Eastern Economy Edition

Second Edition

Introduction to Information Technology

V. Rajaraman

PHI

Introduction to INFORMATION TECHNOLOGY

SECOND EDITION

V. Rajaraman
Honorary Professor
Supercomputer Education and Research Centre
Indian Institute of Science
Bangalore

PHI Learning Private Limited
Delhi-110092
2013

INTRODUCTION TO INFORMATION TECHNOLOGY, Second Edition
V. Rajaraman

ISBN-978-81-203-4731-1

The export rights of this book are vested solely with the publisher.

Eleventh Printing (Second Edition) ... ... April, 2013

Published by Asoke K. Ghosh, PHI Learning Private Limited, Rimjhim House, 111 Patparganj Industrial Estate, Delhi-110092 and Printed by Rekha Printers Private Limited, New Delhi-110020.

Contents

CHAPTER

1 Data and Information

LEARNING GOALS

After reading this chapter, you should be able to:

1. Explain the difference between data and information.
2. Classify different types of data which are processed by computers.
3. Explain the functions of the units of a desktop computer.
4. Describe how data is processed by a computer.

1.1 INTRODUCTION

Information Technology (IT) may be defined as the technology that is used to acquire, store, organize, process, and disseminate processed data which can be used in specified applications. Information is *processed data* that improves our knowledge, enabling us to take decisions and initiate actions.

Example 1.1 Let us take a very simple example. A home-maker who buys vegetables, provisions, milk, etc., everyday would write in a diary the money spent on each of these (see Table 1.1). At the end of each day she adds up the data on money spent for these items. The total obtained is the information which she uses to adjust expenses to spend within her budget. This is illustrated in the block diagram of Fig. 1.1.

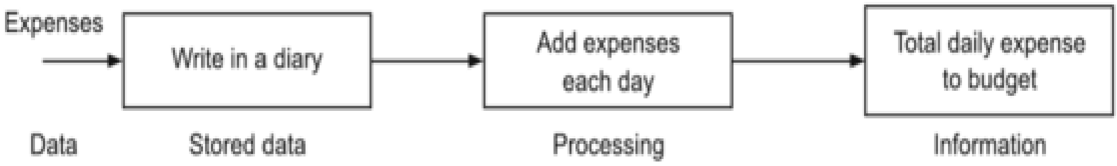


FIG. 1.1 Data and information.

Table 1.1 Daily Expenses

Date	*Expenses in rupees*				
	Vegetables	*Milk*	*Provisions*	*Miscellaneous*	*Daily total*
1.1.2001	25.50	20.00	95.00	150.00	290.50
2.1.2001	30.40	20.00	85.40	250.50	386.30
3.1.2001	15.50	25.00	128.00	80.00	248.50
:	:	:	:	:	:
:	:	:	:	:	:
31.1.2001	19.50	20.00	25.00	15.00	79.50
Total	750.50	650.00	2800.50	2852.50	7053.50

Observe that data is the raw material with which she started, and information is processed data that allows her to initiate action to balance her budget.

The data entered in the diary each day may be processed in other ways too to obtain different information. For example, if the total monthly expense on milk is divided by the monthly income, it gives information on the proportion of the budget spent on milk. This is shown in Fig. 1.2.

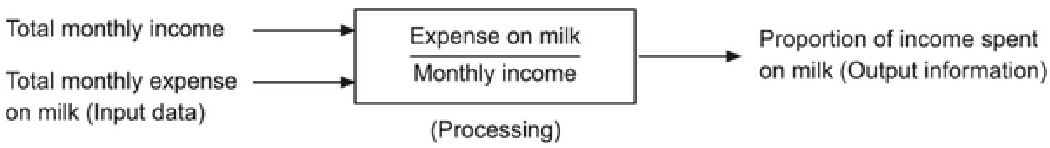


FIG. 1.2 Information as input data.

This information may be useful to manage the family income in a more efficient manner. Observe that the information obtained in Fig. 1.1 is used as data in Fig. 1.2. This illustrates that the distinction between data and information is not always clear. The point to be emphasized is that mere facts and figures about activities do not enable one to take decisions or to initiate actions. Only when they are processed and presented in an effective manner, they become useful.

Example 1.2 As an example of how organizing data enhances our understanding, let us consider the marks obtained by students in an examination. The marks by themselves do not give any immediate idea about the performance of the class. By processing this data, a bar chart may be obtained, which gives the number of students with marks between 100 and 90, 90 and 80, 80 and 70, and so on. This chart (Fig. 1.3) gives the teacher of the class information on the performance of the class that would enable him or her to initiate appropriate action such as which students need special attention.

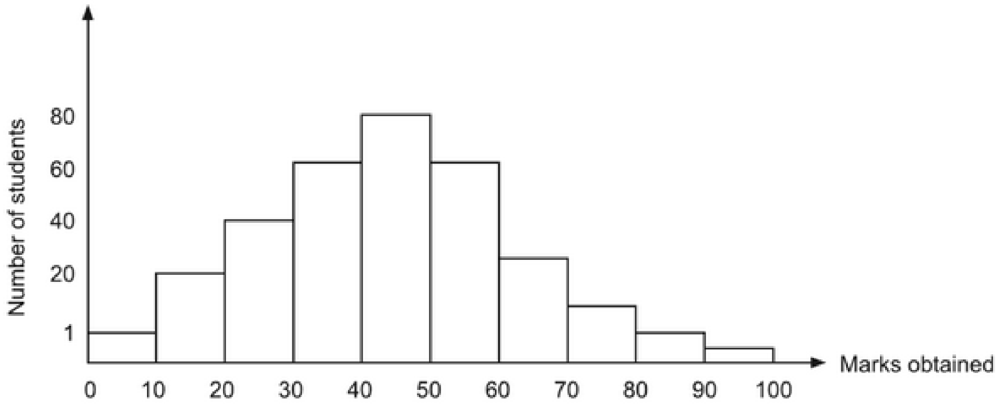


FIG. 1.3 Bar chart giving performance of students in a class.

Example 1.3 Sometimes data becomes meaningful if the context in which it occurs is given. For example, 560012 is data. If one writes Bangalore 560012, then we may infer that 560012 is the postal code of a locality in Bangalore. Thus, one may also define information as data in a specified context. (In fact, we have processed the string "Bangalore 560012" using our common sense knowledge and obtained the information that 560012 is a post code.)

1.2 TYPES OF DATA

In Examples 1.1 and 1.2, the data was numeric. This is the simplest data type and historically the earliest data type to be processed. The versatility of IT comes from the ability to process a variety of data types (Fig. 1.4).

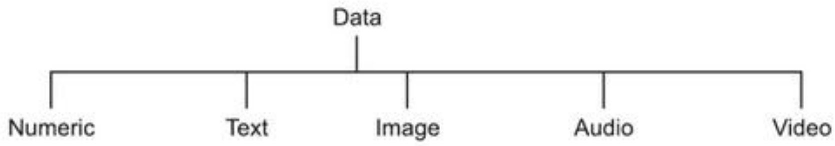


FIG. 1.4 Types of data.

Some additional data types, besides numbers, are:

1. *Text.* For example, a paragraph in this book is textual data.
2. *Picture or image.* For example, your photograph (both black and white and colour) is image data. Other types of images are a map of India, a fingerprint, a line drawing as in Fig. 1.3 above, an image transmitted by a satellite, and an X-ray of your chest. Their common feature is that they are two dimensional objects. In the literature the terms picture and image are used interchangeably.
3. *Audio* or *sound.* For example, speeches, songs, telephone conversations, street noise, etc. Their main property is that they are waveforms whose amplitude vary with time and cause pressure waves in the air which enter our ears and we hear the sound (Fig. 1.5).

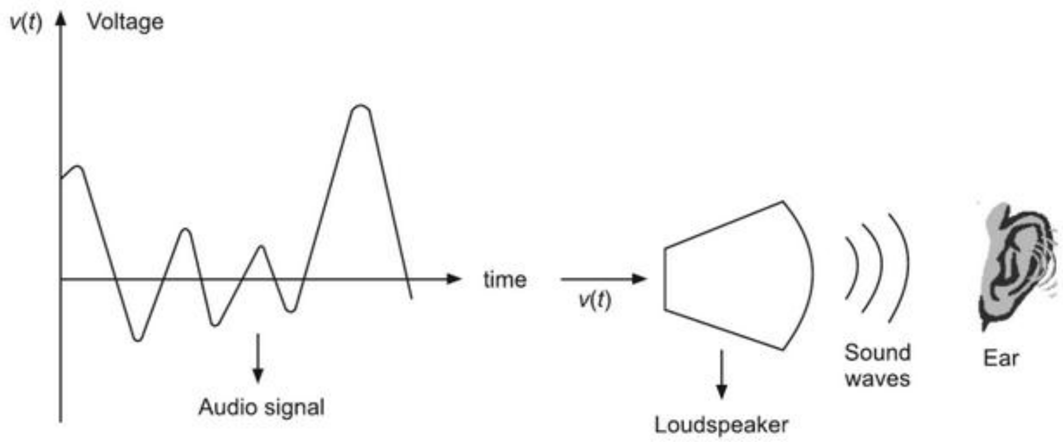


FIG. 1.5 Audio signal to sound conversion.

4. *Video or moving pictures.* When a number of pictures (each one slightly different from the other) are shown one after another at a rate of about 30 to 60 pictures per second, due to persistence of vision, we have an illusion of movement. The most popular examples of video are movies such as silent movies starring Charlie Chaplin in the 1930s. Another example is animation used in computer games. Video is usually combined with audio to give a better effect, for example, computer games with background noises synchronized with the images or current cinemas (called "talkies" in early days when sound was introduced in movies).

In modern Information Technology all these types of data are required to be processed. All these data types taken together are colloquially called "multimedia". We will answer in this chapter and subsequent chapters the following questions:

1. What are these data types?
2. Why should we process these data types? In other words, what are the main objectives of processing these data types?
3. How to represent these data types for processing by computers?
4. How are these data types processed?

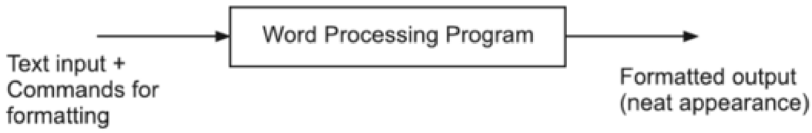
1.2.1 Text Data

In order to process these data types, we should be able to read or enter the data into a computer, process them, and display and distribute (on request) the results of processing to the outside world. We first begin by considering textual data.

Example 1.4 Word Processing

One of the most important applications of IT now is *word processing*. In this application a rough draft of a text (may be an essay) is typed using the keyboard of a computer. A number of commands are also given along with the text to format it, i.e., right-justify the text, arrange the text as paragraphs, italicize some words, check spelling etc. The data type in this example is a string of characters or a text. The output is formatted text which has a neat appearance and is thus easy to read, as illustrated in Fig. 1.6.

A word processor also does many other interesting tasks. For example, it detects spelling errors and (if you command it) corrects them. Some word processors also examine the construction of a sentence and gives suggestions to improve the sentence.



(a)

This is a sample of an unformatted text which is not right justified. It also has speling errors. This text has no special fonts suh as italics and boldface. In Fig. 1.4(b) ths text is formated, high-lighted, right justified and correted.

(b) Unformatted text with errors

This is a sample of an *unformatted text* which is not right justified. It also has spelling errors. This text has no special fonts such as italics and boldface. In Fig. 1.6(c) this text is **formatted**, highlighted, right justified and corrected.

(c) Text formatted and spell checked using a word processor

FIG. 1.6 Text processing, unformatted and formatted text.

Example 1.5 Dictionary

Another interesting aspect of IT is the way it is able to enhance our ability to perform many day-to-day tasks. For example, if you do not know the meaning of a word you refer to a printed dictionary. Words in a dictionary are arranged alphabetically to enable you to search the word quickly. Nowadays, dictionaries are published not only in printed form but also in computer readable form, either on the web, or using a Compact Disk Read Only Memory (CDROM), which is similar to the music CDs you buy in the audio shops. The words are now arranged not only in an alphabetical order but each word is also linked with words with similar meaning. Thus, if you want to find the meaning of a word, you type it using the keyboard of a computer and its meaning is displayed. No manual searching is needed. Not only do you *get* the meaning but also a list of other words which have the same or similar meaning. This will allow you to pick the right word which meets your requirement.

There are also other interesting things which are done by the designers of the computer readable dictionary. Every word is linked not only with words having a similar meaning but also with those that have an opposite meaning. For example, the word "early" is linked with words such as "punctual", "prompt", etc., which have a similar meaning. It is also linked with words such as "late", "tardy", "delay", which have an opposite meaning. Given a word, this linking will allow you to search for words with opposite meaning. Besides this, very complex

1.2.3 Audio Data

Audio or sound data such as music and speech is used by us everyday. Audio data also needs processing.

EXAMPLE 1.8 Flight Data Recorder

Conversations between persons in the cockpit of an aircraft and flight controllers at airports are routinely recorded in a "black box" during every flight of an aircraft. If there is an air crash, this record is converted to digital form, processed to remove the background noise, and improve the clarity of recorded speech to assist investigators.

EXAMPLE 1.9 Digital Audio Recording

Music recorded in old records are re-recorded digitally after removing background noise and enhancing their quality. These records can be compressed to a form called MP3. This enables a few hours of music to be stored on a single CD. Audio players called MP3 players are now available to play this digitized compressed music.

EXAMPLE 1.10 Music Synthesis

Computers are nowadays used to synthesise music using basic notes. The synthesis can be done using many innovative ideas resulting in interesting new tunes.

EXAMPLE 1.11 Internet Telephone

Voice data is now transmitted using the worldwide network of computers, called the Internet, making telephone calls to foreign countries much cheaper than before.

1.2.4 Video Data

As we saw earlier in this chapter, a moving picture or video consists of a sequence of still images which are projected in rapid succession. Due to persistence of our vision we get an illusion that we see rapid motion. The still images must be repeated at least 30 to 60 times a second to get the illusion of a moving picture. Video data is often synchronized with audio data as in a motion picture.

EXAMPLE 1.12 Movie Applications

An example of video data processing is putting colour on old black and white movies. Another example is dubbing a movie with a sound track in a language which is different from the original language in which the movie was taken.

EXAMPLE 1.13 Image Morphing

Morphing of images on a VDU screen is a process by which successive images smoothly change from one frame to the next giving an illusion of transformation of one type of image to another. For example, the image of a young girl gradually becomes old and is

Example 1.11 Digital Audio Recording

Music recorded in old audio cassette tapes are re-recorded digitally after removing background noise and enhancing their quality. These records can be compressed to a form called MP3. This enables a few hours of music to be stored on a single CD. Audio players called MP3 players are now available to play this digitized compressed music. MP3 compression will be described in Chapter 4.

Example 1.12 Music Synthesis

Computers are nowadays used to synthesize music using basic notes. The synthesis can be done using many innovative ideas resulting in interesting new tunes.

Example 1.13 Internet Telephone

Voice data is now transmitted using the worldwide network of computers called the Internet. Telephone calls to foreign countries are much cheaper now. We will discuss this in greater detail in Chapter 15.

1.2.4 Video Data

As we saw earlier in this chapter, a moving picture or video consists of a sequence of still pictures which are projected in rapid succession. Video data is often synchronized with audio data as in a motion picture.

Example 1.14 Movie Applications

An example of video data processing is putting colour on old black and white movies. Another example is dubbing a movie with a sound track in a language which is different from the original language in which the movie was taken. It is even possible to create movies by using computer generated characters. For example, a movie "Toy Story" was produced using this method.

Example 1.15 Image Morphing

Morphing of images on a VDU screen is a process by which successive images smoothly change from one frame to the next giving an illusion of transformation of one type of image to another. For example, the image of a young girl gradually becomes old and is then transformed into that of a lion! This type of transformation of images is often shown in television advertisements.

1.3 SIMPLE MODEL OF A COMPUTER

In Section 1.2 we have seen that we deal with not only numerical data but also text, images, audio and video as shown in Fig. 1.4. Information Technology is concerned with:

1. Acquisition of data
2. Storage of data
3. Organization of data
4. Processing of data

5. Output of processed data, i.e., information
6. Dissemination or distribution of information.

If a machine is to be a versatile data processing machine it should:

- be able to acquire or read data of all the five types, i.e., numbers, text, images, audio and video and also be able to read instructions to process the data.
- have the facility to store and organize data. The amount of data storage required is quite large—trillions of characters nowadays. It should also be able to store the instructions to process data.
- be able to process data. Data is processed by interpreting and executing a set of instructions called a *program*, stored in the machine's primary memory unit.
- have devices to output the processed data. Remember that output data can be numeric, textual, audio, image or video. Sometimes the results may also be stored in the storage unit for future use (see Fig. 1.7).
- also be easily connectable to other computers using communication networks for widely disseminating information.

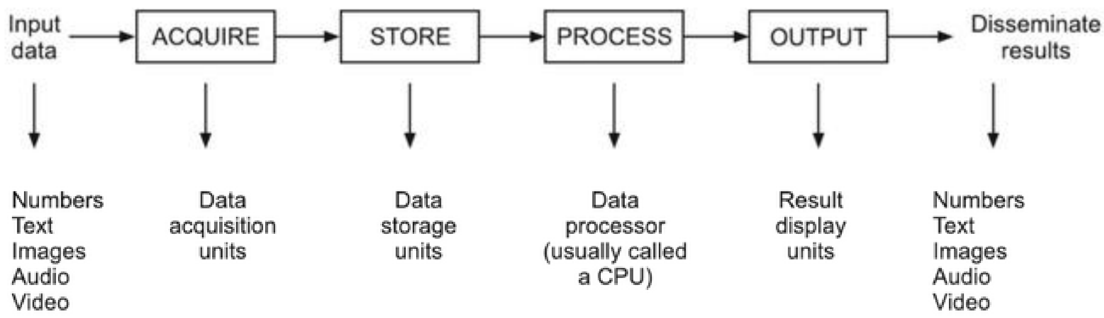


FIG. 1.7 Various steps in data processing.

Such a machine has a structure shown in the block diagram of Fig. 1.8. It has an *input system* which is used to acquire data from the external world and convert it into a form which can be stored in its storage system. There are a variety of input units ranging from a keyboard and mouse to specialized data acquisition systems such as video cameras, microphones and scanners. The collection of input units constitutes the input system.

The machine has a *memory system* where the program for processing data and the data to be processed are stored. The memory system also consists of a variety of units, the most important one being a fast Random Access Memory (RAM), which is the primary memory used to store the instructions and the data to be processed. As the amount of data to be stored is usually very large and the stored data is used for many purposes, the storage system has, besides a RAM, a variety of units such as hard disks, CDROMs, DVDROMs, flash memory (also known as pen drives) and magnetic tape drives. The collection of this storage is called *secondary storage*. RAM is normally volatile, i.e., when power is switched off, the data stored in it is lost. Secondary storage systems are non-volatile. They retain the data stored in them when power is switched off.

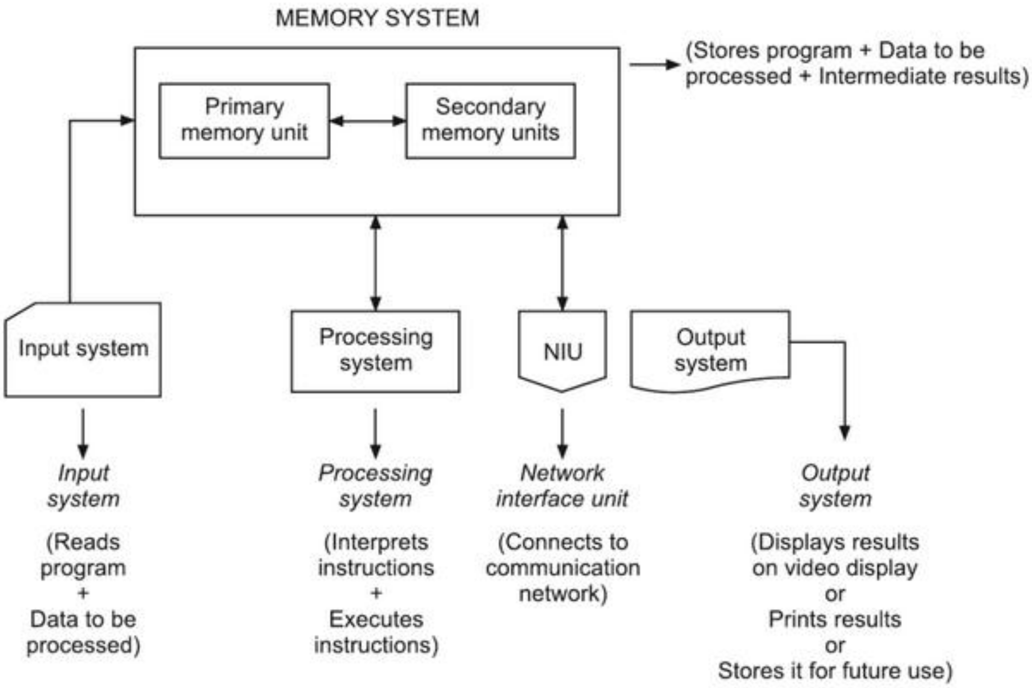


FIG. 1.8 Block diagram of a data processing system (or computer system).

The third important system is the *processing system*. The processing system is the heart of the machine and is designed to interpret and execute the instructions of a program stored in the memory. There are a variety of processing systems. The simplest ones are used to control microwave ovens, washing machines, etc. Faster powerful processors are used for complex mathematical calculations. Some others process audio and video data and are known as Digital Signal Processors (DSP). Many computing systems now have several processing units.

The next important part of a computer is the *output system*. The output system prints or displays the results of data processing. Sometimes the results of processing may not be immediately printed. They may be stored for printing later. Sometimes data may be printed or displayed and also stored for future reference. There are several types of output devices such as video displays, inkjet printers, loudspeakers and laser printers.

The last important part of all recent computers is a Network Interface Unit (NIU) or Network Interface Card (NIC) to connect it to a communication system and through this to other computers. Nowadays it is rare to find an isolated computer not connected to a communication system. Such a connection is essential to widely disseminate data processed by the computer.

1.4 DATA PROCESSING USING A COMPUTER

To fix our ideas we will take an example to show how data is processed to obtain information. The steps followed are:

1. Analyze the given data processing task and understand *what is to be done*.
2. Having understood the task to be performed find a method to do it.

3. Express the method to be followed as a step-by-step procedure which is called an *algorithm.*
4. Express the algorithm using a precise notation called a *programming language* obtaining what is called a *computer program. Programs* can be interpreted and executed by a computer's processing system.
5. Input the program to be executed and store it in the memory of the computer. Keep the data to be processed ready and waiting at the input unit.
6. Order the computer to start executing the program.
7. The computer interprets the program stored in its memory. When a command to "READ" data is encountered (which is normally at the beginning of the program) it reads data waiting at the input unit and stores it in the memory of the computer. It then continues following the program step by step and carries out the data processing task.
8. At the end of the program being executed (or during the execution of a program), an instruction(s) will be found to write the result(s) via the output unit. This is the processed data, that is, the information which is required. It may also send the results to another computer connected to it.

We will now illustrate the steps given above with an example.

Example 1.16 Counting Vowels in a Text

The task to be performed:

Count the number of vowels in a given text.

Method

Read the text character by character. If the character is a vowel, that is, it is a, e, i, o, u or A, E, I, O, U, count it as a vowel. When the end of the text is reached, output the count as the number of vowels.

Algorithm

Step 1 Create a counter to count the number of vowels and store 0 in it.
count-vowels = 0

Step 2 Call vowel-set = {a, e, i, o, u, A, E, I, O, U}
Repeat Step 3 to Step 5 until the end of input text is reached. Go to Step 6 when no more characters are left in the text.

Step 3 Read a character from the string of characters (or text) waiting at the input unit. Store it in input-character.

Step 4 *If* input-character is in vowel-set
then Add 1 to count-vowels and continue
else continue

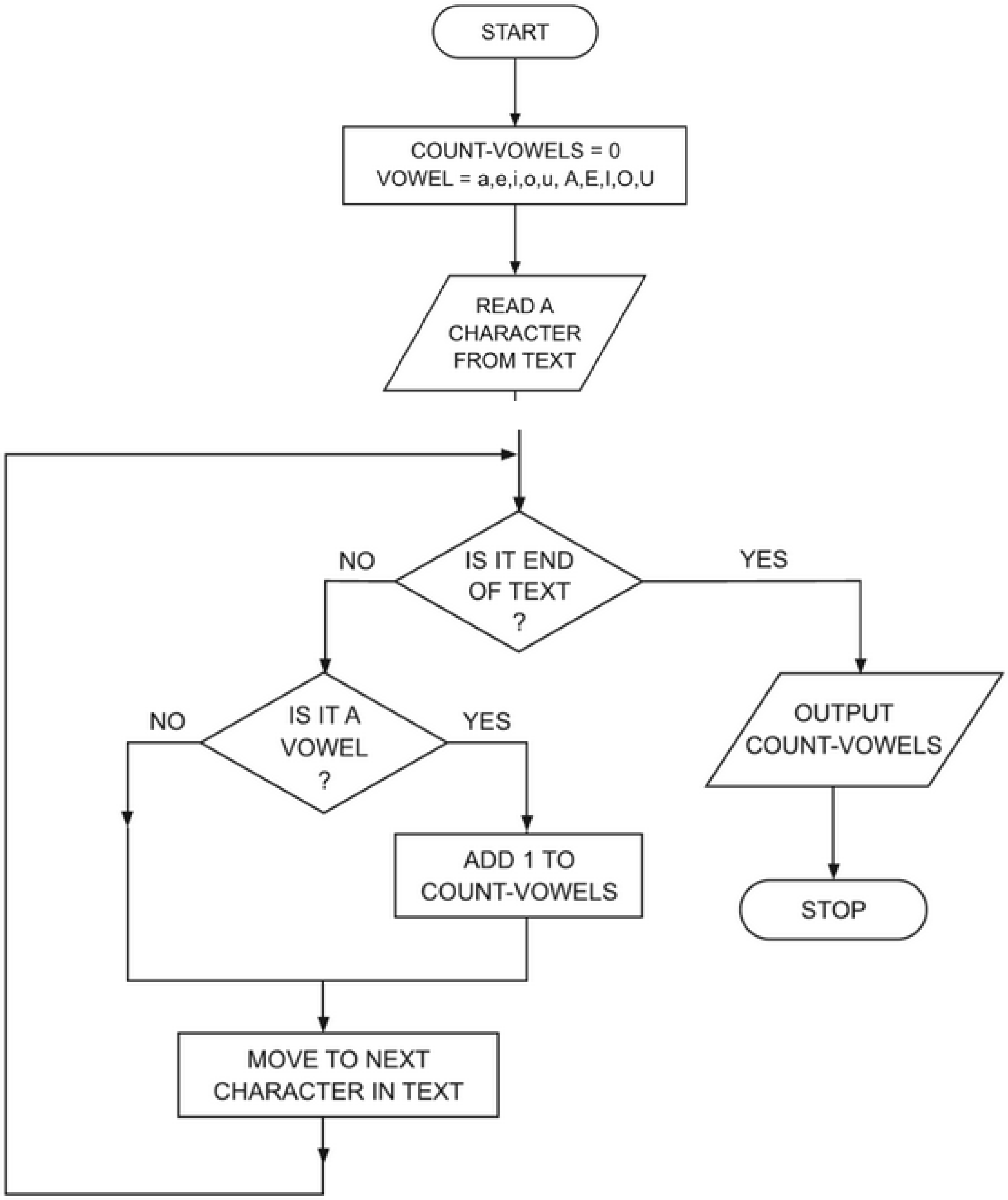
Step 5 Move to next character. *Remark*: We now go back to Step 3

Step 6 Output count-vowels. *Remark*: This step is reached when end of input text is reached.

Step 7 Stop

Observe that the algorithm is independent of the length of the text and what is contained in the text. Given *any* text, the algorithm will find the number of vowels in it. In other words, the method is general and not dependent on specific data input. This is what gives the power to a computer. Once an algorithm is written, it may be used for all similar tasks. This property of an algorithm which makes it independent of input data is called *data independence of algorithm*.

The algorithm is often pictorially shown as a flowchart. Flowchart is easy to understand for beginners. They use a notation standardized by the International Standards Organization. We have shown the algorithm as a flowchart in Fig. 1.9 using the standard notation. Flowcharts are used to document algorithms.



(vi) An electronic circuit called a *graphics card* to display pictures and video. This card is mounted on the motherboard.

(vii) A small electronic video camera external to the PC to capture images and video.

All the above constitute what is known as *hardware*. A computer is useless unless it has a number of ready-made programs to enable easy use of the hardware. This is provided by *software*. The most important software required to use a computer is what is known as an *Operating System* (OS). It is also essential to have a set of *application programs*. An OS coordinates the activities of various parts of a computer, namely, the CPU and the input/output units, manages memory, organizes data and allows you to control the activities of the PC. The most popular OS for PCs is called Windows OS made by Microsoft Corporation. Another OS gaining ground is called GNU/LINUX which is free software.

There are two major sets of application software used today. One set called "productivity programs" is used to improve normal office tasks such as word processing, keeping track of appointments, organizing files, preparing presentations, calculating installments to be paid on hire purchase, etc. This set of programs is called an *Office Suite*. Microsoft office and Open office are two such programs (Open office is free whereas Microsoft office has to be bought). The other set of programs uses the Internet. This set consists of a program called a browser used to access the World Wide Web, emailing programs, programs to search the World Wide Web for information, programs to access social networking services such as Facebook, Twitter, blogs and programs which allow you to video call a friend (e.g., Skype).

1.6 THE ORGANIZATION OF THE BOOK

At the beginning of this chapter, we stated that Information Technology may be defined as the technology which is used to acquire, store, organize, and process data, as well as disseminate processed data which can be used in specified applications. We also said that by data we mean not only numbers but also text, pictures, audio, and video. Thus, in this book, we describe how to acquire, store, organize, process, and distribute data of all five forms (known as multimedia). Nowadays processing of multimedia data is by computers and all computers are digital; that is, internally they store and process data only in a binary form, namely, 0s and 1s. We will explain the reasons for this in the next chapter. Also data processing requires both hardware and software. Thus in this book we discuss both aspects, namely, hardware and software.

In Chapters 2, 3, 4, and 5, we show how to represent, acquire, and convert to digital form numbers, text, pictures, audio, and video data. Acquisition requires both hardware devices and ideas for conversion of continuously varying signals to a digital form. Thus we explain various methods of digitally encoding numbers, text, pictures, audio, and video, as well as reading their binary equivalents and storing them in a computer's storage unit.

In Chapters 6, 7, 8 and 9 we focus on hardware units required to process, store, output, and widely make accessible the processed data via a worldwide interconnected network of computers. We describe, respectively, in these chapters various storage devices, processing units, computer networks, and output devices.

In Chapters 10, 11, 12, 13 and 14, we deal primarily with software. We discuss in Chapter 10 why operating systems are needed, what they do and at an elementary level how they work. In Chapter 11 we show how data of various types can be systematically organized for easy retrieval. We then describe in Chapters 12, 13 and 14 what software is available to process numerical, textual, image, audio and video data. We briefly explain how to use some of these ready-made programs.

In Chapters 15, 16 and 17 we discuss some widely used important applications of Information Technology. Chapter 15 describes various important applications of the Internet, a worldwide network of computers. We explain the use of email and how to use the World Wide Web to disseminate and obtain information of interest to you. Chapter 16 describes the use of computers to process data for business and industry. One of the major applications of IT today is to effectively manage large organizations. We discuss various types of information needed by organizations and how they are obtained by appropriate processing.

IT and the worldwide computing infrastructure provided by Internet have led to a revolution in doing business and running governments. We read every day in the newspapers how you can shop sitting at home. We also read about various initiatives taken in e-governance in which IT is applied by state governments in attempts to provide responsive and transparent administration. In Chapter 17 we show how IT is used in e-commerce. The same principles are applicable to e-governance. Recognizing the spread of IT and its use in commerce, Government of India is one of the few governments in the world to enact legislation to legalize e-transactions and prevent misuse of the worldwide network of computers. We briefly describe IT Act 2000 and its amendments in 2008 enacted by Parliament.

IT has started to affect our society in many new ways, both good and bad. In the last Chapter (Chapter 18), we discuss some of the societal impacts of IT. A number of free services such as email, wikis, and storage on the Internet are now available. Anyone can post pictures, videos, and music clips on the World Wide Web. Social networking sites such as Facebook, Twitter, and LinkedIn are proliferating. This chapter is intended to bring awareness among young students about these new services and the challenges and opportunities which have emerged recently due to Information Technology.

1.7 EPILOGUE

One of the major characteristics of Information Technology is rapid change. In 1965, Gordon E. Moore (one of the founders of Fairchild Semiconductors, U.S.A) predicted, based on data available at that time, that the density of transistors in integrated circuits will double at regular intervals of around 2 years. Based on the experience from 1965 to date, it has been found that this prediction is surprisingly accurate. In fact, the number of transistors per integrated circuit chip has doubled approximately every 19 months. This is called "Moore's Law". In 1974 the largest DRAM chip had 16Kbits, whereas in 2010 it was 100 Gbits, a growth of 6 million times in 36 years (doubling almost every 19 months). The number of transistors in microprocessors has also grown similarly. The clock speed of microprocessors which indicates how fast they process data was 2 MHz in 1974, whereas it was 3 GHz in 2010 (a growth of

15000 times). The growth of secondary storage (namely, disk) has been no less spectacular. In 1974 the storage capacity of a single disk drive was around 2 Mbytes, whereas it was 2 TB in 2010 (a growth of million times in 26 years—doubling of capacity almost every 15 months). This huge increase of processing power and memory capacity has led to the development of several novel applications besides significantly improving existing application software. Another significant event in the last decade is the rapid increase of communication bandwidth of both terrestrial and wireless systems, which has added another dimension to the growth of Information Technology. In the history of technology, no other technology has grown and changed as rapidly as Information Technology. Thus writing a book on this subject is a challenge. Numbers given in this book such as typical memory size, speed of processing of images and speed of scanning of books to create a digital library are valid at the time of writing this book. They will change. The basic principles presented in the book would not change. Thus while reading this book, students should be aware that the number of bytes in the main memory, disks and pen drives is not invariant. They should search the World Wide Web to get the current numbers when they read this book.

SUMMARY

1. Information Technology may be defined as the technology which is used to acquire, store, organize, and process data as well as disseminate processed data which can be used in specified applications.
2. Information is processed data based on which decisions can be taken and actions initiated. It also improves our knowledge which enables us to do our work better.
3. There are five types of data which are processed. They are numbers, text, images (or pictures), audio, and video.
4. Each of these types of data requires a different method of acquisition, storage, processing, and display. A modern data processing system should be able to process all these types of data.
5. The term computer is universally used to name a data processing system.
6. A computer has an input system to read programs and data, a memory system which is used to store data as well as instructions needed to process data, a processor to process data, and an output system to disseminate the processed data to the outside world.
7. All modern computers also have a unit to connect them to a communication network.
8. In order to process data, first the processing task is understood and a step-by-step method (known as an algorithm) is evolved to carry out the task.
9. The algorithm is expressed more precisely using a programming language. This is called a program.
10. The program is stored in the memory and the data for processing is fed via the input unit whenever it is required.

11. The program stored in the memory is interpreted and executed by the processing unit.
12. The results of processing are displayed by the output unit.
13. The idea of storing a program in the computer's memory allows repeated execution of the same set of instructions as they are readily available in the memory.
14. The most important requirement of an algorithm is input data independence. By generalizing an algorithm, it is possible to use it several times when only the input data changes.
15. A desktop computer is nowadays widely used. It has a keyboard and mouse as input units, video display unit (VDU) as output unit, a microprocessor as a CPU, electronic memory as the main memory and a hard disk as secondary storage.
16. A desktop computer requires software called an operating system to coordinate the activities of all its units. It has a set of ready-made application programs for common applications such as word processing, emailing and simple accounting functions.

EXERCISES

1.1 Define Information Technology.

1.2 What is the difference between data and information? Give an example of organizing data which allows human decision-making.

1.3 What are the different types of data? Give an example of each of these data types.

1.4 What is the major difference between audio data and textual data?

1.5 What are the data types which are functions of time?

1.6 What is the difference between picture and video?

1.7 Define multimedia data.

1.8 What is word processing? What are the facilities provided by a word processor?

1.9 What special facilities are provided by a computer readable and processable dictionary as compared to a text-based dictionary?

1.10 Give examples of image or picture data.

1.11 What are the advantages of obtaining a computer drawn plan of a building?

1.12 Give two examples of audio data.

1.13 What is image morphing? Have you seen any morphed images?

1.14 What features should a machine have to perform the tasks required by Information Technology?

1.15 Give the block diagram of a machine capable of processing data and disseminating it. Explain the functions of each of the blocks.

1.16 What is the difference between primary memory and secondary storage unit?

1.17 What are the steps followed in processing data using a computer?

1.18 What is an algorithm? Write an algorithm to count all punctuation marks in a text.

1.19 Write an algorithm to find whether a given number is odd or even. Obtain a flowchart for this algorithm.

1.20 Express the algorithm you developed in Exercise 1.18 as a flowchart.

1.21 Develop an algorithm to count the number of capital letters in a text. Express it as a flowchart.

1.22 What is the main advantage of storing a program in the main memory of a computer?

1.23 What do you understand by data independence of an algorithm? Illustrate the idea of data independence with an example.

1.24 What are the parts of a desktop computer such as a PC? Draw a block diagram of such a computer.

1.25 What is a keyboard used for? Is it an input unit or an output unit?

1.26 What is a VDU? What is its application?

1.27 What is an icon? How is it used?

1.28 What is a mouse? What is its use? Is it an input or an output unit?

1.29 What is a motherboard? What parts of a computer are accommodated in a motherboard?

1.30 What storage devices are included in the secondary storage unit?

1.31 Which are the removable secondary storage units of a PC?

1.32 What is a scanner? What is its use?

1.33 In order to process audio data what units should be provided in a PC?

1.34 What is the application of a graphics card in a PC?

1.35 What software is essential to use a PC?

1.36 What is the main use of an Operating System (OS)? Name two OS for PCs.

1.37 What is the most commonly used set of application software in a PC? What applications does it include?

The cursor now advances to the next line and displays the grand total. The clerk now presses another function key which commands the computer to print the bill (Table 2.3) using a printer attached to the computer. Observe that the bill has company name, etc. preprinted. Only the data about each item, price, quantity, etc. is printed by the PC. The bill can now be given to the customer to receive payment.

Table 2.3 Bill printed by the Computer

ABC GENERAL STORES
285, 6th Main Street
Gandhinagar
Bangalore 560 037

Serial no.	*Item code*	*Item name*	*Price*	*Qty.*	*Total price*
1.	45789	Blue Jeans	585.50	2	1171.00
2.	78494	Black Belt	125.25	1	125.25
3.	94862	Socks	55.25	4	221.00
			Grand Total		1517.25

Thank You. Visit again

In this example we saw that the clerk entered the data using a keyboard. The keyboard is called an *input device*.

The data are entered by the clerk manually using the keyboard. This is called *manual data entry*. The entered data is used by the *processor*, which does the required additions and multiplications, and displays the final bill on the video display. The video display unit is an example of an *output device*. The bill is printed by a *printer* attached to the computer, another example of an *output device*.

In this example, numeric and character *data* were *acquired manually* by a clerk (for processing by the computer) using a keyboard. We will now show how the keyboard is organized. Besides the keyboard, there are many other devices used to input numbers and text. We will describe some of them in the next section.

2.2 INPUT UNITS

2.2.1 Keyboard

A picture of a keyboard is shown in Fig. 2.1. This type of keyboard has been standardized for use in all types of computers such as PC, laptop or a tablet. (There are special keyboards for children, industrial use, etc., and one may get details about them in the list of references at the end of this book). We will be describing only the standard keyboard. This keyboard is also called the QWERTY keyboard as these letters are the first six letters in the third row from top. This arrangement of letters was standardized for mechanical typewriters in the last century but has continued for legacy reasons. The keyboard consists of the following major categories of keys:

FIG. 2.1 A typical keyboard of a computer.

1. *Letter keys.* These are the 26 letters of English alphabet arranged as shown.
2. *Digit keys.* There are two sets of digit keys; one on the second row from the top (see Fig. 2.1) of the keyboard, and the other is a numeric key pad at the right which allows quick entry of numbers with the fingers of one hand.
3. *Special character keys.* These are characters such as <, >, ?, /, {, }, [,], (,), ., ;, ", \, !, @, #, $, %, ^, &, *, –, +, =, _. Most of these are printed when the shift key in the keyboard is pressed down and the key on which it is written is pressed. For example, when shift key and the key with digit 9 in the second row from top are pressed together, (is printed.
4. *Non-printable control keys.* These are used for backspacing, going to the next line, tabulation, moving cursor up or down, insert, delete characters, etc. There is also a *space bar* at the bottom for leaving a space.
5. *Function keys.* These are labeled F1, F2, up to F15 and can be configured to invoke programs stored in the computer when pressed. For example, in the billing problem explained in the last section, after entering quantity and price, the clerk pressed a function key to multiply price by quantity and display the product under the next column of the bill. Similarly, another function key computes the grand total and displays it at the bottom of the last column of the bill (see Table 2.3).

The functions of some of the non-printable control keys are listed below:

Delete key. This key deletes the current character and backs the cursor line to the previous character.

Return key. After typing a line, this key is pressed to go to the next line.

Arrow keys move the cursor up, down, back or forward as indicated by the arrow.

Tab key moves the cursor to the next tab stop.

Shift key. When kept pressed and a letter key is pressed, the capital letter is printed.

When a digit key, for example 5, on top line and the shift key are pressed, the character above the digit, namely, %, is printed. This can be locked by pressing the caps lock key. When the caps key is locked, a light on the keyboard is lit. This indicator light is to remind the user of the locked/unlocked status.

2.2.2 Character Reader

In Example 2.1 we assumed that the label with the item code and price is read by a clerk and the data manually entered by him/her using the keyboard. Manual data entry is slow. A clerk may type the values incorrectly either due to carelessness or the label being not legible. It is thus desirable to eliminate manual data entry. It is done by printing the numbers on the label using a special font which is recognizable by what is known as a *hand-held scanner*. A hand-held scanner is a device similar to a pen which emits light at its tip. When it is slowly dragged over the label, the white parts reflect light and the dark parts do not. This is used by an electronic eye in the device to transfer the image to a memory in the device. This is matched with a set of pre-stored fonts and the digits are recognized. This sequence of digits is sent to the computer which displays it on the screen of the video display and stores it in its main memory. This method is both faster and more accurate as compared to manual data entry.

