the technology it uses. Table 1.1 shows the technology used in the different generations of computers. With advancement in the generation, the performance of computers improved not only due to the implementation of better hardware technology but also superior operating systems and other software utilities.

Table 1.1 Technology used in different generations of computers

Generation number	Technology	Operating system	Year of introduction	Specific computers
1	Vacuum Tube	None	1945	Mark 1
2	Transistor	None	1956	IBM 1401, ICL 1901, B5000, MINSK-2
3	SSI and MSI	Yes	1964	IBM S/360/370, UNIVAC 1100, HP 2100A, HP 9810
4	LSI and VLSI	Yes	1971	ICL 2900, HP 9845A, VAX 11/780, ALTAIR 8800, IBM PC
5	HAL	Yes	Present and beyond	–

1.4 CLASSIFICATION OF COMPUTERS

Most designs of computers today are based on concepts developed by John von Neumann and are referred to as the von Neumann architecture. Computers can be classified in variety of ways on the basis of various parameters such as usage, cost, size, processing power, and so on. The classification of computers is presented below based on their power and their use.

Supercomputer

Supercomputer is the most expensive and fastest type of computer that performs at or near the currently highest operational rate for computers. A Cray supercomputer is a typical example. These are employed for specialized applications that require immense amounts of mathematical calculations such as weather forecasting, nuclear energy research, and petroleum exploration etc.

Mainframe

A mainframe computer supports a vast number of users to work simultaneously and remotely. Apart from providing multi-user facility, it can process large amounts of data at very high speeds and support many input, output and auxiliary storage devices. These computers are very large in size, and expensive. The main difference between a supercomputer and a mainframe is that a supercomputer can execute a single program faster than a mainframe, whereas a mainframe uses its power to execute many programs concurrently. The IBM 370 and IBM 3090 are examples of mainframe computers.

Minicomputers

A minicomputer is powerful enough to be used by multiple users (between 10 to 100) but is smaller in size and memory capacity and cheaper than mainframes. Two classic examples were the Digital Equipment Corporation VAX and the IBM AS/400.

Microcomputers

The microcomputer has been intended to meet the personal computing needs of an individual. It typically consists of a microprocessor chip, a memory system, interface units and various I/O ports, typically resided in a motherboard. There are many types of microcomputers available.

Desktop computer A micro computer sufficient to fit on a desk.

Laptop computer A portable microcomputer with an integrated screen and keyboard.

Palmtop computer/Digital diary/Notebook/PDAs A hand-sized microcomputer having no keyboard. The screen serves both as an input and output device.

1.5 ANATOMY OF A COMPUTER

A computer can accept input, process or store data, and produce output according to a set of instructions which are fed into it. A computer system can be divided into two components which are responsible for providing the mechanisms to input and output data, to manipulate and process data, and to electronically control the various input, output, and their storage. They are known as hardware and software. The *hardware* is the tangible parts of the computer. Whereas, the *software* is the intangible set of instructions that control the hardware and make it perform specific tasks. Without software, a computer is effectively useless.

1.5.1 Hardware

Hardware is the physical components of a computer that includes all mechanical, electrical, electronic and magnetic parts attached to it. A computer consists of the following major hardware components:

- Input and output devices
- Central processing unit (CPU)
- Memory unit and storage devices
- Interface unit

A brief description of the most common hardware found in a personal computer is given in the next few sections.

Input devices

The data and instructions are typed, submitted, or transmitted to a computer through input devices. Input devices are electronic or electro-mechanical equipment that provide a means of communicating with the computer system for feeding input data and instructions. Most common input devices are briefly described below.

Keyboard Keyboard is like a type-writer. A keyboard, normally, consists of 104 keys. These keys are classified into different categories which are briefly described below.

Character keys These keys include letters, numbers, and punctuation marks. On pressing any character key, the corresponding character is displayed on the screen.

Function keys There are 12 functional keys above the key board which are used to perform certain functions depending on the operating system or the software currently being executed. These keys are placed at the top of the key board and can easily be identified with the letter F followed by a number ranging from 1 to 12.

Control keys Alt, Ctrl, Shift, Insert, Delete, Home, End, PgUp, PgDn, Esc and Arrow keys are control keys.

Navigation keys These include four arrows, Page Up and Page Down, Home and End. These keys are normally used to navigate around a document or screen.

Toggle keys Scroll Lock, Num lock, Caps Lock are three toggle keys. The toggle state is indicated by three LEDs at the right-top side of the keyboard. For example, on pressing caps lock, letters typed by the user will appear in upper case. On pressing again, letters are typed on the screen in lower case.

Miscellaneous keys These keys include Insert, delete, escape, print Screen etc.

The keys on the keyboard are placed in a series of rows and columns called the *key matrix*. Each key holds a position with respect to a row and column. When a key is pressed, the key switch in that position closes a circuit, sending a signal to the circuit board inside the keyboard. The keyboard controller uses the x and y coordinates of the matrix position to determine which key was pressed, thereby determining what code is transmitted to the computer by the keyboard.

Mouse A mouse is the pointing device attached to a computer. It is used to move the cursor around the screen and to point to an object (such as icon, menu, command button etc.) on the computer video screen for the purpose of selecting or activating objects on graphical interface provided by the operating system or the software currently being executed and executing various tasks. It has two or three buttons for clicking. The mouse tracks the motion of the mouse pointer and senses the clicks and sends them to the computer so it can respond appropriately.

The mouse can be connected to the system either through a USB connector or wirelessly through infrared radiation. A wireless mouse needs to be powered through batteries.

Scanner A scanner is a device that captures pictures or documents so that they can be stored in storage devices, seen on the video screen, modified suitably, transported to other computers, or printed on a printer. A personal computer with a scanner and printer can function as a photocopier.

Output devices

Output devices mirror the input data, or show the output results of the operations on the input data or print the data. The most common output device is monitor or visual display unit. The printer is used to print the result. A hard copy refers to a printout showing the information. On the other hand soft copy means information stored on a storage device.

Monitor Computer display devices are commonly known as Visual Display Unit (VDU) or monitor. It operates on a principle similar to that of a normal television set. Various technologies have been used for computer monitors. They are also of different sizes. CRT (Cathode-ray tube) and LCD (liquid crystal display) monitors are the two common types which are widely used.

The CRT is composed of a vacuum glass tube which is narrower at one end. One electron gun is placed at this end which fires electrons. The electron gun is made up of cathode (negatively charged) and one anode (positively charged). On the other side it has a wide screen, coded with phosphor. The beam of electron strikes on the surface of screen and produces an image by photo luminance process. There is some vertical and horizontal coil to deflect the electron beam in any position of the screen. An image is formed by constantly scanning the screen. To send an image to the screen, the computer first assembles it in a memory area called a video buffer. The graphics are stored as an array of memory locations that represent the colors of the individual screen dots, or pixels. The video card then sends this data through a Digital To Analog Converter (DAC), which converts the data to a series of voltage levels that are fed to the monitor.