2.2.3 Magnetic Ink Character Reader (MICR)

In this method specially styled characters are printed on documents such as cheques using magnetic ink.

The document is fed to a unit called *magnetic ink character reader*. This recognizes the characters. In a cheque, for instance, the cheque number and the bank code are preprinted with magnetic ink at the bottom of the cheque (see Fig. 2.2). The amount of the cheque is entered later using a magnetic ink character writer. The cheque can now be read by a special input unit which can recognize magnetic ink characters. At the end of a day, all such cheques are collected by a clearing house, processed by a computer and correct amounts are debited from the specified accounts.

FIG. 2.2 A MICR cheque. The line at the bottom of the cheque uses Magnetic Ink and special fonts.

2.2.4 Bar Codes

In Example 2.1 we stated that each item bought by a customer has a label containing the product code and the product price and that this data is entered manually by a clerk. We also stated that such a manual data entry can lead to human errors in entering the data. Thus it is better to eliminate manual data entry. In order to do this, a system called a *bar code* has been developed. The bar code consists of a set of vertical lines of varying widths. Universal product code has standardized bar codes for various items to uniquely identify them. For example, on the right corner of the back cover of this book, the bar code uniquely identifying this book is printed. The unique identifier for books is called International Standard Book Number (ISBN). This number is printed on top of the bar code. Observe that the ISBN digits use a font that is easy for a character recognizer to recognize.

The bar code is scanned using a device called a *bar code scanner*. Hand-held scanners are inexpensive and easy to use. The alignment of the scanner in relation to the bars is not very critical as the bars are quite long. Table mounted scanners are also available and are easy to use for reading labels printed on bottles, cans, etc. Bar codes are printed on gummed labels and stuck to the merchandise. When the bar code is scanned, the item code and name are recognized by the scanner and printed in the bill (Table 2.3) eliminating manual data entry. A table with item code and price of all items stocked by the stores is stored in the computer's disk memory. When the bar code of the item is read, this table is looked up by the computer and the price is retrieved and printed. Apart from reducing manual effort and eliminating errors due to manual data entry, this method has another advantage. When the price of an item changes, it is only necessary to change it in the table stored in the computer's disk memory.

Two dimensional bar code known as *QR (Quick Response) code* has recently become very popular. It consists of black square dots of various sizes arranged in a square pattern printed on white sheets. Each QR code can store a lot of alphanumeric information, much more than a bar code. They can even be read by smart mobile phone cameras. Many vendors use QR code to store their website address known as URL and print them in newspaper advertisement. Thus when a smart mobile camera is pointed to it, the home page of the vendor appears on the phone. (We will explain in detail later the terms website and URL.)

2.3 INTERNAL REPRESENTATION OF NUMERIC DATA

Once a data is fed to a computer it is converted to a form that is efficient to store and easy to interpret by the hardware of the machine. This is called the *internal representation* of data. This should be contrasted with the *external representation* which is the form that is easy for humans to read and understand. For example, we use the decimal system to represent numbers for our use; this is the external representation. Another example of external representation is the bar code we described in the last section. A decimal number is represented in another form called binary for storing in the memory of a computer. This is the internal representation.

At this point we would like to point out that even though we are concerned with processing multiple varieties of data, there is a unifying aspect when we examine their internal representation. All these data types are represented by a string of *binary digits* (zeros and ones). They are all processed by a computer which can process only data represented as zeros and ones. The results obtained after processing are also zeros and ones but in a rearranged form based on the processing rules used. This result is converted back to numbers, text, audio, etc., which can be understood by us. In other words, even though the processing is on binary data and the results are also binary, they are converted back to an external representation, namely, numbers, text, audio, image, or video data as appropriate.

The question that will now be foremost in your minds is: Why all data should be converted to zeros and ones? The answer is that the physical devices used to store and process data in computers (as of today) are *two-state* devices. A switch, for example, is a two-state device. It can be either ON or OFF. Very reliable recording or reading on a magnetic surface (such as the one used in hard disks) is achieved when the surface is magnetized in either one of two opposite directions. The two states in this case are magnetic field aligned 'left to right' (S $\rightarrow$ N) or 'right to left' (N $\leftarrow$ S). Electronic devices such as transistors used in computers function most reliably when operated as switches, that is, either in conducting mode or in non-conducting mode. Thus, all data to be stored and processed using computers are transformed or coded as strings of two symbols, one symbol to represent each state. The two symbols normally used are 0 and 1. They are known as *bits*, an abbreviation for *b*inary dig*its*.

2.3.1 Encoding of Decimal Numbers

The next question is: How do we represent numbers using bits? The simplest way is to represent each digit in a number by a unique string of bits. There are 4 (2 × 2) unique combinations of 2 bits, namely:

00 01 10 11

This is because each bit can be either 0 or 1. There are 2 bits giving 2 × 2 = 4 possible unique strings of 2 bits each. There are 2 × 2 × 2 = 8 unique strings of three bits each, i.e.,

000 001 010 011 100 101 110 111

There are ten decimal digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Thus 3-bits strings are not sufficient to represent 10 digits. We need at least 4 bits. The number of unique 4 bits strings are 2 × 2 × 2 × 2 = 16 and they are:

0000	0001	0010	0011	0100	0101	0110	0111
1000	1001	1010	1011	1100	1101	1110	1111

We can arbitrarily pick any 10 of these 16 strings and assign each of them to represent 0, 1, 2, ..., 9. Any such mapping between 4-bit strings and the decimal digits is known as *binary encoding* of decimal digits. One of these assignments is called *natural binary coded decimal digits* (NBCD) and is shown in Table 2.4.

Table 2.5 Binary Counting Sequence

Binary number	*Decimal equivalent*	*Binary number*	*Decimal equivalent*
0	0	1001	9
1	1	1010	10
10	2	1011	11
11	3	1100	12
100	4	1101	13
101	5	1110	14
110	6	1111	15
111	7	10000	16
1000	8	10001	17

Just as powers of 10 are important in the decimal system of enumeration, powers of 2 are important in the binary system. We give in Table 2.6 powers of 2 and their decimal equivalent. The abbreviation K in Table 2.6 stands for 1024 which is approximately 1000, a Kilo. Thus the notation 16K means 16 × 1024 = 16384. The abbreviation M (Mega) stands for 1024 × 1024 = 1048576, which is nearly a million. The abbreviation G (Giga) is used to represent 1024 × 1024 × 1024, which is nearly a billion, T (Tera) for 1024 × 1024 × 1024 × 1024, nearly a trillion, and P (Peta) for 1024 × 1024 × 1024 × 1024 × 1024, nearly a quadrillion, 10^{15}.

Table 2.6 Powers of 2

Power of 2	*Decimal equivalent*	*Power of 2*	*Decimal equivalent*	*Abbreviation*
2^0	1	2^{10}	1024	1K
2^1	2	2^{11}	2048	2K
2^2	4	2^{12}	4096	4K
2^3	8	2^{20}	1048576	1M
2^4	16	2^{21}	2097152	2M
2^5	32	2^{22}	4194304	4M
2^6	64	2^{30}	1073741824	1G
2^7	128	2^{31}	2147483648	2G
2^8	256	2^{40}	1099511627776	1T
2^9	512	2^{50}	1125899906842624	1P

Example 2.4 What is the value of 3M?

Solution $M = 2^{20} = 1048576$ from Table 2.6

Thus, 3M = 3 × 1048576 = 3145728

2.3.3 Representation of Fractions

So far we have considered decimal and binary integers. Decimal fractions are interpreted as follows:

$$0.235 = 2 \times 10^{-1} + 3 \times 10^{-2} + 5 \times 10^{-3}$$

0.235	$= 2 \times 10^{-1}$	$+ 3 \times 10^{-2}$	$+ 5 \times 10^{-3}$
↑	↑	↑	↑
Decimal point	One-tenth position	One-hundredth position	One-thousandth position

Observe that negative powers of 10 are used as weights to multiply the digits in the fractional part of the number.

A binary fraction is represented by a string of 1s and 0s on the right of a *binary point*. The bits are multiplied by negative powers of 2 to obtain the decimal value of the binary fraction as shown below:

0.1011	=	1×2^{-1}	+	0×2^{-2}	+	1×2^{-3}	+	1×2^{-4}
↑		↑		↑		↑		↑
Binary point	=	1/2	+	0	+	1/8	+	1/16
	=	11/16 = 0.6875 (in decimal)						

We give below some more examples of binary numbers and their decimal equivalents:

Example 2.5

$$\begin{aligned}(111011.101)_2 &= 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &\quad + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \\ &= 32 + 16 + 8 + 0 + 2 + 1 + 1/2 + 0 + 1/8 \\ &= (59.625)_{10}\end{aligned}$$

Note that we have used the subscript 2 to indicate that the number is binary, and the subscript 10 to indicate that the number is decimal. This notation to represent the base of a number is useful to prevent misinterpretation of numbers.

2.3.4 Hexadecimal Representation of Numbers

The binary equivalent of a 10 digit number will be approximately 32 bits long. It is difficult to write such long strings of 1s and 0s and convert them to equivalent decimal numbers without making mistakes. The *hexadecimal* system, which uses 16 as base, is a convenient notation to express binary numbers. This system, by definition, uses 16 symbols, viz., 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F. Note that the symbols A, B, etc. now represent numbers in hexadecimal. As 16 is the fourth power of 2, namely 2^4, there is a one-to-one correspondence between a hexadecimal digit and its binary equivalent. We need 4 bits to represent a hexadecimal digit. Table 2.7 gives a table of hexadecimal digits and their binary and decimal equivalents.

Table 2.7 Binary, Hexadecimal, and Decimal Equivalent

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

A binary number can be quickly converted to its hexadecimal equivalent by grouping together successively 4 bits of the binary number, starting with the least significant bit and replacing each 4-bit group with its hexadecimal equivalent given in Table 2.7. The following examples illustrate this.

Example 2.7

Binary number	0111	1100	1101	1110	0011
Hexadecimal equivalent:	7	C	D	E	3

Example 2.8

Binary number 001000111110000.0010110

Example 2.8

Binary number 001000111110000.0010110

Grouped binary number	0001	0001	1111	0000	·	0010	1100
Hexadecimal equivalent:	1	1	F	0	·	2	C

Observe that, in Example 2.8 above, groups are formed from the least significant bit for the integer part of the number and from the most significant bit for the fractional part. If the number of bits in the integer part is not a multiple of 4, we insert *leading* 0s, as these have no significance for the integer part. If the number of bits in the fractional part is not a multiple of 4, then we introduce *trailing* 0s, as these have no significance in the fractional part.

Conversion from hexadecimal to decimal system is simple. It uses the fact that the base of the hexadecimal system is 16. We give below two examples of hexadecimal to decimal conversion.

Example 2.9

$$\begin{aligned} \text{D6C1} &= \text{D} \times 16^3 + 6 \times 16^2 + \text{C} \times 16^1 + 1 \times 16^0 \\ &= 13 \times 16^3 + 6 \times 16^2 + 12 \times 16 + 1 \times 16^0 \\ &= 53248 + 1536 + 192 + 1 \\ &= (54977)_{10} \end{aligned}$$

Example 2.10

$$\begin{aligned} (\text{F9A} \cdot \text{BC3})_{16} &= \text{F} \times 16^2 + 9 \times 16^1 + \text{A} \times 16^0 + \text{B} \times 16^{-1} + \text{C} \times 16^{-2} + 3 \times 16^{-3} \\ &= (15 \times 256) + (9 \times 16) + (10 \times 1) + 11/16 + 12/256 + 3/4096 \\ &= 3840 + 144 + 10 + 11/16 + 12/256 + 3/4096 \\ &= (3994.7351074)_{10} \end{aligned}$$

2.3.5 Decimal to Binary Conversion

As mentioned at the beginning of this section, decimal representation is used by people whereas binary representation is used by computers. For decimal numbers whose value is not significant in an application, we encode the decimal number by replacing each of the digits by its 4-bit string equivalent given in Table 2.4. However, when the value of the number is important, we have to convert decimal to binary. The method of converting a decimal integer to its binary equivalent uses the fact that any decimal integer may be expressed as a sum of powers of 2 as shown below:

$$d = (23)_{10} = (1 \times 2^4) + (0 \times 2^3) + (1 \times 2^2) + (1 \times 2^1) + (1 \times 2^0)$$
$$= (10111)_2$$

The easiest way to find the coefficients of the powers of 2 is to divide the given number by 2, and each of the successive quotients by 2. Division is terminated when a quotient becomes zero. The binary equivalent of the decimal number is given by the sequence of remainders obtained during division. The least significant bit of the binary number is the first remainder obtained and its most significant bit is the last remainder. The procedure is illustrated now:

Example 2.11 Find the binary equivalent of $(23)_{10}$

Solution

2	23	*Remainder*	
2	11	1	→ Least significant bit
2	5	1	
2	2	1	
2	1	0	
	0	1	→ Most significant bit

$(23)_{10} = (10111)_2$

Example 2.12 Find the binary equivalent of 36

Solution

2	36	*Remainder*	
2	18	0	→ Least significant bit
2	9	0	
2	4	1	
2	2	0	
2	1	0	
	0	1	→ Most significant bit

$(36)_{10} = (100100)_2$

Decimal fractions may also be converted to binary. The method is based on observing that a decimal fraction is expressed as a sum of negative powers of 2. Successive multiplication of the fraction by 2 would give the coefficients of the negative powers of 2.

Example 2.13 Find the binary equivalent of 0.8125

Solution

$$
\begin{aligned}
(0.8125)_{10} &= 0.5 + 0.25 + 0.0625 \\
&= 1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4} \\
2 \times (0.8125)_{10} &= 2 \times (1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4}) \\
&= 1 + (1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}) \\
&= 1 + 0.625 \\
2 \times 0.625 &= 2 \times (1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}) \\
&= 1 + (0 \times 2^{-1} + 1 \times 2^{-2}) \\
&= 1 + 0.25 \\
2 \times 0.25 &= 2 \times (0 \times 2^{-1} + 1 \times 2^{-2}) \\
&= 0 + 0.5 \\
2 \times 0.5 &= 2 \times 2^{-1} \\
&= 1
\end{aligned}
$$

Thus, $(0.8125)_{10} = (0.1101)_2$

It is clear from this example that, if we multiply a decimal fraction by 2, the integer part of the answer will be the most significant bit of the binary fraction. The fractional part of the answer is multiplied by 2 to obtain the next significant bit of the binary fraction. The procedure is continued until the fractional part of the product is 0. The method is illustrated below with examples:

Example 2.14 Find the binary equivalent of $(0.5625)_{10}$

Solution

Product		*Integer part of the product*	
0.5625×2	$= 1.125$	1	⟶ Most significant bit
0.125×2	$= 0.25$	0	
0.25×2	$= 0.5$	0	
0.5×2	$= 1.0$	1	⟶ Least significant bit
$(0.5625)_{10}$	$= (0.1001)_2$		

Data processing using computers requires processing of not only the 26 capital (or upper case) English letters but also the 26 small (or lower case) English letters, 10 digits and around 32 other characters, such as punctuation marks, arithmetic operator symbols, parentheses, etc. The total number of characters to be coded is thus: 26 + 26 + 10 + 32 = 94. With strings of 6 bits each, it is possible to code only $2^6 = 64$ characters. Thus, 6 bits are insufficient for coding. If we use strings of 7 bits each we will have $2^7 = 128$ unique strings and can thus code up to 128 characters. Strings of 7 bits each are thus quite sufficient to code 94 characters.

Coding of characters has been standardized to facilitate exchange of recorded data between computers. The most popular standard is known as *ASCII* (American Standard Code for Information Interchange). This code uses 7 bits to code each character. Besides codes for characters, in this standard, codes are defined to convey information, such as end of line, end of page, etc., to the computer. These codes are said to be for *non-printable control characters*. Table 2.9 gives the ASCII code for both printable and non-printable control characters. Columns 1 and 2 are non-printable codes. The entry CR, for example, indicates carriage return (or end of line) control character. The most significant bits of the code are given in Table 2.9 as column headings and the least significant bits of the code are given as row headings. Thus the code for A, for example, is identified from the table by finding the column and row bits. The column gives bits 100 as bits b_6, b_5, b_4, and the row gives bits 0001 for b_3, b_2, b_1, b_0.

Table 2.9 ASCII Code for Characters

Least significant bits $b_3\ b_2\ b_1\ b_0$	*Most significant bits* $b_6\ b_5\ b_4$							
	000	*001*	*010*	*011*	*100*	*101*	*110*	*111*
0000	NUL	DLE	SPACE	0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	“	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	‘	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	‘	<	L	\	l	|
1101	CR	GS	–	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	–	o	DEL

Thus the code for A is:

b_6	b_5	b_4	b_3	b_2	b_1	b_0
1	0	0	0	0	0	1

The internal representation of the string RAMA J is:

1010010	1000001	1001101	1000001	0100000	1001010
R	A	M	A	SPACE	J

Observe that the blank between RAMA and J also needs a code. This code is essential to leave a blank between RAMA and J when the string is printed.

Observe that in ASCII code digits are encoded using 7-bit codes. Thus if an item code, say, is a combination of letters and digits we use the 7-bit equivalent of each character to encode it. In ASCII code, digits are considered as printable characters rather than numbers with value.

Example 2.17 The licence number of a car is KA02M47. What is its ASCII code?

Solution

1001011	1000001	0110000	0110010	1001101	0110100	0110111
K	A	0	2	M	4	7

In addition to ASCII, another code known as *ISCII* (Indian Script Code for Information Interchange) has been standardized by the Bureau of Indian Standards. The full description of this code is available in the document IS:13194–91 published by the Bureau of Indian Standards. It is an 8-bit code which encodes both English and Indian script alphabets. It retains the standard ASCII code for English. It extends Table 2.9 by adding columns 1010, 1011, up to 1111 (observe that Table 2.9 as shown has columns 0000 to 0111 only). With this addition, it is possible to define 96 more characters.

A string of bits used to represent a character is known as a *byte*. Characters coded in ISCII need 8 bits for each character. Thus a byte, in this case, is a string of 8 bits. A character coded in ASCII will need only 7 bits. The need to accommodate characters of languages other than English was foreseen while designing ASCII and therefore 8 bits were specified to represent characters. Thus a byte is commonly understood as a string of 8 bits.

A coding scheme for characters, called *Unicode*, has been standardized specifically to accommodate a large number of special symbols such as Greek characters α, β, ν, etc., mathematical symbols such as $\Rightarrow$, and non-English characters. It uses 16 bits (2 bytes) to represent each character. As $2^{16} = 65536$, the number of different characters which can be coded using Unicode is very large. Unicode has, however, defined codes from $(0000)_{16}$ to $(FFFD)_{16}$ hexadecimal which is 65534 codes, 2 less than 65536. Thus any character of any language in the world can be represented with this large number. It is important to note that the first 128 characters of Unicode are identical to ASCII codes. Thus Unicode is compatible with existing ASCII and ISCII coded data stored in computers. Unicode described above is now called Unicode 1.0. Unicode has evolved and the current standard is Unicode 6.1 which is a 32 bit code. In theory it can encode a billion characters but has defined only 110,181 characters as of now.

Another coding scheme for characters has become popular as it has been recommended by the World Wide Web consortium. It is called UTF-8. It is a compact efficient Unicode encoding. It encodes ASCII in only one byte and is useful for legacy (i.e., old) systems that require Unicode support because program developers do not have to modify text processing code. Subsequently UTF-16, a 16 bit code compatible with Unicode has been standardized. (For details on UTF-8, and UTF-16 the reader is referred to an article in Wikipedia.)

2.5 ERROR-DETECTING CODES

Errors may occur in recording data on magnetic surfaces due to bad spots on the surface. Errors may also be caused by electromagnetic disturbances during data transmission between units. It is therefore necessary to device methods to guard against such errors. The main principle used is to introduce extra bits in the code to aid error detection. A common method is to use a *parity check bit* along with each character code. A parity check bit is appended to the 7 bits of the code of each character so that the total number of 1s in each character code is *even*. For instance, the ASCII code of the letter E is 1000101. The number of 1s in this string is odd. A parity check bit 1 is appended to this string to obtain a code which is now 8 bits long and has an even number of 1s in it. The code for E with an even parity bit appended is:

10001011
↑
Even parity bit

If the ASCII code of a character has already an even number of 1s in it, then the parity check bit appended is 0. For example, the ASCII code of A is 1000001 and its code with an appended parity check bit is 10000010, where 0 is the parity bit appended as the least significant bit. All characters now have 8-bit codes including the parity check bit. Whenever a character is read from storage or received from a remote location, the number of 1s in its code (including the parity bit) is counted. It has to be even. If it is odd, then at least one bit has changed from 0 to 1 or from 1 to 0. Thus, a *single* error in any of the 8 bits of the code will definitely be detected. Two errors cannot be detected by this scheme as the total number of 1s in the code will remain even when two bits change from 0 to 1 or 1 to 0. As in practice the probability of more than one bit changing is very small, this scheme of using a single parity check bit is considered adequate.

Instead of appending a parity check bit which makes the total number of 1s in the code even, one may choose to append a parity check bit which makes the total number of 1s in the code odd. Such a parity check bit is known as an *odd parity check* bit. This scheme also facilitates detection of a single error in a code.

Codes have also been devised which use more than one check bit to not only detect but also correct errors. These are called *error-correcting codes*. We will not discuss these codes in this book.

In this chapter we studied how to represent and store numbers and characters in a computer. In the next chapter we will study how pictures are represented and stored in a computer.

SUMMARY

1. Data to be processed by computers originate as written documents. It is necessary to convert these to a form that can be stored in a computer. This task is done by input units.
2. A commonly used input unit is the keyboard. Data is entered manually using the keyboard by pressing keys corresponding to the input characters to be stored and processed by computers.
3. Manual data entry using a keyboard is error prone. It is preferable to have devices which will directly read data and store them in a computer's memory. Such devices are now available.

4. For numeric data, optical character readers, magnetic ink character readers and bar code readers are used.
5. Keyboard or an optical character reader is used to input text.
6. As the physical devices used for storage and processing of data in computers are two-state devices, it is necessary to transform or code all data using only two symbols. The two symbols used are, by convention, the digits 0 and 1. These are called "binary digits" or "bits" for short.
7. The number of unique strings possible in a code with 2 bits in each string is 4. These strings are 00, 01, 10 and 11. In general, the number of unique strings with n bits in each string is 2^n. Each string in this set of strings may be used to code or represent one character. There are 128 unique strings with 7 bits each.
8. The American Standards Institution has evolved a standard code to represent characters to be stored and processed by computers. This code, called ASCII, uses 7 bits to represent each character. The ASCII code defines codes for English letters (capital and small), decimal digits, 32 special characters and codes for a number of symbols used to control the operation of a computer. The symbols used for control are non-printable. A string of 8 bits is known as a *byte*. A byte can encode 256 characters.
9. A character representation code known as Unicode has been standardized. Unicode 6.1, the current standard is a 32-bit code and has the capability to represent all languages of the world.
10. Decimal numbers are coded depending on the manner in which they are to be processed by the computer. If they are used merely as symbols without any attached values (e.g., motor car registration numbers, item codes, telephone numbers, etc.), they are equivalent to text and are *coded* using the ASCII code. If values are to be assigned to decimal numbers and if they are used as operands in arithmetic operations, then they are *converted* to equivalent binary numbers.
11. Binary numbers are formed using the positional notation. Powers of 2 are used as weights in the binary number system. A binary number 10111, for example, has a decimal *value* $= 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 23$.
12. A decimal number is converted into an equivalent binary number by dividing the number by 2 and storing the remainder as the least significant bit of the binary number. The quotient is divided by 2 and the remainder becomes the next bit of the binary number. This procedure is continued until the quotient becomes 0. The binary equivalent of 9, for example, is 1001.
13. Large numbers require long sequences of 0s and 1s to represent them. A more concise representation is obtained using the hexadecimal notation. The base of the hexadecimal system is 16 and the symbols used in this system are: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F. Strings of 4 bits have an exact equivalent hexadecimal value. This facilitates conversion of binary numbers to hexadecimal numbers. For example, the binary number 101 1101 1011 1110 0011 0100 is equal to the hexadecimal number 5DBE34.
14. Decimal fractions may be converted to binary fractions.
15. Errors may occur while recording or reading data and when data is transmitted from one unit to another unit in a computer. Detection of a single bit error in the code for a character is possible by introducing an extra bit in its code. This bit, known as the *parity check bit*, is appended to the code. This bit is chosen so that the total number of 1s in the new code is even. If a single bit is incorrectly read, written or

transmitted, then the total number of 1s in the corrupted code would become odd, thereby indicating an error. Note that a single parity bit cannot correct any errors; it can detect only one bit error.

EXERCISES

2.1 What are the steps in data acquisition?

2.2 Explain the data acquisition steps, using as an example, finding the average of marks obtained by students in five subjects in an examination. Assume that there are 20 students in the class.

2.3 What are the different types of keys in a keyboard?

2.4 What is a hand-held scanner? What are its applications?

2.5 Where are MICR characters used?

2.6 What is a bar code? Give an example of use of a bar code.

2.7 What is the difference between internal and external representation of data?

2.8 In what internal form are data stored in a computer? Why is this form used?

2.9 How many symbols can be encoded with 5 bits?

2.10 Encode in binary the decimal number 842369.

2.11 What is the difference between encoding a decimal number in binary and converting it to binary? Encode the number 47. Convert 47 to binary.

2.12 Find the decimal equivalent of the binary numbers (a) 110101110 (b) 111010 (c) 1000000 (d) 10000110.

2.13 What is the decimal value of 32K?

2.14 A computer has 512 Mbytes of memory. Is the number of bytes 512000000 or more? If it is more what is the value?

2.15 What are the binary equivalents of: (a) 25.625 (b) 325.6 (c) 2524.0485.

2.16 How many symbols are used in hexadecimal? What is the hexadecimal equivalent of: (a) 1011 (b) 111001011111 (c) 101.110101

2.17 Find the decimal equivalents of the following hexadecimal numbers:

(a) AE.6FC (b) D123.AB (c) EFF.3DA

2.18 Give the ASCII codes of the following:

(a) aADd (b) HE CAME (c) $623.40

2.19 What is ISCII code? How is it different from ASCII code?

2.20 What is Unicode 1.0? What is the advantage of Unicode compared to ASCII? What is the disadvantage?

2.21 What is an error detecting code? How does it detect an error? Add an even parity bit to the following strings of bits:

(a) 101110 (b) 110011011 (c) 10110101

2.22 How many errors can be detected if odd parity bit is added to a bit string? Explain how an error is detected.

3

ACQUIRING IMAGE DATA

LEARNING GOALS

In this chapter we will learn:

1. How to acquire printed, typed and handwritten text.
2. How to acquire images such as line drawings, black and white pictures, pictures with several grey levels and colour images.
3. How optical character recognition works and why it is needed.
4. The need to compress image files for efficient storage and how it is done.
5. How an image scanner and a digital camera work and how to efficiently store scanned images.

3.1 INTRODUCTION

In the last chapter we saw how numbers and characters are input to a computer and how they are converted to bits for storage and processing. Other important data which need to be acquired (or input) to a computer are printed, typed or handwritten text and pictures. There are different types of pictures which we may have to input to a computer. The simplest are line drawings such as the plan of a house, layout of pipes in a city, block diagram such as Fig. 3.9 of this chapter. Black and white pictures such as cartoons, pencil drawings are somewhat more complex. Highest in terms of complexity are colour pictures. Each one of these requires a different method of inputting. In a page of a book, for example, we may have text combined with line drawings, cartoons and colour pictures. It may be required to store a copy of this page preserving the relative placements of text and pictures. In this chapter we will describe how these types of data are input and stored in a computer.

Let us take two examples of situations in which one needs to acquire texts and pictures.

Case 1: Many libraries have large collections of old periodicals. Examples are law journals, medical journals, science journals and social science journals. They are kept primarily for assisting researchers who read these before they start work on a subject. These journals are bulky and occupy considerable space. The older journals are kept in stacks in libraries which are difficult to access. Further, if a particular journal is issued to a reader, no one can read it till it is returned and put back in the shelf. Misplacement of journals is a perpetual problem in libraries. With the emergence of information technology, librarians have been exploring whether important journals, particularly old volumes can be stored in a computer. By the end of this chapter you will know how it can be done and at what cost.

Case 2: All of us have a large collection of photographs at home. The old pictures are black and white and the more recent pictures are in colour. Photographs continue to accumulate over time and occupy a lot of space in albums. It is also difficult to retrieve a specific photograph you may want. Photographs also fade with the passage of time. Is it possible to store all your photo albums in a computer? How can a particular photograph be retrieved from this collection?

In this chapter we will find out how photographs can be stored in a computer's memory. We will have to wait till we learn about processing and data organization before we can answer the second question.

3.2 ACQUISITION OF TEXTUAL DATA

We will first define what is textual data. A page of this book is textual data. Pages in this book are of different types. There are some pages which have just text. The text, of course, may have section headings in bold face print, some words in italics and other words in a different font. This style is used to enhance readability. There may be other pages which may have pictures along with text and others with tables also. Thus text can be quite general, not merely a string of characters. In Fig. 3.1 we show various types of textual data to be acquired.

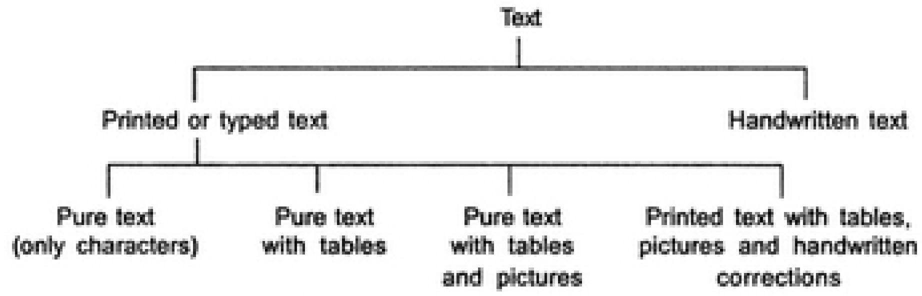


FIG. 3.1 Various types of textual data.

The simplest way of inputting printed, typed and handwritten text is by manually typing them using a keyboard. The text will be stored as a string of bits using either ASCII code or Unicode. The main advantages of this method of acquiring text are:

(i) The number of bits used to store the text are minimal.
(ii) As the coded representation of each character of the text is available in the memory it will be possible to search for specified strings of characters and edit or process them.

The main disadvantages of this method are:

(i) It is slow as input is manual.
(ii) It is expensive as people have to be hired to do the work manually.
(iii) It is also less accurate as errors are made during typing and even with careful proofreading, errors still remain.
(iv) Images cannot be input.

Thus it is necessary to look for better alternatives. The standard method available today is to use an *image scanner*. An image scanner is an opto-electronic device which is used to optically scan a page of text including images and convert it to a string of 0s and 1s and store it in a computer's memory. There are several types of scanners available in the market, and some of which are:

1. Flatbed scanners
2. Sheet-fed scanners
3. Drum scanners
4. Hand-held scanners

Of these, drum scanners are the most expensive and also give the best performance. They are rarely used nowadays. Hand-held scanners are the least expensive and perform the worst. Flatbed scanners are the most popular as their price for a specified performance is the lowest. All the four types of scanners use the same basic principle to scan the text/image. We will now explain in greater detail how the flatbed scanner works.

Besides a scanner, a digital camera may also be used to photograph pages. The resultant images are handled in the manner similar to scanned images. We describe digital cameras in Section 3.6.

3.2.1 Flatbed Scanner

A flatbed scanner has the following parts:

1. An optical system consisting of a light beam, lens and mirrors.
2. An array of solid state "electronic eyes", technically known as CCD (Charge Coupled Device) array, which sense light and give an electrical output, whose value depends on the intensity of the light.
3. An electronic circuit which converts the electrical output given by the array of "electronic eyes" to a set of bits which are stored in the memory of a computer.

The scanner has a flat glass surface over which the document to be scanned is placed with the printed side facing the glass as in a Xerox copier. The paper is now pressed down with a cover. A light source focuses light on an entire line as a thin beam. This is called a *scan line*. Light is reflected from the paper along the scan line. Wherever there is a dark spot (corresponding to a printed character), no light is reflected, and wherever there is no print, that is, it is white, bright light is reflected. This is illustrated in Fig. 3.2. The reflected light is gathered by a row of "electronic eyes", that is, a row of Charge Coupled Devices (CCDs). Dark spots which do not reflect light are sensed and stored as 0s and white spots which reflect light as 1s by the array of CCDs. In many scanners 300 CCDs are arranged per inch on a horizontal line giving 300 bits per inch resolution. After gathering bits from one scan line, the motor moves the light source to the next scan line and the black and white parts of this line are converted to 0s and 1s again. The process continues till the entire page is scanned from top to bottom. The number of scan lines in the vertical or *Y*-direction is determined by the mechanical precision of the motor moving the light beam. It is usually double the horizontal resolution. Thus, if the horizontal resolution is 300 bits per inch, there will be 600 scan lines per inch in the vertical direction. It usually takes between 5 to 20 seconds to scan a full page.

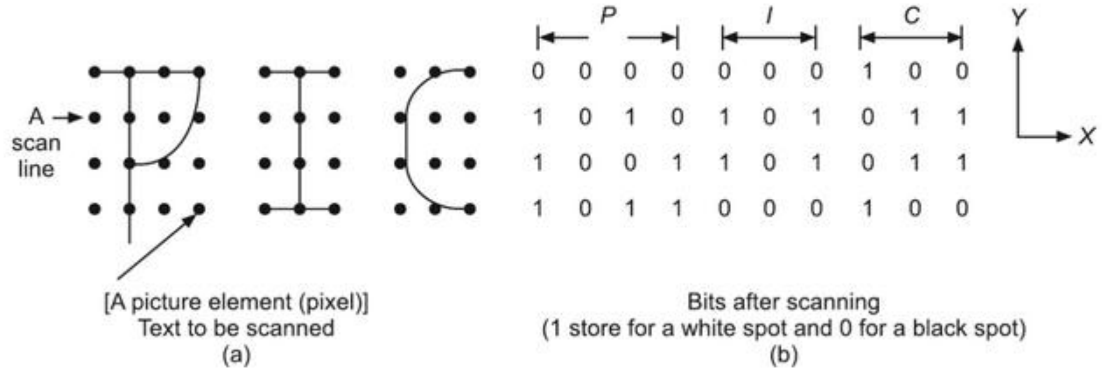


FIG. 3.2 Scanning a text.

Each reflected light beam picked up by the CCD array is called a *picture element* which is abbreviated *pixel* (sometimes known as *pel*). Thus a scanner with 300 CCD elements per inch in the *X*-direction and 600 scan lines per inch in the *Y*-direction will store 300 × 600 = 180,000 pixels per square inch. If the text being scanned is printed characters, then each pixel is represented by 1 bit, a black being 0 and a white 1. If we scan an A4 size paper with text area of (7″ × 9″), the number of bits stored will be (7 × 300 × 9 × 600) bits = 11.34 megabits = 1.4175 MB. (In this book we will use the notation 1 Mb for 1 megabit, 1 MB for 1 megabyte. Similarly, 1 Gb is 1 gigabit and 1 GB is 1 gigabyte). This scanned version of the text is known as a *bit map* form. If the information on the paper is "pure" text, it can be typed and stored as ASCII coded characters. If we assume that there are 80 characters per line and 24 lines per page, the number of characters in a page is 1920. This needs a storage space of 1920 bytes if ASCII code is used. We see that the bit map form requires 738 times the storage space as compared to the ASCII coded form! Nowadays bit mapped data are compressed using compression algorithms (to be discussed in Section 3.5). Even if we compress by a factor of 25, the bit map form will still require 30 times more storage space. Thus storing a text as a bit map is very inefficient. The questions which will naturally occur to you is: "Is it not possible to convert the bit map form to ASCII form?". The answer is a qualified yes. It is possible for typed or printed text but not for handwritten text. The conversion from bit map form to ASCII form is done by a software called *Optical Character Recognizer* (or OCR for short) which we will describe in the next section.

3.2.2 Optical Character Recognition

As we pointed out in Section 3.2.1 a scanner treats a page of text as though it is an image and gives a bit map representation of the page. The main disadvantages of bit map representation are:

- it requires a huge storage space.
- as individual characters are not stored, it is difficult to search the text for character strings.

Thus software has been designed to take a scanned text as input and convert it into its ASCII coded equivalent. This is called Optical Character Recognition (OCR) software. There are four broad categories of text, which need recognition:

1. Printed or typed text using a standard font such as Times Roman 10
2. Handwritten text
3. Preprinted forms with blank spaces which are filled with handwriting. The letters are unconnected.
4. A page containing text with mixed fonts, pictures and tables.

Let us consider each case. In case 1, namely, printed or typed text using a single well-known font, OCR software does an excellent job. It is easy to obtain 99.5% accuracy. Accuracy of even 100% has been reported by some commercial software (see OKI printing solutions—www.bli.com).

Handwritten texts are of two types. One of them is called unconnected handwriting. In other words, individual letters are written separately with a clear handwriting. Good recognition algorithms are available in such a case and accuracy of 85% to 90% has been reported. The other type of handwriting is one in which individual letters are connected. This is called cursive handwriting. The cursive handwriting styles of people widely differ. Thus recognizing scanned cursive handwriting is difficult and has been a topic of research for several years. As of now (2013) recognition accuracy is less than 80%. Better accuracy has been obtained in specialized applications where vocabulary is known in advance or inferred from context (e.g., postal address).

At this stage we have to distinguish between on-line handwriting recognition and off-line recognition (also known as static recognition) using OCR. In on-line handwriting recognition, a person writes on the screen of a tablet computer with a stylus. The movement of the stylus can be sensed while writing which is additional information to the recognition system. A user can train the system to recognize his/her handwriting (by trial and error). Such personalized recognition systems have much better performance with recognition accuracy exceeding 95% in some good systems.

In case 3, namely, pre-printed forms filled by hand using uppercase letters which are unconnected, an accuracy of 80% to 85% has been attained.

In case 4, which is the case where old books, newspapers, etc. are scanned, the material will include pictures, tables, equations, etc. It may also be in multiple columns. The fonts will not normally be uniform. They may include boldface, italics and other non-standard fonts. In such cases, the accuracy depends on the quality of the image, page skew, scanner distortions, type size, font types, etc. In these cases the accuracy is around 85% to 95% under controlled conditions. It is necessary to proofread and manually correct the errors in the recognized text.

The main principle used by the recognition software is to compare each scanned letter with a sample set stored in the computer and find a match. This matching becomes easy if the font used is known to the recognition software and a uniform font is used. Having found matches for letters, a word is formed by stringing letters found between blank spaces in the text. The word is matched with words stored in a dictionary of words. The nearest matching word from the dictionary is selected. Thus, even if one or two letters in a word are incorrectly recognized the matching process selects a correct word. All selected words are put together to form the ASCII coded text which is output. The quality of recognition depends on the size of the stored dictionary and the recognition method. As we said earlier, no OCR software today gives 100% correct output. Thus, it is necessary to manually correct errors, if any, in the output text. The number of corrections may be quite small (1 in 100 words) if the software is good.

An operator "proofreads" the text output by OCR by comparing it with the original scanned text. This is time consuming but is much faster than retyping an entire page.

The time taken by OCR software to convert a scanned text to ASCII form depends on the processing speed of the computer which is used for recognition and also the efficiency of the OCR software. For scanning resolution of 300 dpi and using currently available desktop computers (3 MHz clock, 4 core CPU), the time varies from 2 seconds to 4 seconds per page. This does not include the time needed to manually proofread a page after it is recognized. The proofreading time depends on the number of errors in the output and also the experience of the proofreader in reading and correcting the English text. An average proofreader may take from 2 to 5 minutes to read and correct errors in a page of text containing around 3200 characters. We will now give a few numerical examples of time and storage space needed to store scanned text.

Example 3.1

(i) How much time will be needed to scan a 1000 page bound book, convert it to ASCII and correct the output?

(ii) How much storage space is needed to store it in the computer's memory?

Solution

Case 1—The binding is cut and the individual pages are scanned

(i) If an inexpensive flatbed scanner is used, each page is to be put on the glass plate and cover closed (just like in a Xerox copier) before scanning commences. This will take approximately 15 seconds per page.

(ii) The time taken by the light beam assembly to move and scan the whole page depends on the speed of the motor which moves the beam. It varies between 3 seconds to 9 seconds. We will take an average of 6 seconds/page.

(iii) After scanning, the OCR will take time to recognize characters and convert the bit map to ASCII form. The time needed depends on the quality of OCR software and the computer used. It also varies from about 2 seconds to 4 seconds. We will take an average of 3 seconds per page.

47-51

is required for this assuming a scanner with 300 × 600 resolution is used and 8 bits are used per pixel?

Solution

$$\text{Storage needed for 1 photo} = 300 \times 600 \times 3.5 \times 5.5 \times 8 \text{ bits}$$
$$= 3.465 \text{ MB}$$
$$\text{Storage for 1000 old photos} = 3465 \text{ MB}$$
$$= 3.465 \text{ GB}$$

which requires many CDs to store!

From the above example, we see that considerable amount of storage is needed to store pictures. If a large photo archive is to be preserved, it is essential to reduce the storage needed without sacrificing quality. This is possible and is done by what are known as *compression algorithms* which remove redundant pixels without sacrificing quality. From the compressed picture the original picture is reconstructed by another algorithm without losing fidelity. We will describe the idea used by such algorithms later in this chapter.

3.3.4 Acquiring Colour Pictures

Most scanners available in the market today scan black and white as well as coloured documents. Red, green and blue are the primary colours and our eyes perceive different colours by adding these primary colours in different proportions. Thus the principle used in colour scanning is to measure the intensity of red, green and blue colours (called RGB) in the light reflected by a picture element. The RGB components in reflected light are found by a rapidly rotating disk which has R, G, B filters (see Fig. 3.7) and picking up R, G, B components separately. The intensity of each of these is measured by a CCD to 8 significant bits (in most scanners). If R, G, B are combined, we have 24 bits to represent different colours. Thus, 2^{24} = 16 million colours can be represented. All 1s represent white and all 0s black, and each bit pattern one colour. If more colours are

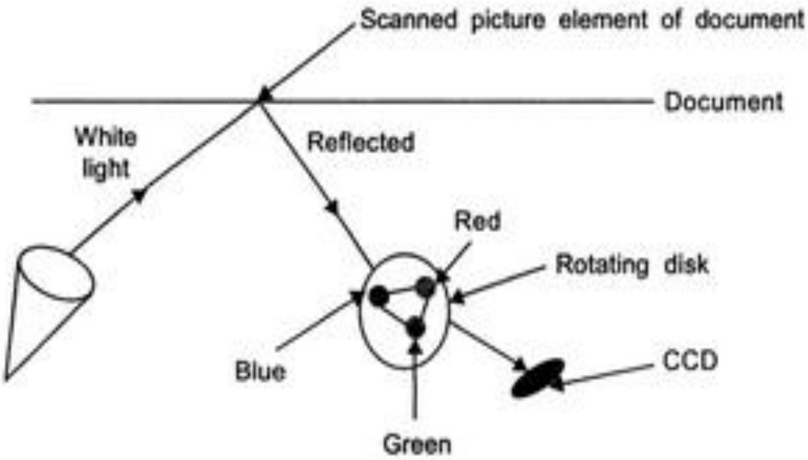


FIG. 3.7 Finding RGB components in reflected light in a colour scanner.

required, then the number of bits used to represent R, G, B intensities must be increased. If 30 bits are used, 2^{30} = 1 billion colours can be represented (in theory).

Scanners are advertised as having 30-bit colour, but due to errors in sensing, they normally use the "best" 24 bits to represent each colour while storing the image. Because colour scanners use 24 bits to represent each pixel, they require 3 times the storage needed for black and white pictures. (Remember that 8 bits are used to represent greys in a black and white photograph.)

EXAMPLE 3.4 Creating a Digital Colour Atlas

An atlas has coloured maps of different countries. It has 200 pages. Each page has a map drawn in 8″ × 11″ area. If the maps are digitized and stored, how much storage is needed?

Solution. Assume that a high resolution scanner with 600 CCDs per inch is used and that the scanner scans 1200 lines per inch. Such a high resolution is used because details in a map are important. The number of bytes required to store the atlas is:

$$\text{Bytes per page} = 8 \times 11 \times 600 \times 1200 \times 24 \text{ bits}$$

$$= 190.08 \text{ MB}$$

(We have used 24 bits per pixel)
As the Atlas has 200 pages, it needs

$$200 \times 190.08 \text{ MB} = 38.016 \text{ GB}$$

Again, we see the enormous storage space needed, which should be reduced for cost-effective storage.

3.4 STORAGE FORMATS FOR PICTURES

When we scan a picture, we represent it by a set of pixels. A pixel is coded by 1 bit for black and white images, 8 bits for monochrome pictures with 256 grey levels, and by 24 bits for colour pictures. The representation of pictures by an equivalent array of bits is called a *bit-map* representation. A file which stores the bit-map representation of a picture is called a *bmp file*. Another standard method of storing digitized image is called tagged image file format (*tif* file). "Tag" refers to the information about the image stored in this format. The information is the height and width of the image in pixels and a table of colours used (see Fig. 3.8). The main problem with bmp and tif files is their huge size. We saw that a 7″ × 9″ picture will need 7 × 9 × 300 × 600 bits = 1.4175 MB if stored as a bit map. If 24 bits colour per pixel is used, the storage needed is 24 times, that is, 34.02 MB. This size is very large. Thus it is essential to compress an image file into one that requires less number of bits without lowering the quality of the picture. For example, if we are able to use (on the average) 8 bits per pixel for colour images instead of 24 bits per pixel, we achieve a compression of 1:3. This is called *compression ratio*.

There are many methods available to compress image files. The main idea used to compress a file is to use the processing power of a computer and perform a number of

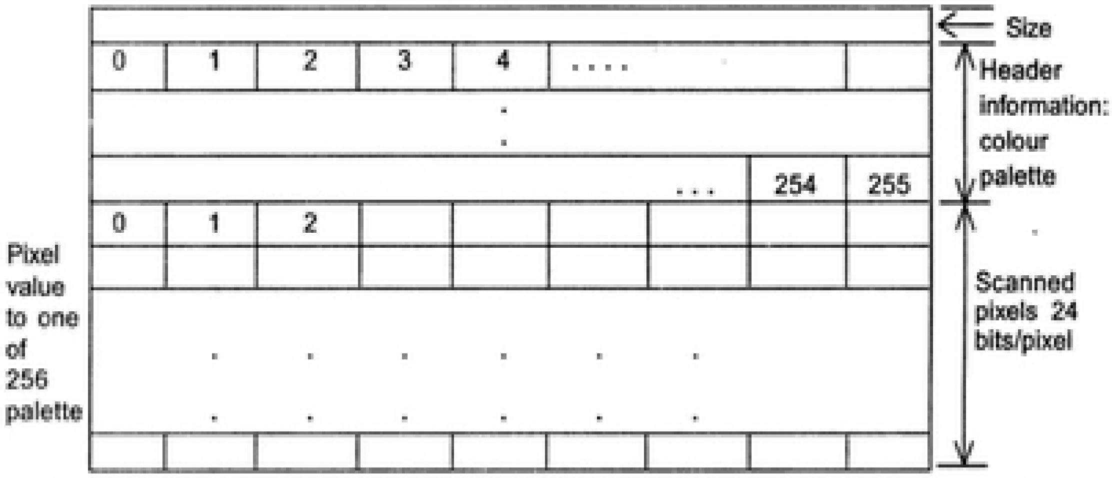


FIG. 3.8 Tagged image file format. Observe that 24-bit colour is mapped to one out of 256 colours. The header gives data on number of pixels in rows and columns of image.

processing operations on the file which results in a file that uses less storage. Thus we spend processing time in order to save storage (see Fig. 3.9). As storage requirements are very high, in practice reduction of storage space is essential. It is cheaper to store a compressed file. If an image is to be transmitted using a communication line, it should be compressed and sent. This reduces the time and cost of communication. At the receiving end the file should be decompressed, that is, expanded, again using processing power to view the original image on a VDU or print it using a printer. Decompression also requires time. The decompressed image must be a truthful reproduction of the original image without distortion, colour change, etc. Another important point to remember is that the time taken to compress an image file may be comparatively high, for example, it may take a minute to compress a file. Decompression time, however, has to be short. People get impatient if they have to wait for more than 10 seconds for a compressed image to be expanded for viewing.

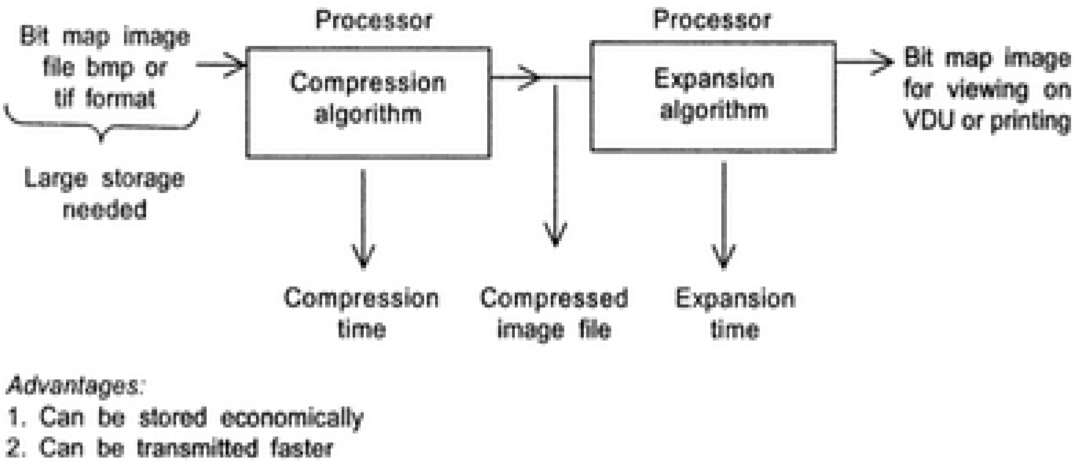


FIG. 3.9 Compression and expansion of image file.

Many algorithms have been proposed to compress image files. There are two general principles used by these algorithms. They are:

(i) *Redundancy:* In a picture, neighbouring pixels are highly correlated. Given a pixel, it is easy to guess its neighbours.

(ii) *Irrelevancy:* Our eyes do not perceive very high resolution of pictures and too many colours. It is thus sufficient to display only what we can see.

The main criteria to be taken into account to compare compression algorithms and compressed files are:

- By how much does the image file get compressed?
- How much time does it take to compress the image file?
- How much time does it take to expand the compressed file and get back the image for viewing?
- Is the quality of the image lowered due to compression/expansion?
- What is the type of image for which it works well?
- Is the format of compressed files widely accepted so that such files can be exchanged among users? Such standardization is becoming very important in this internet age.

There are two major file formats used for compressed image files. These are:

- Graphics Interchange Format (GIF) and
- Joint Photographic Expert Group (JPEG) Format.

We will now describe these.

Graphics Interchange Format (GIF)

It is standardized for 256 colours (8 bits per pixel) of maximum size 65535×65535 pixels. It is widely used for world wide web images. The compression algorithm replaces frequently occurring groups of bits by short bit patterns (see Table 3.4 on p. 59). This reduces the number of bits by a factor of 2 to 3. The algorithm used is called LZW compression. As a first step, 24 bits for colours per pixel given by a scanner are reduced to 8 bits per pixel. This group of 8 bits is used to look up in a table one out of 256 colours. One of the colours can be called *transparent* which is usually the same as the background display colour being used in the VDU. Reduction of 24 bits of colour to 8 bits compresses the image file by 67%. This colour reduction is not a great sacrifice from the point of view of quality as most people cannot perceive more than 256 colours. One of the features of GIF which has made it popular is called *interlacing*. It transmits every alternate line of an image first during decompression so that a user sees a blurry image quickly which is preferred by viewers who are normally impatient to see an image. The full image with all lines follow next and is of better quality. A compression ratio of 1:6 to 1:10 is achieved in GIF by combining LZW and colour compression.

GIF is the preferred file format for monochrome grey scale images as such images use only 8 bits per pixel. The GIF compression method works very well if the picture is a line drawing, cartoon, handwritten text or an X-ray, which do not have the richness of colour and details of a colour photograph. This format is a preferred format for

Graphics Interchange Format (GIF)

It is standardized for 256 colours (8 bits per pixel) of maximum size 65535 × 65535 pixels. The compression algorithm replaces frequently occurring groups of bits by short bit patterns (see Table 3.4 on p. 60). This reduces the number of bits by a factor of 2 to 3. The algorithm used is called LZW compression. LZW stand for Lempel-Ziv-Welch after its inventors. It is a *lossless compression algorithm*, that is, no information is lost due to compression. As a first step, 24 bits for colours per pixel given by a scanner are reduced to 8 bits per pixel. This group of 8 bits is used to look up in a table one out of 256 colours. One of the colours can be called *transparent* which is usually the same as the background display colour being used in the VDU. Reduction of 24 bits of colour to 8 bits compresses the image file by 67%. This colour reduction is not a great sacrifice from the point of view of quality as most people cannot perceive more than 256 colours. One of the features of GIF which has made it popular is called *interlacing*. It transmits every alternate line of an image first during decompression so that a user sees a blurry image quickly which is preferred by viewers who are normally impatient to see an image. The full image with all lines follows next and is of better quality. A compression ratio of 1:6 to 1:10 is achieved in GIF by combining LZW and colour compression.

GIF is the preferred file format for monochrome grey scale images as such images use only 8 bits per pixel. The GIF compression method works very well if the picture is a line drawing, cartoon, handwritten text or an X-ray, which do not have the richness of colour and details of a colour photograph. This format is a preferred format for transmitting medical images where loss of information is not desirable.

GIF uses LZW compression algorithm which is patented. GIF also limits display of 256 colours. Thus a new file format which also supports lossless compression, increases the number of colours and has no patent restrictions was suggested by a group in 1996. It was later approved and published as an ISOL/IEC standard in 2004. This file format is called PNG (Portable Network Graphics) and uses file extension .png. The advantage of PNG is that it allows a much wider range of colour depths including 24-bit and 48-bit colour. Apart from the advantages noted above, PNG images render faster on the screen than GIF and are thus good for use in the World Wide Web. However, PNG has not superseded GIF as GIF is more widely supported by web browsers. GIF intrinsically supports animated images whereas PNG does not.

Joint Photographic Experts Group (JPEG) Format

It is a compression method pronounced as Jay-peg standard. It is a standard arrived at by an international group of experts and is very popular for compressing colour photographs with rich details. JPEG supports up to 24 bits colour per pixel and image sizes up to 65535 × 65535 pixels. The standard provides a range of quality factors which allow a user to sacrifice quality for higher compression of an image file. Compression performance is best specified by stating the image quality (after compression/decompression cycle) and the corresponding bits per pixel obtainable by the compression algorithm. For a set of colour pictures it is stated in the literature that JPEG gives the results given in Table 3.2.

Table 3.2 JPEG Compression and Quality for Colour Pictures Using 24 Bits per Pixel

Compressed image quality	*Bits per pixel after compression*	*Average compression ratio*
Moderate quality image	0.25 to 0.5	1:64
Excellent quality image	0.75 to 1.5	1:21
Indistinguishable from the original	1.5 to 2	1:14

The method of compression used by JPEG is said to be *lossy* as it throws away a number of bits in the picture which are not relevant for the quality of the image as seen by the human eye. These bits cannot be recovered. This works very well for pictures with rich detail such as photographs, but poorly for line drawing, colour cartoons, photographs, and generally for pictures with sharp transitions.

For such pictures GIF is better both in terms of compression ratio and quality. Unlike GIF which combines the compression algorithm and the file format for the compressed image, JPEG standard specifies only the compression method and not the file format. BMP or TIFF files are compressed using JPEG algorithm and the compressed file is stored using a file format known as JFIF (JPEG File Interchange Format). JFIF format is being standardized by the International Standards Organization (ISO) as JPEG 2000 Part 5. In Table 3.3 we summarize the discussions in this section.

Table 3.3 Image File Formats

File format	*Applications*	*Advantages*	*Remarks*
Bit map (.bmp)	Uncompressed output of scanner	All information in image preserved	High storage. Should be compressed.
Tagged image file format (.tiff)	Exchanging image files	Widely used	Not usually compressed but can be compressed.
Graphical interchange format (.gif)	Exchanging compressed files. Used for transmitting medical images, line drawings, cartoons, and monochrome images	Reasonable compression. Compression speed high. Lossless compression of 1:6 to 1:10	Limited colour palette. Widely supported.
JPEG compression (.jpg) JPEG file format (.jfif)	Compressing colour pictures with rich details. Most common in World Wide Web transmission of colour photos	Can specify high compression for low quality and low compression for high quality. For overall compression, see Table 3.2	Complex compression algorithm. Should not use for line drawings, cartoons and monochrome images, lossy compression.

3.5 IMAGE COMPRESSION FUNDAMENTALS

You may be wondering how images are compressed. The following techniques are used for compression. Each one uses a property of an image.

Table Look-up and Truncation (CLUT)

A scanner may obtain 24 bits per pixel and provide 16 million possible colours. These cannot be used if an image is to be viewed on a VDU which uses only 8 bits for colour. Our eyes cannot also perceive 16 million colours. 24 bits are mapped to 8 bits (i.e., 256 colours) by picking a group of similar colours and using "nearest" colour. This is called CLUT compression (colour lookup table). This is used in gif compression. Another method is to throw away the least significant bits in a pixel representation. This is called *truncation*. Yet another truncation method uses the fact that the human eye is more sensitive to variation in brightness than to variations in colour.

This implies that colour components may be dropped in a limited way without sacrificing quality.

Run Length Encoding (RLE)

In many images there is a uniform background. For example, a picture may have a sky blue background. Thus a whole scan line may all be 0s. Instead of storing 300 zeros, it is enough to specify how many times 0 recurs. In binary it would be coded as shown in Fig. 3.8.

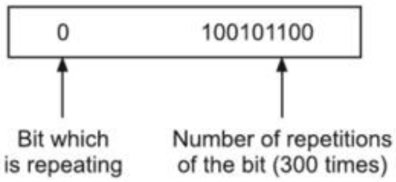


FIG. 3.8 Run length encoding of a bit's repetition.

Thus, 300 zeros are now represented by 10 bits. Repeating patterns of bits can be detected and replaced by the pattern and the number of times the pattern repeats. In Fig. 3.9 we give an example. Instead of storing $(606 \times 8) = 4848$ bits the RLE format stores only 18 bits—a big saving. This method exploits the repetition of a particular pixel in either an entire scan line or part of a line.

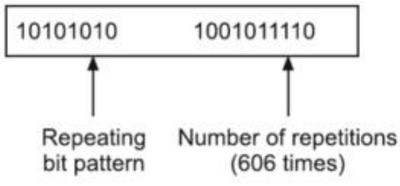


FIG. 3.9 Run length encoding a pattern of bits.

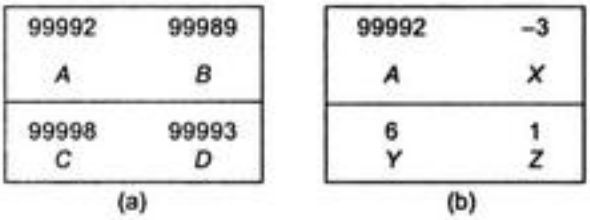


FIG. 3.13 Transforming picture data.

$C = Y + A$ and $D = Z + A$. This example is just to illustrate the idea of transform method of compression and use of inverse transform for expansion. In practice, transforms operate on large blocks of pixels and they perform more complex transformations. In JPEG compression a transformation method, called Discrete Cosine Transform (DCT), is used for compression.

Statistical Compression

The general idea here is to find out how many times a particular string of bits are repeated in a picture. For example, if we take an image having 120000, 4 bit pixels and find out how many times each combination of 4 bits occur, we can obtain a frequency table as shown in Table 3.4.

Table 3.4 Frequency of Occurrence of Pixel Strings

	Pixel	*Frequency of occurrence*	*Code*
0	0000	10000	10
1	0001	15000	1
2	0010	6000	011
3	0011	3000	110
4	0100	4000	101
5	0101	8000	001
6	0110	17000	0
7	0111	8000	000
8	1000	9000	11
9	1001	11000	01
10	1010	13000	00
11	1011	3000	111
12	1100	6000	010
13	1111	4000	100
14	1110	2000	0000
15	1111	1000	0001

By inspecting this table we find that the frequency of occurrence of different pixels is different. If we assign new codes in such a way that the code length of the most

5.3 COMPRESSION OF VIDEO DATA

We saw in Section 5.2 that an enormous amount of storage is required to store digitized video. Apart from the storage required, the enormous size of data leads to difficulty in display and transmission of images. If the image is stored in a secondary memory, the bandwidth needed to display one colour picture frame within (1/30)th of a second is $(640 \times 480 \times 30 \times 24)$ bits/s, which is 221.184 Mbps. No secondary store can provide such a large bandwidth with current technology. If the picture is to be transmitted over a communication line from a video server to a VDU, the same bandwidth is needed from the communication line, which again is impractical. Thus we need compression of video for storage and transmission. We should, of course, expand it before it is displayed.

In Chapter 3 we saw how digitized images can be compressed in various ways. One of them is JPEG compression algorithm which achieves a compression by a factor of about 20. This compression is good for still pictures but for video we look for further possibility of compression as we are now transmitting a group of pictures one after another in a short time. There is not much variation between pictures in a sequence, and this should be exploited. For example, in a scene where a bus is going on a road, most of the background such as the sky, hills, etc., will not change; only a small part of the video corresponding to the moving bus will change from frame to frame. In such a case, only the changing part needs to be stored or transmitted for display. If the bus occupies (1/8)th of the image area, then we need not store or transmit (7/8)th of image in each frame. Only the first frame and the *changes* in subsequent frames need to be stored.

This is the main principle used in a compression algorithm standardized by an international committee known as the Motion Picture Experts Group (abbreviated as MPEG-pronounced em-peg). In MPEG compression a sequence of 12 to 20 frames are considered a package, known as a *Group of Pictures (GOP)*. This GOP is kept in a temporary storage in the video camera and compressed by a local processor. The compressed pictures are stored on the video tape, thereby reducing storage needs. The MPEG compression algorithm works as follows:

A JPEG compressed frame is used to begin the process. It is called an I picture or intra coded picture. It is a key picture which is used to compress subsequent frames. The I picture is picked such that it is a beginning of a sequence of pictures. I picture is used to index and retrieve a video sequence. For example, in a marriage video it may be the beginning of mangalsutra tying ceremony. The I picture is divided into "tiles" of 16×16 pixels. The corresponding tiles are taken in the next frame and the difference in pixel values of these tiles is calculated (see Fig. 5.5). For most of the tiles, this difference will be very small as the background such as the sky or hills remain stationary. Only a small part of the picture will be different. The difference values will typically require very few bits; instead of storing 24 bits per pixel, the difference may be only 2 to 4 bits, thus reducing the number of bits stored by a factor of 12. When the difference in pixel values start changing too much, we take another I picture to start the process again. The actual MPEG compression algorithm uses this idea and what is known as motion prediction to account for change of position of moving objects. It is thus more complicated but gives maximum compression. If JPEG compression is used for I pictures and motion compression

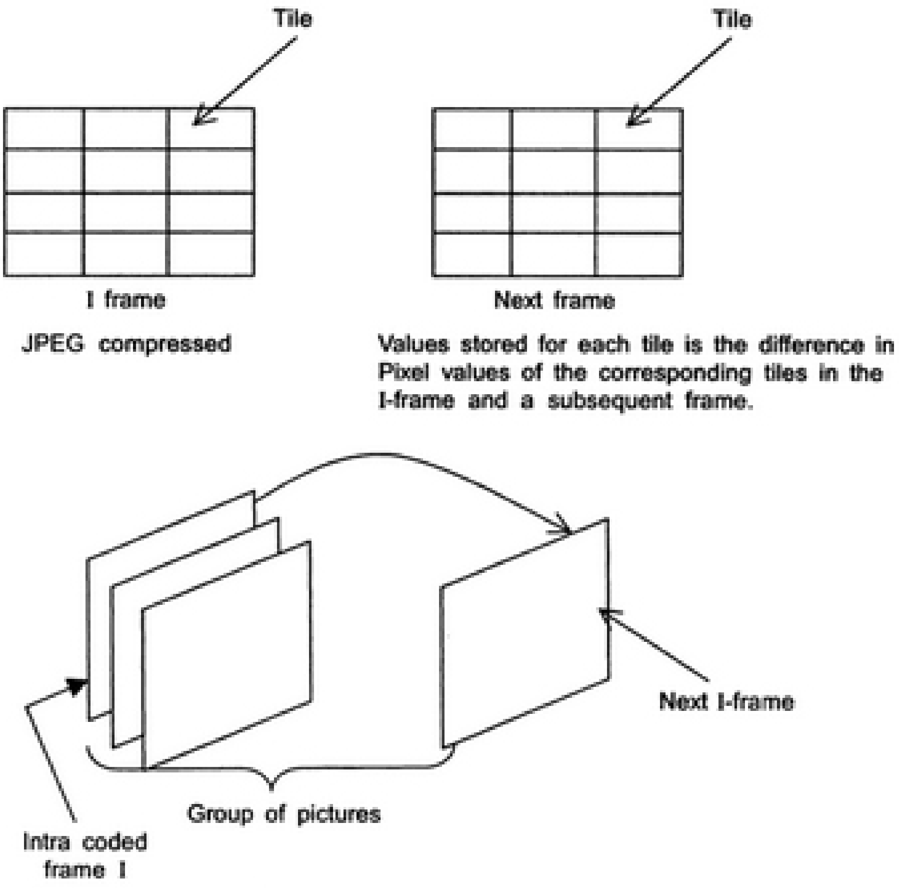


FIG. 5.5 MPEG compression principles.

for a group of pictures using MPEG, we get a reduction between 100 to 150 in the bits stored for the video. Thus a two-hour video will require 2 GB storage instead of 200 GB for the uncompressed video. It is possible to store 2 GB in an optical disk known as DVD-ROM (Digital Versatile or Video Disk Read Only Memory) which has the same physical size as CD-ROM used in audio systems. We will describe how DVD works in Section 6.8.2.

We saw that the bandwidth required to transmit one uncompressed image frame within (1/30)th second is 221.184 Mbps. After MPEG compression, the bandwidth needed will go down by a factor of 100 to 150. Thus, to transmit compressed video we will require a bandwidth of around 2 Mbps, which is feasible with current technology.

In Fig. 5.4 we showed the basic steps in digitizing video analog tape. We should now add one more block, namely, the MPEG compressor as shown in Fig. 5.6, to reflect the fact that this compression is now standardized.

The MPEG compressor is a complex piece of digital hardware which must have a memory to temporarily store a group of pictures and a fast processor to implement

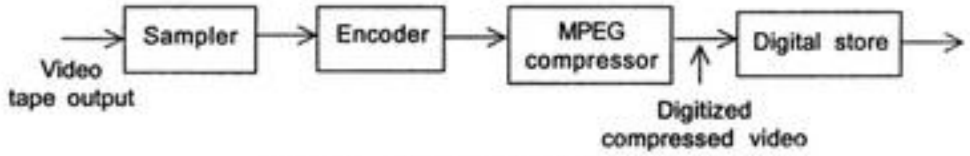


FIG. 5.6 Compressing video tape output.

compression algorithm. The hardware must be fast enough to compress video in real time, that is, while the video is being played. The decompressor also should have sufficient speed. With the improvement of speed of PCs, decompression can be performed by software running on a PC and pictures can be displayed on a VDU.

5.4 MPEG COMPRESSION STANDARD

Just as JPEG standard gives the trade-off between quality of image to be seen and the amount of compression, MPEG's aim is also to provide flexibility, allowing appropriate features based on application. The major features of MPEG compression standard are:

1. The compression method gives 1 to 1.5 Mbps playback rate with acceptable quality of at least (320 × 240) pixels per frame.
2. It is possible to randomly access any specified frame in the compressed video. The random access points are coded as I frames.
3. In addition to normal forward playback, fast forward and fast reverse to specified points are possible.
4. Audio/video synchronization is maintained.
5. Compression, decompression delays can be controlled.
6. Compressed video can be edited.
7. Low-cost processors and integrated circuit chips can compress the video in real time.

MPEG is not a single standard but a series of standards. MPEG-1 is a standard which is for TV quality video and low resolution VDU of a PC (640 × 480 pixels resolution). MPEG-2 supports High Definition TV (HDTV). HDTV provides a 16:9 aspect ratio and 60 frames/second. MPEG-4 addresses the problem of transmitting video on slow communication lines such as telephone lines. It allows intergrating animations with camera captured video. Telephone lines with low speed are used in video conferencing and applications using the world wide web. MPEG-2 is emerging as the de-facto standard for standard TV broadcasts.

A video tape has audio synchronized with video. MPEG-1 and MPEG-2 also define audio compression standard and synchronization issues. MPEG-1 gives CD quality audio. MPEG-2 audio compression standard supports six channel surround sound specified in Dolby AC-3 standard for the so called surround-sound available in recent home audio systems. We will now revisit the two cases we described in Section 5.1 and describe how they are done.

EXAMPLE 5.1 A marriage video is three hours long and it is taken in colour. It is desired to store the video in compressed form in a Personal Computer. How much storage is required for it?

Solution. If we want to display the video on the VDU of a PC, each frame requires (640 × 480) pixels. The frames have to be repeated 30 times a second in order to see a continuous video. Each pixel requires 24 bits to represent all colours. Thus one second of video needs 640 × 480 × 30 × 24 bits. Three hours of video requires:

$$\underbrace{3 \times 60 \times 60}_{\text{No. of seconds}} \times \underbrace{640 \times 480 \times 30 \times 24}_{\text{bits/second}} = 23.887872 \times 10^{11} \text{ bits}$$

$$= 2.985984 \times 10^{11} \text{ bytes} \equiv 298 \text{ Gbytes}$$

This is too high for a PC's secondary storage. If it is compressed using MPEG-2 compression, a 100-fold compression is possible. The compression is normally done by professionals with special hardware and the resulting compressed digital video occupies around 3 GB, which is normally stored on a DVD disk. The DVD disk can be read and the content decompressed by a program in the PC to view the video.

You may now wonder why you should go through this complicated process of digitizing the video. All of us watch a marriage video tape using an analog video cassette player without any fuss, though a little bored!. The main advantage of digitizing is the possibility of accessing specific scenes in the marriage by indexing the data on the disk. Editing of the information in the video disk is also easy. We will discuss how these can be done after we learn more about organizing data in storage later in this book.

EXAMPLE 5.2 We described a Video-on-Demand system in Section 5.1. Suppose a video store wants to store 100 video movies to cater to its customers. How much storage must it provide to store these video movies in digital form? What should be the bandwidth of the cable required to transmit the video on demand to a customer? Assume that compressed video is sent to a customer and that the customer has a small electronic system known as a set top box which is kept above his TV set. The set top box decompresses the compressed video and displays it on the TV screen. A high-quality digital TV broadcast normally provides 1280 × 720 pixels per frame.

Solution. With the higher resolution the bits per second of video is

$$1280 \times 720 \times 24 \times 30 = 633.552 \text{ Mbps} = 82.944 \text{ MB/s}$$

For a three hour video (most Tamil/Hindi movies) the storage needed is:

$$3 \times 60 \times 60 \times 82.944 \text{ MB} = 895.8 \text{ GB}$$

If we assume MPEG-2 compression with a 150 fold compression the storage needed for a three hour video $\simeq$ 5.972 GB.

To store 100 videos the storage needed is 597.2 GB. It is possible to store this in an array of very large disks with current technology.

and also proprietary interests within the family of MPEG standards. This was accomplished by standardizing only the structure of the compressed bit stream emanating from the MPEG compressor. The compression algorithm is not standardized, which allows different companies freedom to innovate. The de-compressor is implicitly defined by the standard compressed bit stream (see Fig. 5.7).

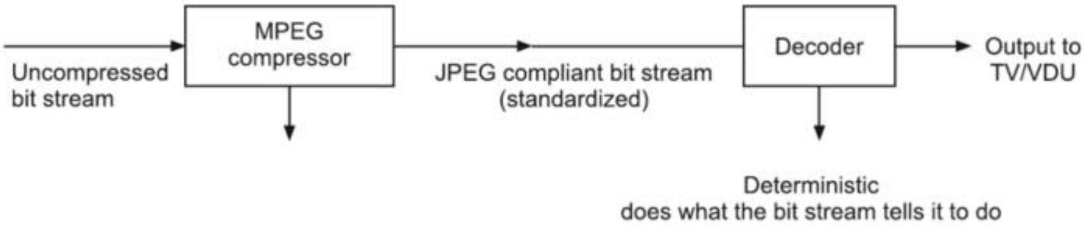


FIG 5.7 MPEG Compression.

Just as JPEG standard gives a trade-off between quality of image to be seen and the amount of compression, MPEG's aim is also to provide flexibility, allowing appropriate features based on application. It is also a lossy compression. The major features of MPEG1 and 2 compression standards are:

1. The compression method gives 4 to 1.5 Mbps playback rate with acceptable quality of at least (352 × 288) pixels per frame and 25 frames per second.
2. It is possible to randomly access any specified frame in the compressed video. The random access points are coded as I frames.
3. In addition to normal forward playback, fast forward and fast reverse to specified points are possible.
4. Audio/video synchronization is maintained.
5. Compression, decompression delays can be controlled.
6. Compressed video can be edited.
7. Low-cost processors and integrated circuit chips can compress and decompress the video in real time.

MPEG is a family of standards starting from MPEG1, which was standardized in 1991. It was followed by MPEG2 in 1994 and MPEG4 in 2000. MPEG1 and 2 standards consist of five major parts: system which specifies multiplexing and synchronizing audio and video streams, video, audio, conformance tests to ensure compliance of the compressed bit stream with the standard and a complete software implementation of the decompressor (also called decoder) and a sample software implementation of the compressor (also called coder). MPEG1 standardized TV quality videos storable in Video Compact Disks (VCDs). Minimum video quality is for (352 × 288) pixels PAL with 25 frames per second. The audio standard is MP3 stereo. The bit rate of compressed digital stream is 1.5 Mbps. MPEG2 standard is for digital broadcast TV and for recording on standard DVDs. The standard is used for high resolution pictures (1920 × 1152) pixels at 60 frames per second requiring transmission speed of 80 Mbps and for High Definition TV (HDTV) (1440 × 1152) pixels at 60 frames per second requiring transmission bandwidth of 60 Mbps. At the low end TV (352 × 258) pixels at 25 frames per

EXERCISES

5.1 What is the difference between still images and video? Why does digitized video require a large storage space if it is to be stored?

5.2 What are the advantages of digital video as compared to analog video on tape?

5.3 How much storage does a (640 × 480) pixel monochrome image video of 2-hour duration need? How much is the storage if it is a colour video?

5.4 Why is it necessary to compress digital video data? Briefly explain the priniciple of the method used for MPEG-2 compression.

5.5 What is I frame? What are the basic requirements to be met by a frame to be picked as an I frame?

5.6 Explain how video is acquired using a digital video camera. Which is the most cost-effective compression method? Justify your answer.

5.7 What is the compression ratio achieved using MPEG-2 file format?

5.8 Explain the difference between acquiring monochrome video and colour video.

5.9 Explain how a video is acquired using a digital video camera.

5.10 What are the major objectives of MPEG compression method? What are the series of MPEG standards and what is each one intended for?

5.11 An English colour movie of 2-hour duration is to be digitized and stored in DVD. How much storage is needed in DVD?

5.12 What are the advantages of digitizing a movie?

5.13 What is a Video-on-Demand (VoD) service? What facilities should be provided by a VoD service?

5.14 A VoD service wants to provide a service with a library size of 150 movies each of average three hours duration. How much disk space will be needed? What compression method is normally used?

Solution With (1440×1152) resolution to transmit uncompressed video, the cable speed needed is

$$1440 \times 1152 \times 24 \times 60 = 2388 \text{ Mbps}$$

We have assumed 24 bits/pixel for the colour video. If we assume MPEG4 compression with bit stream speed of 60 Mbps, the compression ratio is nearly 40. Thus a 3-hour video will need storage of $(3 \times 60 \times 60 \times 2388)/(40 \times 8)$ Mbytes = 80.6 Gbytes. Hundred movies will require a storage of 8060 Gbytes, or 8.06 Terabytes.

If the movie bit stream is to be transmitted on a 5 Mbps channel, we have to sacrifice quality by compressing by a factor of 150. In this case, a 3-hour video will need 21.5 GB store.

We saw in this chapter how to acquire video data and store it. We will discuss in the next chapter the various types of storage devices available for use with computers.

SUMMARY

1. A video camera acquires a moving picture by exposing a scene to a CCD array and scanning the array from left to right starting from the top left corner and ending at the bottom right corner. The scanning is repeated at least 30 times per second to represent motion.
2. For good VDU quality image, at least (640×480) pixels per frame are needed. Each pixel needs 3 bytes coding to represent colour. As the image is repeated 30 times per second, a total of $(640 \times 480 \times 30 \times 3) = 9.216$ MB are generated each second. A 2-hour video needs 200 GB.
3. Besides the enormous disk storage, bandwidth of 9.216 MB/second is needed to read the video from the disk to the display device. It is not economical to read from secondary storage device data at this high rate.
4. It is thus essential to compress the digital video data.
5. In video compression each image frame is first compressed using JPEG compression discussed in Chapter 3.
6. In most videos, a large portion of the background image does not change; movement is limited to a small part of an image. This fact is exploited in video compression.
7. A group of pictures (between 8 and 16) are taken. A starting image is used as a reference for the group. The difference in pixel values of corresponding points in the group of images is much smaller than the number of bits used per pixel. This idea is used to encode other images in a group, thereby compressing the video. Compression ratio by a factor of 70 to 150 has been achieved using MPEG2.
8. A set of standards called MPEG (MPEG1, MPEG2, MPEG4) are used to compress video using the general principle stated in point 7 above. MPEG1 is used for normal TV and VDU. MPEG2 is used for High Definition Digital TV. MPEG4 has several versions. It can be used for high quality video transmission on 80 Mbps channels to low speed data transmission paths such as telephone lines.

Step 2: The clerk enters the data from the form given by you using the keyboard of the terminal. This data is temporarily stored in a memory of the terminal as it is required for immediate reference for booking.

Step 3: Using the data on train number, date of travel, origin and destination stations, the computer retrieves data on availability of berths in the specified train on specified date from a large storage device called a *secondary storage*. This data is stored in a secondary storage as the data on reservation on trains is very large. Data on over 500 trains are stored and for each train data is needed for 60 days as reservations can be done 60 days before the date of journey. If 1000 berths are there in each train, the data will be: No. of trains × No. of days × No. of berths per day × No. of bytes needed for information on each berth = $500 \times 60 \times 1000 \times 100 = 3$ GB.

Step 4: Using the data stored in step 2, a program retrieves data on the specified train on the specified day from the secondary memory and brings it to the main memory of the computer which processes the reservation request. The reason it is brought to the main memory is to allow the processor to compare the seat status with your request for a berth.

Step 5: The berth availability data is compared with your berth request by the processor (which uses an appropriate program stored in the main memory). If berth is available, a ticket is printed with a unique identification number (called PNR) for later reference. The berth is marked as booked and the new information on available berths is written back in the secondary memory.

We thus see from this case that there is a need for a main memory from which data is taken for processing. The amount of data stored in the main memory is relatively small. The main memory is also essential to store program which processes the data. We also saw that there is a need for much larger secondary memory for storing data which needs to be accessed frequently and a portion brought to the main memory for processing. Data in the secondary memory should be "on-line", that is, it should be available whenever needed within a short period for processing while the customer is waiting. By a short period we mean a few thousandth's of a second, that is, a few milliseconds.

There is a need for one more level of memory for the railway reservation system. We now explain why. The reservation data is very important and should not be accidentally lost from the secondary memory. Accidental loss may occur due to hardware fault, software failure, or human error. If this occurs it is essential to restore the data. If at the end of each day all the data from the secondary storage unit is "backed up", that is, a copy is made and stored in another memory device which can be removed from the computer and kept in a safe place such as a steel almirah then it is possible to restore the data in the secondary storage using this backed-up data. Such a storage unit which is not required on-line but is needed to save data for use when required is called an *archival memory unit*.

Apart from the need to back up data from secondary store, there is another need for archival storage. In the railway reservation system, the data on berths stored in the secondary memory is not required after the arrival of a train at its destination. It should thus be removed from the secondary memory. A question which arises is, should it be removed and kept elsewhere or just deleted? It should not be deleted but kept at least for some time due to the following reasons:

1. It may be needed to refund cost of tickets to customers if the train is cancelled.
2. It may be needed for legal purposes.
3. The archival data may be used later for further analysis to find out information such as seasonal variation of traffic and length of waiting list to enable adding coaches, etc.

The archival storage must be able to store large amount of data in a compact removable storage device at a low cost per byte. The archival store must be able to store data for a long period; the data in it should not vanish after a period of time.

This case has shown us the need for three types of memory, namely:

1. The main memory which has a very short access time (tens of nano seconds) and relatively small size (usually millions of bytes).
2. A secondary memory which has storage of the order of gigabytes. Its access time may be longer, around hundreds of milliseconds.
3. An archival memory which has a large storage capacity (several hundred gigabytes), is removable, compact and low cost per byte of storage.

Case 2: All of you would have seen washing machines advertised as "intelligent" or motor car advertisements claiming to use 16 bit processors. These are called embedded computers and are permanently programmed. In other words, the programs are stored in a non-erasable small memory and used to control the operation of systems such as motor cars or washing machines. These are called Read Only Memories which will be described in Section 6.5.

6.2 MEMORY CELL

Memories in computers are made using what we call ***memory cells***. A variety of memory cells are used to fabricate memories. A common characteristic of all cells used in today's computers is that they can be in one of two ***stable states***. By a stable state we mean that the cell remains in that state unless it is intentionally disturbed. A simple example of a system with two stable states is a seesaw you find in children's playground. It is shown in Fig. 6.1. The seesaw will remain in state 1 as long as a child sits as shown. If a heavier child sits on the opposite end of the seesaw, it will go to stable state 2.

FIG. 6.1 A seesaw in one of the two stable states.

6.3.2 A Flip-Flop Storage Cell

A flip-flop is a storage device which uses four semiconductor switches to store either a 0 or a 1. In Fig. 6.4 we show a memory cell which uses a flip-flop to store a 0 or a 1.

FIG. 6.4 A memory cell using a flip-flop.

If a 1 is applied to input and a write signal is applied to the Write line, S1 closes and a 1 is written in the flip-flop. The flip-flop will remain in this state storing a 1 unless it is disturbed. If the data stored in the flip-flop is to be read, a read signal is applied to the Read line which closes S2, and the data stored in the flip-flop appears on the output line. If a 0 is fed to input and a write signal is applied, then a 0 is written in the flip-flop.

The major advantages of a flip-flop compared to a capacitor are:

- It takes lesser time to store a bit in it as compared to a capacitor memory cell.
- The data stored in a cell is not lost with passage of time. There is no need to refresh the memory.
- The readout from a flip-flop is non-destructive. In other words, when data is read from a flip-flop, it is not erased.

The major disadvantages of a flip-flop as compared to a capacitor are:

- Flip-flops are more expensive.
- Flip-flops occupy more space as they use four semiconductor switches.
- They need continuous application of power to maintain their state. When power fails, the stored data is lost. It is thus *volatile*.

Flip-flop based memory cells are used to fabricate what is known as Static Random Access Memory (SRAM). We will describe SRAM in Section 6.4.

6.3.3 A Magnetic Storage Cell

The physical device used as a magnetic storage cell is a magnetic recording surface. The method used to store data (1 or 0) on a magnetic surface is shown in Fig. 6.5.

If a 1 is to be written on the magnetic surface, a current is sent through a coil wound on a magnetic write head. This current creates a magnetic field (as shown in the figure) in the gap in the write head. A plastic or metal surface is coated with a ferromagnetic material. This surface is kept very close to the write head. The field in the gap of the head magnetizes the surface in the same direction as the field as shown in the figure. Observe that a right to left magnetization (←) is taken as a 1. If a 0 is to be written, the current in the coil is sent in the opposite direction. In this case the magnetic field in the gap is from left to right (→), and the surface also is magnetized in the same direction which we call a 0. In order to write a sequence of 1s and 0s, the magnetic surface is moved (in one direction). The direction of the current through the coil of the write head is adjusted as needed to write a 1 or a 0. Bits are thus recorded on the surface as shown in Fig. 6.5(b).