CRT monitors are too bulky and consume high power. Apart from these, users are very much concerned about potentially damaging non-ionizing radiation from CRT monitor.

Nowadays, LCD monitors are replacing CRTs and becoming the de-facto choice to the users because of its size**,** display clarity**,** low radiation emission and power consumption**.**

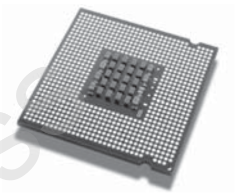
An LCD display produces an image by filtering light from a series of cold cathode fluorescent lamps (CCFLs).through a layer of liquid crystal cells. Gradually, CCFL backlighting technology is being replaced by low-power light-emitting diodes (LEDs). A computer screen that uses this technology is sometimes referred to as an LED display.

Wide-carriage dot matrix printer. Courtesy: **Dale Mahalko** (This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license; http://creativecommons.org/licenses/by-sa/3.0/deed.en)

Printer The printer is a device that prints any data, report, document, picture, diagrams, etc. Printers are categorized based on the physical contact of the print head with the paper to produce a text or an image. An *impact printer* is one where the print head will be in physical contact with the paper. In a *non–impact printer*, on the other hand the print head will have no physical contact with the paper. The Dot matrix printer is considered as an Impact printer and Laser printer is considered as Non-impact printer.

In a *dot matrix printer*, the printer head physically ‘hits’ the paper through the ribbon which makes the speed of the printer relatively slow. The printer head consist of some two dimensional array of dot called ‘dot matrix’. Every time when it strikes the paper through ribbon its dots are arranged according to the character which is going to be printed. The ink in the ribbon falls on the surface of the paper and thus the character get printed. In *inkjet printer*, instead of a ribbon one ink cartridge holds the ink in it. They are placed above the inkjet head. The printing head takes some ink from the cartridge and spreads it on the surface of the paper by the jet head. This ink is electrically charged. An electric field is created near the paper surface. Thus the small drops of ink are arranged in the surface according to the character it prints. These printers are fast and capable of printing good quality graphics. The *laser printer* uses a laser beam to create the image.

Central processing unit (CPU)



Central Processing Unit or CPU can be thought of as the brain of the computer. Most of the processing takes place in CPU. During processing, it locates and executes the program instructions. It also fetches data from memory and input/output devices and sends data back.

Physically, it is an integrated circuit (IC) silicon chip, mounted on a small square plastic slab, surrounded by metal pins. In the world of personal computers, the term microprocessor and CPU are used interchangeably. It is more accurate to describe it as a CPU on a chip because it contains the circuitry that performs processing.

The CPU itself can be divided into different functional units which are described below-

Registers These are high-speed storage devices. In most CPUs, some registers are reserved for special purposes. For example, the Instruction Register (IR) holds the current instruction being executed. The Program Counter (PC) is a register that holds the address of the next instruction to be executed. In addition to such and other special-purpose registers, it also contains a set of general-purpose registers that are used for temporary storage of data values as needed during processing.

Arithmetic logic unit (ALU) It is the part of the CPU that performs arithmetic operations, such as addition and subtraction as well as logical operations, such as comparing two numbers to see if they are the equal or greater or less.

Control unit (CU) The control unit coordinates the processing by controlling the transfer of data and instructions between main memory and the registers in the CPU. It also coordinates the execution of the arithmetic logic unit (ALU) to perform operations on data stored in particular registers. It consists of

- an *instruction decoding circuit* that interprets what action should be performed.
- a *control and timing circuit* directs all the other parts of the computer by producing the respective control signals.

Nowadays, a high-speed memory, called *cache memory*, is embedded with the CPU chip. This improves the computer performance by minimizing the processor need to read data from the slow main memory.

The CPU's processing power is measured in terms of the number of instructions that it can execute per unit time. Every computer comprises of an internal clock, which emits electronic pulses at a constant rate. These pulses are used to control and synchronize the pace of operations. Each pulse is called a *clock cycle* which resembles a rectangular wave with a rising half of the signal and a falling half. In other words, a full clock cycle is the amount of time that elapses between pulses of the oscillating signal. Each instruction takes one or more clock cycles to execute. The higher the clock speed, the more instructions are executed in a given period of time. Hertz (Hz) is the basic unit of computer clock frequency which is equal to one cycle per second. CPU speed has been improved continuously. It is typically measured in megahertz (MHz) or gigahertz (GHz). One megahertz is equal to one million cycles per second, while one gigahertz equals one billion cycles per second.

Nowadays, multiple processors are embedded together on a single integrated-circuit chip, known as multi-core processor e.g. a *dual-core processor* has two CPUs and a *quad core processor* has four CPUs.

Note

- An integrated circuit, or IC, is a matrix of transistors and other electrical components embedded in a small slice of silicon.
- A microprocessor is a digital electronic component with miniaturized transistors on a single semiconductor integrated circuit (IC). One or more microprocessors typically serve as a central processing unit (CPU) in a computer system or handheld device allocating space to hold the data object.

Memory unit

Components such as the input device, output device, and CPU are not sufficient for the working of a computer. A storage area is needed in a computer to store instructions and data, either temporarily or permanently, so that subsequent retrieval of the instructions and data can be possible on demand. Data are stored in memory as binary digits, called *bits*. Data of various types, such as numbers, characters, are encoded as series of bits and stored in consecutive memory locations. Each memory location comprises of a single byte which is equal to eight bits and has a unique address so that the contents of the desired memory locations can be accessed independently by referring to its' address. A single data item is stored in one or more consecutive bytes of memory. The address of the first byte is used as the address of the entire memory location.

CPU uses registers exclusively to store and manipulate data and instructions during the processing. Apart from registers, there are mainly two types of memory that are used in a computer system. One is called *primary memory* and the other *secondary memory*.

Primary memory Primary memory is the area where data and programs are stored while the program is being executed along with the data. This memory space, also known as *main memory*, forms the working area of the program. This memory is accessed directly by the processor.

A memory module consists of a large bank of flip-flops arranged together with data traffic control circuitry such that data can be stored or read out on or from a set of flip-flops. A flip-flop can store a binary digit. These flip-flops are grouped to form a unit memory of fixed length and each of which is identified by a sequence number known as a memory address. These type are called Random Access Memory, or RAM, where any location can be accessed directly, and its stored contents get destroyed the moment power to this module is switched off. Hence, these are volatile in nature. Primary memory devices are expensive. They are limited in size, consume very low power, and are faster as compared to secondary memory devices.

There is another kind of primary memory increasingly being used in modern computers. It is called *cache memory* (pronounced as "cash"). It is a type of high-speed memory that allows the processor to access data more rapidly than from memory located elsewhere on the system. It stores or caches some of the contents of the main memory that is currently in use by the processor. It takes a fraction of the time, compared to main memory, to access cache memory. The management of data stored in the cache memory ensures that for 20 per cent of the total time, during which the cache is searched, the data needed is found to be stored in cache. As a result the performance of the computer improves in terms of speed of processing.