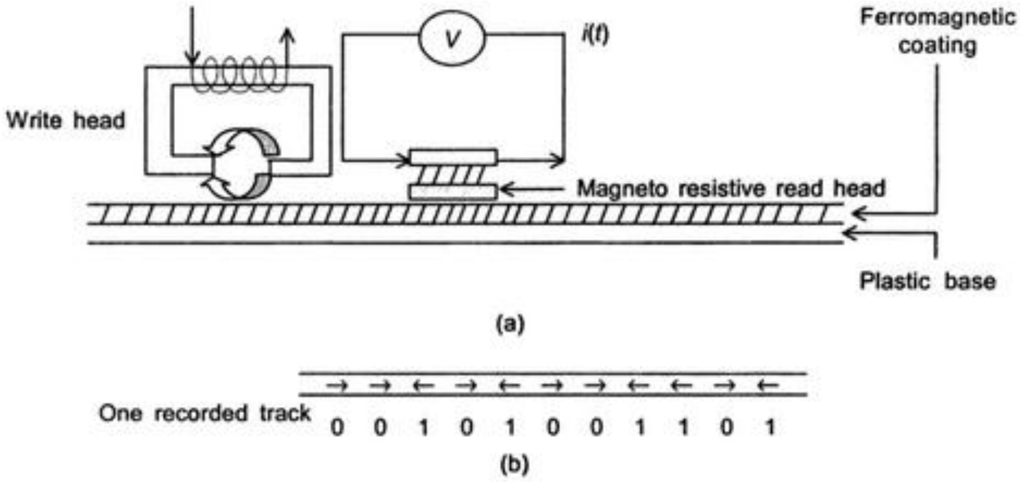


FIG. 6.5 Magnetic storage cell

To read data stored on the surface, a read head is used. The read head is called a magneto resistive head. The resistance of this head increases if a magnet with S → N alignment is placed below it and the resistance decreases if the alignment of a magnet below it is opposite, namely, N ← S. The current *i*(*t*) flowing through the head is proportional to the resistance of the head. Thus, if the current increases we say that a 1 is read (1 is N←S), and if it decreases then a 0 is read. Recording on a magnetic surface is used in magnetic hard disk, floppy disk, and magnetic tape. We will describe these storage systems in Sections 6.6.11, 6.7 and 6.9.

6.3.4 A Polycarbonate Cell

This type of cell is used in what are known as *laser disks* or *compact disks*. The surface of a thin polycarbonate substrate is coated with a reflecting material, usually aluminium. Over this, a protective layer is coated (see Fig. 6.6). To write bits on this surface a laser beam is used. Wherever a 1 is to be written, the beam is turned on and burns a 'pit' up to the reflective layer. Wherever a 0 is to be written, the laser beam is defocused and no 'pit' is burnt. A sequence of cells, each cell being a pit or no pit (no 'pit' is called a 'land') are traced along a spiral track on the surface as the disk is rotated. Reading is achieved by rotating the disk and moving a laser beam along a track. Wherever there is a 'land' light reflects from the reflective layer and no light gets reflected from a 'pit'. The reflecting light is sensed by an electronic light sensing device and converted to an electrical signal representing a 0 or 1.

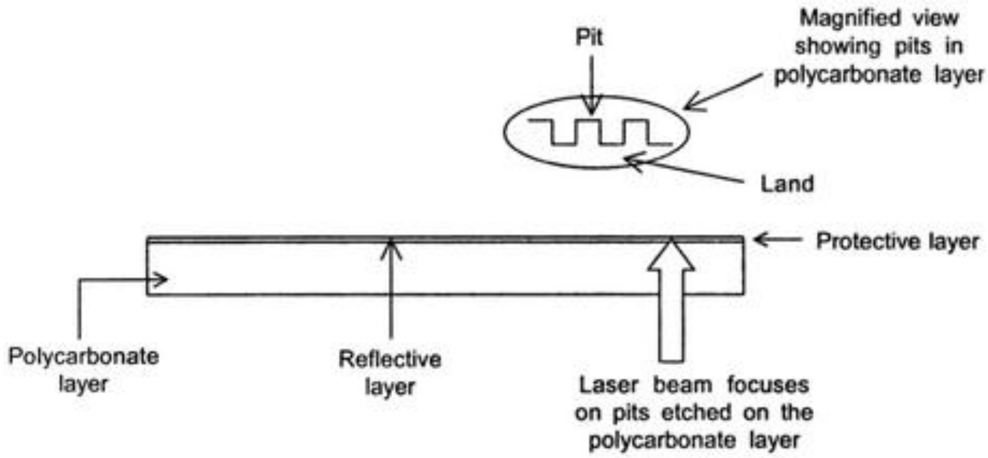


FIG. 6.6 Memory cell on polycarbonate layer.

6.4 RANDOM ACCESS MEMORY

The main memory of a computer is made using either capacitor storage cells or flip-flops (as physical device) to store data. A set of cells are strung together to form a unit called a *word.* As each cell stores a bit, a word consists of a set of bits. The number of bits in a word is called *word length*. A word is always treated as an entity and is written into or read from a memory as a unit. Words to be written are first entered in a small temporary storage unit called a *register*. This temporary storage is known as the Memory Data Register (MDR, for short). When a word is read from memory, it is also stored in MDR, overwriting its contents.

The location in the memory where a word is to be stored is called the *address* of the word. When a word is to be stored in a memory, it is first stored in MDR. The address where this word is to be stored is entered in another register called Memory Address Register (MAR, for short). The memory control circuits issue a write signal and the contents of MDR are written in the address specified by MAR. If a word is to be read, its address is entered in MAR and a read signal is issued by the memory's control circuits. This results in the contents of the word whose address is in MAR to be copied from the memory to MDR.

It is thus clear that the memory unit described above is organized as a number of addressed locations, each location storing a word. The address normally starts at 0, and the highest address equals the number of words that can be stored in the memory. If a memory has 4096 locations, each location storing a byte (8 bits), the address starts with 0 and ends with 4095. The MAR should have 12 bits, the address starting with 0 and end with 111111111111 or FFF (Hex). The MDR should have 8 bits to accommodate 8 bits of a word.

Figure 6.7 depicts the block diagram of the memory which stores 4096, 8-bit words. The figure shows a MAR which is 12 bits long, and MDR which is 8 bits long. The

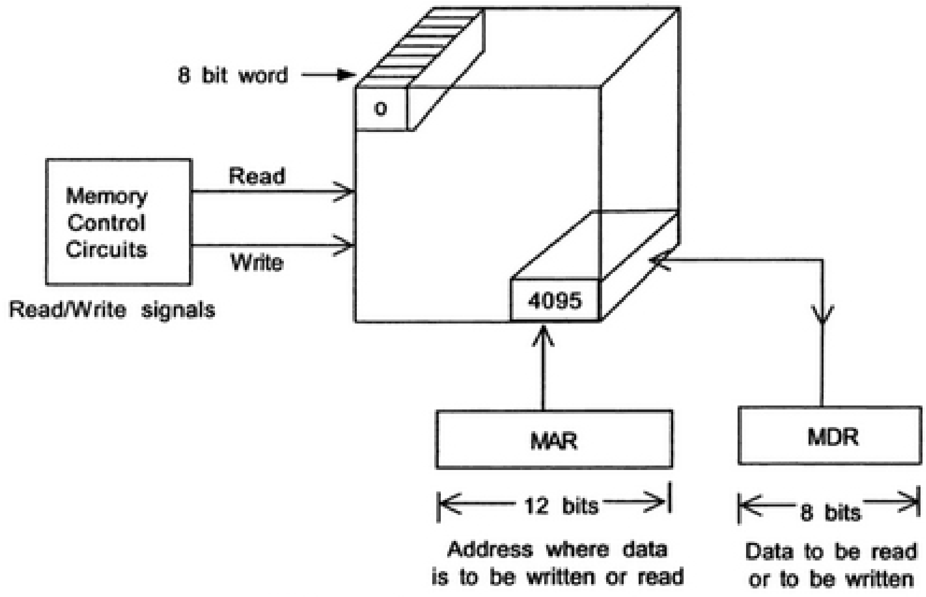


FIG. 6.7 A Random Access Memory.

memory control circuits generate read/write signal. Two parameters of great interest to users are (i) the size of the memory, and (ii) the speed of the memory.

EXAMPLE 6.1 A RAM has 256 MB memory and 16 bits are accessed in each memory read/write. What is the size of MAR and MDR of this memory?

Solution: The capacity of memory is 256 MB. If any byte is to be accessed it should be possible to address any byte. As there are 256 MB = 2^{28} bytes, the number of bits in MAR is 28 bits. As 16 bits are retrieved in each memory cycle, the size of MDR must be such as to accommodate all the bits retrieved. Thus the size of MDR is 16 bits.

The size of a memory is specified using bytes as the unit. In Example 6.1 the memory size is 256 MB. The speed of a memory is specified by three parameters: *access time, write time* and *cycle time*. Assume that the address from which data is to be read is placed in MAR and at time t_0 a read signal is issued. If the required data is available in MDR at t_1, the interval $(t_1 - t_0)$ is known as *access time* or *read time* (see Fig. 6.8a). Assume the address at which data is to be written is placed in MAR and data to be written is placed in MDR and a write signal issued at time t_2. If the data is written in the specified address at time t_3, then $(t_3 - t_2)$ is called the *write time*. The interval between t_0 at which read/write signal is issued to a memory and the time t_4 at which next such instruction can be safely issued, namely $(t_4 - t_0)$, is known as the *cycle time* of the memory.

If the time to read/write a word in a memory is *independent of the address*, then this memory is called a *Random Access Memory* (RAM). The main memory used to store data and programs is normally a RAM. The speed of data processing is directly dependent on

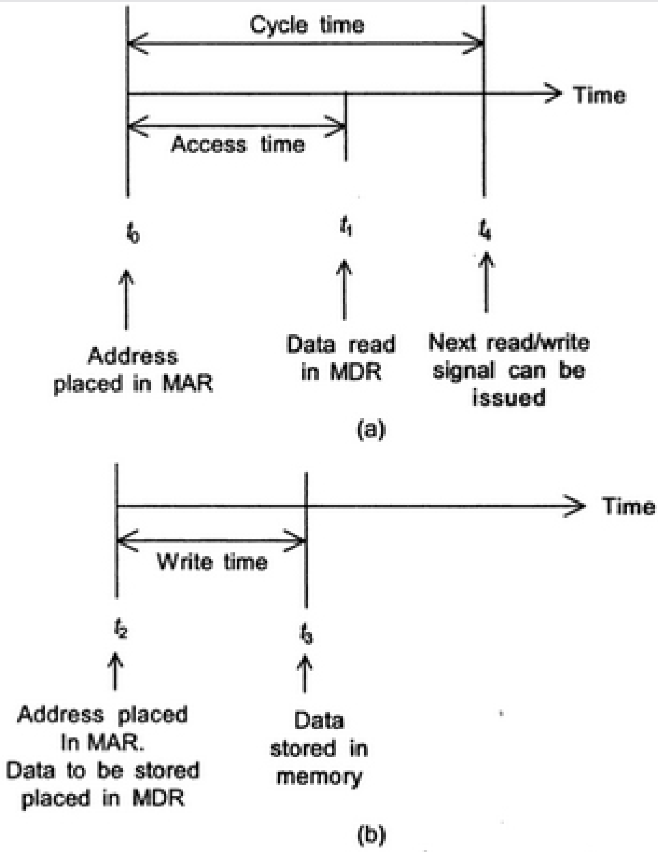


FIG. 6.8 Defining access time, write time and cycle time of RAM.

the speed of RAM. Thus the higher the speed, the shorter will be the processing time. If the access time of a word is dependent on where it is placed, then a knowledge of where instructions of a program are placed would be essential to write efficient programs (i.e., programs which require minimum time to execute). This problem is eliminated if the main memory is a RAM. Thus all computers now use RAM as main memory.

There are two types of RAMs used in today's computers. One of them is called Dynamic Random Access Memory (DRAM), and the other Static Random Access Memory (SRAM). The storage cell of DRAM is a capacitor. Thus DRAM will lose the data stored in it unless it is periodically refreshed, that is, data in the entire memory is rewritten. This rewriting may have to be done after each read/write cycle. This will, besides complicating the memory circuits, also slow down the memory. The storage cells used in SRAM is a flip-flop. A flip-flop does not lose data stored in it with passage of time and therefore does not need refreshing. The cycle time of SRAM is around 5 times shorter than that of DRAM. Thus SRAM is preferable as the main memory. However, the cost per bit of storage is higher for SRAM. Further, SRAM capacity is much lower compared to DRAM for a given integrated circuit chip area. Thus DRAMs are used as the main memory of PCs in most computers. In order to take advantage of SRAM's higher speed, it is used as *cache memory*. Cache memory size is smaller. A portion of data and instructions immediately needed to run a program is brought and put in cache from which it is taken by CPU for processing. We will describe cache memory in greater detail later in this book.

If you see advertisements for a PC, the main memory is usually specified as SDRAM or, in more recent PCs, as RDRAM. SDRAM stands for Scynchronous DRAM. It uses a technology to design DRAM in which the time taken to read/write the first word is long but if read/write is to neighbouring addresses, it takes much shorter time. RDRAM stands for Rambus DRAM where Rambus is the name of a company which has developed this technology. The data path in RDRAM for reading memory is designed with the aim of reducing cycle time per word. Both DRAM and SRAM memories are *volatile*. A memory is said to be volatile if data stored in it is lost when power supply to it is turned off.

There is a class of SRAM which uses a technology called CMOS which consumes very little power. These can be operated on batteries for reasonably long period and are not volatile.

Another terminology used in describing RAMs is whether they are "destructive read out". By destructive read out we mean that the contents of the memory are erased by the process of reading. Hence it must be rewritten as soon as it is read. SRAM is non-destructive readout, whereas DRAM is a destructive readout memory.

6.5 READ ONLY MEMORY

In Section 6.4 we saw how semiconductor storage cells are used to fabricate the main random access memory of computers. We saw that RAMs are used for storing programs and data used during computation. RAMs are erasable and rewritable and are large. In Section 6.1 (case 2), we saw the need for Read Only Memories (ROM) in certain types of applications. In this section we will describe the structure of ROMs and their applications.

Like a RAM, ROM also is an addressable random access memory. In Fig. 6.9 we depict a 4 KB ROM with 8 bits per word. ROM uses semiconductor memory cells which can be permanently placed in one of two states, 0 or 1, by a process called *fusing*. It is a non-destructive, non-volatile memory. In other words, when we read data stored in it, a copy of the data is delivered; the original remains in the memory. It is non-volatile, that is, when power is turned off, the data stored in it is not lost. With passage of time

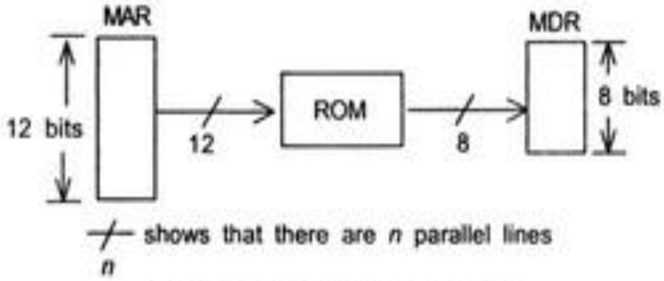


FIG. 6.9 Structure of a ROM.

also the data stored in a ROM is not lost. In other words, the data stored in a ROM is retained indefinitely. Just like in a RAM, to read a data stored in ROM, the address of the data is placed in MAR and a read signal is issued. A copy of the data in the addressed location appears in MDR from which it can be used for an application.

An important application of ROM is to store tables which do not change. As an example, let us see how a seven-segment display (shown in Fig. 6.10) commonly used in digital watches and calculators can be designed using a ROM.

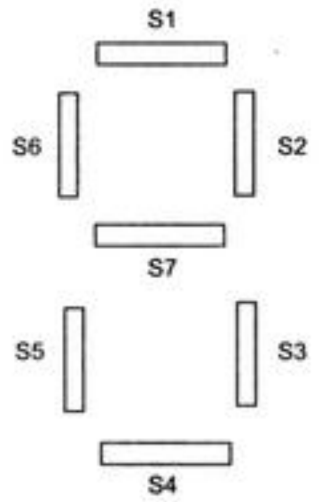


FIG. 6.10 A seven segment display.

A table showing numbers (their binary equivalents) as input and segments of the display to be lighted for each number is given in Table 6.1.

Table 6.1 Contents of ROM for Seven Segment Display

Digit	*Input*	*Output*						
		Segments						
		S1	S2	S3	S4	S5	S6	S7
0	0000	1	1	1	1	1	1	0
1	0001	0	1	1	0	0	0	0
2	0010	1	1	0	1	1	0	1
3	0011	1	1	1	1	0	0	1
4	0100	0	1	1	0	0	1	1
5	0101	1	0	1	1	0	1	1
6	0110	0	0	1	1	1	1	1
7	0111	1	1	1	0	0	0	0
8	1000	1	1	1	1	1	1	1
9	1001	1	1	1	1	0	1	1

S1, S2, S3, S4, S5, S6, S7 are segments (a light emitting device) which can be either lighted or not lighted.

Let us examine line 1 of Table 6.1. When the input is 0, all the lights except S7 in the seven segment display should light up. This is shown by putting a 1 for the segment to be lighted and 0 for the segment which is to remain dark. This table can be stored in a ROM. When 4 bits corresponding to the number to be displayed is placed in the MAR of a ROM, the seven bits corresponding to S1, S2, S3, ..., S7 appear in the MDR. These bits are used to light up the appropriate lamps. For example, if we place 0 1 0 1 in the MAR (corresponding to the number 5), the output bits are S1 = 1, S2 = 0, S3 = 1, S4 = 1, S5 = 0, S6 = 1, S7 = 1. Thus the segments S1, S3, S4, S6 and S7 will light up as shown in Fig. 6.11, and thus 5 is displayed.

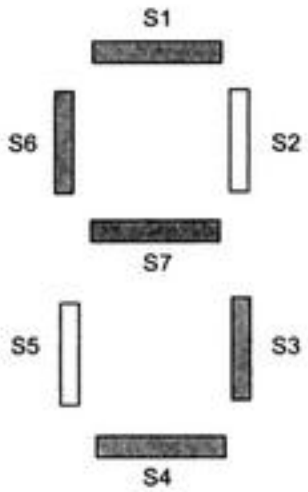


FIG. 6.11 Displaying 5 with seven segment display.

Another application of ROM is for storing small programs used to control devices such as washing machines, building lifts, etc. A ROM which has data written in it during manufacture in a factory is known as *factory programmed ROM*. Factory programmed ROMs are used when large quantities of the ROM are needed. The ROM with a seven-segment table will be factory programmed. ROM with program for a washing machine will also be factory programmed.

In many applications a user may like to store his own special tables or programs. For example, an instrument may require a conversion table for calibrating it, which may vary from instrument to instrument. Some common programs stored are for controlling the operation of custom designed appliances such as a security system. In such cases a ROMs in which data can be written in the field are required. Such ROMs are available and are called Programmable ROM (PROM). Programming requires special equipment to write. The time taken to write may be a few seconds as compared to reading time which may be a fraction of a microsecond.

PROM is not flexible as the data stored in it cannot be altered. If an error is made while writing, the PROM has to be thrown away. One would thus like to have an erasable PROM in which data which has been written can be erased and fresh data written in the field (i.e., in your laboratory/office). Erasable and programmable (i.e., rewritable) ROMs have been made. These are known as Erasable Programmable ROM (EPROM).

There are two types of EPROM known as UVEPROM and EEPROM. UVEPROM is erased by shining ultraviolet light on a window on the EPROM chip. It can then be reprogrammed and new data stored. In EEPROM, electrical pulses are used to erase the contents of PROM; thus the name Electrically Erasable PROM. After erasing the contents, new data can be written. It is thus seen that ROM is not a read/write memory like the main memory of computer. Even though ROM is also a random access memory, the write time is much higher than the read time in ROM. Writing also needs special circuitry in ROM.

In Fig. 6.12 we summarize the information on semiconductor memories.

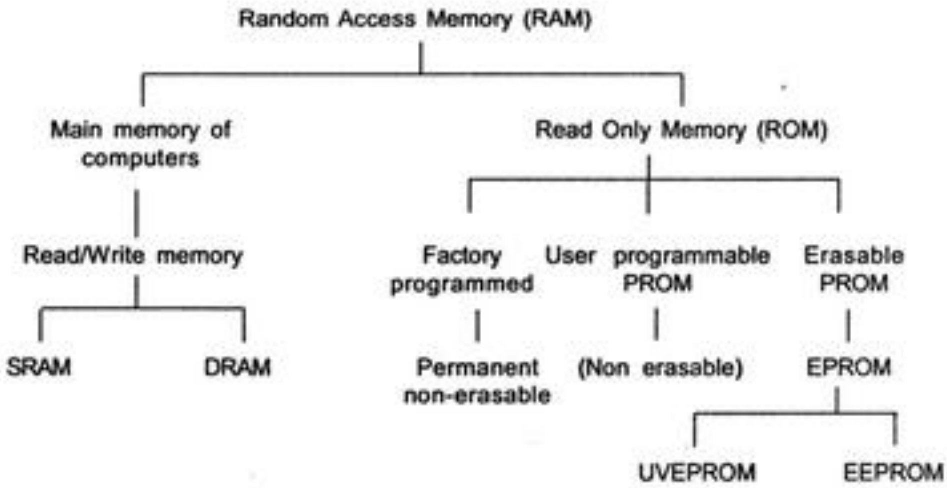


FIG. 6.12 Varieties of semiconductor Random Access Memories.

6.5.1 Flash Memory

A small variant of Electrical Erasable and Programmable Read Only Memory (EEPROM) is called a flash memory. Flash memories are semiconductor memories. They use one transistor per memory cell and come in capacities ranging from 1 MB to 1 GB. They are non-volatile. They will retain data stored in them without a power source. Thus they are attractive as a storage device to replace magnetic storage such as floppy disk. The read time of flash memory is much smaller (tens of nanoseconds) as compared to write time (tens of microseconds). They are currently used to store images from digital cameras. Recently, some computer manufacturers have announced that they will be replacing the floppy disk drive with flash memory cards. Flash memory cards are very compact (1″ × 1″) and store several megabytes. They are currently more expensive than floppy disks. Their cost is, however, going down.

6.6 SECONDARY MEMORY

We saw in Section 6.1 that, besides the main memory which is a random access memory, we need secondary memory which has a much higher capacity but is relatively slow. One should, however, be able to access any data in it arbitrarily. In other words, it should be directly addressable memory.

By a direct access memory we mean it is a memory in which data is stored in specified addresses. Data can be read directly if the address is given. In this sense it is similar to a Random Access Memory (see Section 6.4). In RAM, the time taken to store/retrieve data is *independent of address* but in a direct access memory the time to store/retrieve data is *dependent on the address*. In other words, the time taken to retrieve data in a direct access memory depends on where it is stored. The most common direct access memories used in computers are magnetic disks and CD ROM disks. We will describe in this section magnetic disks which are used as secondary memory in today's computers.

6.6.1 Magnetic Disk Drives

Magnetic disks are smooth metal plates coated on both sides with a thin film of magnetic material. A set of such magnetic plates are fixed to a spindle one below the other to make up a disk pack (see Fig. 6.13). The disk pack is sealed and mounted on a disk drive. Such a disk drive is known as a Winchester disk drive. The disk drive consists of a motor to rotate the disk pack about its axis at a speed of around 7200 revolutions per second. The drive also has a set of magnetic heads mounted on arms. The arm assembly is capable of moving in or out in radial direction (Fig. 6.13). Data is recorded on the surface of a disk by the magnetic heads while the disks rotate about their common axis. Thus recording is on concentric *tracks* on each disk surface. A set of corresponding tracks on all surfaces of a disk pack is called a *cylinder* (see Fig. 6.13). If a disk has n platters, there are $2n$ surfaces and thus $2n$ tracks per cylinder. A track is divided into *sectors*. A sector is the smallest addressable unit in a disk. Thus, Writing/Reading of a disk is in sectors. If a string of bytes to be written is less than the capacity of a sector then the last byte is duplicated to fill up available sector space and the result is stored in a sector. If a disk has p sectors per track and a sector stores s bytes, then the storage in each track is $p \times s$ bytes. If the disk has t tracks per surface and there are m surfaces, the total capacity of the disk is:

$$\text{Total capacity} = p \times s \times t \times m \text{ bytes}$$

EXAMPLE 6.2 A hard disk has 5268 tracks on each surface. It has six plates. How many tracks are there per cylinder? What is the total number of cylinders in the disk? How many tracks are there in the disk?

Solution: There are six plates and hence 12 surfaces. The corresponding tracks in all the surfaces constitute a cylinder. Thus there are 12 tracks in each cylinder. Total number of cylinders = Tracks per surface = 5268. Total number of tracks = Tracks per surface × number of surfaces = 5268 × 12 = 63216.

lighted and 0 for the segment which is to remain dark. This table can be stored in a ROM. When 4 bits corresponding to the number to be displayed are placed in the MAR of a ROM, the seven bits corresponding to S1, S2, S3, ..., S7 appear in the MDR. These bits are used to light up the appropriate lamps. For example, if we place 0 1 0 1 in the MAR (corresponding to the number 5), the output bits are S1 = 1, S2 = 0, S3 = 1, S4 = 1, S5 = 0, S6 = 1, S7 = 1. Thus the segments S1, S3, S4, S6 and S7 will light up as shown in Fig. 6.11, and 5 is displayed.

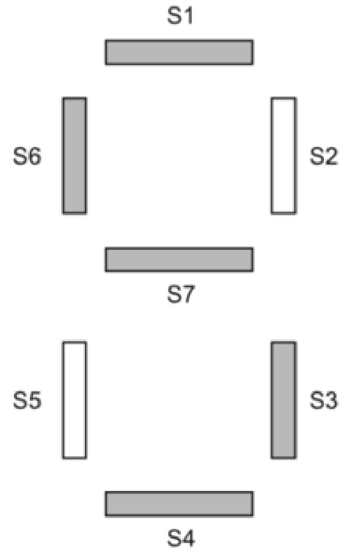


FIG. 6.11 Displaying 5 with seven-segment display.

Another application of a ROM is for storing small programs used to control devices such as washing machines, microwave ovens, etc. A ROM which has data written in it during manufacture in a factory is known as *factory programmed ROM*. Factory programmed ROMs are used when large quantities of the ROM are needed. The ROM with a seven-segment table will be factory programmed. ROM with program for a washing machine will also be factory programmed.

In many applications a user may like to store his own special tables or programs. For example, an instrument may require a conversion table for calibrating it, which may vary from instrument to instrument. Some common programs stored are for controlling the operation of custom designed appliances such as a security system. In such cases ROMs in which data can be written in the field are required. Such ROMs are available and are called Programmable ROM (PROM). Programming requires special equipment to write data in the ROM. The time taken to write the required data in a PROM is normally a few seconds as compared to reading time which is usually less than a microsecond.

PROM is not flexible as the data stored in it cannot be altered. If an error is made while writing in a PROM, it has to be thrown away. One would thus like to have a PROM in which data which has been written can be erased and new data written in the field (i.e., in your laboratory/office). Erasable and programmable (i.e., rewritable) ROMs (EPROMs) are now available.

There are two types of EPROMs known as UVEPROM and EEPROM. UVEPROM is erased by shining ultraviolet light on a window on the EPROM chip. It can then be reprogrammed and new data is stored in it. In EEPROM, electrical pulses are used to erase the contents of PROM, and thus the name Electrically Erasable PROM. After erasing the contents, new data can be written. It is thus seen that ROM is not a read/write memory like the main memory of computer. Even though ROM is also a random access memory, the write time is much higher than the read time. Special circuitry is needed to write data in a ROM.

In Fig. 6.12 we summarize the information on semiconductor storage systems.

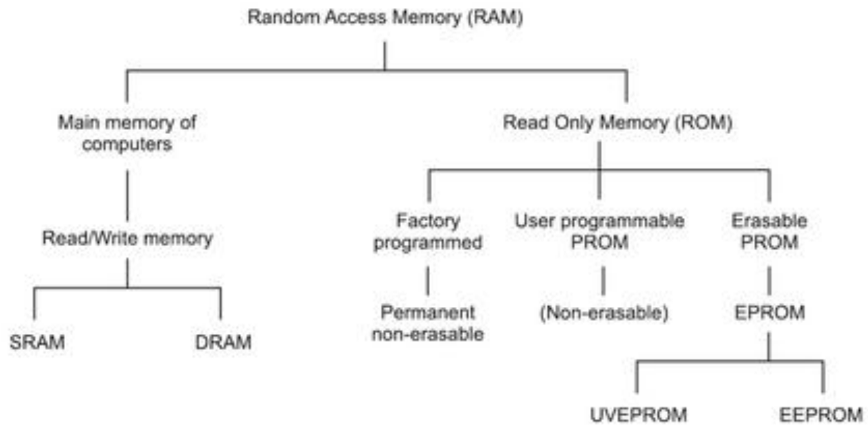


FIG. 6.12 Varieties of Semiconductor Random Access Memories.

6.5.1 Flash Memory

A small variant of the Electrically Erasable and Programmable Read Only Memory (EEPROM) is called a flash memory. Flash memories are semiconductor memories. They use one transistor per storage cell and come in capacities ranging from 16 MB to 256 GB. They are non-volatile. They will retain data stored in them without a power source. Thus they are attractive as a storage device to replace magnetic storage such as a disk. Data is written and read in large blocks of 512 to 4096 bytes in flash memories unlike EEPROM. Thus it is similar to a disk rather than a RAM. Unlike a ROM in which the time needed to read data is much smaller than the time required to write data, there is no difference between read and write times in flash memories. Speed is specified as the data transfer rate from and to the flash memory and is quoted as nx, where x is 150 KB/second and n is variable. The value of n is around 100 and it is continually increasing. Digital cameras were the first devices to use flash memories. Currently, they are used in several applications. They have replaced floppy disks as portable storage. There are two shapes which are common. A stick shape called a *pen drive* plugs into a USB port of a computer or a laptop. A flat disk shape 1.5 cm × 1 cm can be inserted in cameras and tablet computers. They are also being used as a replacement of hard disk drives in laptops and tablet computers. The flash drives used as replacement of hard disks are large, 128 GB to 256 GB, and can transfer data at 1500 MB/second. They are much more expensive than hard disks but are faster. Unlike hard disks they do not have any mechanical moving parts and thus no translational and rotational delays. They are also rugged, can withstand shock and consume very little power. Another use of flash memories is as cache memory of disks to reduce latency delay.

6.6 SECONDARY STORAGE

We saw in Section 6.1 that, besides the main memory which is a random access memory, we need secondary storage which has a much higher capacity but is relatively slow. One should, however, be able to access any data in it arbitrarily. In other words, it should be directly addressable storage.

By a *direct access storage* we mean it is a storage in which data is stored in specified addresses. Data can be read directly if the address is given. In this sense it is similar to a Random Access Memory (see Section 6.4). In RAM, the time taken to store/retrieve data is *independent of address* but in a direct access store the time to store/retrieve data is *dependent on the address*. In other words, the time taken to retrieve data in a direct access storage depends on where it is stored. The most common direct access storage used in computers are magnetic disks, CDROMs and flash storage. We will describe in this section magnetic disks which are used as secondary store in today's computers.

6.6.1 Magnetic Disk Drives

Magnetic disks are smooth metal plates coated on both sides with a thin film of magnetic material. A set of such magnetic plates are fixed to a spindle one below the other to make up a disk pack (see Fig. 6.13). The disk pack is sealed and mounted on a disk drive. Such a disk drive is known as a Winchester disk drive. The disk drive consists of a motor to rotate the disk pack about its axis at a speed of around 7200 revolutions per second. The drive also has a set of magnetic heads mounted on arms. The arm assembly is capable of moving in or out in radial direction (Fig. 6.13). Data is recorded on the surface of a disk by the magnetic heads while the disks rotate about their common axis. Thus recording is on concentric *tracks* on each disk surface. A set of corresponding tracks on all surfaces of a disk pack are called a *cylinder* (see Fig. 6.13). If a disk has n platters, there are $2n$ surfaces and thus $2n$ tracks per cylinder. A track is divided into *sectors*. A sector is the smallest addressable unit in a disk. Thus sectors are read and written on a disk. If a string of bytes to be written is less than the capacity of a sector, then the last byte is duplicated to fill up available sector space and the result is stored in a sector. If a disk has p sectors per track and a sector stores s bytes, then the storage in each track is $p \times s$ bytes. If the disk has t tracks per surface and there are m surfaces, the total storage capacity of the disk is:

$$\text{Total storage capacity} = p \times s \times t \times m \text{ bytes}$$

An interface can control several peripherals such as disk, CDROM, printer, etc. There are four standard interfaces marketed by vendors. They are:

1. ATA or EIDE (Enhanced Integrated Drive Electronics)
2. SCSI (Small Computer System Interface pronounced SCUZZY)
3. Ultra wide SCSI
4. Fibre Channel Interface.

These interfaces are compared based on at what rate data can be transferred from the disk to main memory and back, the highest speed drives they can support and the number of drives that can be controlled by the interface.

Of these interfaces, ATA is the preferred interface used with desktop PCs with windows operating system. This is due to its low cost and performance which is sufficient for desk top PC applications. SCSI is a more expensive interface and is used with expensive faster computers which support more than four peripherals such as several hard disks, CDROM, scanner, printer, etc. High speed disks also require SCSI interface. If a computer is used by many users and higher transfer rate disks are needed, ultra SCSI is appropriate. Fibre channel has the highest performance. It is used to interface several disks/peripherals and arrays of disks. It can support up to 127 peripherals. In Table 6.3 we summarize the characteristics of these interfaces.

Table 6.3 Characteristics of Interfaces

Interface	*Maximum transfer rate of disk*	*Maximum disk drive speed*	*Number of drives that can be controlled*
ATA	10 MB/s	4500 rpm	4
SCSI	40 MB/s	7200 rpm	8
Ultra Wide SCSI	80 MB/s	10000 rpm	8
Fibre channel	100 MB/s	10000 rpm	127

So far we have discussed at considerable length the hardware details of hard disks which is the most important secondary memory and what we might call the "work horse" in applications. In fact, this is the primary "on line" memory. By "on line" we mean data in it is available as soon as it is requested by the computer's processor and delivered to the CPU within a few milliseconds. In the example of railway reservation system we described at the beginning of this chapter, we saw that when a passenger reserves a berth, this data is stored on a disk. The location where it is stored must be unique. The location is related to a number, namely PNR, which uniquely identifies the reserved ticket. If a passenger wants to, say, cancel a ticket, the clerk enters the PNR of the ticket on her keyboard. The CPU translates this PNR to the address in disk where data relating to this PNR is stored. The time to retrieve the ticket details from the address and display it on VDU screen will be a few tens of milliseconds. It has to be small (less than a second) as it is done while a passenger is waiting. If the data is stored off line, that is, in a storage medium not connected to the computer, it has to be manually mounted and will take minutes and is obviously unacceptable to a person waiting in a queue. Thus "on-line" availability of data for such applications is essential and a magnetic disk is the most important on-line secondary memory.

6.7 FLOPPY DISK DRIVE

Another secondary memory which is popular is a floppy disk drive. The capacity of a floppy disk is thousand times smaller than that of a hard disk. Further, the time to access data stored in a floppy drive is at least 10 times slower compared to a hard disk. The cost of a floppy drive is however lower. Nowadays floppy disk is not used as an on-line memory as it is slow and also because data in it may be lost due to wear and tear. The wear and tear is due to the fact that the read/write head is in physical contact with the disk surface during its operation.

Floppy disks are made of magnetic oxide-coated thin Mylar sheets—thickness of a sheet of paper. This flexible sheet is cut into circular pieces 3.5 inches (or 8.9 cm). As the material used for storage of data is not a hard plate but a flexible circular sheet, it is called a "floppy disk". The floppy disk is packaged in a 3.5 inch square hard plastic envelope with a long slit for read/write head access, and a hole in the centre for mounting the disk drive hub (see Fig. 6.16). The floppy disk along with the envelope is slipped into the drive mechanism. The mechanism holds the envelope, and the flexible disk is rotated inside the envelope by the drive mechanism. The inner side of the envelope is smooth and permits free rotation of the magnetic sheet. The read/write head is held in physical contact with the floppy disk. The slit for read/write remains closed until the disk is inserted into the drive. The slit opens when the disk starts spinning. The head is moved radially along the slit. Data is recorded in concentric tracks. The capacity of a floppy disk is 1.44 MB and the transfer rate is around 40 KB/s. For reading or writing on the disk,

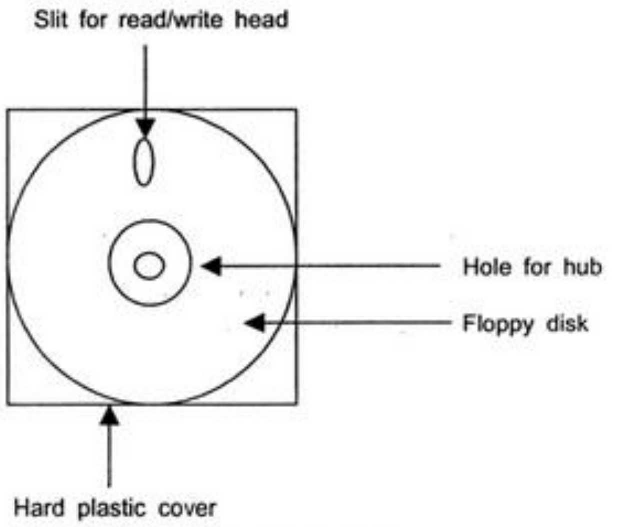


FIG. 6.16 A floppy disk.

the head is to be in contact with the disk surface. Thus both the disk and the head wear out. Manufacturers guarantee 2 million read/write operations on a floppy disk. You should contrast this with a hard disk drive where the read/write head floats on a thin layer of air and it is called non-contact recording and there is no wear and tear. The read/write speed of a hard disk is also much higher.

We must emphasize again the need to distinguish between a drive mechanism and the medium in which data is recorded. In the case of a floppy disk, the medium is a floppy disk which is removable from the drive. As the disk is removable, it can be used to record data which can be sent by mail or by courier. For example, if you write an article for a magazine it can be stored in a floppy disk and mailed to the magazine. This is called *soft copy*. The printed text is called the *hard copy*.

In contrast to this, hard disks are sealed inside an enclosure and fixed to the drive; they are not removable. One major disadvantage of floppy disk is its limited storage capacity. Floppy disk technology, however, is making rapid strides. Some of the technology used in hard disks such as non-contact recording and magneto-resistive heads are being used in the design of recent floppy disk drives. Sony and Fuji have a new recording format called HiFD for storing data on a floppy disk. With this technology it is possible to store 200 MB on a 3.5 inch floppy disk and it is backward compatible with the more common 1.44 MB floppy disk. The major application of a floppy disk nowadays is as a medium in which data can be downloaded from a computer and sent by mail. The limited capacity makes it difficult to download multimedia such as audio or video files on a floppy. Thus floppy is becoming obsolete and being replaced by flash memories.

6.8 COMPACT DISK READ ONLY MEMORY (CDROM)

In Section 6.3 we described how data is stored and retrieved from a polycarbonate cell used in laser disk. In this section we will see how this cell is used in a storage system called CDROM. The drive is called a *CDROM drive* and the medium in which data is recorded is called *CDROM disk*. A CDROM disk is a circular shiny metal-like disk whose diameter is 5.25 inches. It can store 650 MB which is equivalent of 3,00,000 pages of printed text. As the name implies, we assume that data on a CDROM is pre-recorded and it is read using a CDROM drive.

The data on a CDROM is recorded as a sequence of lands and pits along spiral tracks. (see Fig. 6.17). The data stored on CDROM is read by inserting the disk in the drive. A drive in the motor rotates the disk at a speed of 360 revolutions per second. A sharply focused laser beam senses pits and lands as the disk rotates. This is converted to 1s and 0s by an electronic interface unit.

CDROM disk speed is indicated using the notation nx, where x is an integer indicating the original nominal speed of 150 KB/s and n is the factor by which it is to be multiplied. Thus a CDROM drive specified as 52x has a speed of 52×150 KB/s = 7.8 MB/s.

A standard has been evolved for storing data on CDROM. Such a standard is essential to widely distribute recorded CDs which can be read by different manufacturers' computers which may also use different operating systems. The current standard is called ISO 9660,

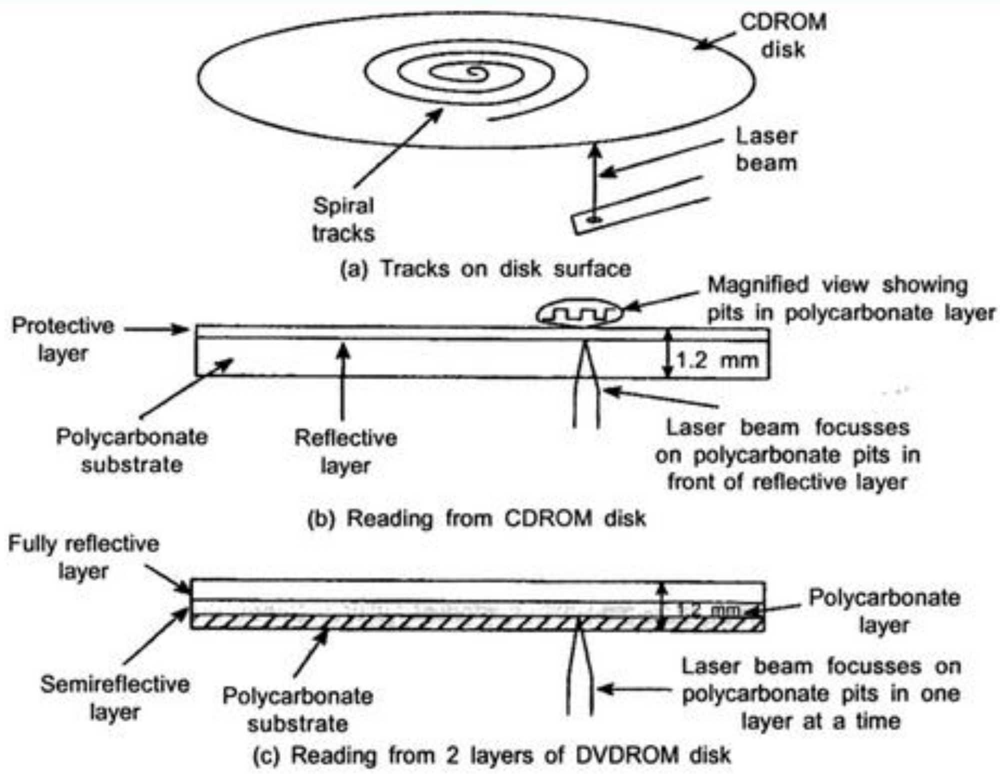


FIG. 6.17 CDROM and DVDROM.

standardized by the International Standards Organization which has been accepted by all CDROM vendors. This standard defines a Volume Table of Contents (VTOC) which allows opening any one of 140,000 files with a single seek using directories in the CDROM. A software developed by Microsoft Corporation known as MSCDEX (Microsoft CD Extension) allows reading of CDROMs on a PC. MSCDEX makes a CDROM look like a large hard disk to a programmer. Nowadays most software is distributed on CDROMs by vendors and thus it is essential to have a CDROM drive on a PC or any other computer.