Secondary memory Secondary memory provides large, non-volatile, and inexpensive storage for programs and data. However, the access time in secondary memory is much larger than in primary memory. Secondary storage permits the storage of computer instructions and data for long periods of time. Moreover, secondary memory, which is also known as *auxiliary memory*, stores a huge number of data bytes at a lesser cost than primary memory devices.

Note

- The memory unit is composed of an ordered sequence of storage cells, each capable of storing one byte of data. Each memory cell has a distinct address which is used to refer while storing data into it or retrieving data from it.
- Both RAM and cache memory are referred to as primary memory. Primary memory is comparatively expensive, and loses all its data when the power is turned off. Secondary memory provides less expensive storage that is used to store data and instructions on a permanent basis.

Memory operations There are some operations common to both primary and secondary memory devices. These are as follows.

Read During this operation, data is retrieved from memory.

Write In this operation, data is stored in the memory.

Using read and write operations, many other memory-related functions such as copy and delete are carried out.

Unit of memory The memory's interface circuit is designed to logically access a byte or a multiple of a byte of data from the memory during each access. The smallest block of memory is considered to be a byte, which comprises eight bits. The total memory space is measured in terms of bytes. Thus, the unit of memory is a byte. The capacity of memory is the maximum amount of information it is capable of storing. Since the unit of memory is a byte, the memory's capacity is expressed in number of bytes. Some units used to express the memory capacity are as follows:

Kilobyte (KB)	= 1024 bytes
Megabyte (MB)	= 1024 Kilobytes
Gigabyte (GB)	= 1024 Megabytes
Terabyte (TB)	= 1024 Gigabytes
Petabyte (PB)	= 1024 Terabytes
Exabyte (EB)	= 1024 Petabytes
Zettabyte (ZB)	= 1024 Exabytes
Yottabyte (YB)	= 1024 Zettabytes

The size of the register is one of the important considerations in determining the processing capabilities of the CPU. *Word size* refers to the number of bits that a CPU can manipulate at one time. Word size is based on the size of registers in the ALU and the capacity of circuits that lead to those registers. A processor with a 32-bit word size, for example, has 32-bit registers, processes 32 bits at a time, and is referred to as a 32-bit processor. Processor's word size is a factor that leads to increased computer performance. Today's personal computers typically contain 32-bit or 64-bit processors.

Memory hierarchy The various types of memory used in a computer system differ in speed, cost, size, and volatility (permanence of storage). They can be organized in a hierarchy. The memory hierarchy in the computer system is depicted in Fig. 1.1.

Figure 1.1 shows that on moving down the hierarchy, the cost per bit of storage decreases but access times increases (i.e., devices are slow). In other words, from top to bottom, the speed decreases while the capacity increases and the prices become much lower.

Of the various memories specified in the hierarchy, those above the secondary memory are volatile and the rest are non-volatile. While designing a computer system, there must always be a balance on all of the above factors, namely speed, cost, volatility, etc. at each level in the hierarchy.

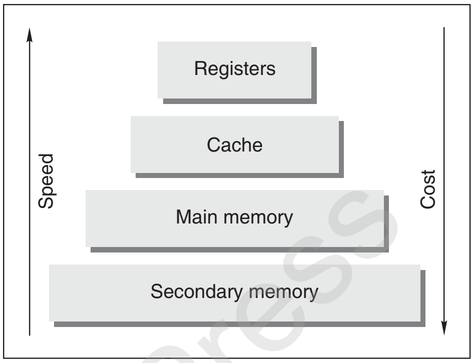


Fig. 1.1 Memory hierarchy

The devices in a computer system other than the CPU and main memory are called **peripherals.** Popular peripheral devices include printers, digital cameras, scanners, joysticks, and speakers.

Interface unit

The interface unit interconnects the CPU with memory and also with the various input/output (I/O) devices. The instructions and data move between the CPU and other hardware components through interface unit.

It is a set of parallel wires or lines which connects all the internal computer components to the CPU and main memory. Depending on the type of data transmitted, a bus can be classified into the following three types:

Data bus The bus used to carry actual data.

Address bus memory or Input/output device Addresses travel via the address bus.

Control bus This bus carries control information between the CPU and other devices within the computer. The control information entails signals that report the status of various devices, or ask devices to take specific actions.

A model of the bus-based computer organization is shown in Fig. 1.2.

Most of the computer devices are not directly connected to the computer's internal bus. Since every device has its own particular way of formatting and communicating data, a device, termed *controller*, coordinates the activities of specific peripherals. The processor reads from the input devices or writes on the output devices with the help of the device controllers. Each input device or output device has a specific address. Using these addresses, the processor selects a particular I/O device through the associated device controller for either transferring data or any control commands.

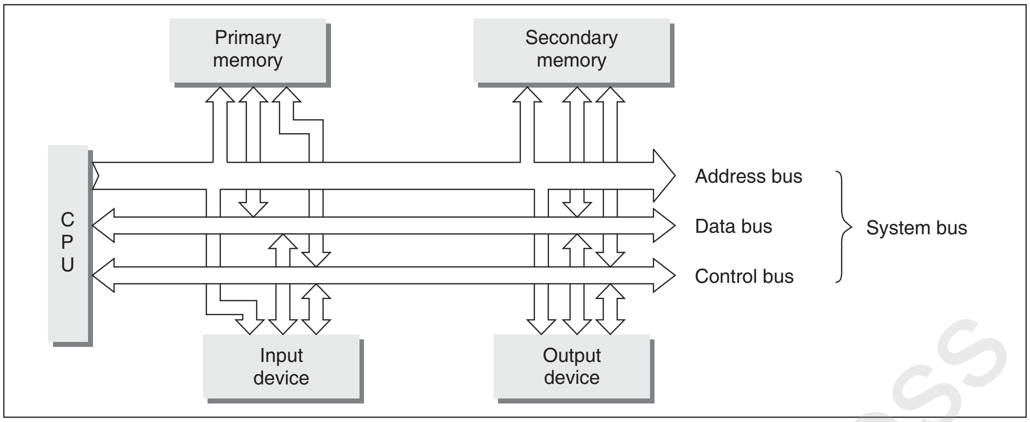


Fig. 1.2 Bus-based computer organization

Motherboard

All the components in the computer system are mounted and connected together by an electronic circuit board called motherboard or main board.

To make all these things work together the motherboard provides some kind of physical connection among them. (See Fig. 1.3).

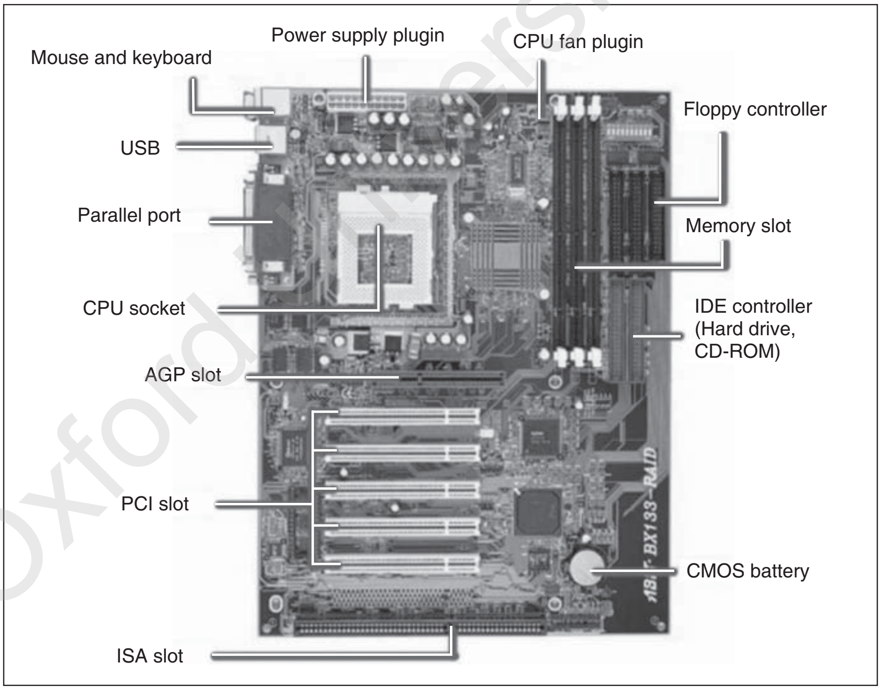


Fig. 1.3 Motherboard

In general, a motherboard consists of the following.

CPU socket This holds the central processor which is an integrated chip along with the system clock, cache, cooling fan, etc.

Memory sockets These sockets hold the RAM card that contains RAMs.

Interface module This is for the hard disk, floppy disk, and CD-ROM drives.

ROM integrated chip This is embedded with the basic input/output system software.

Ports and expansion slots A port is used to connect a device with the bus. Physical ports include serial and parallel ports, to which peripheral devices such as printers, scanners, and external storage devices can be attached. The slots are used to attach accessories such as graphics (video) cards, disk controllers, and network cards. There are two different standards for expansion slots: ISA (Industry Standard Architecture) and PCI (Peripheral Component Interconnect). Most common types of ports and slots are briefly described below.

ISA slots These are for connecting ISA compatible cards.

PCI slots These are for connecting I/O devices.

Advanced graphics port (AGP) Video card is inserted into this slot.

Parallel port The parallel port is also known as the printer port, or LPT1. It is capable of sending eight bits of information at a time.

Serial ports these are sometimes called communication ports or COM ports. There are two COM ports, COM1 and COM2. size of COM1 is larger than that of COM2. COM1 has 25 pins and used for connecting Modems. COM2 is 9 pin port used for interfacing serial mouse. D-type connectors are used to with there ports.

USB (universal serial bus) This is also a serial port but data rate is more than the serial port. USB is used as a general-purpose communication channel in Personal Computers. Many different devices, such as mouse, keyboards, hard disk drives, portable CD-ROM/DVD drives, pen-drives, scanners, cameras, modems and even printers are usually connected to these ports.

CMOS The CMOS in a Personal Computer stands for Complementary Metal Oxide Semiconductor memory. It is a type of RAM that stores the necessary attributes of system components, such as the size of the hard disk, the amount of RAM, and the resources used by the serial and parallel ports etc. Since RAM loses its content when the power is switched off, a small battery, on the motherboard, powers the CMOS RAM even when the computer power is switched off thereby retaining its stored data.

System unit

The System Unit holds all the system components in it. It is sometimes called cabinet. The main components like motherboard, processor, memory unit, power supply unit, and all the ports to interface computer's peripherals. Inside the unit all the components work together to give the service that the user needs. Based on its use, cabinets are of two types.

(i) AT cabinets (or mini-tower)
(ii) ATX cabinets

AT cabinets are smaller and cheaper than ATX cabinets and are popularly called mini-tower cabinets. They are used for older processors and smaller motherboards. ATX cabinets, on the other hand, are marginally larger in size than AT cabinets and are more expensive as they come with more features such as powered sliding front panels and extra disk storage compartments.

Note

- The motherboard is a printed circuit board which contains the circuitry and connections that allow the various components of the computer system to communicate with each other. In most computer systems, the CPU, memory, and other major components are mounted to wiring on the motherboard.
- The input, output, and storage equipment that might be added to a computer system to enhance its functionality are known as *peripheral devices*. Popular peripheral devices include printers, digital cameras, scanners, joysticks, and speakers.

1.5.2 Software

Software provides the instructions that tell the hardware exactly what is to be performed and in what order. This set of instructions is sequenced and organized in a computer program. Therefore, a program is a series of instructions which is intended to direct a computer to perform certain functions and is executed by the processor. In a broader sense, software can be described as a set of related programs. But software is more than a collection of programs. It refers to a set of computer programs, which provide desired functions and performance, the data which the programs use, data structures that facilitate the programs to efficiently manipulate data and documents that describe the operation and use of the programs.
A comparison between computer program and software is listed below (Table 1.2).

Table 1.2 Comparison between computer program and software

Computer program	Software
Programs are developed by individuals. A single developer is involved.	A large number of developers are involved.
Small in size and have limited functionality	Extremely large in size and has enormous functionality.
The user interface may not be very important, because the programmer is the sole user.	For a software product, user interface must be carefully designed and implemented because developers of that product and users of that product are totally different.

Nowadays, most of the software must be installed prior to their use. Installation involves copying several files to computer memory or requires a series of steps and configurations depending on the operating system and the software itself so that it can be run or executed when required.

Software is generally categorized as system software or application software or utility software.

System software

System software is designed to facilitate and coordinate the use of the computer by making hardware operational. It interacts with the computer system at low level. Examples of such software include language translator, operating system, loader, linker, etc. However, the most important system software is the *operating system* which is a set of programs designed to control the input and output operations of the computer, provide communication interface to the user, and manage the resources of the computer system, such as memory, processor, input/output devices etc. and schedule their operations with minimum manual intervention. Other programs (system and application) rely on facilities provided by the operating system to gain access to computer system resources. The *loader* is the system software which copies a executable program from secondary storage device into main memory and prepares this program for execution and initializes the execution.

Hardware devices, other than the CPU and main memory, have to be registered with the operating system by providing a software, known as *device driver*, for communication between the device and other parts of the computer. This type of system software is used by printers, monitors, graphics cards, sound cards, network cards, modems, storage devices, mouse, scanners, etc. Once installed, a device driver automatically responds when it is needed or may run in the background.

Modern operating system recognizes almost all connected hardware devices and immediately begins the installation process. Such a device, for which the operating system automatically starts the installation process, is called a *plug-and-play* device. However, there are few hardware devices for which the user has to manually initiate the installation process.

Application software

Application software is designed to perform specific usages of the users. Microsoft Word, Microsoft Excel, Microsoft Power Point, Microsoft Access, Page Maker, Coral Draw, Photoshop, Tally, AutoCAD, Acrobat, WinAmp, Micro Media Flash, iLeap, Xing MP3 Player etc. are some of the examples of application software.