Major applications of CDROMs are:

1. Distribution of software.
2. Distribution of large texts such as encyclopedias, big manuals, conference proceedings, etc.
3. Distribution of multimedia files—that is, text, audio, graphics and video clips used in education and entertainment.

If a CDROM is properly handled, it is estimated that it can preserve stored data for 100 years. The only problem is that with development of technology, the models of CDROM readers will change and old CDROMs may not be compatible with new readers.

6.8.1 CDROM—Recordable

Hitherto the primary method of producing and distributing CDROM has been to store the data to be recorded on a CDROM on a magnetic tape in ISO 9660 format and send it to a CDROM producer who creates a CDROM master using expensive equipment. This master is used to produce multiple copies which are tested and labelled. This is economical for producing a large number of copies. For producing a single copy or a few copies, a writable CD drive is required. When data is written using such a drive, it should be readable on any standard CDROM reader. Such a CD writer is available now.

The medium used as a writable CD is of standard CD size (5.25 inches diameter disk) made of polycarbonate with a groove cut on it. The groove is covered by a photosensitive dye above which there is a layer of gold (called reflective layer) and finally a protective layer. The writable CD is mounted on the CD writer and is rotated by a motor. A laser beam is used to write data on the disk. For writing a 1, the laser beam is turned on. The laser beam fuses the dye to the substrata forming a pit. When a 0 is to be written, the laser beam is not turned on, thus leaving a "land". When this disk is inserted in a CD reader, it will sense "lands" and "pits" and correctly interpret the sequence of 0s and 1s written on the CDROM.

Data to be written on a CD is first stored in a hard disk using ISO 9660 format. It is streamed at the rate of 300 KB/s to the CD writer which steers it to the write head which is a laser beam. The beam is modulated depending on whether a 0 or a 1 is to be written. The writer's laser drive head writes the data. Writing has to be done in one session, that is, a whole CDROM has to be recorded in one go. The recorded disk can be read on any ordinary CDROM drive. Once a blank CD is written, it cannot be erased and rewritten. Thus, CDROM is not a read/write memory like a floppy disk. Recently, rewritable CDROM media have emerged. In this type of CDROM if a mistake is made in writing, the entire CD contents has to be erased and new data will have to be rewritten.

6.8.2 Digital Versatile Read Only Memory (DVDROM)

DVDROM uses the same general principle as a CDROM. Pre-recorded DVDROMs are more common. Recordable DVDROMs are slowly appearing in the market but as of now they are not widely used. Pre-recorded DVDROMs are primarily used to store large video files as their capacity is very high compared to CDROM.

A DVDROM reader also uses a laser beam to read pits and lands as in a CDROM. The wavelength of the laser beam is however much smaller. Data is recorded on two different layers on the disk. The pits are much smaller as a shorter wavelength laser beam gives a smaller focused spot. This allows more data to be recorded on a track and also allows tracks to be closer to one another. The capacity of DVDROM is thus 8.5 GB which is much higher than the 650 MB capacity of CDROM. Another variety of DVDROM has data recorded on both sides. This, of course, requires the DVDROM to be reversed to read the reverse side. With double sided data recording, the capacity is double that of a single side record, that is, 17 GB.

The 1x speed of DVDROM is 1.38 MB/s as against 150 KB/s of CDROMs. The high of computer will use almost all of these memory systems—each for a different purpose. The requirement to use all these types of storage is similar to the requirements an office manager has while dealing with a variety of files in his day-to-day work. The most urgent files needing his immediate attention are kept in front of him on the table. The space on the table is limited and only files needing his immediate attention are put there. An SRAM, cache memory is similar to this. At the next level are the files which the manager should process during the day. They are kept in the in-tray on the desk which can be quickly accessed. This is analogous to the main RAM in a computer whose access time is small. Files which need not be accessed immediatetly but may have to be accessed quickly when needed are indexed and kept in filing cabinets in his personal assitant's office adjoining his office. These files take longer to retrieve as compared to those from the in-tray. This storage is similar to secondary memory of a computer (hard disks/floppy disks). Files which have been disposed off and need rarely to be accessed are kept in big indexed files in the head office (or the so called old records units in some offices). These are similar to archival storage, e.g., CDROMs or magnetic tapes, e.g., HIC or DAT. Finally, legal books, reference books and office procedure manuals are kept in a library. CDROMs and DVDROMs are used for a similar purpose in computers. In Table 6.5 we summarize the characteristics of various storage devices discussed in this chapter.

Table 6.5 Comparative Characteristics of Storage Devices

Storage type	*Average capacity*	*Technology*	*Average time to access a byte*	*Applications*	*Relative cost per byte*
Cache memory	512 KB	Static RAM (Semiconductor flip-flops).	1 ns	Data and instructions immediately required is stored.	10
Main memory	512 MB	Dynamic RAM (Capacitors in integrated circuits).	10 ns	Program and data to be processed by program.	1
Read only memory	64 KB	Semiconductor integrated circuit.	10 ns to read byte. Writing done occasionally.	Storing tables, permanent small programs.	1/10
Addressable direct access memory (Hard disk)	60 GB	Magnetic recordable/ readable surface on hard disk.	6 ms	On-line secondary memory for data and program store.	1/100
Addressable direct access memory (Floppy disk)	2 MB, 200 MB (High density)	Magnetic recordable/ readable surface on flexible sheet.	100 ms	Archiving small amounts of data. Exchanging text files.	1/1000

(Contd.)

Table 6.5 Contd.

Storage type	*Average capacity*	*Technology*	*Average time to access a byte*	*Applications*	*Relative cost per byte*
CDROM	650 MB	Factory written laser disk.	100 ms	Store large texts, pictures and audio. Software distribution.	1/1000
DVDROM	8 GB	Factory written multilayer laser disk.	100 ms	Video files.	1/100000
CDROM-R CDROM-R/W	650 MB	Writable laser disk.	100 ms	Large text files. Graphics, audio files.	1/10000
Serial access memory (magnetic tape–non addressable) Half-Inch cartridge	400 GB compressed 200 GB native	Half inch width magnetic tape in a cartridge.	Serial access 30 MB/s	Archiving large files. Back up of on-line data on disk.	1/1000
Digital Audio Tape (DAT)	4 GB	4 mm width magnetic tape in a cartridge.	Serial access 360 KB/s	do	1/10000

SUMMARY

1. A storage system is an important part of a data processing system.
2. A storage system consists of a number of different units—each unit satisfying a specific purpose in data processing.
3. One of the most important storage units in a computer is its main memory which is used to store the program currently being executed and the data needed by this program.
4. As the main memory supplies instructions to be executed and the data for these instructions to the central processing unit, the speed at which these are delivered to CPU should match the CPU's speed. It should thus be fast.
5. As data and programs are stored in the same memory and programs can have jump statements, the time needed to retrieve instructions or data from the main memory should be independent of where they are stored. Such a memory is called a Random Access Memory (RAM).

6. A cache memory is a very fast but small memory used as a buffer between CPU and main memory. As CPU is much faster than main memory, often CPU will be waiting for data to be processed. This is alleviated by a cache memory which is a fast, small memory in which instruction/data required immediately is copied from the main memory and kept.
7. The cost of memory is proportional to its speed. The higher the speed, the greater is the cost of the memory. Thus the size of the main memory cannot be increased indiscriminately.
8. There are many applications requiring very large quantity of data to be processed. All this data cannot be kept in the main memory as the cost will be too high. We thus need a lower cost memory with much higher capacity which can be directly accessed and data in it retrieved during data processing. This is called a direct access secondary memory.
9. The most popular direct access secondary memory in a computer is a hard disk whose capacity is tens of GB and the access time a few milliseconds.
10. Another popular secondary memory is a floppy disk drive. It has a much lower capacity compared to a hard disk. The floppy disk medium is, however, a removable medium. The floppy disk can be mailed by removing it from the drive and mailing it.
11. We must clearly distinguish between a disk drive and the magnetic disk which is the medium where data is recorded. In a hard disk the medium is fixed to the drive whereas in a floppy drive the medium is removable.
12. It is necessary to back up data stored in an on-line disk store to insure against accidental loss of data.
13. Backing up of data is done in what is known as an archival storage medium. Floppy disk is appropriate for archiving small files, but not large ones.
14. Apart from backing up data, massive amounts of processed data have to be preserved often for legal reasons or for detailed analysis. These data are also stored in archival store.
15. Magnetic media known as Half-Inch Cartridge (HIC) Tape and Digital Audio Tape (DAT) are used as archival store. HIC can store 400 GB and DAT around 4 GB.
16. Another archival storage is Compact Disk Read Only Memory (CD ROM). In this data is stored as lands and pits in a polycarbonate layer by laser beams. The disk is also read by a laser beam. The writable version of CDROM is used as a backup store of processed data for archiving.
17. Pre-recorded CDROMs are used for distributing software, large text files and multimedia files (graphics and audio). They can store 650 MB.
18. DVDROM stores data on multilayers of polycarbonate and the data is read with a laser beam. Pre-recorded DVDROMs can store 8 GB and are used to distribute video files.

19. Storage units are classified as volatile and non-volatile. Volatile memories lose data stored in them when power is switched off. Main memory is volatile whereas disks, tapes, CDROM and DVDROM are non-volatile.

20. Storage unit are also classified as addressable and non-addressable. In an addressable store each data record is stored in a unique address and stored/retrieved by specifying the address. Main memory, disk (hard and floppy), CDROM, DVDROM are addressable. In a non-addressable store data is written sequentially and read in the order in which it is written. No address is used to store or read data. Tape storage is non-addressable.

21. If time to store/retrieve data is independent of the address where it is stored, such a memory is known as Random Access Memory (RAM). The main memory of a computer is a RAM. Disk, CDROM are directly addressed memory and cannot be randomly accessed.

22. Memories are made up of individual memory cells where each cell stores one bit. Different technologies used as storage cells are: a capacitance, a semiconductor logic unit called flip-flop, surface recording on a magnetic surface, and pits/lands on a polycarbonate layer.

23. Main memories use semiconductor cells. Disks and tapes use magnetic surface recording, and CDROMs and DVDROMs pits/lands on polycarbonate layer.

24. Read Only Memory (ROM) is a non-volatile semiconductor memory used to store small programs to control instruments/appliances and for storing useful tables. There are four varieties of ROM: factory programmed, user programmed, ultra violet erasable electrically reprogrammable, and electrically erasable programmable.

25. A Variation of Electrically Erasable Programmable ROM is called a flash memory. It is non-volatile and can store upto 512 MB at reasonable cost. It is used to store images in digital cameras and are replacing floppy disks in PCs.

EXERCISES

6.1 What is the purpose of memory in a computer?

6.2 What are the main characteristics of a memory cell?

6.3 What is a non-volatile memory?

6.4 What is destructive reading of a memory cell?

6.5 Explain how a capacitor can be used as a storage cell. Is readout from capacitor cell destructive?

6.6 Why do you have to refresh a memory which uses capacitor as storage cell?

6.7 Explain how a flip-flop memory cell is used. How is it different from a capacitor memory cell? Is reading from this cell destructive?

19. Storage units are classified as volatile and non-volatile. Volatile memories lose data stored in them when power is switched off. Main memory is volatile whereas disks, tapes, CDROM and DVDROM are non-volatile.

20. Storage unit are also classified as addressable and non-addressable. In an addressable store each data record is stored in a unique address and stored/retrieved by specifying the address. Main memory, disk (hard and floppy), CDROM, DVDROM are addressable. In a non-addressable store data is written sequentially and read in the order in which it is written. No address is used to store or read data. Tape storage is non-addressable.

21. If time to store/retrieve data is independent of the address where it is stored, such a memory is known as Random Access Memory (RAM). The main memory of a computer is a RAM. Disk, CDROM are directly addressed memory and cannot be randomly accessed.

22. Memories are made up of individual memory cells where each cell stores one bit. Different technologies used as storage cells are: a capacitance, a semiconductor logic unit called flip-flop, surface recording on a magnetic surface, and pits/lands on a polycarbonate layer.

23. Main memories use semiconductor cells. Disks and tapes use magnetic surface recording, and CDROMs and DVDROMs pits/lands on polycarbonate layer.

24. Read Only Memory (ROM) is a non-volatile semiconductor memory used to store small programs to control instruments/appliances and for storing useful tables. There are four varieties of ROM: factory programmed, user programmed, ultra violet erasable electrically reprogrammable, and electrically erasable programmable.

25. A Variation of Electrically Erasable Programmable ROM is called a flash memory. It is non-volatile and can store upto 512 MB at reasonable cost. It is used to store images in digital cameras and are replacing floppy disks in PCs.

EXERCISES

6.1 What is the purpose of memory in a computer?

6.2 What are the main characteristics of a memory cell?

6.3 What is a non-volatile memory?

6.4 What is destructive reading of a memory cell?

6.5 Explain how a capacitor can be used as a storage cell. Is readout from capacitor cell destructive?

6.6 Why do you have to refresh a memory which uses capacitor as storage cell?

6.7 Explain how a flip-flop memory cell is used. How is it different from a capacitor memory cell? Is reading from this cell destructive?

6.8 Explain how bits are stored and read from a magnetic surface. Is readout from magnetic surface destructive? Is this memory volatile?

6.9 How are bits stored on a laser disk?

6.10 What do you understand by a Random Access Memory?

6.11 Draw a block diagram of a 16 KB, byte addressable RAM.

6.12 What are the sizes of MAR and MDR of 1 MB word addressable memory where each word is 32 bits long?

6.13 Define the terms access time, cycle time and write time of a RAM. Is cycle time larger or smaller than access time?

6.14 What do you understand by DRAM? What type of memory cells are used in DRAM?

6.15 What do you understand by SRAM? What type of memory cells are used in SRAM? Is SRAM faster or slower than DRAM?

6.16 Do main memories of computers use SRAM or DRAM?

6.17 Where are SRAMs used and why?

6.18 What is a ROM? Is ROM a random access memory?

6.19 What are the main applications of ROM?

6.20 A seven segment display is to be used to display octal digits (0 to 7). Design a ROM for it.

6.21 List various types of ROM. What is an EEPROM?

6.22 What is a flash memory? What are its applications?

6.23 Explain how data is organized on a hard disk. A disk has 16 surfaces, 6676 cylinders and 1024 sectors per track and 512 bytes/sector. What is the capacity of the disk? How do you address a sector in this disk?

6.24 What do you understand by seek and latency times of hard drives? What are their typical values in currently available disks?

6.25 If a disk rotates at 10000 rpm, what is its maximum latency delay?

6.26 What are the functions of a disk interface? List the names of four standard interfaces. Which is a popular interface used with PCs?

6.27 What do you understand by aerodynamically floating head? Why should rooms where disks operate be kept dust free?

6.28 What is the capacity of a typical floppy disk? What are the applications of a floppy disk? Are floppy disks used for software distribution nowadays?

6.29 What is the capacity of a typical CDROM today? What are its applications?

6.30 What do you understand by the term 52x speed CDROM? What is the standard used to store data on a CDROM?

To see why you need a CPU in a computer and what it is expected to do we will review the problem of how a computer is used in a railway reservation system. To reserve tickets by a train, e.g. (Secunderabad Express) to go from, say, Bangalore to Secunderabad you follow the steps given below:

1. A form is filled by you with details of tickets to be purchased by you. A sample form is given in Table 7.1.
2. The form is given to the booking clerk at the reservation counter.
3. A blank form is displayed on the Video Display Unit of the terminal in front of the clerk.
4. The clerk types the data given in your form using the keyboard of the terminal.
5. After typing the data the clerk presses an enter key on the keyboard. This signals the end of data entry to the Central Processing Unit (CPU) to which the keyboard is connected.
6. A sequence of instructions to process the data and to reserve accommodation is stored in the main memory of the computer. We call this the berth reservation program.
7. When the end of data entry signal is received by the CPU, it seeks the assistance of the berth reservation program.
8. The CPU now retrieves the first instruction from the memory, interprets it and executes it. In this example the instruction reads the train number, date of travel and berths requested.
9. Using the data on train number and date of travel, the CPU retrieves from the disk memory the relevant reservation data.
10. The next instruction is now retrieved from the main memory and placed in CPU.
11. The CPU interprets this instruction which is to compare the request for reservation with the data on available berths. The comparison is done.
12. If the result of the comparison says that requested berths are available. The CPU retrieves an instruction which commands the printer to print the ticket. Else it branches to another command which displays that berths are not available.

The above sequence of operations have been greatly simplified to illustrate the main point that a CPU is essential to retrieve instructions from memory, interpret them and carry them out.

The description of the problem given above, namely, reserving a ticket in a train is an example of what is known as *general purpose computing*. The desk top PC and larger computers used by big organizations are examples of general purpose computers. The CPUs used in such computers are called general purpose processors and are the ones which are most familiar to us as we see advertisements for such processors in newspapers, TV, etc.

Besides CPUs used in general purpose computers, a much larger number of CPUs are used in systems we use every day, such as washing machines, microwave ovens, cell phones, VCD players, and motor cars. Machines using such CPUs are called "intelligent" in the newspaper advertisement as they are programmable and thus more flexible to use. In fact, the applications of CPUs for such purposes is increasing day-by-day. Such CPUs are known as *embedded processors* as they are built into products such as washing machines, motor cars, etc.

The structure of CPU is similar for both general purpose computers and specialized CPUs of embedded processors. Their main differences are the richness of instructions they can carry out, their speed, and the way in which they use memory.

All CPUs nowadays are microprocessors, popularly known as microchips. They are called microprocessors as the entire processor is fabricated on one silicon chip of size around 5 mm × 5 mm. Each chip has millions of switches which are wired together to carry out various operations on strings of bits. Over the years, microprocessors have become more and more powerful (doubling their performance every eighteen months), while the cost remains the same or lower.

When you see advertisements of computers they use terms such as "1.5 GHz Intel 32 bit CISC processor with 64 MB SDRAM, 80 GB disk, 1.22 MB floppy drive, Intel mother board, 2 parallel and 3 USB ports, PCI bus, ..., etc. You will be wondering what these terms mean. By the end of this chapter you will know what the terms 1.5 GHz, 32 bit CISC processor, motherboard, USB ports, etc., mean.

7.2 THE STRUCTURE OF A CENTRAL PROCESSING UNIT

A block diagram of a computer is given in Fig. 7.1. It is similar to the figure in Chapter 1 but with some more details added. The main purpose of a CPU is to retrieve instructions from memory, interpret them, and carry out the operations specified in the instruction. In order to perform the operation, it will need cooperation from the other units of a computer, namely, input unit, output unit, main memory and secondary memory. Thus CPU has the ability to send control signals to all these units (shown as dotted lines in Fig. 7.1) and "order" them to carry out the operations required by the programs.

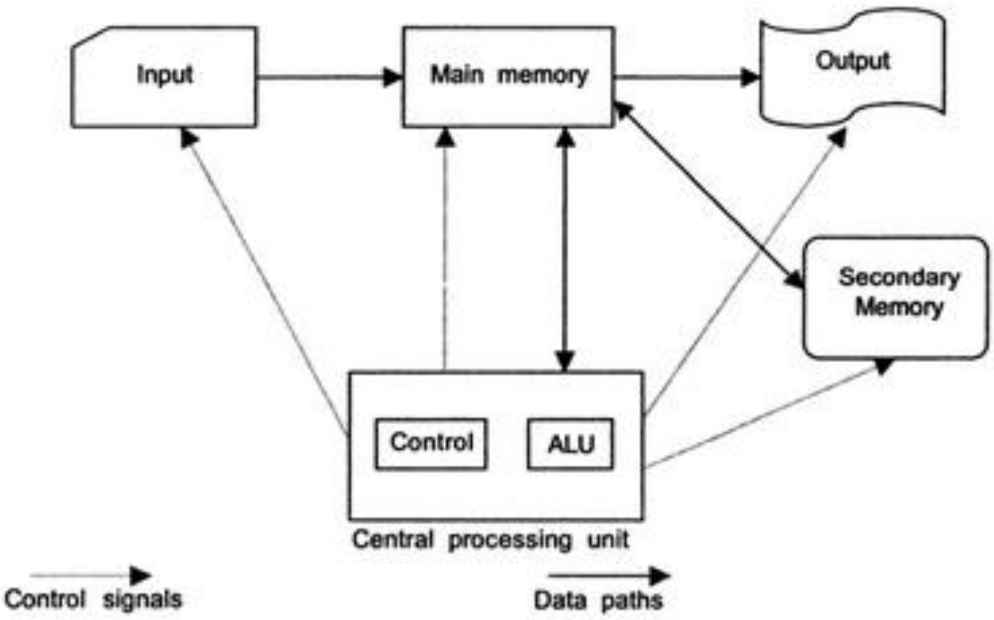


FIG. 7.1 Organization of CPU and other units of a computer.

The CPU is designed to interpret and carry out a number of operations, called its instruction set. The instruction set consists of a variety of operations required to solve problems. A representative set of instructions is given in Table 7.2.

Table 7.2 Representative Instruction Set Interpreted by a CPU

Type of operation	*Examples*	*Mnemonics*
Arithmetic	Add, Subtract, Multiply, Divide, Increment, Decrement	ADD, SUB, MUL, DIV, INC, DEC
Logical	Compare, Bit operations (AND, OR, NOT)	COM, AND, OR, NOT
Control	Jump, Jump if negative, Jump if equal, Stop	JMP, JNE, JEQ, STOP
Memory	Load register from memory, Store register in memory.	LOAD, STO
Input/Output	Read, Display, Print	READ, DIS, PRT

Normally an instruction should specify the following:

1. The operation to be performed.
2. The location in which the operands are stored in the main memory. The place in main memory where the operand is stored is called the *address of the operand*. The number of addresses specified in an instruction depends on the type of CPU. Some processors specify only one address, whereas others specify two or three operand addresses.
3. From which address in main memory the next instruction is to be taken. Normally, instructions are stored in successive addresses and instructions are executed sequentially. The normal flow is changed in some programs based on some inputs or results of previous operations.

The CPU hardware is designed to interpret operation codes which are strings of bits. Addresses are also specified using binary numbers. If there are 64 operations we will need 6 bits to code them (remember $2^6 = 64$). If the memory has a capacity of 16 MB, we will need 24 bits to specify an address in memory (as $2^{20} = 1$ M and $2^4 = 16$). An instruction (specifying only one address) will be 6 + 24 = 30 bits long. For example if an addition operation is coded as 000011 and the address of the operand to be added is 1011 1111 0010 1101 0111 0110, the instruction will be:

000011 1011 1111 0010 1101 0111 0110

This is called a *machine language* instruction. It is obvious that it is very difficult for people to correctly write such a sequence of bits. Thus instructions are usually written using meaningful abbreviations to represent operation codes (known as mnemonics) and strings of characters to represent operand addresses. We will describe the working of CPU using this idea. We will start with a very simple example of adding two numbers. A program (using mnemonics of Table 7.2 for operation codes and letters to represent addresses in memory) to add two numbers stored in addresses A and B is given as Table 7.3. These instructions are stored in memory.

Table 7.3 A Program to Add Two Numbers

Address where an instruction is stored in memory	*An instruction*		*Explanation*
	Operation code	*Address of operand*	
01	READ	A	Read number from input and store in address A of memory
02	READ	B	Read number from input and store in address B of memory
03	LOAD	A	Take number from address A and put it in CPU register
04	ADD	B	Add to CPU Register number stored in address B of memory
05	STO	C	Take the sum from CPU register and store in C of memory
06	DIS	C	Display on VDU the number stored in C
07	STOP		Stop execution

The addresses where these instructions are stored are also shown in Table 7.3. In Fig. 7.2 we show how these instructions are placed in memory. We will use this to explain how this program is executed by the CPU. The sequence of steps followed by the CPU are:

1. Read and store the program in the main memory of the computer. One instruction is stored in each address of memory.
2. Place the address of first instruction in an instruction counter in CPU.
3. Retrieve the instruction from the main memory. The address from where it should be retrieved is specified in the instruction counter (here it is 01). Place the instruction in a register called Instruction Register of CPU (A register is a temporary store in which data can be placed. When new data is placed in a register, the earlier contents are automatically erased and new value is stored.
4. Add 1 to instruction counter so that it now has the address of the next instruction in memory.

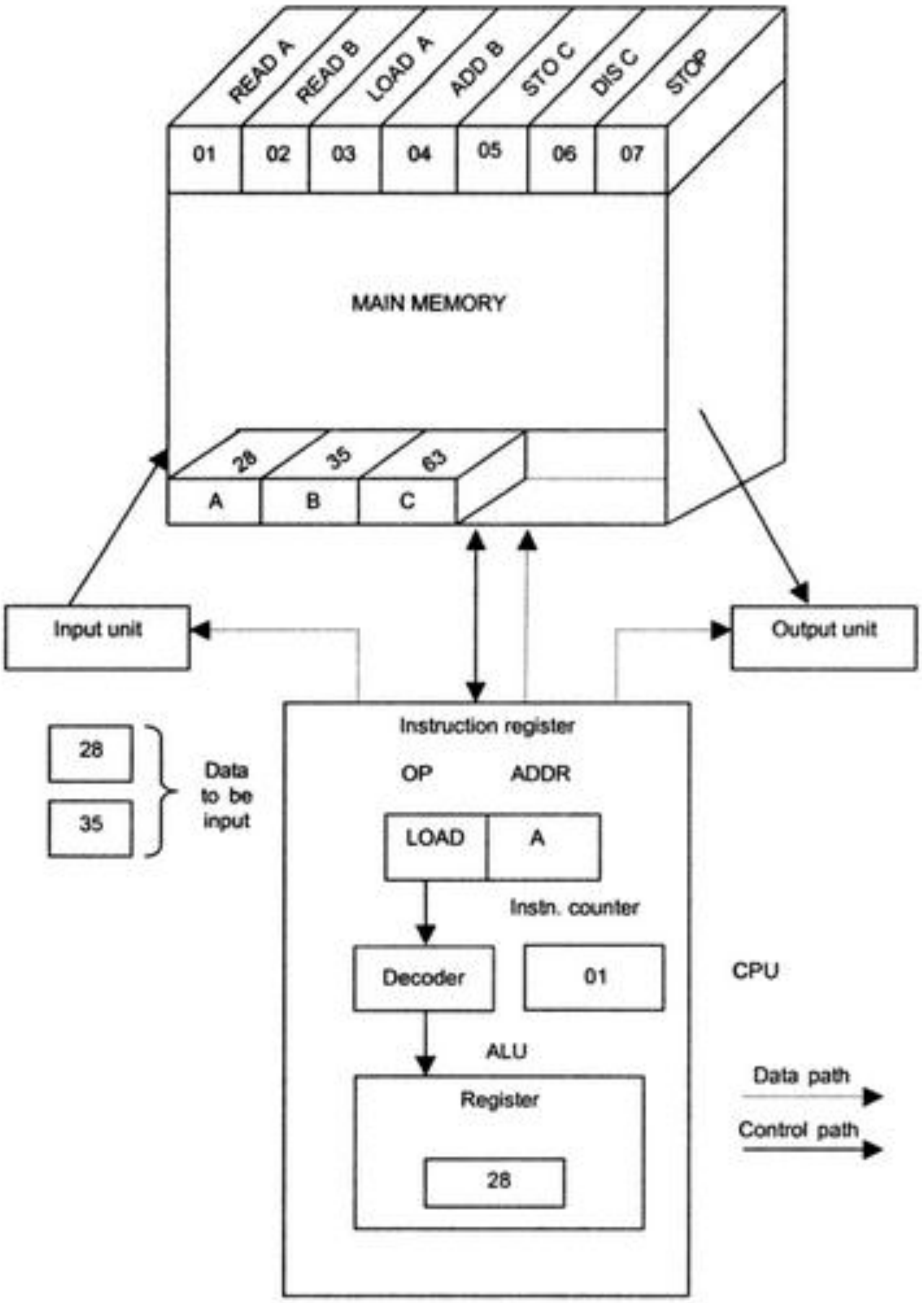


FIG. 7.2 Instruction storage in main memory and logical units in CPU for executing instructions.

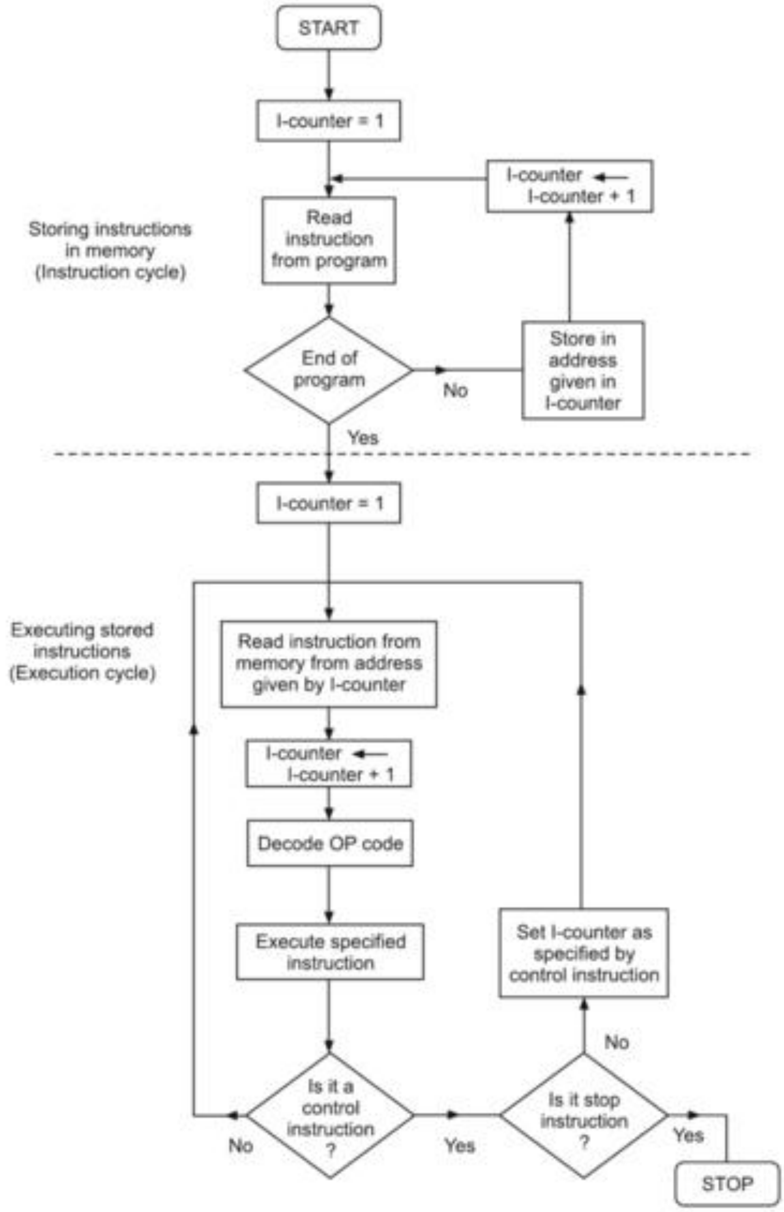


FIG. 7.3 Details of execution of a program by CPU.

Table 7.4 A Program to Find Larger of Two Numbers

Address where an instruction is stored in memory	*An instruction*		*Explanation*
	Operation code	*Address of operand*	
01	READ	A	Read first number and store it in A.
02	READ	B	Read second number and store it in B.
03	LOAD	A	Take contents of A and place it in ALU register.
04	SUB	B	Subtract from the ALU register number stored in B.
05	JNE	08	If the contents of ALU register are less than 0, take next instruction from address 08. Else execute next instruction.
06	DIS	A	If the first number is larger, this is the instruction executed next. It displays A which is a larger number.
07	STOP		Job over, stop.
08	DIS	B	If the second number is larger, this instruction is executed and B is displayed.
09	STOP		Stop computation.

2. The number of bits in an instruction is called its ***word length***. The word length of CPUs has been steadily increasing ever since microprocessors (i.e., CPU on a single integrated circuit chip) were made by a company called Intel. Intel began making a processor named 4004 in 1971. It was followed by 8008 and 8080 in 1972 and 1974 which was first used in a PC manufactured by International Business Machines (IBM). The 4004 had a 4-bit word and 8008 had an 8-bit word. Progressively the word length increased with Intel 8086, 80286 (16 bits), 80386 and 80486 (32 bits) and Pentium Series (32 bits). In Table 7.5 we give a comparison of these processors. This series of Intel processors starting with 8080 are said to belong to a "family of processors" as machine language programs written for 8080 can run without any change in the latest processor. These processors are also said to be ***upward compatible***. This compatibility is essential to allow time and money spent to design software for earlier models of a series to be executed on the newer models. Upward compatibility implies that the machine language can be run as is with no modifications. Thus all the machine instructions of earlier models have to be there in the new model. Newer models are faster. Thus they will execute old programs faster. Newer models will have

Table 7.5 Evolution of Intel Family of Microprocessors

Microprocessor	*Year of manufacturer*	*Clock speed MHz*	*Number of transistors in chip*	*Word length*	*Memory size (Maximum)*
4004	1971	0.108	2300	4	640 bytes
8080	1974	2	6000	8	64 KB
8086	1978	10	29000	16	1 MB
80286	1982	12	134000	16	16 MB
80386	1985	16	275000	32	4 GB
80486	1989	25	1.2 Million	32	4 GB
Pentium	1993	66	3 Million	32	4 GB
Pentium II	1997	300	7.5 Million	32	64 GB
Pentium IV	2000	1500	100 Million	32	64 GB

several new instructions which are powerful but will not be used by the older programs. This in turn implies that programs written for the newer models will not run on older models. That is the reason why old machines have to be "retired". Useful and powerful programs written and available on newer models cannot be used if you have an older model processor.

3. The architecture i.e., the internal structure of a processor. There are three broad classes of CPU architecture. One of them is RISC (Reduced Instruction Set Computer). As the name implies, it has a very small number of instructions. Its control unit is simple. This processor is used in many high performance computers. The second one is called CISC (Complex Instruction Set Computers). This type of processor is used in PCs. Most of the processors designed by Intel which are used in PCs (e.g., the latest processor is called Pentium 4) are CISCs. They have over 300 machine instructions and a complex control unit. The structure of an instruction is also complex. Some instructions specify three operand addresses, some only one address and some no address, making the length of instructions variable unlike RISC processors where the structure of the instruction is simple. The third class of CPU emerged in 2000. It is called VLIW (Very Long Instruction Word) Processor. In this type of processor, about three instructions which can be executed simultaneously are packed as one composite instruction. The number of bits used per instruction is around 128 bits. Contrast this with the length of instructions used in CISC and RISC, which is currently 32 bits. It is expected that VLIW processor will be used in the next generation of PCs.
4. Another important parameter used to describe a CPU is its clock speed or frequency. The clock synchronizes all the operations carried out by a CPU. It is actually a periodic sequence of pulses and the unit used is hertz (Hz). For example, the

clock frequency of Pentium 4 CPU is 1.7 GHz. The clock speed is an important parameter as it is an indicator of the rate at which a CPU can execute instructions. Even though higher clock rates give higher processing speeds, this is not always true. It is true only for a given microprocessor family. For example, Intel 80486 operated with 25 MHz clock whereas Pentium II had a clock frequency of 300 MHz. One may thus say that Pentium II is almost 12 times faster than 80486. However, a RISC processor manufactured by another company (for example, Sun Microsystems) with a clock frequency of 300 MHz may not be much slower than a Pentium IV operating at 1.5 GHz as a RISC may execute several instructions in one clock cycle.

5. Another parameter used to describe the speed of CPU is called its *mips* (million instructions per second) rating. Normally, it is specified as hundreds of mips. It is not a very reliable measure as the time taken by different instructions is different and the mips rating of RISC is normally higher than CISC processors using the same clock frequency. Thus another measure called SPEC rating is used to compare speeds of CPUs. The expansion of SPEC is System Performance Evaluation Consortium. It is a consortium of computer manufacturers who selected a set of common application programs (called a benchmark set) which are run on a computer to assess its speed. Higher SPEC marks imply that it is a faster computer. The advantage of SPEC marks is that the performance measure is based on the time taken to run a variety of complete programs written in a high level language such as C and thus realistically reflects the performance of not only the CPU but also the language used and the operating system.

7.4 INTERCONNECTION OF CPU WITH MEMORY AND I/O UNITS

We have already seen that instructions to be executed by a CPU are retrieved from the main memory interpreted by it and executed. CPU coordinates the execution of instructions. The CPU is connected to the main memory by a set of parallel wires called a *bus*. Bits are transferred from memory to CPU and back via a bus. There are two buses which connect CPU to memory. One is the *address bus* which carries the address bits to the MAR of memory from the CPU and the other is called the *data bus* which carries the data/instructions from CPU to the MDR of memory. Besides these two buses, there is also one more bus called *control bus* which controls a large number of operations such as Read/Write from or to memory and also input/output operations. The number of parallel wires constituting a bus is called *bus width*. If MAR has 24 bits (16 MB memory), then the address bus width is 24. Normally we retrieve an instruction from memory and place it in CPU in one memory read cycle. Thus the size of the data bus from the memory to CPU equals the number of bits in an instruction also called CPU word length. Most current CPUs, for example, Intel processors used in PCs, have a word length of 32 bits and hence a data bus width of 32. The instruction width of the new Very Long Instruction Word (VLIW) processor (called *IA-64* by Intel) is 128 bits and therefore the data bus width is 128 bits. The number of bits in the control bus depends on the signals required to coordinate input/output units as well as the main memory. It is normally around 16. The interconnection of CPU, memory and I/O units with buses is shown in Fig. 7.4.

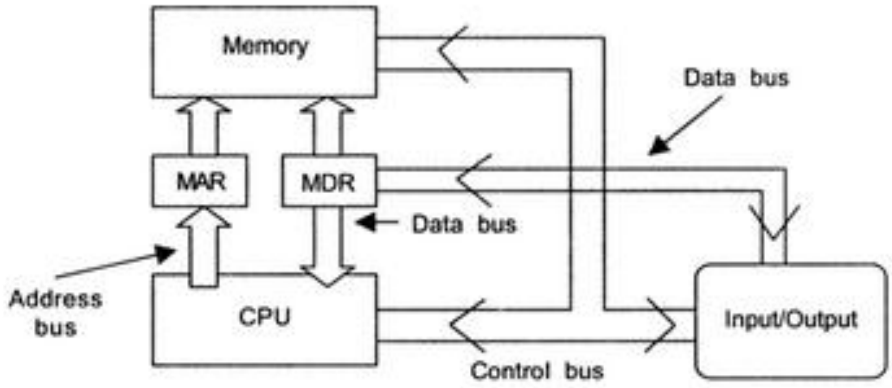


FIG. 7.4 Configuration of buses connecting CPU with main memory and input/output units.

The three buses, namely, data bus, address bus, and control bus emanate from the CPU chip and are used to connect the CPU to both memory and I/O units. This collection of buses is called a *system bus*. The structure of these buses, i.e., the number of wires, speed of transmission of data and length are normally standardized by an industry group. This standardization is essential to allow diverse manufacturers of various peripheral devices such as CDROM, VDU, keyboard to design devices which can be easily connected to a PC. This is a great boon to consumers as competitors can build peripherals at reasonable cost. Further, many new device manufacturers, such as say a digital video camera manufacturer can design and incorporate an interface in his device so that it can be plugged to a standard PC and data recorded in it downloaded to PC for editing, storage, etc.

The bus standards also evolve with time as PCs become more powerful. An early standard was ISA (Industry Standard Architecture) for 8 bit PCs of the first generation. It evolved into EISA (Extended ISA) which is a faster 32 bit bus. Another standard proposed by Intel is PCI (Peripheral Component Interconnect) bus which allows 32 and 64 bit implementations.. It can be run at 33 MHz and one can connect to 16 peripheral devices. Advertisements for PCs mention the standard used by the particular PC manufacturer.

The CPU, Memory and integrated circuits necessary to connect I/O units to the CPU and main memory are all mounted on what is called a *motherboard* (Fig. 7.5). Motherboards are assembled and sold by reputed vendors such as Intel. There are also many competitors who make motherboards which are bought by vendors who assemble PCs. The motherboard also has a ROM where a program called BIOS (Basic Input Output System) is stored to control all the peripheral devices connected to a computer.

A motherboard has a set of connection points called ports to connect units such as disk, VDU, keyboard, etc. Besides that, it has a set of parallel wires (bus) connected to a pluggable connector to attach I/O units such as floppy, CDROM, etc. The motherboard is mounted in a case and constitutes a "PC box". Devices such as VDU, keyboard and mouse are outside this box. At the back of PC box or cabinet, sockets called *ports* are placed to connect peripherals as shown in Fig. 7.6.