There are two categories of application software, *custom software* and *pre-written software packages*. Software that is developed for a specific user or organization in accordance with the user's needs is known as *custom software*.

A pre-written software package is bought off the shelf and has predefined generic specifications that may or may not cater to any specific user's requirements. The most important categories of software packages available are as follows:

- Database management software, e.g. Oracle, DB2, Microsoft SQL server, etc.
- Spreadsheet software, e.g. Microsoft Excel.
- Word processing, e.g. Microsoft Word, Corel Wordperfect and desktop publishing (DTP), e.g. Pagemaker.
- Graphics software, e.g. Corel Draw.
- Statistical, e.g. SPSS and operation research software, e.g. Tora.

Note

- Without any software, the computer is called a bare machine, having the potential to perform many functions but no ability to do so on its own.

1.6 MEMORY REVISITED

The different types of memories available for a computer are shown in Fig. 1.4.

1.6.1 Primary Memory

All modern computers use semiconductor memory as primary memory. One of the important semiconductor memories used in desktop computers is ***Random Access Memory (RAM)***. Here "random access" means that any storage location can be accessed (both read and write) directly. This memory is faster, cheaper, and provides more storage space in lesser physical area. These very large-scale integrated semiconductor memory chips are mounted on pluggable printed circuit boards (PCBs). Enhancement or replacement of memory with such PCB memory modules is easy. These characteristics have made semiconductor memory more popular and attractive. The only drawback of semiconductor memory is that it is volatile, i.e., it loses its contents whenever power is switched off. RAM holds the data and instructions waiting to be processed by the processor. In addition to data and program's instructions, RAM also holds operating system instructions that control the basic functions of a computer system. These instructions are loaded into RAM every time the computer is turned on, and they remain there until the computer is turned off. There are two types of RAM used in computer systems– *dynamic* and *static*.

Dynamic RAM (DRAM) is a type of RAM that employs refresh circuits to retain its content in its logic circuits. Each memory cell in DRAM consists of a single transistor. The junction capacitor of the transistor is responsible for holding the electrical charge that designates a single bit as logical 1. The absence of a charge designates a bit as logical 0. Capacitors lose their charge over time and therefore need to be recharged or refreshed at pre-determined intervals by a refreshing circuitry.

A more expensive and faster type of RAM, *Static RAM (SRAM)*, does not require such type of refreshing circuitry. It uses between four to six transistors in a special 'flip-flop' circuit that holds a 1 or 0 while the computer system is in operation. SRAM in computer systems is usually used as processor caches and as I/O buffers. Printers and liquid crystal displays (LCDs) often use SRAM to buffer images. SRAM is also widely used in networking devices, such as routers, switches, and cable modems, to buffer transmission information.

Both dynamic and static RAM are volatile in nature and can be read or written to. The basic differences between SRAM and DRAM are listed in Table 1.3.

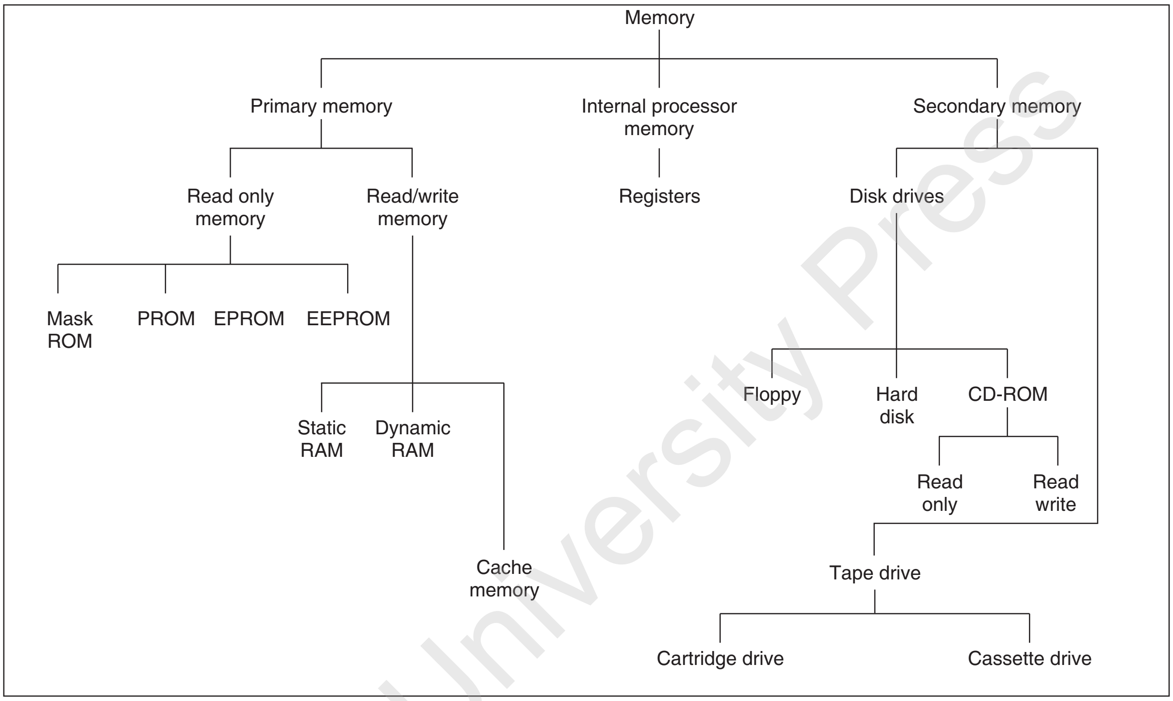


Fig. 1.4 Types of memory

Table 1.3 Static RAM versus dynamic RAM

Static RAM	Dynamic RAM
• It does not require refreshing.	• It requires extra electronic circuitry that "refreshes" memory periodically; otherwise its content will be lost.
• It is more expensive than dynamic RAM.	• It is less expensive than static RAM.
• It is lower in bit density.	• It holds more bits of storage in a single integrated circuit.
• It is faster than dynamic RAM.	• It is slower than SRAM, due to refreshing.

There are several popular types of dynamic RAM used in computers. They are SDRAM (Synchronous Dynamic RAM), RDRAM (Rambus Dynamic RAM) and DDR RAM (Double Data Rate RAM).

The SDRAM used to be the most common type of RAM for personal computers. It was reasonably fast and inexpensive. It is no more used in the present day for personal computers as much improved RAMs are available now.

The RDRAM was developed by Rambus Corporation and is its proprietary technology. It is also the most expensive RAM and is used mostly in video interface cards and high-end computers that require fast computation speed and data transfer. RDRAMs are preferred for high-performance personal computers.

The DDR RAM is a refinement of SDRAM. DDR stands for *Double Data Rate*. It gives faster performance by transmitting data on both the rising and the falling edges of each clock pulse. DDR 2, DDR3 are other higher-speed versions of DDR RAM.