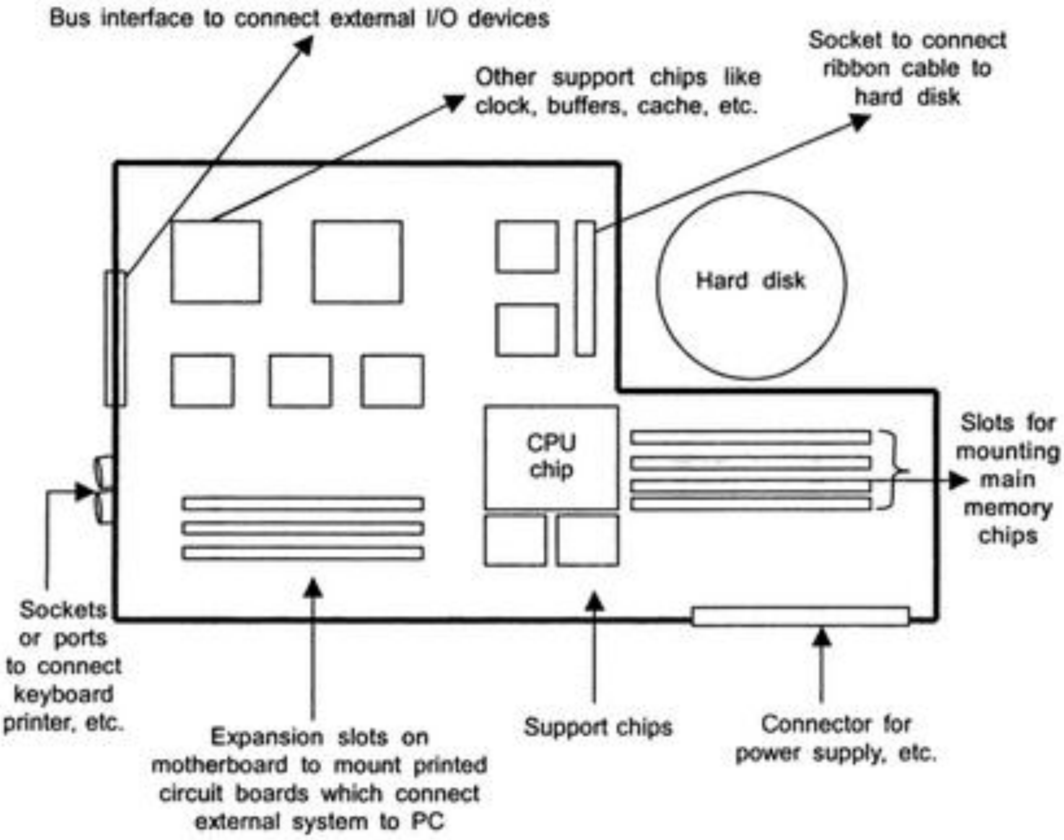


FIG. 7.5 Sketch of motherboard of a PC showing important components.

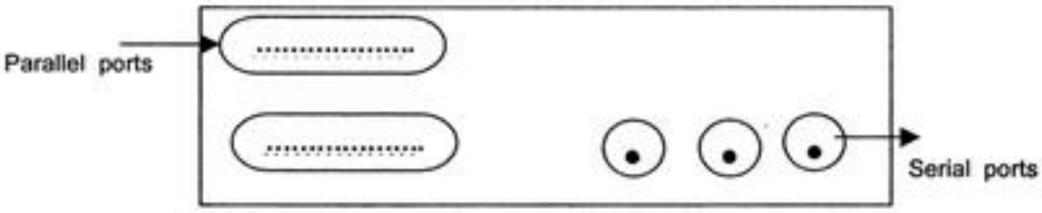


FIG. 7.6 Parallel and serial ports at the back of PC cabinet.

In a *parallel port* data on parallel lines coming from the motherboard are connected. They are transmitted to peripherals via a set of parallel wires (called ribbon cables). As all bits travel in parallel, these ports carry 16 to 32 bits simultaneously. They are thus faster compared to *serial ports* which transmit single bits serially, that is, one after another. Faster peripherals such as hard disk are connected to parallel ports. Serial ports which carry one bit at a time from the motherboard are connected to slower devices such as keyboard, printer, etc. A standard serial bus is known as *Universal Serial Bus* (USB)

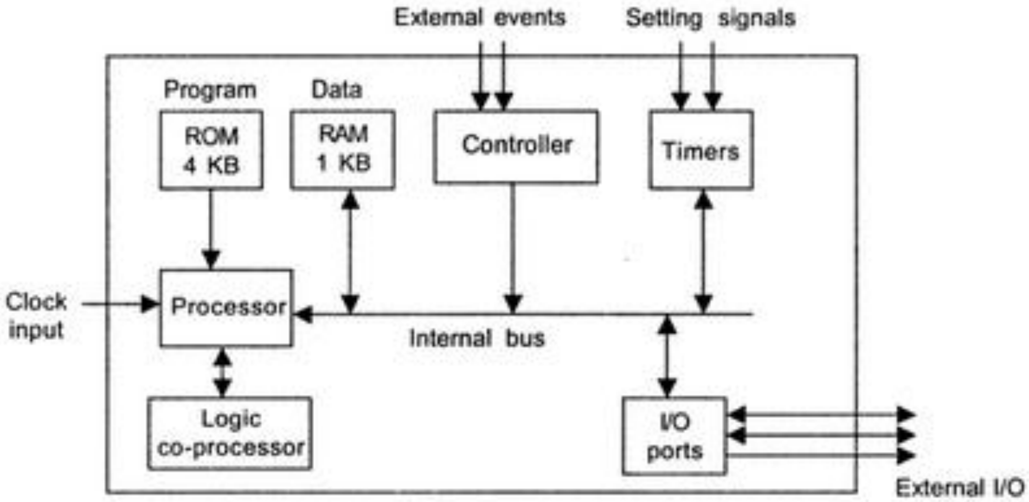


FIG. 7.8 Internal structure of a single chip microcontroller.

Currently the most popular microcontroller is Intel's model 8051. The cheapest model has 4 KB ROM for programs and 128 bytes of RAM. The program is factory programmed for each specified application. Before burning the ROM, a prototype with EEPROM is used to test the system. Versions of microcontrollers which consume very little power are now available for use in portable systems.

Development systems with necessary software are available for various models of microcontrollers. Development systems are used to create bug free application programs before microcontrollers are integrated into systems such as washing machines.

7.5.2 Digital Signal Processors

Digital Signal Processors (DSP) are optimized to process *real-time* audio and video signals. By real-time we mean the time we measure with our watches. For example, cricket match shown on TV "live" is real-time whereas the replay shown later on is not. A real-time processor accepts input signals while it is being generated, for example, a telephone conversation. Digital signal processors are widely used in cost-sensitive consumer products such as hi-fi audio systems, TVs, etc., and thus their cost must be much lower than that of general purpose processors such as Pentium. Whereas general purpose processors cost around Rs. 5000, DSPs cost around Rs. 500. In Fig. 7.9 we show how a DSP is connected to external input/output. The input signal to DSP is analog, that is, it is a continuously varying signal such as speech, music, video or electrocardiogram. The *analog to digital* (A/D) *converter* samples the analog signal at appropriate intervals depending on the bandwidth of the signal as explained in Chapter 4 and converts it to a sequence of binary numbers. This sequence of bits is input to DSP. Observe that the input is fed to DSP in real-time. The DSP processes the input using a program stored in its memory and a sequence of bits stream out as output in real-time. It is converted back to a continuous waveform by a *Digital to Analog* (D/A) *converter*. The A/D and

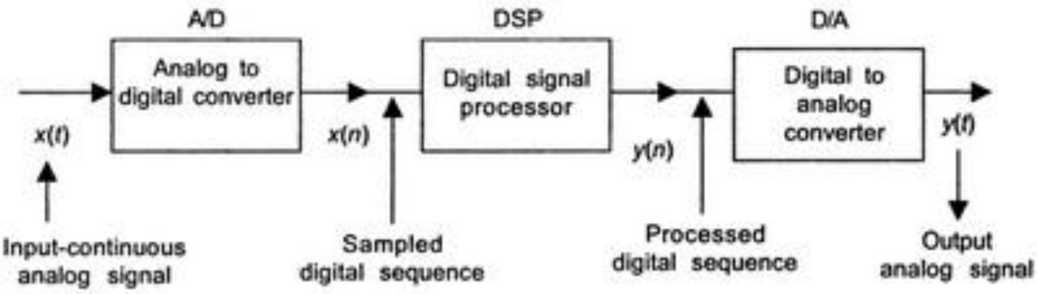


FIG. 7.9 Analog input/output to DSP.

D/A converters are integrated circuit chips. These are often part of the DSP itself, that is, it forms a part of DSP chip and are not external to it.

We saw in Chapter 4 that if a DSP is to process audio signals with up to 20 kHz frequency, we saw in we should sample the signal at intervals less than $(1/40 \times 10^3)$ second = 25 microseconds. Thus we can take a sample every 24 microseconds. If each sample is 16 bits, then 16 bit samples appear every 24 microseconds at the input of the DSP. This is equivalent to $(16/24 \times 10^{-6})$ = 666.6 Kbits per second. The DSP must be able to process at least 666.6 Kbps for this type of audio signals.

Most DSP applications for audio signal processing perform operations using formula of the type

$$y(n) = x(1)h(n-1) + x(2)h(n-2) + x(3)h(n-3) + \dots + x(m)h(n-m)$$

where $y(n)$ is the nth sample output, $x(1)$, $x(2)$, $x(3)$, ..., etc., are input samples of $x(t)$ (a function of time) at equal intervals, and $h(n-1)$, $h(n-2)$, ..., etc., are called weights whose values are adjusted based on the type of processing. Observe that we need to multiply and add. A unit called *Multiply Accumulate Unit* (MAC) is thus an integral part of DSPs. The arithmetic is to be done while the input data is streaming in and no input sample should be missed. Thus DSPs have fast arithmetic unit with two ports. One port continuously reads while the other port simultaneously outputs the processed stream of bits. DSPs normally have two different memories on chip—one for data and the other for program. There are two independent paths from the two memories to CPU so that an instruction can be fetched from the instruction memory while a result is being written in the data memory. To summarize the most important units of a DSP are:

- An A/D converter
- A D/A converter
- A Multiply Accumulate Unit
- A fast ALU which handles real-time arithmetic
- An on-chip data memory
- An on-chip program memory

A block diagram giving the internal structure of a DSP is given in Fig. 7.10.

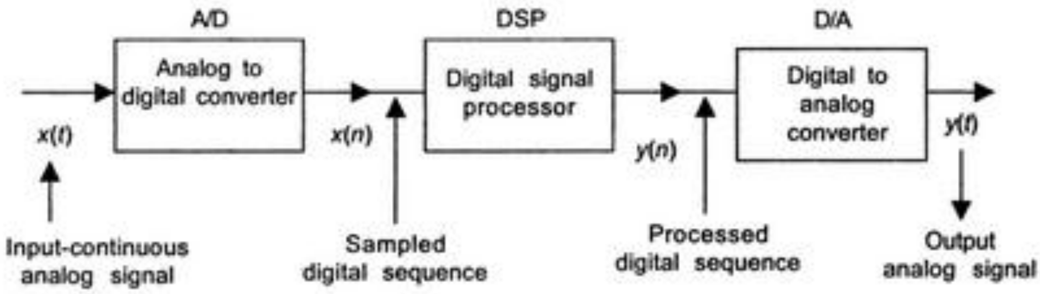


FIG. 7.9 Analog input/output to DSP.

D/A converters are integrated circuit chips. These are often part of the DSP itself, that is, it forms a part of DSP chip and are not external to it.

We saw in Chapter 4 that if a DSP is to process audio signals with up to 20 kHz frequency, we saw in we should sample the signal at intervals less than $(1/40 \times 10^3)$ second = 25 microseconds. Thus we can take a sample every 24 microseconds. If each sample is 16 bits, then 16 bit samples appear every 24 microseconds at the input of the DSP. This is equivalent to $(16/24 \times 10^{-6})$ = 666.6 Kbits per second. The DSP must be able to process at least 666.6 Kbps for this type of audio signals.

Most DSP applications for audio signal processing perform operations using formula of the type

$$y(n) = x(1)h(n-1) + x(2)h(n-2) + x(3)h(n-3) + \ldots + x(m)h(n-m)$$

where $y(n)$ is the nth sample output, $x(1)$, $x(2)$, $x(3)$, ..., etc., are input samples of $x(t)$ (a function of time) at equal intervals, and $h(n-1)$, $h(n-2)$, ..., etc., are called weights whose values are adjusted based on the type of processing. Observe that we need to multiply and add. A unit called *Multiply Accumulate Unit* (MAC) is thus an integral part of DSPs. The arithmetic is to be done while the input data is streaming in and no input sample should be missed. Thus DSPs have fast arithmetic unit with two ports. One port continuously reads while the other port simultaneously outputs the processed stream of bits. DSPs normally have two different memories on chip—one for data and the other for program. There are two independent paths from the two memories to CPU so that an instruction can be fetched from the instruction memory while a result is being written in the data memory. To summarize the most important units of a DSP are:

- An A/D converter
- A D/A converter
- A Multiply Accumulate Unit
- A fast ALU which handles real-time arithmetic
- An on-chip data memory
- An on-chip program memory

A block diagram giving the internal structure of a DSP is given in Fig. 7.10.

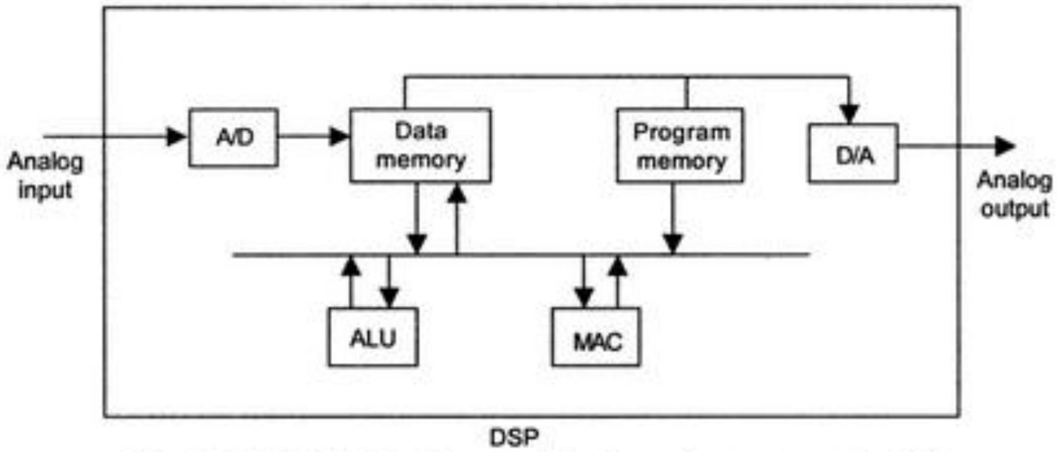


FIG. 7.10 Basic block diagram of the internal structure of a DSP.

Observe the differences between this and general purpose processor structure. Video signals have 1000 times higher bandwidth. Thus the processing rate must be 1000 times faster. With better integrated circuit technology, real-time processing of video signals with DSP is also becoming feasible.

7.6 CONCLUSIONS

In this chapter we have given an overview of how the central processing unit—which is the ''brain'' of a computer—works. The speed at which processors carry out instructions is continuously increasing—almost doubling every two years. Processors are also getting cheaper. This is due to continued improvements in integrated circuit technology. The number of switches which can be put in a single chip which was around 100 in 1960s is around a billion now; an increase of 10 million times in 40 years! The relentless increase in power of processors has led to many new uses of processors. Computers which were originally used to process numbers now process not only numbers but sound and motion pictures. These applications are now possible due to increase in speed of processors and also due to improvement in programming methods. The increase in speed of processors coupled with the lowering cost has led to many innovative applications of processors. In fact, some of the most powerful processors are used in computers designed for playing video games. These gaming computers have excellent audio-video capability and need to respond in real time. Other innovative applications of processors are in toys, robot dog, portable CD players and satellite radios. Processor capability is no more an issue if you have an innovative idea to use digital processing power.

SUMMARY

1. The Central Processing Unit (CPU) is the "brain" of a computer and processes data fed to it from the main memory.
2. The CPU consists of circuitry to perform arithmetic operations such as add, subtract, multiply and divide. It can also compare characters, numbers, etc., and perform other logical operations.
3. The rules for processing data are specified as a series of instructions. An instruction has two parts: a part which says what operation is to be performed, and a part which specifies the address(es) in the main memory where the operand(s) would be found.
4. The instructions which can be directly interpreted by the CPU hardware are known as machine language instructions.
5. The speed at which instructions can be processed by a CPU is an important parameter.
6. The speed of processing is determined by the frequency of the clock which drives all the circuits of the CPU, the width (number of bits) which can be processed in each clock cycle and the number of operations which can be carried out simultaneously in one clock cycle.
7. The processor speed is often quoted by vendors as mips (million instructions per second). This is not a very good measure as the amount of work done to perform different instructions may vary widely.
8. A more realistic measure of the speed of processors is what is known as its SPEC rating. SPEC is an abbreviation of System Performance Evaluation Consortium, which is a group of professionals and processor manufacturers. This rating is based on time taken to execute a set of what are known as *benchmark programs* on a processor. Higher speed processors have higher SPEC marks.
9. Processor manufacturers continue to increase the clock speed and the variety of instructions which a processor can execute. However, they maintain upward compatibility of processors in a processor family. This ensures machine language compatibility of processors in a processor family. Thus machine language programs can be executed without modification on all models of processors in a family.
10. There are two broad classes of processor architecture. They are called CISC (Complex Instruction Set Computer) and RISC (Reduced Instruction Set Computer). RISC processors have a simpler internal structure and are often used in faster, more expensive computers. Intel processors (the 8086, 8026, Pentium) are CISC processors.
11. The number of different instructions which can be interpreted by a CPU is called its instruction repertoire.

8.29 What is a router? What are its functions?

8.30 What are the functions of TCP? Is TCP essential in Internet? Why?

8.31 How does TCP recover lost packets? How does it sequence packets in the right order?

8.32 What is DNS? Why is it necessary?

8.33 Who assigns names to Internet addresses? Give a typical address and explain different parts in it.

8.34 How is DNS organized? Why is this organization chosen? How does it ensure flexibility in assigning email ids?

8.35 What is the current protocol used in Internet? What are its disadvantages? What are its advantages?

8.36 Compare IPv4 and IPv6 protocols. How does IPv6 ensure flexibility of Internet?

ЫЕ

CHAPTER

9 Output Devices

LEARNING GOALS

After reading this chapter, you should be able to:

1. List the types of devices used with computers to output processed data.
2. Choose output devices for various typing/printing applications.
3. Explain how an audio output unit works and its applications.

9.1 INTRODUCTION

Once data is processed, it is necessary to present the results in a form that can be used in an application. We saw that a computer processes all types of data—numbers, characters (text), images, audio and video. We thus need output units to display text and images and to present audio and video signals. There are two broad classes of devices for displaying numbers, text and images. One class displays the results of processing on a screen for viewing, and the other prints on paper the processed data. Printed output is known as a *hard copy*. If the output is recorded on a removable storage such as pen drive or CDROM, it is called a *soft copy*. Display devices use *pixels* (picture element) as the basic unit to represent text, pictures, etc. The area available for display can be imagined to be overlaid on a graph sheet. Each square in the graph sheet can be thought of as representing a pixel. The number of squares per square inch specifies the *resolution* of the display. Greater the resolution, the better is the picture quality.

Audio data processed by a computer results in a sequence of 8-, 12- or 16 bit numbers, depending on the amplitude resolution required in the application. These digital quantities are converted to continuously varying voltages by a device called a *digital to analog converter* (D/A converter). The continuous voltage waveform given by D/A converter is fed to a loudspeaker which emits sound waves.

Video output may be viewed on video display units (also called a *monitor)* or on television screens attached to a computer. We will describe optical output devices in the next section.

can be built to specifications and if there is a mass market, its cost will come down continuously.

You would have seen newspaper advertisements for computers which would specify all the hardware details such as CPU type, memory size, disk size, etc. Besides that, it may say "Microsoft Windows 7 operating system preloaded" or "Linux operating system", and you may be wondering what it is. It is software known as an *Operating System* which is essential to make the computer you buy a useful device. If there is no Operating System, the computer will not work.

You would also have seen advertisements by computer education companies which say "Become a computer expert—learn C++ and Visual Basic". C++ and Visual Basic are programming languages. Of course, there is more to becoming a computer expert than knowing two programming languages. As we pointed out at the beginning of this chapter, programming language is nothing but a precise notation to express algorithms. It is useful to know a language, but it is not essential knowledge to use a computer. Operating system and programming languages are part of what is known as *system software*. We will learn more about these later in this chapter. Besides these types of advertisements, you probably would have seen advertisements saying, "Buy Microsoft Office Home edition to make the PC in your home or office assist you in all your office work". Microsoft Office is what is known as *packaged software*. We will discuss this in Chapters 12 and 13.

You would have reserved accommodation either in a train or in a bus. The clerk uses a keyboard of a computer to enter your request, and the reservation status, ticket, etc. are displayed on the video screen. The entire process of reservation is written as a program which is stored in the computer's memory. Such programs written for specific applications are known as *application programs*.

In this section we saw that there are three broad classes of programs, namely,

- System software
- Packaged software
- Applications software

We will begin our discussions by examining system software which provides the environment to develop applications software. There are following three types of system software:

- Operating system
- Programming languages and their translators
- Utility programs

10.2 OPERATING SYSTEM

Bare computer hardware is similar to a TV without a remote controller. A computer needs an Operating System (OS) to control its operation and make it usable. An OS consists of two parts: one part is called the BIOS (Basic Input Output System) which is stored in a ROM. The other part which provides most of the services is stored on the hard disk. A user interacts with a computer using the OS as it provides many important facilities. An OS may be defined as a set of system programs that control and coordinate the operation of a computer system. Some of the major facilities provided by an operating system are:

1. Starting the operation of a computer when the power is first turned on.
2. Storing users' programs in memory and scheduling them for execution in an orderly fashion.
3. Invoking programming language translator programs when necessary.
4. Controlling input and output operations.
5. Managing the use of main memory.
6. Managing and manipulating (i.e., editing) of users' files.
7. Easy interaction between users and computers.
8. Providing security to users' jobs and files.
9. Keeping accounts of resource usage.

Thus, a user of a modern computer uses a machine whose hardware is hidden by layers of software with features which include, besides those provided by its processor (e.g., arithmetic and logic circuits, memory, etc.), a number of functions provided by the OS listed above (see Fig. 10.1).

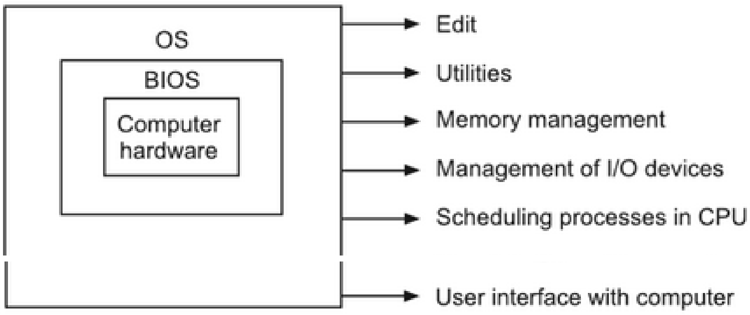


FIG. 10.1 Functions provided by an OS to a user.

Let us now begin our discussion by examining the functions of BIOS.

10.2.1 Basic Input-Output System (BIOS)

As already pointed out, BIOS is a small program to start and control a computer. It is stored in a ROM (non-erasable memory). When a computer is switched on, BIOS instructions are retrieved and start executing. It first tests the memory and displays on the screen the available memory. Now it prepares the computer to start interpreting and executing users' programs. This process is called *booting* of the system.

The major functions of BIOS are:

- Interpreting your keystrokes on the keyboard and storing data typed by you in the main memory.
- Controlling the display and the printer.
- Enabling input and output of data via other ports.

One of the advantages of BIOS is that when new I/O devices come in the market, they can be easily made to work with an existing computer by adding new programs known as *device driver programs* to the BIOS. Device drivers control the operation of I/O devices.

Software, which is permanently stored in a ROM and cannot be modified by a user, is called *firmware* in the literature. Thus BIOS is a firmware.

10.2.2 Functions of an Operating System

An OS is a program which is permanently stored in a part of the main memory and is protected from accidental or intentional intrusion by users' programs. The non-changing part of OS is called a *kernel*. The main function of an OS is to control and optimally use the resources of a computer system. It is similar to the functions of the captain of a cricket team. The captain decides who will bowl and when, who will bat and the batting order and also controls the placement of fielders. He tries to optimally use the resources under his command such as bowlers, fielders, batsmen, etc. Similarly, an OS allocates and controls the use of the resources of a computer, namely, the main memory, the CPU, and the I/O units. It allocates memory to various programs being executed in a computer, monitors their progress and schedules the use of CPU to these programs.

Besides resource management, the other major functions of an OS are:

1. *Command processing*, that is, interpreting users' commands and executing these commands. Typical commands are open a file, copy a file, and close a file.
2. *Detection of errors*, if any, in the functioning of hardware or system software and gracefully shutting down the system if it is a fatal error. If the error is not serious, it will report it, and if possible, correct the error and continue processing.
3. *Managing power consumption*. For example, in a PC if there is no activity on the keyboard/mouse for a specified time, the CPU is taken to a "*sleep mode*", shutting down many of its functions.
4. *Managing files*. For example, users of Microsoft's OS place related files in what is known as a folder. The file itself is given a name and there is an extension stating the type of file. For example, the identifier of a file is biodata.doc. The *file name* is *biodata*, and the *extension* is doc which specifies that it is a document prepared using a word processor. Another example is myphoto.bmp specifying that the file myphoto is a bit mapped file. Various storage devices connected to the PC are given single letter codes to indicate to the Microsoft OS where the file is stored. For example, in Microsoft OS, A and B may indicate pen drive, C hard disk, D CDROM and E may represent a tape drive. Files are organized in folders. For example, the file myphoto.bmp may be stored in hard disk in a folder named photos. The path to the file is specified by the notation C:\photos\myphoto.bmp. The OS manages directories of files and retrieves specified files when the path is specified. Another function is keeping a "ghost" copy of a file after it is deleted by a user. This file is kept in what is known as *recycle bin* so that a user may get it back if he/she has accidentally erased it.

9. The major function of an OS is to manage and optimally use all the hardware resources to efficiently execute users' programs.
10. Besides resource management, the other major functions of an OS are: interpret and execute user commands, detect hardware and/or software errors, and report and take appropriate action, manage files and file directories, format disks, defragment disk storage, and relocate data from bad sectors on disk.
11. Many types of OSs are available in the market and they are designed for a variety of uses and for different types of computers. In Tables 10.1 and 10.2 we have summarized this.
12. Programming languages for computers are developed with the primary objective of making it easy for a large number of persons to use computers without the need to know the hardware details of a computer. In fact, most programming languages are machine independent.
13. Assembly languages are low level programming languages which have one-to-one correspondence with the hardware instructions of a computer. They are used where efficiency of code is very important.
14. High level machine independent programming languages (called HLL) are designed depending on application areas. They are called procedural when a detailed step-by-step algorithm (for solving a problem) is required before using the language.
15. Procedural languages are normally matched to application areas. An exception is C which is an application independent general purpose language.
16. Non-procedural HLLs require a precise specification of the problem to be solved.
17. Problem oriented HLLs are designed to solve a narrow class of problems. These languages use ready-made programs for solving specific applications. Only data needed for a specific problem is to be input in a well-defined form.
18. A classification of programming languages is given in Fig. 10.2.

EXERCISES

10.1 What do you understand by the terms computer hardware and computer software? How do you distinguish between the two?

10.2 Define the terms: algorithm, programming language and computer program.

10.3 What is system software? Give an example.

10.4 What is packaged software? Cite an example.

10.5 What is application software? Give an example.

10.6 What are the three types of systems software?

10.7 What is an operating system? Why is it required?

10.8 What are the different facilities provided by an operating system?

10.9 What is BIOS? What are its functions?

10.10 What is a device driver? Why is it required?

10.11 Where is BIOS normally stored? Why?

10.12 What is a firmware? How is it different from software?

10.13 What is a file allocation table (FAT)? Why is it required?

10.14 What is file fragmentation? What is defragging?

10.15 Enumerate different types of OS.

10.16 What is multitasking? What are its advantages? Which OS provides multitasking?

10.17 What is a multiuser OS? Which OS provides multiuser facility?

10.18 What is multiprogramming? What is the advantage of multiprogramming?

10.19 What is the difference between a machine dependent and a machine independent OS? Give an example of a machine independent OS?

10.20 What is an open source OS? Give an example.

10.21 What is a high level machine independent programming language? Give an example of such a language.

10.22 What is an assembly language? Is assembly language machine independent?

10.23 What are the advantages of an assembly language? What are its disadvantages?

10.24 Give some applications of assembly language.

10.25 What is a procedural language? Give an example.

10.26 What is a non-procedural language? Give an example.

10.27 What features does a procedural language normally have? Give a sample of each of these language features.

10.28 What is an interpreter? When are interpreters used?

10.29 What is a compiler? What is the difference between an interpreter and a compiler?

10.30 What is an object oriented language? Give an example.

10.31 What is Java? What are the advantages of using Java?

10.32 What is a scripting language? Give an example of a scripting language. What is the major application of a scripting language?

10.33 What type of language is LISP? For what applications is LISP used?

10.34 What is a problem oriented language? Give an example.

10.35 Classify programming languages based on their application. Give the name of a programming language used for each of these applications.

CHAPTER

11

Data Organization

LEARNING GOALS

After reading this chapter, you should be able to:

1. Explain the need to systematically organize data for storage in a computer's secondary memory.
2. Explain what a database management system is and why it is needed.
3. Explain how a database is organized as relations and describe the parts of a relation.
4. Organize non-text databases such as audio files and image files for easy retrieval.
5. Explain the need to back up databases to recover data resources lost or damaged due to accidents.

11.1 INTRODUCTION

In our day-to-day life, we keep a number of organized data. The simplest example is an address book which has the name, address, telephone number, and email id of all your friends and relatives. We enter the names, address, etc. in an address book and arrange them in alphabetical order. If the list is small and does not change very often, a simple handwritten address book is sufficient. However, if it is large or changes frequently or if we want to extract addresses and print them in letters or use them for several applications, a manually maintained address book is not sufficient; it is preferable to use a computer to store and maintain the address book. The data should be methodically organized to retrieve the required data. The method of organization will also depend on the type of data and its purpose.

In the case of a simple address book, we may have data such as the name of the person, street address, city, pin code, telephone number, and email address. We may observe that the telephone number and email address may be missing in some cases. This collection of data has some homogeneity in the sense that they all contain addresses. A collection of related data having a similar purpose is called a *database*.

11.26 Assume that you have a large library of 500 English Pop songs on audio cassette tapes. Explain the steps you will use to convert this collection to MP3 format for storing in a database.

11.27 What is the major difference between an audio and video database? You have a collection of 100 movies (each around 2 hours duration). Is it possible to store it as an on-line database? If yes, how much disk space is required to store the movies?

11.28 What is data archiving? Why is it necessary to archive data?

11.29 What do you understand by the term 'file backup'? In what storage medium do you back up files? Why should you back up files?

11.30 What do you understand by daily, weekly and monthly backups? Why are they needed?

made programs available in a statistical library of programs are used rather than writing programs from the beginning.

- *Solving science and engineering problems:* These involve operations of calculus such as differentiation, integration, and solutions of ordinary and partial differential equations. As computers are designed only to do arithmetic, that is, add, subtract, multiply and divide, it is necessary to convert operations of calculus to that which use only these operations. This is a vast subject called *numerical analysis* whose aim is to develop numerical computer programs implementing most commonly used numerical algorithms. There are also problem oriented languages such as MATLAB to solve many problems in science and engineering. Many science and engineering problems require *simulation*—that is, making a computer-based model of a machine or a system to be built which is "almost identical" to the actual machine or system. Often the simulation is useful to select the most appropriate values for some of the parameters which can be adjusted. To take an example, nowadays the shape of the front of high speed train engines has a smooth conical form rather than being flat. The front part of the engine is called its nose. A smooth conical nose reduces the friction caused by air when the train travels at a high speed and saves the power required to pull the train. A simulation program is written for a computer to mimic various shapes for the nose and pick that shape which, for a given range of speeds of the engine, minimizes the frictional force.
- *Simulation in business and economics:* Simulation is of interest not only to engineers and scientists but also to those in business and economics. A very simple example is to predict the quantity of various medicines to be kept in a medical shop. The simulation uses data on past demands for the medicine over a period of time, the time taken to get the supply from the wholesaler and loss due to storing unsold medicines, particularly when they have a short expiry period beyond which they have to be thrown away. Simulation uses estimated demands and supply delays to determine the quantities of various medicines to be stocked to maximize profit. Another simpler example is to find out monthly payments to be made for purchasing a car given the price of the car, down payment, and interest rate charged by a bank. You would like to compare offers made by various banks and pick the one most suitable for you.

All the above are examples where numerical computations are involved. As this book has been written for students of all disciplines, namely, science, arts, commerce, medicine, biology, etc., we will describe only the tools most useful for all.

12.2 USE OF SPREADSHEETS

Businesses have been using ledgers for keeping accounts. Each page of a ledger has rows and columns containing numbers. Columns and rows are normally given numerical labels. An example is given in Table 12.1.

Table 12.1 A Portion of a Ledger Page Maintained by a Shopkeeper

Item name	*Quantity in stock*	*Quantity sold since first of month*	*Price per unit*	*Qty. when re-order is to be initiated*	*Qty. to be re-ordered*
1	2	3	4	5	6
Sugar	100 kg	50	15.00	20	150
Urad Dal	50 kg	35	30.50	10	100
Sona rice	500 kg	280	22	300	1500

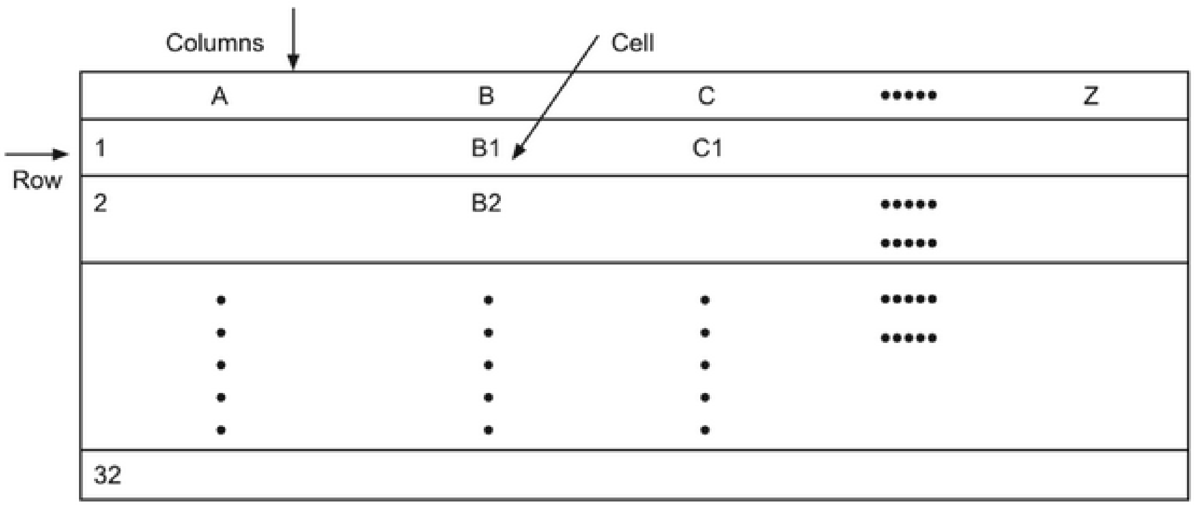
If the shopkeeper wants to find out the total earnings during the month, the value in column 3 is multiplied by that in column 4 at the end of the month. These individual products are added for all the items to obtain the grand total. This calculation is easily done with a ready-made program called a *spreadsheet.*

Another example is to find the monthly payments to be made in a hire purchase scheme in which the cost of the item bought is given, the interest charged is given in per cent per year, and the number of installments is to be found out. One can, of course, write a program to do it but a spreadsheet simplifies the calculation as no program is to be written; only data and the formula for calculation are to be given. Another advantage of a spreadsheet is its ability to answer what is known as "what-if" question. In the above example you can ask the question "what happens to my monthly installments if I am able to pay a bigger amount as initial payment". It is also possible to compare the hire purchase schemes offered by different vendors and find out the real interest rate each one is charging. Many other interesting questions relating to business decisions can be asked once a specified situation is modeled as a spreadsheet. Thus, a spreadsheet can also be thought of as a modeling tool which models different scenarios in a business and provides answers to pertinent questions. We will discuss some of these problems in the rest of this section.

12.2.1 Overview of Spreadsheets

Spreadsheet consists of many worksheets; each worksheet is displayed on the screen of a monitor. Each worksheet consists of rows and columns as shown in Table 12.2.

Table 12.2 Spreadsheet Format



in the dictionary if your spelling is incorrect. In fact, it will underline all words not found in the dictionary. For example, if you write Dosa, it will not find the word in the dictionary and will highlight it. Some word processors have an auto-correct feature which highlights a word not in the dictionary as soon as you type it and allow you to correct the word immediately but it can be slow. Spell checkers do not correct incorrect usage of words or errors in grammar. For example, if you write: he come house, it will not correct the sentence as all the words are correctly spelled.

Another feature available in many word processors is a list of synonyms and antonyms. Such a list is known as a Thesaurus. There are many shades of meanings for a word and you may like to pick one which is appropriate in the context of your sentence. For example, the synonyms of the word hot are: burning, feverish, scorching, sizzling. One of these words may be the best one to use in a given context. The antonyms of hot are cold, chill, icy, frigid. Such a facility allows you to polish your manuscript.

Lastly, many word processors also have grammar checkers. These programs highlight grammatical errors and also suggest corrections. It is up to you to accept or reject the suggestion. For example, if you type: He come house, which is grammatically wrong, it will point out the error. The possible alternatives are: He came home or he will come home depending on the context. Only a good grammar checker will find this error. Grammar checkers have not yet become very good as English has words with a number of possible usages.

Personalization of Business Correspondence

Very often you would like to write a common letter to a large number of persons. For example, a letter may be:

```
Dear <Kichu>
There is a meeting arranged in Hotel Rama on 30 October
2012 to discuss our club's activities. We will be very
happy if you and <Kamala> can attend.
Looking forward to seeing you
Yours sincerely
Ramu
```

In the above letter, the words within angular brackets, namely, <Kichu> and <Kamala> are the names of two persons, those of husband and wife. Thus the letter is of the form:

```
Dear <name>
There is a meeting arranged in Hotel Rama on 30 October
2012 to discuss our club's activities. We will be very
happy if you and <wife> can attend.
Looking forward to seeing you
Yours sincerely
Ramu
```

You can now generate two files. One has the form letter as shown above and the other name and wife name as shown in Fig. 13.1 and then input them to the Merging Program. The output will be a file of personalized letters.

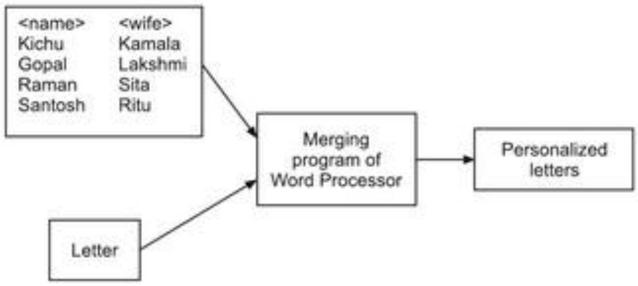


FIG. 13.1 Personalizing letters.

After personalizing a letter if address labels are to be printed and merged, these labels can be placed in a file and merged as shown in Fig. 13.2. This is sometimes called *mail merge*.

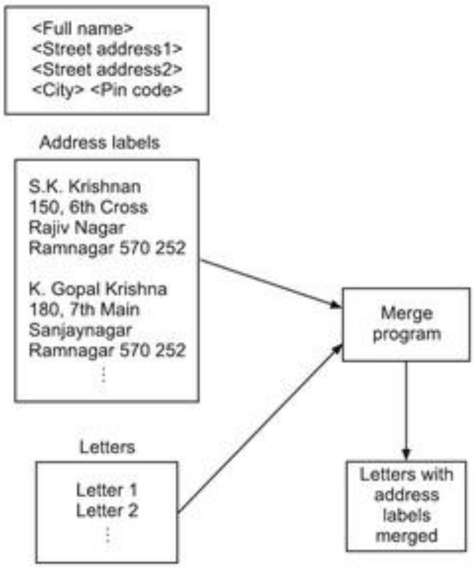


FIG. 13.2 Merging address labels to letters.

Creating Composite Documents

You may have used a spreadsheet program to create an invoice. If you want to incorporate this invoice at an appropriate place in a word document, it is possible. The invoice spreadsheet

CHAPTER

14 Processing Multimedia Data

LEARNING GOALS

After reading this chapter, you should be able to:

1. Differentiate between image processing, image generation and image recognition.
2. Explain image morphing and animation.
3. Differentiate between audio data editing, audio data generation and audio data recognition.

14.1 INTRODUCTION

All of you may have photo albums containing a large number of photographs taken at various times. We saw in Chapter 3 how these photographs can be digitized and stored in computers. Once they are stored you may like to erase backgrounds such as foliage and enhance the faces of people in the photo. You may like to cut off parts of a group photograph and take a portion and enlarge it. These are some applications which are of interest in day-to-day work. Apart from editing existing photographs, you may like to artificially create, using the computer, "paintings", drawings, etc. These are also possible. The aim of this chapter is not to make you an expert in doing all these jobs but give you a flavour of what tasks are possible with today's computers.

We have seen in Chapter 4 how music may be digitized and stored in a computer's disk. You may like to select songs and create an album with a desired sequence of songs. Another operation you may like to do is to enhance the lower frequencies and add special effects such as an echo which gives one a feeling of a concert hall. These are some operations which one can perform with digitized audio files.

Finally we will also examine in this chapter what operations are appropriate on video data. Video data are of two types. One of them is movies taken with a video camera, and the other are cartoons that are artificially created using computers. Videos taken with cameras are digitized and edited using computers. In fact, many video recordings of marriages add special effects, introduce background music and zoom on specified individuals in certain scenes. These are all done using computers.

The primary objective of this chapter is to make you aware of the operations that can be performed on graphics, video and audio data and their applications.

14.2 GRAPHICS PROCESSING

So far in this text we have come across various types of graphics objects. In Chapter 3 we studied how to acquire image data and how to store them in compressed form. The processing of image data may be divided into three broad categories. When the data are acquired from photographs, X-ray pictures, etc., they may require modification to improve their quality to highlight specific features. This is called *graphics data processing* or, more commonly, *image processing*.