Another type of RAM, termed Video RAM (VRAM), is used to store image data for the visual display monitor. All types of video RAM are special arrangements of dynamic RAM (DRAM). Its purpose is to act as a data storage buffer between the processor and the visual display unit.

There is a persistent mismatch between processor and main memory speeds. The processor executes an instruction faster than the time it takes to read from or write to memory. In order to improve the average memory access speed or rather to optimize the fetching of instructions or data so that these can be accessed faster when the CPU needs it, cache memory is logically positioned between the internal processor memory (registers) and main memory. The cache memory holds a subset of instructions and data values that were recently accessed by the CPU. Whenever the processer tries to access a location of memory, it first checks with the cache to determine if it is already present in it. If so, the byte or word is delivered to the processor. In such a case, the processor does not need to access the main memory. If the data is not there in the cache, then the processer has to access the main memory. The block of main memory containing the data or instruction is read into the cache and then the byte or word is delivered to the processor.

There are two levels of cache.

Level 1 (Primary) cache This type of cache memory is embedded into the processor itself. This cache is very fast and its size varies generally from 8 KB to 64 KB.

Level 2 (Secondary) cache Level 2 cache is slightly slower than L1 cache. It is usually 64 KB to 2 MB in size. Level 2 cache is also sometimes called external cache because it was external to the processor chip when it first appeared.

Read Only Memory (ROM)

It is another type of memory that retains data and instructions stored in it even when the power is turned off. ROM is used in personal computers for storing start-up instructions provided by the manufacturer for carrying out basic operations such as bootstrapping in a PC, and is programmed for specific purposes during their fabrication. ROMs can be written only at the time of manufacture. Another similar memory, Programmable ROM (PROM), is also non-volatile and can be programmed only once by a special device.

But there are instances where the read operation is performed several times and the write operation is performed more than once though less than the number of read operations and the stored data must be retained even when power is switched off. This led to the development of EPROMs (Erasable Programmable Read Only Memories). In the EPROM or Erasable Programmable Read Only Memory, data can be written electrically. The write operation, however, is not simple. It requires the storage cells to be erased by exposing the chip to ultraviolet light, thus bringing each cell to the same initial state. This process of erasing is time consuming. Once all the cells have been brought to the same initial state, the write operation on the EPROM can be performed electrically.

There is another type of Erasable PROM known as Electrically Erasable Programmable Read Only Memory (EEPROM). Like the EPROM, data can be written onto the EEPROM by electrical signals and retained even when power is switched off. The data stored can be erased by electrical signals. However, in EEPROMs the writing time is considerably higher than reading time. The biggest advantage of EEPROM is that it is non-volatile memory and can be updated easily, while the disadvantages are the high cost and the write operation takes considerable time.

Note

- RAM holds raw data waiting to be processed as well as the program instructions for processing that data. It also stores the results of processing until they can be stored more permanently on secondary storage media. Most important point to be noted is that RAM holds operating system instructions which are loaded at start-up and time to time as and when required.
- Dynamic RAM is less expensive, consumes less electrical power, generates less heat, and can be made smaller, with more bits of storage in a single integrated circuit. Static RAM provides faster access with lower bit density and are more expensive than dynamic RAM.
- ROM contains a small set of instructions that tell the computer how to access the hard disk, find the operating system, and load it into RAM. After the operating system is loaded, the computer can accept input, display output, run software, and access data.
- The programmable read-only memory (PROM) is non-volatile and can be reprogrammed only once by a special write device after fabrication. An erasable programmable ROM (EPROM) can be erased by ultraviolet (UV) light or by high-voltage pulses.

1.6.2 Secondary Memory

There are four main types of secondary storage devices available in a computer system:

- Disk drives
- CD drives (CD-R, CD-RW, and DVD)
- Tape drives
- USB flash drives

Hard disk, floppy disk, compact disk (CD), Digital Versatile Disk (DVD) and magnetic tapes are the most common secondary storage mediums. Hard disks provide much faster performance and have larger capacity, but are normally not removable; that is, a single hard disk is permanently attached to a disk drive. Floppy disks, on the other hand, are removable, but their performance is far slower and their capacity far smaller than those of hard disks. A CD-ROM or DVD -ROM is another portable secondary memory device. CD stands for Compact Disc. It is called ROM because information is stored permanently when the CD is created. Devices for operating storage mediums are known as *drives*. Most of the drives used for secondary memory are based on electro-mechanical technology. Mechanical components move much more slowly than do electrical signals. That's why access to secondary memory is much slower than access to main memory.

The **floppy disk** is a thin, round piece of plastic material, coated with a magnetic medium on which information is magnetically recorded, just as music is recorded on the surface of plastic cassette tapes. The flexible floppy disk is enclosed inside a sturdier, plastic jacket to protect it from damage. The disks used in personal computers are usually 3½ inches in diameter and can store 1.44 MB of data. Earlier PCs sometimes used 5¼ inch disks. The disks store information and can be used to exchange information between computers. The floppy disk drive stores data on and retrieves it from the magnetic material of the disk, which is in the form of a disk. It has two motors one that rotates the disk media and the other that moves two read-write heads, each on either surface of the disk, forward Floppy Disk Drive or backward.

A **hard disk** is a permanent memory device mounted inside the system unit. Physically, a hard disk consists of one or more metal (sometimes aluminum) platters, coated with a metal oxide that can be magnetized. The platters are all mounted on a spindle, which allows them to spin at a constant rate. Read/write heads are attached to metal arms and positioned over each of the platter surfaces. The arms can move the read/write heads radially inwards and outwards over the surfaces of the platters (see Fig. 1.5). Data and programs are stored on the hard disk by causing the write heads to make magnetic marks on the surfaces of the platters. Read heads retrieve the data by sensing the magnetic marks on the platters. The surface of each platter is divided into concentric rings called tracks. The tracks form concentric circles on the platter's surface. Each track is divided into a certain number of sectors. A sector is capable of generally 512 bytes or sometimes 1,024 bytes of data. The head is mounted on an arm, which moves or seeks from track to track. The vertical group of tracks at the same position on each surface of each platter is called a cylinder. Cylinders are important, because all heads move at the same time. Once the heads arrive at a particular track position, all the sectors on the tracks that form a cylinder can be read without further arm motion. The storage capacity of a hard disk is very large and expressed in terms of gigabytes (GB). The data that is stored on the hard disk remains there until it is erased or deleted by the user.

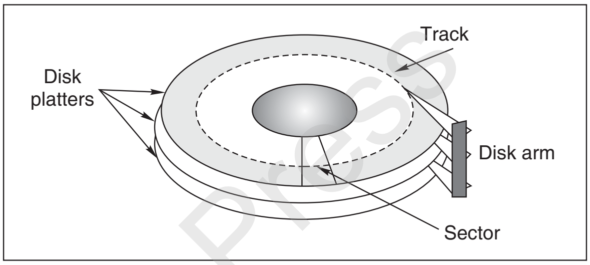


Fig. 1.5 Hard disk organization

The hard disk drive provides better performance and become mandatory for computer systems for the following reasons:

- Higher capacity of data storage
- Faster access time of data
- Higher data transfer rates
- Better reliability of operation
- Less data errors or data loss

A **CD** is a portable secondary storage medium. Various types of CDs are available: CD-R and CD-RW. CD-RW drives are used to create and read both CD-R and CD-RW discs. Once created (i.e. when it has been "burned"), data stored on CD-R (CD-Recordable) disc can't be changed. On the other hand, a CD-Rewritable (CD-RW) disc can be erased and reused. This disk is made of synthetic resin that is coated with a reflective material, usually aluminum. When information is written by a CD-writer drive, some microscopic pits are created on the surface of the CD. The information bit on a CD-ROM surface is coded in the form of ups and downs (known as pits and dumps), created by infrared heat. There is one laser diode on the reading head. The bits are read by shining a low - intensity laser beam onto the spinning disc. The laser beam reflects strongly from a smooth area on the disc but weakly from a pitted area. A sensor receiving the reflection determines whether each bit is a 1 or a 0 accordingly. CDs were initially a popular storage media for music; they were later used as general computer storage media. Most personal computers are equipped with a CD-Recordable (CD-R) drive. A CD-Rewritable (CD-RW) disc can be reused because the pits and flat surfaces of a normal CD are simulated on a CD-RW by coating the surface of the disc with a material that, when heated to one temperature becomes amorphous (and therefore non-reflective) and when heated to a different temperature becomes crystalline (and therefore reflective).

1.7 INTRODUCTION TO OPERATING SYSTEMS

A computer system has many resources such as the processor (CPU), main memory, I/O devices, and files. The operating system acts as the manager of these resources and allocates them to specific programs and uses them as and when necessary for the tasks.

An operating system may be defined as a system software which acts as an intermediary between the user and the hardware, an interface which isolates the user from the details of the hardware implementation. It consists of a set of specialized software modules that makes computing resources (hardware and software) available to users. Thus, the computer system is easier to use with the operating system in place than without it. Some of the operating systems used nowadays are Mac, MS Windows, Linux, Solaris, etc.

The common functions of an operating system includes –

Process(or) management The process abstraction is a fundamental mechanism implemented by the operating system for management of the execution of programs. A process is basically a program in execution. The operating system decides which process gets to run, for how long and perhaps at what priority or level of importance.

Memory management Operating system is responsible for keeping track of which parts of the memory are currently being used and by whom. It organizes and addresses memory; handle requests to allocate memory, frees up memory no longer being used, and rearranges memory to maximize the useful amount. Often several programs may be in memory at the same time. The operating system selects processes that are to be placed in memory, where they are to be placed, and how much memory is to be given to each.

Device management The operating system allocates the various devices to the processes and initiates the I/O operation. It also controls and schedules accesses to the input/output devices among the processes.

File management A file is just a sequence of bytes. Files are storage areas for programs, source codes, data, documents etc. The operating system keeps track of every file in the system, including data files, program files, compilers, and applications. The file system is an operating system module that allows users and programs to create, delete, modify, open, close, and apply other operations to various types of files. It also allows users to give names to files, to organize the files hierarchically into directories, to protect files, and to access those files using the various file operations.

Apart from these functions, an operating system must provide the facilities for controlling the access of programs, processes, memory segments, and other resources.

The *kernel* is that part of operating system that interacts with the hardware directly. The kernel represents only a small portion of the code of the entire OS but it is intensively used and so remains in primary storage while other portions may be transferred in and out of secondary storage as required. When a computer boots up, it goes through some initialization functions, such as checking the memory. It then loads the kernel and switches control to it. The kernel then starts up all the processes needed to communicate with the user and the rest of the environment.

The user interface is the portion of the operating system that users interact with directly. Operating systems such as MS-DOS and early versions of UNIX accepted only typed-in text commands. Now most operating systems provide users a graphical user interface for their interactions with the system. Operating systems such as Microsoft Windows, Solaris and Linux allow the user to interact with the operating system through icons, menus, keyboard and mouse movements. The user interface and way of interactions vary widely from one operating system to another.

1.7.1 Loading an Operating System

In some digital devices like controllers of small appliances, hand-held devices and videogame console, the operating system is relatively simple and small and is stored in ROM (Read-Only Memory). The operating system is also present in a ROM for systems such as industrial controllers and petrol-filling equipment. In such a system, it gains immediate control of the processor, the moment it is turned on.

In personal computer, the operating system is usually stored on hard disk. Because size of the operating system is large enough, it cannot be placed entirely in RAM. The kernel, the core part of the operating system, is loaded into RAM at start-up and is always present in memory. Other parts of the operating system are loaded into RAM as and when required. It is to be noted that there is no operating system resident in a new computer. The operating system is usually sold on a CD or DVD media and has to be permanently transferred from a CD or DVD media to the hard disk by expanding compressed files and initializing the whole system for use.

Booting is the general term for the process that a computer or other digital device follows from the instant it is turned on until the operating system is finally loaded and ready for use.

The Basic Input Output System (BIOS) is a small set of instructions stored on a PROM that is executed when the computer is turned on.

When the computer is switched on, the ROM circuitry receives power and begins the boot process. At first, an address is automatically loaded into the Program Counter (PC) register. This is done by hardware circuitry. The address given is the location of the first executable instruction of the BIOS. The code in the BIOS runs a series of tests called the POST (Power On Self Test) to make sure that system devices such as main memory, monitor, keyboard, the input/output devices are connected and functional. During POST, the BIOS compares the system configuration data obtained from POST with the system information stored on a Complementary Metal-Oxide Semiconductor (CMOS) memory chip located on the motherboard. The BIOS also sets various parameters such as the organization of the disk drive, using information stored in a CMOS chip. This CMOS chip gets updated whenever new system components are added and contains the latest information about system components.

The BIOS then loads only one block of data, called the *Master Boot Record*, from a specific and fixed place (the very first sector at cylinder 0, head 0, and sector 1) of the bootable device and is placed at a specific and fixed place of main memory. The master boot record is of 512 bytes in size and contains machine code instructions, called a *bootstrap loader*. Then the boot loader program starts the process of loading the OS and transfers control to the OS itself which completes the process.

Note

- *Cold boot* describes the process of starting the computer and loading its operating system by turning the power on. If the computer is running, one can carry out cold boot by first switching it off and then back on.
- *Warm boot* describes the process of restarting the computer and loading its operating system again without switching it off after it has already been running.

1.8 OPERATIONAL OVERVIEW OF A CPU

Any processing executed by central processing unit is directed by the instruction. The processing required for a single instruction is called an *instruction cycle*. The four steps which the CPU carries out for each machine language instruction are *fetch, decode, execute, and store* (Fig. 1.6).

The steps involved in the instruction cycle while executing a program are described below.