It is also possible to synthesize pictures either by using complex mathematical models to represent them or by "painting" pictures on the screen of a monitor using built-in tools to draw and colour them. This is called *generative graphics*. Lastly, a common requirement is to classify image data based on some specific properties, for example, identifying certain features of fingerprints, recognizing peoples' faces, etc. This type of processing involves understanding the contents of a picture by using specific characteristics. Such processing is called *cognitive graphics*. In Table 14.1 we summarize the above discussion.

Table 14.1 Types of Graphics Processing

Type of graphics processing	*What is the main objective of this type of processing?*	*Example applications*
Image processing	Input is a picture as a collection of pixels. Editing picture by changing colour of certain parts, erasing parts, cutting off portions, combining two pictures. Improving quality of pictures	Enhancing and/or cleaning up an acquired picture
Generative graphics (image generation)	Synthesizing pictures from primitive components using graphic tools	Icons in GUI, computer art, Computer Aided Design
Cognitive graphics (image recognition)	Understanding contents of an image and classifying images automatically	Fingerprint classification, face recognition

14.2.1 Image Processing

The starting point in image processing is a picture in bitmap form. There are several ready-made tools marketed by vendors to edit pictures, enhance the visual effects of pictures, change colours, cut and paste images, enlarge or contract images, and perform similar operations. The most popular among these are: Paint—a Microsoft program for use with Windows OS, Adobe's

Photoshop, and Corel Photo-Paint, all of which work using an image file in bitmap form. The operations which are typically available with all these programs are:

Fill with colour, erase, change colour, clip, use brush, use pencil, cut and paste, magnify, insert text, enlarge, contract, and similar operations.

Using all these primitive operations, you can create pictures, modify existing pictures in various ways, combine pictures and in general 'doctor' pictures.

14.2.2 Image Generation

In image processing, we begin with an existing photograph which we digitize for processing. In image generation, technically called generative graphics, we want the computer to create or generate a picture which is as realistic as possible to a photograph of a real person or of a real object, scenery, etc. Here again, there are several ready-made tools available to generate pictures. We will classify pictures as *geometrical objects* which can be drawn using line segments, circles, curves, ellipses, triangles, etc., and as *natural objects* such as an elephant and human faces. Geometrical objects are easy to generate using the primitive operations available on a PC. Generating figures using line segments, circles, etc. is known as *vector graphics.* Vector graphics programs are used extensively to create engineering design drawings, generally called Computer Aided Design (CAD) tools. One of the most popular CAD tools is called AUTOCAD which allows viewing different aspects of an object such as top view, side view, frontal view, and a 3D perspective. The major advantages of tools such as AUTOCAD are:

- Precise drawings to scale and control of colour
- Blending colours and shading
- Ability to slowly rotate a 3D object view on a video screen and make alterations as needed

We start drawing a 3D object using what is known as a *wire frame* [Fig. 14.1(a)]. Wire frame is a 3D drawing of the outline of an object using primitives such as straight lines, circles, ellipses, etc. It is the skeleton of an object. The wire frame is then shaded and coloured with appropriate tools to obtain a realistic view. Manipulation of a wire frame is easy and quick. Thus, if an object's look is to be changed, the wire frame model is changed. The altered wire frame model is then shaded and coloured. You may wonder what is shading. *Shading* is the process of giving the appearance of a surface by using appropriate colour or grey tone to one side of a wire frame. This gives the appearance of a 3D object for a wire frame. This is illustrated in Fig. 14.1(b).

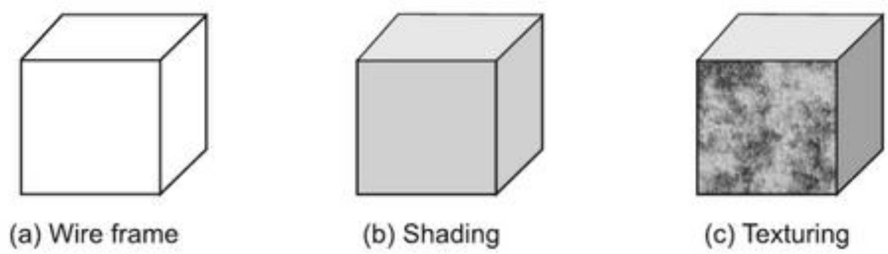


FIG. 14.1 Wire frame, shading and texturing.

block diagram of a text-to-speech system is given in Fig. 14.3. It uses the idea of putting together sounds of letters called *phones*. For example, 'c' in 'cat' and 'p' in 'pat' have two different phones.

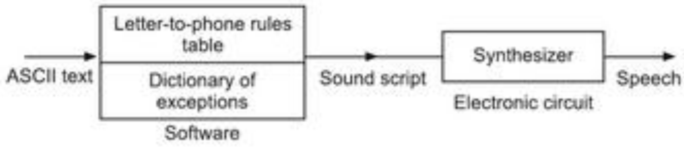
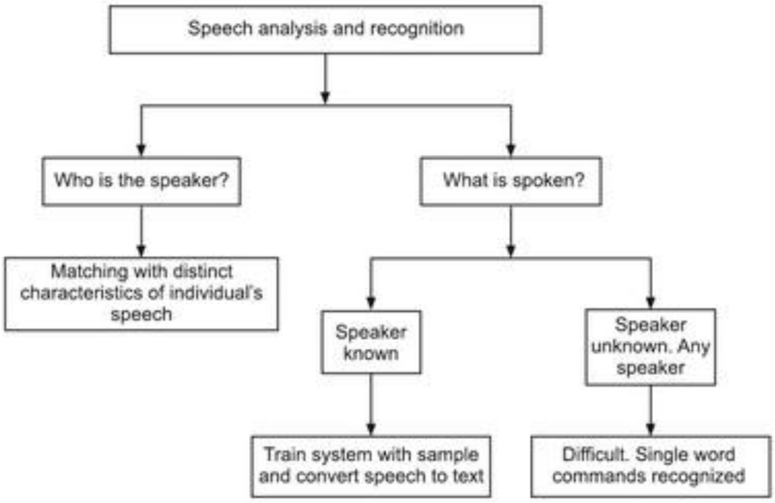


FIG. 14.3 A text-to-speech system.

Given a word, one may attempt to put together individual phones to pronounce that word. English is not a phonetic language, and thus such a simple idea does not always work. For example, k in knife is silent while k in kill is not. Thus a dictionary of exceptions is needed. Once individual letter sounds are generated, they have to be connected to form sounds for words. This is done by using *morph* which is the smallest speech unit that carries a meaning. For example, war is a morph but not warship. The synthesizer creates morphs and combines them to obtain speech. More sophisticated synthesizers introduce prosody and attempt to produce natural sounding speech. The current state-of-the-art produces excellent text-to-speech systems which sound very natural. There are even versions which use British pronunciation of sentences, which is different from American pronunciation.

14.3.4 Speech Analysis and Recognition

In Fig. 14.4 we give an overview of problems related to analysis and recognition of speech. A simple application of speech analysis is recognizing a speaker from a sample of his speech.



by the sender. As email is very fast, normal postal mail is often called 'snail mail' by email users! The use of email has expanded with provisions to attach almost any file to it. Spread sheets, PPTs, images, audio and video files may be attached.

Another major application of the Internet is known as the *World Wide Web* or the *web*, for short. The web has become the most important information infrastructure in the world and has billions of users. At the touch of a terminal button, you can now get information from all corners of the globe. It has also enabled many other services such as e-commerce, e-governance, e-publishing, etc. We will describe what a web is, and how it is created and used. The web became useful with the creation of software called a browser. A browser is used for many purposes: to search for information desired by you, to access any website using a unique address called its URL (Uniform Resource Locator), download files, login to use remote computers and to send and receive email. Internet can also be used for downloading music files and for sending and receiving what is known as voice mail. Recently, services have emerged, called *Internet radio, phone and T.V.* We will describe how they work. Finally, we will examine how images and video can be transmitted using the Internet.

15.2 EMAIL

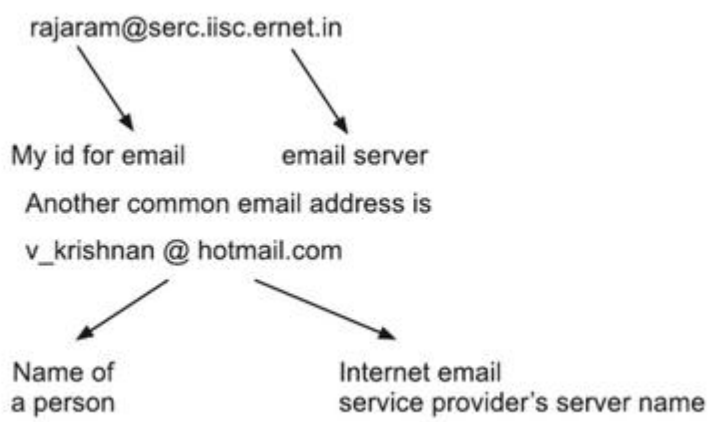
Current email systems provide the following facilities:

- Send a message to a specified recipient or a group of recipients.
- Send a message that includes as an attachment text, spread sheet, power point presentations, audio, images, or video file.
- Send a message along with a program which can be executed at the recipients' computer.
- Forward a message received from a person to another person or persons.

In order to be able to send or receive email, you need a unique address. It is similar to the need to have a postal address if you want to receive letters. The email address has the form:

```
yourname@name of your email server assigned by the domain
name server
```

For example, my email address is:



In order to send/receive email, you need an application program which should be installed in your computer. It allows you to compose a message, edit a message, specify the recipient or recipients by giving their email addresses and specify names of attachment files (if any). It also provides an inbox for you to receive email and an outbox (or drafts) to store email to be dispatched from your machine.

A typical email looks as shown in Fig. 15.1.

```
From    :     ramu6@ee.iitk.ac.in
To      :     rajaram@serc.iisc.ernet.in
Date    :     10 Jan 2003  11.02.04 IST
Subject:      Thank you

Dear sir,

Thank you for responding to my mail and answering
my question.

Kind regards

                                   Sincerely
                                   Ramu
```

FIG. 15.1 A typical email format.

If you want to send an email from your home, you require the following (Fig. 15.2):

1. A telephone connection.
2. A modem in your computer. Most recent computers have a built-in modem which can transmit/receive data at the rate of 56.6 kbps. Nowadays telephone companies provide a modem called DSL (Digital Subscriber Line) which virtually splits your telephone line into two independent telephone channels: one is used for your telephone conversations and the other carries data. Data can be transmitted at the rate of at least 256 kbps. This connection is called *broadband*. Broadband is essential to transmit audio and video data.

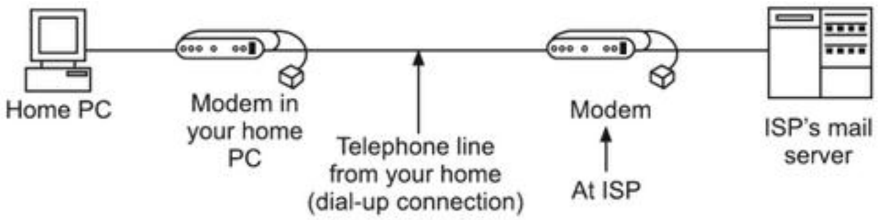


FIG. 15.2 A dial-up line from a home PC to ISP.

3. An account with an Internet Service Provider (ISP) who will maintain for you in his server a mailbox to receive your email (called inbox) and another box for the email you send (called outbox). His server will be on 24 hours a day and 7 days a week as email may come from any part of the world at any time. There are several ISPs in India and they have a variety of subscription plans.

4. The ISP will provide you an email id and allow you to use a secret password of your own to access your mailbox.
5. The ISPs will also normally provide you with a telephone number to connect to their server for sending/receiving email. This telephone number is usually stored by the email application program in your computer. Thus, when you want to send email, you connect to ISP's server by dialing ISP's telephone number. If you have a broadband connection, there is no need to dial the ISP. You are automatically connected when you switch on your modem.
6. It is possible to have files attached to your email (just like enclosures in a typed letter). These could be any file such as ASCII, jpeg, tiff, mp3, ppt, mpg4, etc.
7. Each call to ISP will cost you one local call charge which your telephone company will charge you. Thus it is preferable to compose a number of emails you want to send, keep them ready and send all of them once a day. At the same time, you can also receive the email stored in your mailbox at your ISP. With broadband connection there are no call charges. There is a monthly subscription for the service.
8. Several companies such as Google, Microsoft and Yahoo provide free *email websites* with large size mailboxes and several user-friendly features to send, receive, store and search emails. Most users nowadays use these rather than the ISP's service.
9. Emails can be sent from laptops, tablet computers or high-end mobile telephones. Laptops and tablets are normally Wi-Fi enabled. The email is received by a wireless hot spot and routed to the telephone line (broadband connection). The other alternative is to connect a wireless transceiver (normally in the form of a stick similar to pen drive) to a USB port of your laptop. This connects the computer to the mobile communication network for sending/receiving email. High-end mobile phones (such as Blackberry or iphone) provide email service.
10. DSL modems at homes also have a Wi-Fi connection and a router. Thus laptops or tablet computers can access the Internet using the Wi-Fi enabled modem cum router to send/receive email.

If you are connected to a LAN in your office (or college), then there will be a mail server in the LAN and your mailboxes will normally be in that server. There is no need to dial up ISP. The organization's LAN will be connected to the ISP through a router and possibly a leased line of high bandwidth. The mail server will be on 24 hours a day, 7 days a week and will be sending the mail almost immediately to the destination address (see Fig. 15.3).

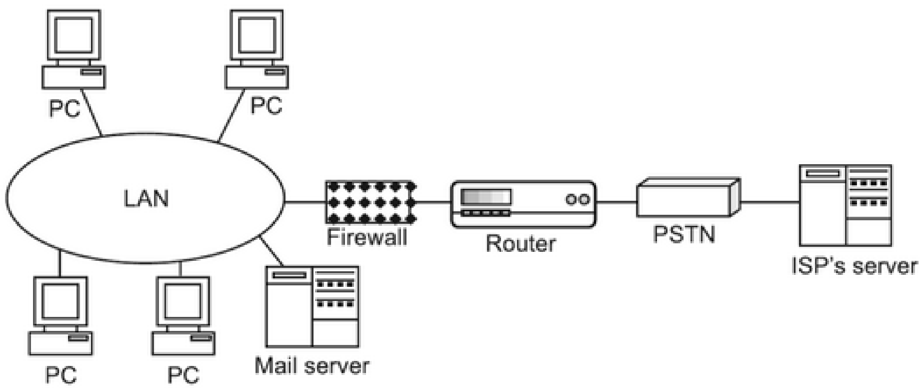


FIG. 15.3 Email via a LAN.

15.3.2 Hypertext

Assume that a text such as that of Fig.15.5 is given. The text gives a brief write-up about Bangalore in which a number of *keywords* are shown using a different font. These keywords are linked to other pages which give information in more detail about the keywords.

About Bangalore

Bangalore is the capital of Karnataka. It is called the Silicon Valley of India as it has a large number of `InfoTech industries`. Leading among them are `INFOSYS`, `WIPRO`, `TCS` and `SONATA`. It is also the home of the `Indian Institute of Science`, a pioneering educational and research institution which was founded in 1907 by Sir J.N. Tata. It also has many `engineering colleges`. It is also called the garden city as it has one of the largest and oldest botanical gardens of India called `Lalbagh`. Lalbagh has several `flowering trees` and some `rare birds` with wonderful chirping sounds.

FIG. 15.5 A document with keyword linking it to other documents (keywords are shown in a different font).

This idea of embedding selectable keyword in a text, which links it to other documents facilitating retrieving information, is called a *hypertext*. The hypertext facilitates what may be called nonlinear reading, i.e., while reading a document if some topic interests you, it is possible for you to click on the keyword corresponding to that topic and go to that document. After reading the document, you could return to the original document to the same point where you left it. In the computer's memory (normally, disk), the documents are linked as shown in Fig. 15.6. This figure is given just to satisfy your curiosity.

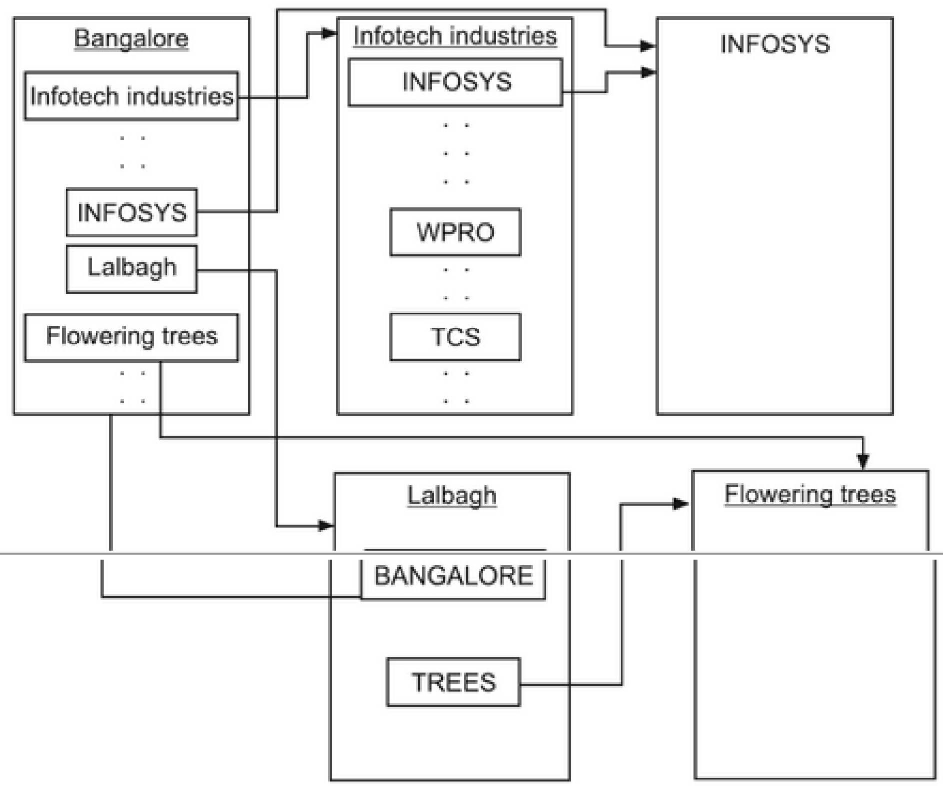


FIG. 15.6 Hypertext links between documents.

15.3.3 Universal Resource Locator

We saw in the last section that a hypertext is organized by specifying a number of keywords and the links from these keywords to other relevant documents. The major innovation in the World Wide Web is that it specifies the address of the server where the document, also called a web page, is kept. This address is used by the browser.

The World Wide Web consists of a number of *web pages*. Web pages use hypertext with a set of clickable (i.e. selectable) keywords which are normally highlighted in a different colour. A collection of related web pages belonging to an organization is called its *website*. Websites are stored in servers (also called hosts) belonging to various organizations with the primary intention of allowing free access to their website to any person anywhere in the world. The host is normally given the name *www* by every organization which wants to be a participant in the World Wide Web of sites. A web page in the World Wide Web is accessed by specifying what is called a Universal Resource Locator (URL). A typical URL has the form shown in Fig. 15.7.

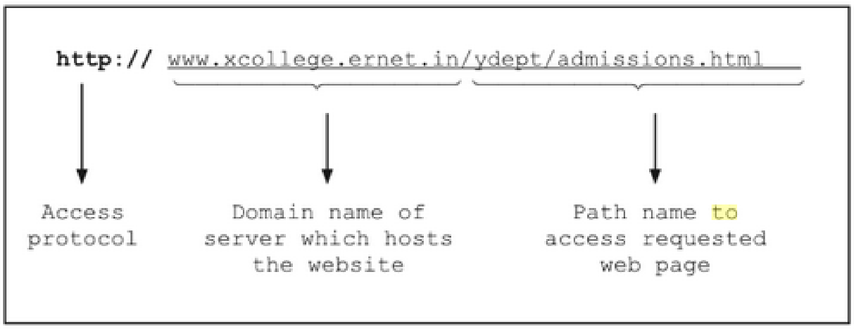


FIG. 15.7 A Universal Resource Locator (URL) of a hypothetical website.

In the URL of Fig. 15.7 **http** specifies the rules agreed on between a client (in which a browser runs) and a server to transfer messages and documents. This protocol is now a universal standard and is often omitted while entering the URL. The second part following ://, namely, `www.xcollege.ernet.in`, is the domain name of the server where the website is stored. Following this is `/ydept/admissions.html` which gives the path to the page stored in the website which you want. It says that a folder named `/ydept/admissions.html` has the page required by you and it is an `html` document.

As we have seen, a browser is used to access web pages. When you enter a URL in the address bar of a browser, a web page is displayed. A sample page is shown as Fig. 15.8. In many cases, you may not know the explicit path to a specific page. Usually, all websites have a home page which contains a table of contents. You can click on any of the items you are interested in (displayed in the home page) and the browser will automatically go to the relevant document, display it and also show the URL in the address field of the page. This is illustrated in Fig. 15.9. If you click on Departments on this page, it will take you to a page with all departments listed. You can click on the specific department you are interested in and it will display information on that department.

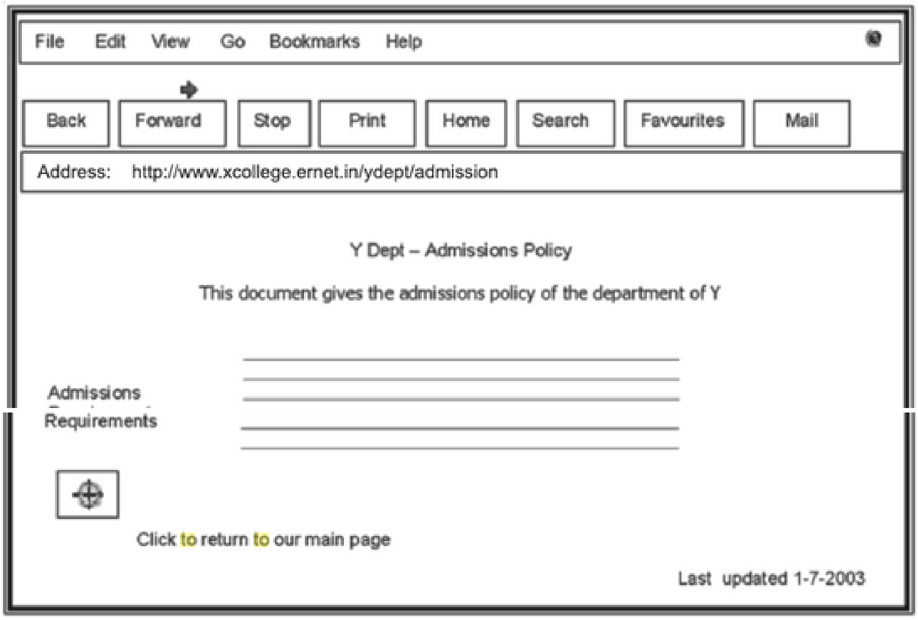


FIG. 15.8 Display of requested URL.

(We have shown a typical screen you will see. It is not that of any particular browser.)

FIG. 15.9 Display of home page of X college. Any of the underlined items may be clicked to go to the relevant page.

15.4 INFORMATION RETRIEVAL FROM THE WORLD WIDE WEB

As of today, there are over 300 million hosts on the World Wide Web serving over two billion users all over the world. (These numbers are estimates which increase every month.) The web is thus a vast mine of useful information. It is so huge (several petabytes, i.e., several million gigabytes) that finding the information you require using it is like looking for a needle in a haystack. You should therefore use a systematic method of searching for information. Many tools are available in the Internet to assist you to obtain the information you require and you must know how to use them. In addition, if you know search techniques, it will be an added advantage.

One of the best ways of searching for information on the web is to start with a known website (i.e. known URL) which will lead you to other similar sites via hypertext links. One link may lead to another and you may forget which intermediate pages interested you. Thus it is advisable to store these URLs using *bookmarks*. Bookmarking tool is available with all browsers.

15.4.1 Resource Directories

If you do not know any URL, it is a good idea to use what are known as *Internet resource directories*. These directories normally categorize resources by subjects (e.g., Astronomy, Biology, etc.) and resource type (e.g., institutions, patents, etc.). Many of these employ human experts for indexing and classifying resources. There are two types of resource directories: those that cover all subjects and those that are for a specific subject. An example of a general purpose directory is *yahoo.com*. Subject specific directories are maintained by professional societies in the subject, for example, the American Mathematical Society (ams.org) and American Chemical Society (acs.org). Scholarly Internet resources selected by librarians of the University of California system is another directory (infomine.ucr.edu).

A good general purpose subject directory is maintained by dir.yahoo.com. This site uses many methods to create subject directories and keep them updated. Some of the methods are:

- Submissions of their URLs by developers of websites with subject keywords.
- Human indexers to search the web and add important keywords to the subject index.
- Use of websites of special interest groups who gather information on the area of their interest.
- Use of programs called *web crawlers* which systematically and automatically visit many websites and add URLs to the subject index.

The subject index maintained by this site (dir.yahoo.com) can be used by you to obtain the information you seek on the subject. Queries called *Boolean queries* are allowed to search the index. By Boolean query we mean a query of the type: [keyword1 (AND) keyword2]. We interpret this as pick all websites which deal with *both* keyword 1 *and* keyword 2. Either one is not sufficient; both should occur. For example, if you are interested in the seventeenth century history of India, you can pose the Boolean query:

Seventeenth century AND History AND India

If you wrote just History or India, a lot of irrelevant information will be retrieved. You should formulate your query so that *only* relevant information is retrieved and *all* relevant information is retrieved. In other words, *no* relevant information should be missed and the information retrieved is *all* relevant and useful. This is not always easy. Even if your query is well formulated, the tool you use may not be very good. Thus it is also important to pick good tools.

Boolean queries allow what is known as NOT as well as OR operation. For example, if you want a list of hotels in Bangalore which are not very expensive, you may write:

> (Bangalore *AND* Hotels) *AND NOT* (Five Star Hotel). This will remove all five star hotels from the list.

The search using a Boolean query in yahoo will display documents subjectwise with the first few lines of the document. This will enable you to decide whether or not to get the full document.

15.4.2 Search Engine-Based Retrieval

Currently, the most popular method of retrieving documents from the World Wide Web is to use what are known as search engines. Four popular search engines are:

- Google.com
- Altavista.com
- Ask.com
- Bing.com

We will examine Google.com, a popular search engine, and describe the facilities available in it to retrieve information. It is important to know that periodically newer search engines become available which are superior to existing ones. The site *infopeople.org* continuously updates information on available search engines.

Google allows you to write a request for information in normal language. For example, if you want to gather information on the seventeenth century history of India, you can just type it in the Google's search box. The search engine will retrieve ALL web pages where *any one* of these words, seventeenth, century, history, India, appears. You will thus get millions of results (called hits) and will waste a lot of time going through the results. Google allows search queries using phrases, i.e. combination of words. For example, "Seventeenth century history India" is a phrase which can be used in a query. In this case, it will retrieve only documents in which all these words appear which will drastically reduce the number of documents retrieved. Google automatically inserts AND as a Boolean operator between phrases. For example, if you write:

"Seventeenth Century" "History" "India", the AND connective is inserted between phrases. For example, "in URL" will limit the search to the specified URL only. Other qualifiers can be "in title", meaning the phrase must appear in the title of a document.

Ask.com search engine answers queries of the type: What is the capital of Serbia? It expects similar queries.

1. *Sound cards on both the computers.* A sound card has the electronic circuit to convert continuously varying audio signals to bits. This is called A/D conversion. It can also convert digital data received to audio. This is called D/A conversion. If you want to simultaneously send and receive audio data, the sound card must have what is known as *full duplex capability*. For telephone conversations this facility is essential.
2. *Microphone and loudspeakers.* Often a microphone is built into the chassis of the computer. Else a small microphone which you can clip on to your dress is needed. For receiving good quality audio, you need two good quality loudspeakers connected to the PC's sound card.
3. *Software for compression/decompression.* If you want to send/receive MP3 compressed audio.
4. *Software called plug-ins to your browser.* Popular plug-ins are: Real Network Player, MS Windows media player and Apple Quick Time. These facilitate listening audio streams and viewing video.

Once you have this infrastructure and an Internet connection, you can send/receive audio on your PC to/from anyone connected to the Internet. There are four different applications we describe below:

1. Voice attachment to email.
2. Downloading audio, for example, music files from a music server to your PC.
3. Listening to a radio broadcast using the Internet. Many Internet radio stations are now operating.
4. Carrying on a conversation with another person whose PC is connected to the Internet. This is called *Internet telephony* or *IP telephony*.

15.6.1 Voice Attachment to email

We saw in Section 15.6.3 that the current email protocol (Multipurpose Internet Mail Extension) allows multimedia files to be attached to email. Voice attachment to email is sending and receiving audio messages between a sender and a receiver. The only extra facilities needed are those given at the beginning of this section, namely, sound card, microphone and loudspeakers.

A sender can record a small audio message (a few minutes) and attach it as a binary file to an email message. The major advantage of a voice attachment is the fact that you can convey a lot of information when you talk compared to a typed message. As we saw in Chapter 4, one minute audio message (speech) will require 375 KB. Three minutes message will require 1.125 MB and is a manageable file size. If it is a long message, we may use MP3 compression, in which case the message size will be reduced by about 12 to 14. As you will be hearing the voice attachment off-line after disconnecting from the ISP server, time taken by the audio file to travel on the Internet is irrelevant.

15.6.2 Downloading Audio Files on the Internet

There are several sites which store audio files and allow you to download them. These files will be usually MP3 compressed binary files. You can use a file transfer protocol and download it using your browser. The only problem is the large size of these files. We saw in Section 4.4 that the size of a music file of 1 hour is 50 MB (MP3 compressed). If your Internet connection is via a dial-up telephone line with a 56.6 kbps modem, the average speed you will get will be about 30 kbps. The time taken to download 1 hour music (50 MB) will be $(50 \times 10^6 \times 8)/(30 \times 10^3) = 3.7$ hours! Obviously you cannot hear the music while it is downloading. A dial-up telephone connection is not suitable to download music. You require a broadband connection. Telephone companies provide broadband connection using a DSL modem. With such a connection you can get an average bandwidth of 250 kbps. There are also wireless modems which you can connect to your PC or laptop. They provide a bandwidth ranging from 250 kbps to 3 Mbps. Such a connection is essential to download music. Assuming 1 Mbps connection, you can download 50 MB (1 hour music) in $(50 \times 10^6 \times 8) /(10^6 \times 60) = 6.6$ minutes. The cost of download depends on the plan you have subscribed to. Several service providers in India advertise unlimited download. However, beyond a pre-set limit, they reduce the speed of your connection! Downloading music implies that the music files are stored in your hard disk and can be played whenever you want. You can also burn (i.e., record it on) a CD. Downloading is safe only if the music is not copyrighted. It is illegal to download copyrighted music without the permission of the owner of the copyright.

15.6.3 Streaming Audio Files

There are websites which store music and you can use your browser to stream or in other words play in real-time music on your computer. The browser must have appropriate plug-in software such as MS windows media player. The Internet is used to send the music file. We have explained in Section 8.5.4 that the Internet is packet switched network and uses a protocol called TCP/IP. IP routes the packets whereas TCP ensures reliability of transmission by putting the packets in their right sequence and requesting the sender to retransmit lost packets. Reliability is essential when text files or program files are transmitted. This protocol slows down data transmission and the network delay is variable depending on the congestion in the network. The variability in speed is not relevant when downloading music as we download the entire file and listen later. For hearing music in real-time TCP/IP is not suitable. There is another transport protocol called UDP (*User Datagram Protocol*). This protocol delivers packets (called datagrams) from source to destination fast but may have errors. Some packets may be out of order and some lost in transit. In other words, UDP is unreliable but fast. Unlike TCP/IP, UDP allows multicasting data, i.e., sending data to several recipients simultaneously. UDP is more appropriate for transmitting data in real time such as music or video streaming. When you hear music or see video, a small break of a few milliseconds (corresponding to lost packets) will not be noticed by you as your ears and eyes interpolate the missing data from surrounding data. If packets are out of order, you will notice it. Thus putting the packets in right order (i.e. time synchronizing) is necessary. This is done by another protocol called Real Time Protocol (RTP) which rides on UDP. Real Time Protocol has the following essential information:

- Time stamp put by the sender giving the time of dispatch of the packet.
- Encoding format used by the payload of the packet.

Time stamp is necessary to put the received packets in the right order. Encoding format is used by the browser to select the appropriate player. In addition, RTP has a companion protocol called Real Time Control Protocol (RTCP) which is used to measure the *Quality of Service* (QoS) of the network and feed this information to the sender for it to take corrective action if the quality is bad (usually in due course). QoS parameters may include packet loss, variations in network speed and number of out of order packets.

In order to receive streaming audio, UDP along with RTP is used. The bandwidth of your Internet connection should exceed 200 kbps for adequate QoS. As the Internet speed is variable and packets may arrive out of order in UDP, buffer storage called *jitter buffer* capable of storing around 5 seconds of streamed audio should be provided at the receiving station. The application at the receiver will use the time stamp information found in RTP to correctly order the packets in the buffer and deliver music at a steady rate to the listener. This solution is analogous to what you would do to ensure continuous flow of water when your water supply is irregular, that is, small trickle sometimes and full flow at other times. You will use a tank and let is fill up for some time. After it fills up, you will draw water from the tank at a constant rate, independent of the rate at which water flows into the tank. As long as you do not draw the water too fast, thereby emptying the tank, the flow from the tank will be steady. The size of the tank should be sufficient to ensure that with expected variability of water flowing in, output rate is constant. Streaming of music for listening (and not storing permanently) does not infringe copyright.

15.6.4 Internet Radio

There are a large number of radio stations which broadcast music, news, etc. on the Internet. The advantage of Internet radio is worldwide coverage. In an Internet radio station while a broadcaster talks on the microphone or plays a CD, the audio is digitized by the sound card on his server, compressed to MP3, and stored in the high speed disk memory of the station's server. If you want to receive the broadcast, you should log on to the URL of the radio station. The station will then connect you to the server. You can then listen to the broadcast by streaming the contents of the server using the Internet.

As Internet radio uses streaming, a broadband connection of at least 200 kbps is required. The rest of the technology is the same as audio streaming. We use UDP/RTP for streaming and a small buffer at the receiving computer to even out delay in the Internet. Thus the broadcast will be heard after a delay of a few seconds. You can listen to broadcast free assuming you have an unlimited download plan with your ISP.

15.6.5 Internet Telephony

By Internet telephony (also known as Voice over IP (VoIP) we mean carrying on a conversation between two parties in real time using the Internet infrastructure. There are several requirements to be met by such a system. They are:

- Voice communication should be in real time. A delay of more than 300 milliseconds in hearing a caller (called mouth to ear delay) is uncomfortable. Plain Old Telephone differentiated services). The quality of telephone service in such cases is not much poorer than circuit switched telephones.

Before we conclude this section, we list the following disadvantages of IP phones compared to circuit switched phones:

- Poorer quality of conversation, particularly when the bandwidth of your "broadband service" is low.
- If the Internet fails, your phone will not work.
- If power fails, your IP phone will not work unlike normal phone.
- VoIP faces the same security threats as your PCs. Such threats are viruses, denial of service, call tampering, etc.

15.6.6 Skype Video Telephony

Skype is a proprietary system which has become highly popular among Internet users. As of December 2012, there were over 3200 million Skype accounts. Skype provides the following services using computers (Desktops, Laptops, Tablet computers, Smart phones, etc.) connected to the Internet.

1. Voice and video calls
2. Video conferencing among (up to) five users
3. Instant messaging
4. File transfer between users

All the above services are free. Besides these, there is a paid service which allows calling analog (i.e., ordinary telephones or mobile phones) from a Skype enabled computer and receiving calls from analog or mobile phones. It is a strong competitor to commercial Internet telephone service due to the fact that it is free, provides video and the quality of voice is reported by most users to be better than Internet telephone. To use Skype, you require:

- Broadband Internet connection (at least 200 kbps)
- Headphone and microphone with your computer
- Duplex sound card
- A video camera (if you want to use video calls or conferencing)
- A Skype account. To get this, you should log on to Skype.com and register
- The computer of the person you want to talk to should be switched on and he/she should be logged on to his/her computer and should have a Skype account

The technology and protocol used by Skype is proprietary and not publicized. The system does not use SIP protocol. It is surmised that it uses a peer-to-peer VoIP model to initiate and carry out a conversation. As opposed to client-server relationship, a peer-to-peer relationship in a computer network implies that any computer may act as a client or as a server depending on the circumstances. Thus peers may provide or use services of other peers. When a user logs on to Skype server to obtain a user id and password, Skype loads a program in the user's computer to identify it as a peer in the Skype community. The program also creates a directory of other reachable peers. The directory has one entry containing the IP address and port number

16.3 WHY SHOULD WE USE COMPUTERS IN BUSINESSES?

When businesses were small and had limited operational goals, it was possible to manually process data of the businesses and run them reasonably well. In fact, even today most small shops with a single owner who manages the shop/establishment do not use computers. However, with the reduction in the cost of computers and their ease of use, it is perceived that a computer can be an invaluable assistant even for small businesses. As we saw in the medical store example, a computer reduces the time to prepare bills, thereby allowing the store owner to attend to more customers. It provides up-to-date data on daily sales and assists him in controlling his inventory to reduce the cost of running the store. As long as his business is small, some of these operations can be done manually (although suboptimally).

As organizations grow, use of computers becomes essential to efficiently manage them. Some of the reasons are:

- As organizations grow, the volume of data they have to process becomes large. It becomes difficult to process them manually. For example, some examinations are taken by more than a hundred thousand students. Tabulating the results manually, finding out the classes in which students have passed, those eligible for supplementary examinations, etc. will take too long and lead to delay in declaration of results. Routine calculations can be done better and faster by using computers.
- Computer-based processing enables the same data to be processed in many ways based on specific needs, thereby allowing managers to look at the performance of an organization from different angles. As we saw, operational data can be processed in many ways to obtain tactical information.
- Markets are becoming competitive with globalization. It is thus essential to produce goods at minimum cost, maintaining quality. Computers have proven to be invaluable aids in efficiently producing goods and managing organizations.

Today, all medium and large organizations have corporate intranets, and computers have become an integral part in their daily operations. Even government offices and government services such as land registration are being computerized all over India, reducing delays and providing hopefully less corrupt governance.

16.4 MANAGEMENT STRUCTURE AND THEIR INFORMATION NEEDS

The medical store owner described in Section 16.2 performs many functions by himself. He works as a purchaser, storekeeper, accountant and salesman. As long as it is a small business (a one-room shop), he may be able to do all the above functions without any assistants. As he expands and is very successful, he may decide to run a chain of medical stores under his overall management. It will now be essential to delegate responsibility to specialists in each area of operation and hold them accountable for efficiently carrying out the assigned responsibilities.

In Fig. 16.1 we give a typical structure of an organization and its functional responsibilities. Observe that the management structure is hierarchical. At the top of the organization is the

President or Chief Executive Officer (CEO). He has reporting to him a number of *middle level managers*, usually designated Vice-Presidents, each responsible for a major functional area such as marketing, materials, finance, human resources, etc. Reporting to the middle level managers are line managers, each one responsible for day-to-day operations of their respective functional areas.

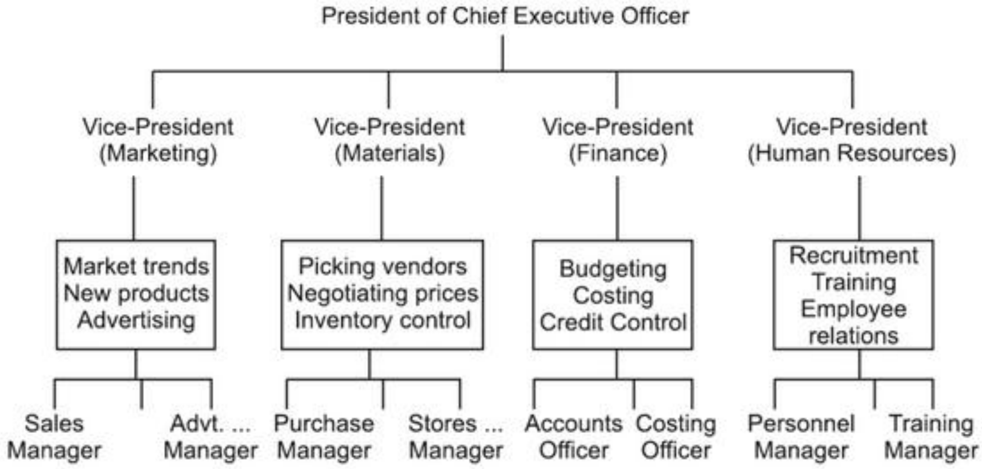


FIG. 16.1 Management structure and functional areas of management.

In this hierarchy, the top level manager, namely, the President, is responsible for long range strategic decisions and he should be provided summarized information useful for him to take strategic decisions. Middle level managers (such as Vice-Presidents) take tactical decisions which need tactical information. This is obtained by abstracting information such as trends from operational data. Finally at the last level are *line managers* who are responsible for day-to-day operations. They use operational data. We summarize this discussion in Fig. 16.2.

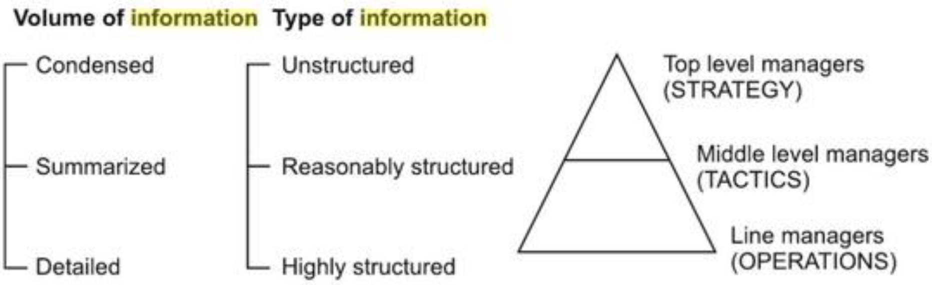


FIG. 16.2 Levels of managers and type of information needed by them.

We saw that operational information is voluminous and we need appropriate summarization to obtain information of relevance for tactical decision-making which is important for middle level managers. The system which uses operational data and applies a number of rules and summarizes them to obtain information relevant for taking tactical decisions is known as a *Management Information System* (MIS).