The Program Counter (PC) is the register that keeps track of what instruction has to be executed next. At the first step, the instruction is *fetched* from main memory and loaded into Instruction Register (IR), whose address is specified by PC register. Immediately the PC is incremented so that it points to the next instruction in the program. Once in IR, the instruction is *decoded* to determine the actions needed for its execution. The control unit then issues the sequence of control signals that enables *execution* of the instruction. Data needed to be processed by the instructions are either fetched from a register from RAM through an *address register*. The result of the instruction is *stored* (written) to either a register or a memory location. The next instruction of a program will follow the same steps. This will continue until there is no more instruction in the program or the computer is turned off, some sort of unrecoverable error occurs.

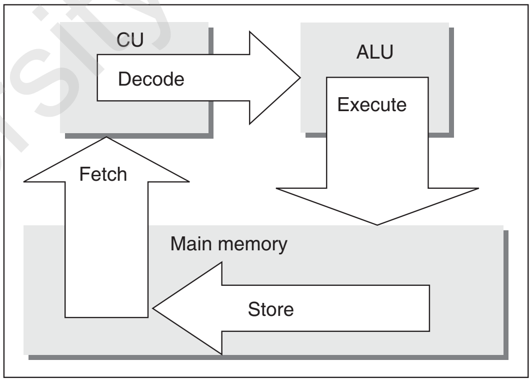


Fig. 1.6 A simplified view of an instruction cycle

Note

A register is a single, permanent storage location within the CPU used for a particular, defined purpose. CPU contains several important registers such as

- The *program counter(PC)* register holds the address of the current instruction being executed.
- The *instruction register (IR)* holds the actual instruction being executed currently by the computer.

To access data in memory, CPU makes use of two internal registers:

- The memory address register (MAR) holds the address of a memory location.
- The memory data register (MDR), sometimes known as the memory buffer register, will hold a data value that is being stored to or retrieved from the memory location currently addressed by the memory address register.

SUMMARY

A computer is defined as 'an automatic electronic apparatus for making calculations or controlling operations that are expressible in numerical or logical terms'.

Starting from the days of the abacus, the concept of developing a computing machine has led to the development of the modern electronic computing machine. There are five generations of computers. Today computers are available in various forms such as personal computers, laptop, palmtop, and mainframes. The electronic computer, of all sizes, perfected through years of development, has become a powerful machine capable of being employed in a variety of applications. A computer has a CPU, a fast-access primary memory (RAM), a non-volatile high storage capacity secondary memory (HDD), an easy-to-use keyboard, a video color monitor console with a graphic pointer device such as mouse and a non-impact printer.

Thus, broadly, the basic computer system consists of a CPU, memory, and input and output devices. Memory can be classified into primary, secondary, and internal processor memory. Cache memory is a part of the primary memory and normally resides near the CPU. The rest of the primary memory consists of various types of ROMs and RAMs.

A PC consists of hardware and software. Software can be classified into system software and application software. The most important system software is the operating system that manages all resources of the computer system and acts as an interface between hardware and software. When the personal computer is switched on, a power on self test (POST) is executed and the operating system is loaded.

KEY TERMS

ALU The Arithmetic Logic Unit (ALU) performs arithmetic and logical operations on the data.

BIOS Basic Input-Output System (BIOS) is a small set of instructions stored in ROM which runs every time when the computer is switched on. BIOS is responsible for Power On Self Test to make sure every immediately required device is connected and functional and finally loading the core part of the operating system into RAM.

Cache memory It is a special high-speed memory that allows a microprocessor to access data more rapidly than from memory located elsewhere on the system board.

CMOS The Complementary Metal Oxide Semiconductor (CMOS) chip in the computer stores information about various attributes of the devices connected to the computer.

Control unit It interprets each instruction and determines the appropriate course of action.

Computer It is programmable device that can store, retrieve, and process data.

CPU It is an Integrated circuit chip which is the ultimate controller of the computer, as well as the place where all calculations are performed.

Hardware It refers to the physical components of a computer.

RAM Random Access Memory (RAM) is a volatile memory that is used to store data and instructions temporarily. It holds raw data waiting to be processed as well as the program instructions for processing that data. It also holds operating system instructions, which control the basic functions of a computer system.

ROM Read Only Memory (ROM) is permanent and nonvolatile memory. It is the place to store the "hard-wired" startup instructions of a computer. These instructions are a permanent part of the circuitry and remain in place even when the computer power is turned off.

Software It refers to the set of computer programs and to the data that the programs use.

FREQUENTLY ASKED QUESTIONS

1. What is a microprocessor?

A microprocessor is an integrated circuit chip that contains all of the essential components for the central processing unit (CPU) of a microcomputer system.

2. What is a chip?

A chip is a small, thin piece of silicon onto which the transistors making up the integrated circuits, e.g. microprocessors have been imprinted.

3. What is a chipset?

In personal computers a *chipset* is a group of integrated circuits that together perform a particular function.

4. What is booting?

The sequence of events that occurs between the time that a computer is turned on and the time it is ready for use, is referred to as *booting*.

5. Where is the operating system stored?

In some digital devices— typically handhelds and videogame consoles— the entire operating system is small enough to be stored in ROM (read-only memory). For most other computers, the operating system program is quite large, so most of it is stored on a hard disk. During the boot process, the operating system kernel is loaded into RAM. The *kernel* provides essential operating system services. Other parts of the operating system are loaded into RAM as and when they are needed.

6. What is a plug-and-play device?

A device for which the installation process starts automatically by the operating system and which usually does not require any human intervention, is called a plug-and-play device.

7. If a computer contains RAM, why does it need ROM too?

Normally, the instructions and data are stored in a secondary storage devices permanently. In addition to data and program instructions currently being processed, RAM also holds operating system instructions that control the basic functions of a computer system. These instructions are loaded into RAM every time when the computer is booted, and they remain resident until the computer is turned off. But RAM is a volatile memory i.e. its content will be lost when the power is turned off. Now ROM plays the important role. ROM contains a small set of instructions called the BIOS (Basic Input Output System). These instructions access the hard disk, find the operating system, and load it into RAM. After the operating system is loaded, the system is ready to be used.

EXERCISES

1. Write full forms of the following:

ENIAC, ALU, CU, RAM, ROM, EPROM, EEPROM, BIOS, POST, MIPS, CMOS

2. Briefly describe the functions of the different components of a conventional digital computer with the help of a suitable block diagram.

3. What is a CPU? What is its function? Mention its several components.

4. Explain the different memory units.

5. Discuss the memory hierarchy within a computer system.

6. What is cache memory? Why is it necessary?

7. Give three examples of system software.

8. Briefly state the role of the operating system in a computer system.

9. What is BIOS? Describe its functions.

10. What is meant by POST?

11. What is the boot sector?

12. Describe the bootstrap process.

13. Distinguish between the following:

(a) Compiler and interpreter

(b) System software and application software

(c) RAM and ROM

(d) Primary memory and secondary memory

(e) Bit and byte

(f) Hardware and software