Let us take an example of a big retail store which sells a large number of items and has a huge customer base. Its sales are high and margins are low. Many such large retail stores are now coming up in urban areas of India. In such a store the middle management will be required to take decisions such as: at what level of stock of each item should it be reordered? How much should be reordered? If multiple vendors can supply an item, with whom should the order be placed? The answer to these questions will depend on data such as the rate at which items are sold, the time taken from the time an order is placed to the time it is delivered, the transport cost, the storage cost, the shelf life, and the loss incurred if an item is not available when requested by a customer. All these data have to be collected separately over a period of time, often using data from an operational data processing system. In this example, the daily sales of various items can be used to compute the average sales per day of each item. Data on delivery times, transport cost, etc. can be separately collected. Using these data, well known management tools (such as operations research) can be applied to determine the optimum stock level of each item at which reorder should be initiated and the quantity to be ordered.

Operational data collected over a long period is called *data archives* and the process of collecting operational data is called *data archiving*. With the availability of huge disks at reasonable cost in which terabytes (10^{12} bytes) of data can be stored, it has become feasible to analyze the archived data. Analyzing archived data to observe patterns which assist in management decision-making is called *data analytics*. A store manager, based on his experience, thinks that in the months of October, December and April, the sale of sugar is very high compared to other months. This guess may be verified by data analytics (i.e., by analyzing archived data on previous sales). In data analytics, a rule is formulated, which may state that in October the sale of sugar is 1.5 times the average, in December 1.3 times the average and in April 1.4 times the average. This rule may be verified by examining the data archive. If the rule turns out to be true, the manager can appropriately decide on ordering and keeping necessary stock. If not, he can try other alternate rules to find out which fit the data reasonably accurately and use these rules.

In general, management information systems use models of business to arrive at appropriate tactical decision rules. The data required to test these rules are obtained very often using data from an operational information system.

Strategic information for decision-making by top management is obtained by what is known as *Decision Support Systems* (DSS). In the example of retail stores, strategic decisions would be whether to discontinue the sale of certain items and replace them with other items, whether to start a new branch in a new suburb, whether to rent more area to expand the store, etc. Decisions such as these require providing top managers with a variety of data transformations and representations.

Strategic information is often unstructured. Strategic decisions are taken after trying to answer questions such as "what will be the profit if I take a particular decision and what opportunities will be lost if the decision is not taken?" In complex decisions, many parameters will be involved. Identifying these and predicting their impact on a decision requires judgment coupled with analysis. For example, taking a strategic decision of whether to open a new branch or not would require the following information:

- Availability of finance
- Projected sales in the new branch
- Competition in the new location

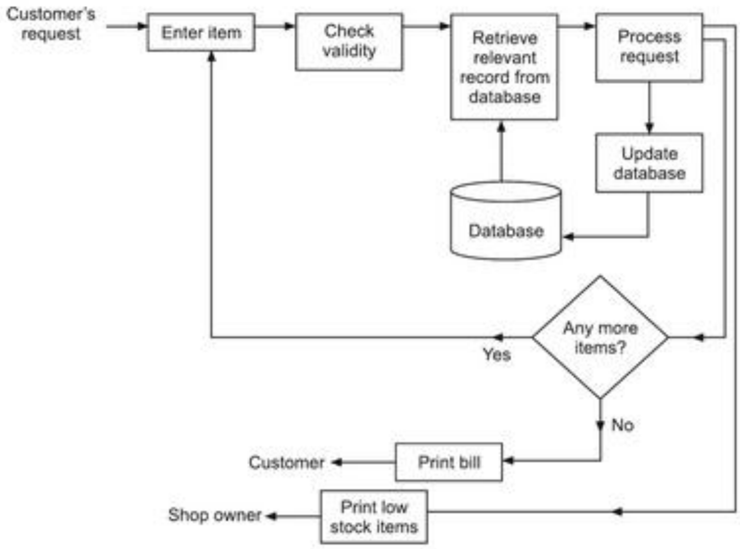


FIG. 16.3 Information system of a medical shop.

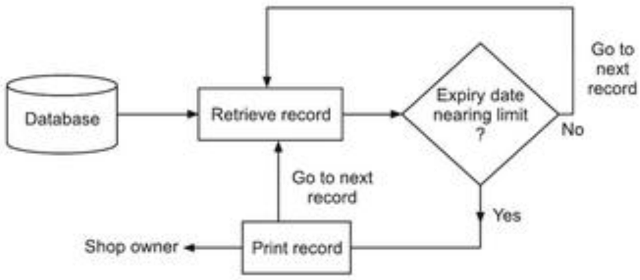


FIG. 16.4 Finding medicines nearing expiry date.

16.6 SYSTEM LIFE CYCLE

Developing an information system for businesses is a complex job. It is necessary to break up the development process into a number of phases. The terminology "System Life Cycle" is widely used to indicate the various phases of system development. Each phase has a conclusion, and a report is written when the phase is completed. It thus gives a convenient milestone to review the progress of the project and make modifications if necessary. We describe below the various phases in a system life cycle model.

16.7 COMPUTER SYSTEM FOR TRANSACTION PROCESSING

Earliest transaction processing systems used a centralized large computer called a *mainframe computer*. The mainframe computer was connected to a number of terminals, called dumb terminals (Fig. 16.6). These terminals were used to enter transactions and display results. The terminals were called dumb as they did not process data. This type of computing is now known as *client-server* computing. In the current systems, the server is either a mainframe or a powerful PC with multiple processors. The main advantage of such a system is easy maintenance of the server hardware and software, and better security. The main disadvantage is centralization. If the server fails, all work comes to a standstill. This type of structure is used for what are known as *back office functions* in organizations. An example is a bank. Banks have two functions: one is to cater to customers while they are waiting and the other is to balance the accounts at the end of the day. The second function is called a back office function as it is not interactive and is not as time critical as customer transactions. Banks and insurance companies which have to process vast quantities of data use mainframe computer as hosts. Mainframes are manufactured typically by IBM. They are expensive but highly reliable with reliable hardware and stable operating systems such as UNIX.

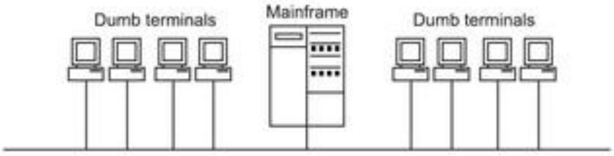


FIG. 16.6 Client-server computing.

With the advent of low cost PCs, there was a shift away from the mainframe—dumb terminals model. The cost of PCs is not much higher than that of dumb terminals. Further, PCs have reasonable computer power. Networks are also fast and can accommodate a large number of PCs. It was thus felt that some of the functions of organizations can be performed by PCs used as clients and others by a powerful PC used as a host or server. Let us take as an example a computerized branch of a small bank. The computers in the branch consist of PCs on the desk of each clerk and a server computer kept in the back office. All the computers are connected to a Local Area Network. The server holds the database with the accounts of all the account holders in the branch. This database is critical and has to be kept securely. It is also backed up every day in an archival store such as CDROM or tape. When a customer presents a cheque for encashment to a clerk, she will enter the account number and the amount given in the cheque in her PC. The PC will subtract the amount requested from the current balance and if the balance is above the minimum balance required for the account, accept the cheque. If not, it will reject the transaction. Updating the record of the customer (if the cheque is accepted) is done by the client. The updated record is sent to the server to be stored. It is thus seen that the server is relieved of doing some of the simple accounting functions, thereby reducing its load. Another client may have a passbook printer as a peripheral. When an account holder gives his passbook for updating, this client will extract the necessary data from the server and print it in the passbook. With the advent of "any branch banking", the design has changed. The account details of all the customers of all the branches are kept in a centralized server, normally a large mainframe in a central office. The client computers of all the branches are connected to this server using a high speed communication system. Thus a customer may present a cheque for encashment at any branch of the bank. This is convenient but if the communication line fails, the whole system fails. It is thus important for all branches to provide redundant communication lines to the server.

Thin Clients

Another trend in client-server computing is to replace PCs as clients by diskless intelligent devices. These devices have an excellent graphical user interface (GUI) similar to the GUIs of Windows PCs. All the computation is performed by the server. Thin clients are cheaper than PCs. If the server is very powerful and the communication speeds are high, these thin clients behave almost like PCs. Many organizations prefer thin clients instead of PCs as the system is more secure. Thin clients are cheaper to maintain.

16.7.1 Distributed Computing

In distributed computing, a number of servers are connected to a network. Databases are also distributed. In distributed computing, the general idea is to reduce the concentration of all power and data in one server. If there are several functional areas in a company and they are reasonably independent, each function and the related data are put in a server. All these servers are interconnected using a network. The functions must be identified so that there is very little need for exchanging data between servers. This allows faster overall transaction processing. The only care to be taken is to prevent ambiguity which may arise while updating databases by multiple clients simultaneously.

16.7.2 Data Archiving

When a business computerizes its operations, all paper records are discontinued. All records are in disk storage, either in a centralized server or in several computers if it is a distributed processing system. These records are precious. If they are corrupted either due to hardware/ software error or due to intentional tampering, the business will be badly affected. Imagine your account in a computerized branch being lost or tampered with. It is thus the responsibility of all computerized businesses to protect their data and restore them if they are accidentally lost. There are two approaches. One is to copy all data in the disk at the end of each day and keep it in a removable storage such as CDROM or tape. This is called a *backup*. As and when a transaction takes place during the day, these are also separately stored. Thus, using the database at the end of the previous day, and the transactions of the day, corrupted data can be restored. Special precautions are taken during design to prevent unauthorized change of data. A trail is kept of the persons who changed the data so that responsibility may be fixed. The other method is to have complete redundancy, that is, keep two identical disks in the system

6. The accounts office electronically pays for items accepted and sends the payment information to the vendor and the purchase office. Electronic payment is made by the accounts office by informing its banker (electronically) to debit its account by the specified amount and credit it to the vendor's bank account (which may be with a different bank). This is known as *Electronic Funds Transfer* (*EFT*), which is also an important aspect of e-commerce. Observe that it is not necessary to print a cheque and mail it.

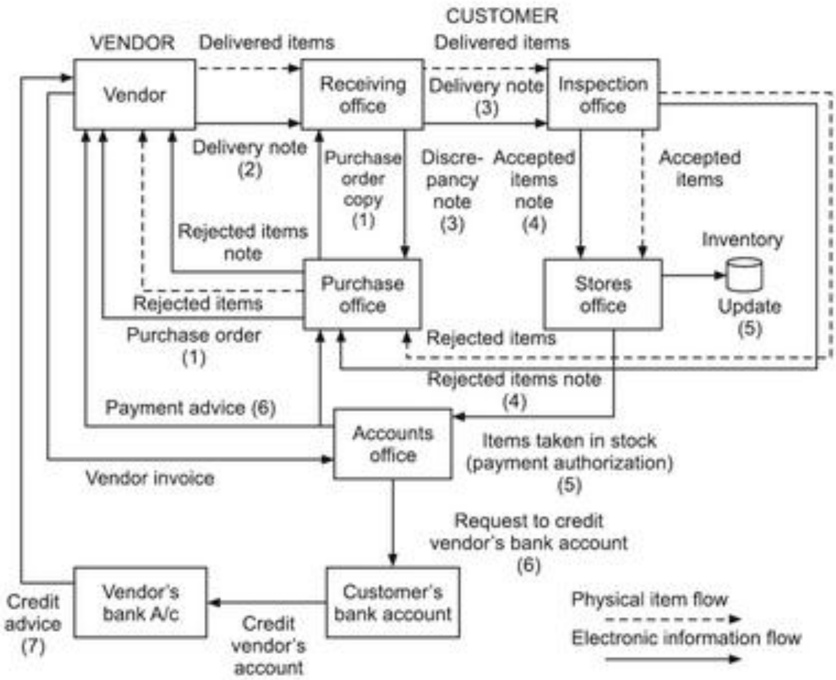


FIG. 17.2 Block diagram of document, item flow and funds transfer in B2B e-commerce.

From the above example, we see that the following hardware and software would be needed for establishing e-commerce between businesses:

1. Each of the businesses must have a LAN interconnecting its respective offices, and the offices themselves should have computers for data entry/receipt, comparison, etc. The system may be a distributed client/server type system with each office being a client and the databases being stored in appropriate servers. The internal system architecture is not a major issue. However, the protocol used by the LAN is usually the same as that used by the Internet, namely, TCP/IP. The organizational computer network using this protocol is called an *Intranet*. Besides using TCP/IP protocol, intranets also have one or more World Wide Web servers.

6. The accounts office electronically pays for items accepted and sends the payment information to the vendor and the purchase office. Electronic payment is made by the accounts office by informing its banker (electronically) to debit its account by the specified amount and credit it to the vendor's bank account (which may be with a different bank). This is known as *Electronic Funds Transfer* (*EFT*), which is also an important aspect of e-commerce. Observe that it is not necessary to print a cheque and mail it.

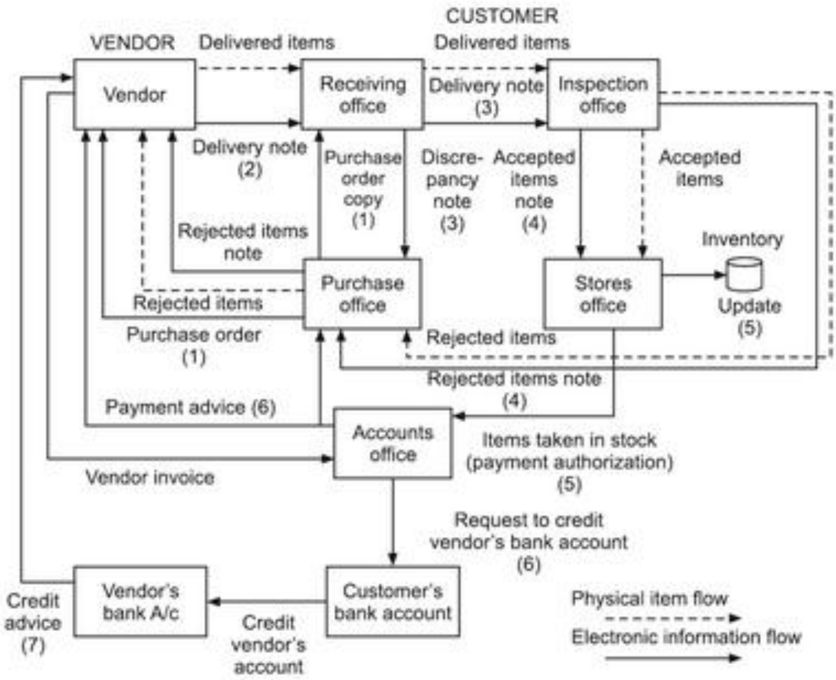


FIG. 17.2 Block diagram of document, item flow and funds transfer in B2B e-commerce.

From the above example, we see that the following hardware and software would be needed for establishing e-commerce between businesses:

1. Each of the businesses must have a LAN interconnecting its respective offices, and the offices themselves should have computers for data entry/receipt, comparison, etc. The system may be a distributed client/server type system with each office being a client and the databases being stored in appropriate servers. The internal system architecture is not a major issue. However, the protocol used by the LAN is usually the same as that used by the Internet, namely, TCP/IP. The organizational computer network using this protocol is called an *Intranet*. Besides using TCP/IP protocol, intranets also have one or more World Wide Web servers.

Table 17.1 Some Popular E-commerce Websites in India (Contd.)

Product	*Web address*	*Service provided*
Music/books	fabmart.com smashits.com phindia.com	Music, books, other goods Music Books by PHI Learning
Indian Railways	irctc.com	Reservation of seats/berths, Railway information
Jobs	monster.com	Placement service
Matrimonial	shadi.com	Find your partner
Real estate	magicbricks.com	Houses/flats for rent or sale
Business to Business	indiamarkets.com seekandsource.com	Listing of businesses, suppliers Industrial goods/sell
Free services	gmail.com formsindia.com	Free email site All types of forms for tax returns, etc.
Customer to Customer	ebay.com	Auction site
Education	nptel.iitm.ac.in	Free courses from IITs and IISc

17.5 ADVANTAGES AND DISADVANTAGES OF E-COMMERCE

We saw in Sections 17.2–17.4 that a large number of goods and services are now being sold/bought using the World Wide Web. The major advantages to a customer using e-commerce are:

1. One can buy/sell items from anywhere using one's computer provided an Internet connection is available.
2. Shopping can be done 24 hours a day, 365 days in a year—an Internet-based shop never closes.
3. One can avail of services such as financial services, legal services, medical advice, railway reservation, etc. from appropriate websites.
4. A wide variety of goods (particularly items like books and music) are accessible easily without one spending time and money in physically visiting and searching for the goods in many shops.
5. Anonymous friendly advice may be available on items one may like to buy/rent.

The advantages which accrue to a business by using e-commerce are:

1. With a website a business can reach out to a worldwide customer base at a very low cost.
2. Order processing cost is reduced as manual data entry is reduced. Business is also carried out faster as all documents are exchanged electronically.
3. Inventory size is reduced as transaction time is reduced.
4. Funds transfer is faster.

Procedure used by S to send a signed document X to R

Step 1 S encrypts the document X with R's public key and sends it to R after verifying R's public key certificate.

Step 2 S prepares a compressed version of the document X (called a message digest). The compression method must be such that, given the compressed document, it cannot be expanded back to the original document. It is also unique for each document. In other words, even if one character of X is changed, its compressed equivalent will be different. Let us call the compressed document CX.

Step 3 S encrypts the compressed document CX with his *private key* (thereby signing it) and sends it to R.

Step 4 R decrypts the compressed document CX using S's *public key*. R obtains S's public key certificate to make sure it is from S.

Step 5 R compresses the document X received from S in step 1 using the same compression method which S used. Let us call it CXS.

Step 6 R compares CXS with CX obtained in Step 4. If CXS = CX, then the document X is signed by S and is not tampered with. If CXS is not identical to CX, then failure is reported. It may be a forgery.

In Fig. 17.8, we show the above steps as a block diagram. In a computer-based system, all the above steps are automated. If S wants to send a signed document X to R, he enters X in his computer and invokes a dialog box to sign. This box asks the email address of R and his public key certificate. When they are entered (by a database lookup), it asks whether X is to be signed. If S clicks: Yes, then it is signed using Steps 1, 2 and 3. When R receives it, he will ask his computer to verify the signature. It will invoke Steps 5 and 6 to authenticate the signature or say it is a forgery.

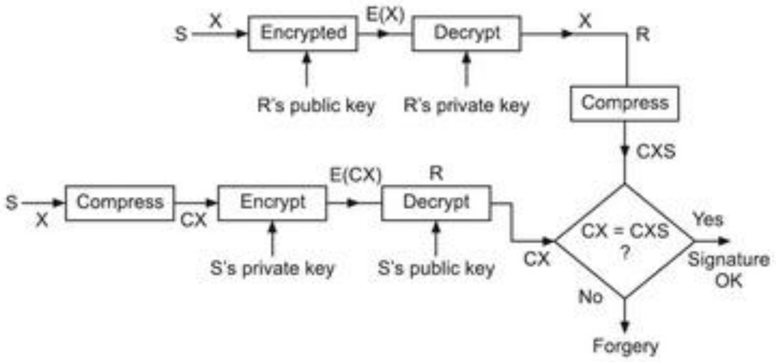


FIG. 17.8 Digital signature.

In Sections 17.6 and 17.7, we saw how transactions in e-commerce are protected and authenticated. Such a protection is essential to establish mutual trust between customers and account. You should check your bank account and send an email to PayPal acknowledging the receipt of the two amounts. This protocol is used by PayPal to authenticate your bank account details and your email address. After authentication your credit card or bank account will be used by PayPal to make payments. Once you create a PayPal account, you can authorize payment to any other PayPal account holder who may be an individual, or a merchant. When you want to pay to an individual, you log on to the PayPal website with your password. A screen will appear on your monitor which will ask you to specify the (PayPal registered) email address of the person or merchant to whom you want to pay and the amount to be paid. Once you enter these and click the pay button, the amount will be transferred from your account to that of the recipient you specified. You will receive an email confirmation of the transaction. The main advantage of using PayPal is that you do not expose your credit card or bank account details to strangers. You expose it only to PayPal which you trust as it is a reputed company. There is no charge for opening an account. PayPal charges the recipient a percentage of the amount received by him. PayPal is owned by eBay, a well known auction site used in C2C e-commerce. PayPal is an international company and payment may be made in different currencies subject to local laws. For example, you may pay for a purchase in dollars up to a limit set by the Reserve Bank of India. PayPal will take money from your rupee account at the existing exchange rate plus commission.

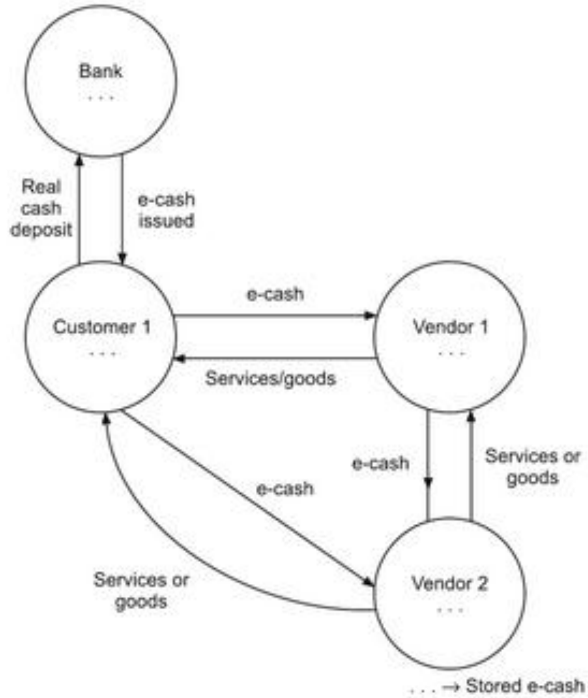


FIG. 17.12 Circulation of e-cash.

Step 2 The vendor decrypts the information using its private key, checks the purchaser's certificates, signature and payment advice, attaches its deposit slip with its account number, and endorses the deposit attaching its public key certificates. This is encrypted and sent to its bank.

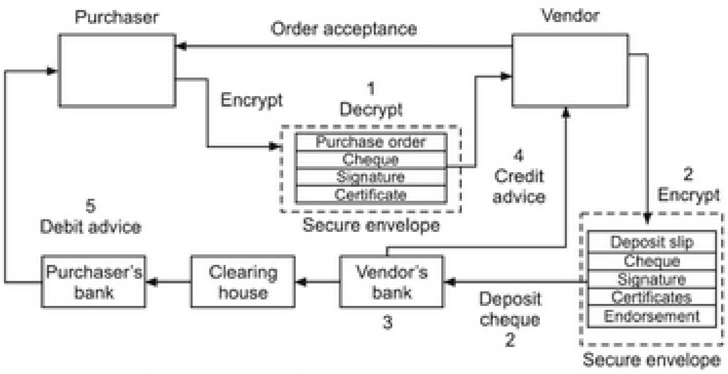


FIG. 17.10 Electronic funds transfer in B2B e-commerce.

Step 3 The vendor's bank checks the signatures and certificates and sends the pay order for clearance to the clearing house.

Step 4 When the pay order is cleared, the amount is credited to the vendor's account and a credit advice is sent to it.

Step 5 The purchaser gets a consolidated debit advice periodically.

We have not described the signing process in detail as it has been described already.

17.10 CASH TRANSACTIONS IN E-COMMERCE

In the previous sections we discussed payment by credit cards and electronic funds transfer in e-commerce. The amount to be paid in each transaction using credit card/electronic funds transfer is quite large and the cost of processing the transaction is high. Thus this procedure is not appropriate for low value transactions. We normally use cash for many transactions, particularly low value transactions. The currency system as we know it today evolved over centuries and is time tested. The major advantages of cash are:

- It is guaranteed by the government and does not normally lose its value in the short range (except when there is hyperinflation).
- It is universally recognized as having value and accepted as legal tender.
- It can be carried around.
- Any person having cash can exchange it for goods or services without the help of a third party such as a bank.
- It is anonymous. Traders cannot normally say who gave a particular currency note.
- Privacy of transactions is ensured because of anonymity.

The major disadvantages of cash are:

- It is not safe to carry around (especially if it is a big amount). If you lose your purse, you will be lucky to get your cash back.
- It is bulky. Governments do not normally print large denomination currency notes to prevent criminals from stealthily transporting large amounts of cash. Government of India has only recently started printing ₹1000 currency notes. Twenty years ago the largest denomination currency note was only ₹100 to reduce black money transactions.

The purpose of electronic cash (abbreviated e-cash) is to mimic cash transactions with all its advantages without its disadvantages. The major problems, however, are:

- Who will issue *e-cash*? If it is a private agency like a bank, will it guarantee the safety of one's e-cash? Deposits in scheduled banks are normally insured by Reserve Bank (up to a specified limit).
- Will anonymity of e-cash transactions be ensured? Should it be ensured?
- Will e-cash issued by a private agency be universally acceptable? Can two individuals exchange e-cash without a third party entering into a transaction? (This is a major advantage of cash. I can borrow hundred rupees from a friend without the Reserve Bank knowing it!)
- How will it be possible to detect forgery?
- A person who "buys" e-cash should not be able to spend the same cash again. Once it is exchanged for a service, it must lose its value or be taken out of circulation.
- How is the cost of handling e-cash recovered? In the case of physical cash, cost is recovered by governments.

Currently the technology of cash transactions in e-commerce has not matured. Many competing systems are being tried out, each with its own advantages and disadvantages. Governments do not want to allow large cash transactions to take place electronically.

We will now describe a simple method which has been used for e-cash transactions. It is being used by some banks in the United States and Europe. No such system is in place in India as of now. It is primarily intended for small cash transactions. The procedure is as follows (see Fig. 17.11).

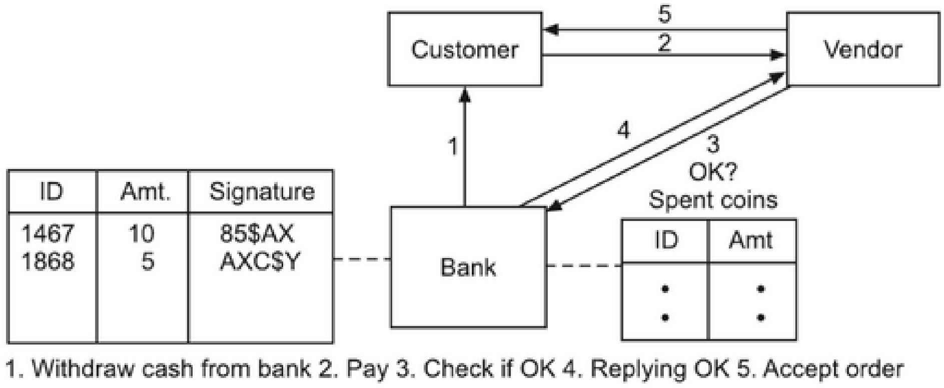


FIG. 17.11 Electronic cash payment.

Step 1 A customer withdraws *e-coins* in various denominations from the issuing bank (or financial institution) and stores it in his PC. The withdrawal takes place by the customer giving a unique identification number and denomination of each coin and requesting the bank to digitally sign it. The bank signs a coin by encrypting (id#,denomination> with its private key. The signed e-coins are of the form <id#,denomination, bank's signature>

Step 2 The customer pays a vendor for goods ordered using the signed e-coins.

Step 3 The vendor sends the e-coins to the issuing bank for authorization.

Step 4 The bank checks whether the e-coin is signed by it and whether it has not been spent already. If it is a valid e-coin, it OKs the transaction and credits the amount to the vendor's account. It puts the e-coin details in a spent e-coin database so that if the e-coin is presented again, it can dishonour it.

Step 5 The vendor supplies the goods ordered.

The communications between customer, vendor and the bank should be encrypted as the Internet is used. As the amounts involved are small, symmetric cryptography is used for these communications as it is faster. There are two points which need clarification. The first is the cost of servicing e-coins. Normally banks will charge a small commission for the service from vendors. The second point is whether a vendor who receives an e-coin from a customer can use it to purchase goods from another vendor. This is not possible as the issuing bank has to authenticate the e-coin and, while doing it, it marks the coin as "spent". Thus, it is not really like good old cash! To mimic real cash, e-cash should have features shown in Fig. 17.12.

The simple protocol used above does not preserve the anonymity of cash. The bank will know which customer and vendor are involved in the cash transaction and can link the two. There is another protocol called transaction blinding in which it is possible for a customer to get e-coins issued by a bank without revealing his identity. The protocol called Chaum's blinding protocol is complicated, and as of now, not used widely.

17.11 PAYMENT IN C2C E-COMMERCE

In C2C e-commerce payment is made between individuals using the Internet. A customer cannot receive credit card payment or payment by electronic funds transfer as he will normally not have the infrastructure to process the payment. Thus an intermediary is required. Such a service is provided by a company called PayPal.com. There are other companies; we consider PayPal as an example. In order to use PayPal, both parties who want to make or receive payments must have a PayPal account. Creating a PayPal account is simple. You log on to PayPal's website and enter your email id, name, address, and credit card or bank account details. Besides this you have to create a password for future use. PayPal will check your credit card details with the issuing bank and if it is OK, it will send you a confirmatory email with a unique id. You acknowledge this mail giving the id sent to you by PayPal. If you gave your bank account details and not the credit card information, PayPal will deposit two small amounts in your bank

account. You should check your bank account and send an email to PayPal acknowledging the receipt of the two amounts. This protocol is used by PayPal to authenticate your bank account details and your email address. After authentication your credit card or bank account will be used by PayPal to make payments. Once you create a PayPal account, you can authorize payment to any other PayPal account holder who may be an individual, or a merchant. When you want to pay to an individual, you log on to the PayPal website with your password. A screen will appear on your monitor which will ask you to specify the (PayPal registered) email address of the person or merchant to whom you want to pay and the amount to be paid. Once you enter these and click the pay button, the amount will be transferred from your account to that of the recipient you specified. You will receive an email confirmation of the transaction. The main advantage of using PayPal is that you do not expose your credit card or bank
that of the recipient you specified. You will receive an email confirmation of the transaction. The main advantage of using PayPal is that you do not expose your credit card or bank
account details to strangers. You expose it only to PayPal which you trust as it is a reputed company. There is no charge for opening an account. PayPal charges the recipient a percentage of the amount received by him. PayPal is owned by eBay, a well known auction site used in C2C e-commerce. PayPal is an international company and payment may be made in different currencies subject to local laws. For example, you may pay for a purchase in dollars up to a limit set by the Reserve Bank of India. PayPal will take money from your rupee account at the existing exchange rate plus commission.

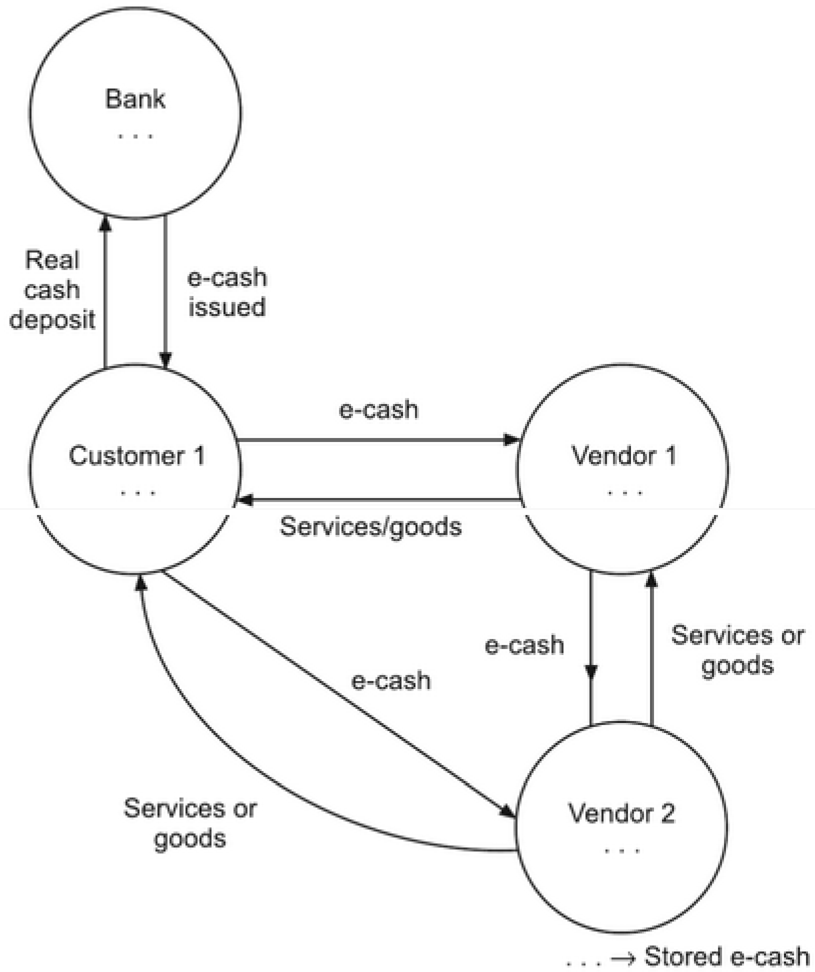


FIG. 17.12 Circulation of e-cash.

17.12 ELECTRONIC DATA INTERCHANGE

In Section 17.3, we stated that in Business to Business (B2B) e-commerce, electronic documents are exchanged between business partners using either a private network or a public switched network. We also stated that in order to interpret them correctly, we need a standard notation which is agreed upon by both parties. This is called Electronic Data Interchange (EDI). EDI is defined as the exchange of business documents between organizations in standardized electronic form which can be interpreted and used directly by application programs. The major advantages of using EDI are:

1. Handling of paper documents is eliminated.
2. There is no need to manually re-key data in documents such as purchase orders, invoices, etc. by participating businesses.
3. Elimination of manual data entry reduces cost, improves accuracy and ensures reliability.
4. Time is saved due to elimination of manual handling and also due to direct application-to-application movement of data at electronic speeds.

17.12.1 Electronic Data Interchange Standards

We will now describe the steps a company A should follow to establish an EDI partnership with company B (see Fig. 17.13).

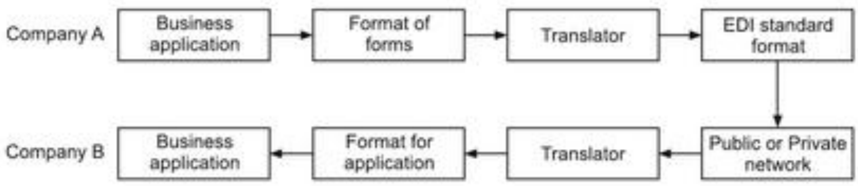


FIG. 17.13 Steps in electronics data interchange.

1. The first step is to agree on a standard format for commonly used documents such as purchase order, invoice, payment advice, delivery note, etc. The formatting information or Data Type Definition, as it is called, should include description of various fields used such as quantities, price, currency used, delivery date, field lengths, character type, ordering of fields in the document, units used, etc. As companies may transact business with many partners, it is desirable to have a universally agreed standard form for all business documents. This realization led to industry groups such as automobile industry, shipping and transport industry to adopt standards for intercompany transactions. These standards later evolved into national and international standards. The two standards are ANSI X.12 standard adopted by the American National Standards Institute for electronic transactions in the United States of America and EDIFACT (Electronic Data Interchange for Administration,

17.12.3 XML for EDI

As we have pointed out in Section 17.12.1, implementing and operating EDIFACT or ANSI-based EDI system is inflexible and expensive. Thus, most businesses, particularly the small ones which would like to participate in B2B e-commerce, require a cheaper alternative which is easy to implement and which uses the Internet rather than a VAN for communication. The rapid growth of the Internet with increase in bandwidth and availability of faster processors has made efficiency less important as compared to flexibility, ease of understanding, and good documentation. This has led to the use of XML format for document exchange between businesses (see Chapter 13 for description of XML).

When a company uses XML to describe business documents, it will also give a set of statements which define the syntax of the XML program. It is called a *Document Type Definition* (DTD). This will be published in the company's website so that any application program that wants to use the XML document can download and interpret the XML document correctly.

17.13 INTELLECTUAL PROPERTY RIGHTS AND ELECTRONIC COMMERCE

The advent of the Internet, which allows easy copying and distribution of all types of data, text, audio and video, has made publishers of books and producers of music and films anxious. The copyright issue has also plagued the spread of digital library movement. The major problem with content available in digital form is the ease with which they can be copied. Digital data can flow across national boundaries freely at electronic speed and it is practically impossible to monitor and control illegal copying. Many content providers encrypt material they store in the web to prevent easy access and copying. Such encryption impedes free flow of data and the "fair use doctrine" which governs the existing copyright laws for print media, CDs, and DVDs. Hackers have been trying to find methods of decrypting encrypted material. In a new landmark law enacted by the United States Congress in October 1998, called the Digital Millennium Copyright Act, it has been made illegal to circumvent access controls used by copyright owners to protect their work and *even to develop technologies* which may be used to circumvent protection. There is raging debate on this issue as it seems to prevent "fair use", which is the basis of agreements arrived at in the World Intellectual Property treaty. The copyright issues get highly complicated when it comes to computer software. The issues are not yet fully resolved and it is still being argued by lawyers and ethicists.

17.14 INFORMATION TECHNOLOGY ACT

India's Parliament passed the Information Technology Act 2000, which provides the legal infrastructure for e-commerce in India. It was amended in 2008 to rectify some loopholes in the earlier act passed in 2000. The object of the Act has been stated as:

To provide legal recognition for transactions carried out by means of EDI and other means of electronic communication, commonly referred to as e-commerce, which involves the use of alternatives to paper based methods of communication and storage of information, to facilitate electronic filing of documents with the Government agencies and further to amend the Indian Penal Code, the Indian Evidence Act, 1891, the Banker's Book Evidence Act, 1891 and the Reserve Bank of India Act 1934, and for matters connected therewith or incidental thereto.

This Act is a landmark Act which now gives legal status to email correspondence and soft copies of documents. The major interesting aspects of the Act are:

1. Email correspondence has legal status and therefore it can be used in evidence. Electronically signed documents are now recognized.
2. A controller of public key certifying authorities has been appointed by the government. This office will recognize certifying authorities who will have the authority to issue public key certificates and verify digital signatures.
3. All applications to government bodies can be filed in electronic form. Government can issue licences, permits, sanctions, approvals, etc. on-line in electronic form.
4. Many archival documents which companies and government departments are required, by law, to be kept for a specified period can now be stored in CDROM or tapes, saving precious space in buildings and enabling easy retrieval. Care must be taken to see that such electronically stored documents keep details which will identify the origin of the document, date and time of creation, dispatch, or receipt.
5. The IT Act provides statutory remedy to companies whose networks are illegally accessed and stored data is stolen or damaged. Monetary claims can be made against intruders.
6. The Act provides punishment to a person who commits the following computer related offences:
   - (i) downloads, copies or extracts data from a database without the permission of the owner;
   - (ii) introduces any soft-contaminant or computer virus into any computer or computer network;
   - (iii) damages programs or data residing in a computer or network or illegally copies them;
   - (iv) disrupts a computer or a network;
   - (v) denies access to a computer or a network by authorized persons;
   - (vi) charges for services availed of by a person to another person by tampering with or manipulating accounts in a computer or a network;
   - (vii) sends out bulk unsolicited email known as spam; and
   - (viii) distributes pornography or material considered generally objectionable.

Hacking or, more accurately, a computer related offence such as creating or distributing a computer virus or defacing a website has now been classified as a crime under the Indian Penal Code. Punishment for such an offence is imprisonment of up to 3 years or fine up to ₹2 lakhs, or both. Teenagers who commit such offences "for fun" should realize that they will have fun in jail up to 3 years!

27. Cheque payments are made between organizations using digitally signed cheques and public key certificates issued by a certifying authority.

28. Payment for small transactions is made using digital coins issued by banks to customer after debiting the customer's account. A digital coin consists of amount, identification number and bank's signature. These coins are given in exchange for goods. The bank reimburses the vendor after checking its signature and ensuring that the coin has not been spent earlier.

29. When multimedia data is stored in the World Wide Web several copies can be made easily. If it is done without the author's permission it infringes his/her intellectual property rights. If it is encrypted free access is denied.

30. Information Technology Act was passed in 2000 to promote e-commerce. It recognizes public key certification, digital signature and soft documents. Persons who hack into other's web pages, distribute viruses or otherwise vandalize the internet can be sent to jail up to three years on criminal charges under the Act.

EXERCISES

17.1 Define e-commerce. What are the different types of e-commerce?

17.2 Explain B2B e-commerce using an example of a book distributor who stocks a large number of books, which he distributes via a large network of book sellers. Assume that the distributor has stocks of books of a large number of publishers and book sellers order books as and when their stock is low. Distributors give one month's time to booksellers for payment.

17.3 Explain B2C e-commerce of a customer reserving airline tickets from his home or place of work.

17.4 Explain C2C e-commerce with an appropriate example.

17.5 List the advantages and disadvantages of e-commerce.

17.6 Explain the system architecture of e-commerce by looking at it as a set of layers with the physical network as the bottom layer and applications as the top layer.

17.7 What do you understand by EDI? Is EDI used in B2C or B2B e-commerce? Why is EDI important in e-commerce?

17.8 What are two major EDI standards used in e-commerce? Which is the standard accepted for government transactions in India?

17.9 How can XML be used for EDI?

17.10 What is the advantage of using XML as compared to using EDIFACT standard.

17.11 What is VAN? What services do VANs provide? What are the advantages and disadvantages of VAN?

17.12 If internet is to be used for EDI, which mail standard is used?

17.13 If e-mail is to be used to exchange EDI between two businesses, what are the points on which they should agree?

17.14 Why is security important in e-commerce? What are the security issues to be taken into account while designing a security system for e-commerce?

17.15 What do you understand by symmetric key cryptography? What are the main advantages and disadvantages of symmetric key cryptography?

17.16 What is public key encryption? In what way is it different from private key encryption? Why is it important in e-commerce?

17.17 What are the main differences between DES based encryption and RSA based encryption? Is it possible to combine these two systems? If so, explain how.

17.18 What is a digital signature? Why is it necessary in e-commerce? What are the necessary conditions a compression function used in digital signature should satisfy?

17.19 Give a block diagram of a system for transmitting a signed purchase order from business 1 to business 2.

17.20 What is a public key certifying authority? Why is a certifying authority required in e-commerce? How does a certifying authority perform its tasks?

17.21 What types of electronic payment systems are required in e-commerce? Why are there different types of payment systems? Explain the necessary characteristics of each type of payment system and give an example each of where it is used.

17.22 Explain SET protocol, used in credit card transactions. What is the main interesting aspect of SET protocol which gives confidence to customers transacting business using the internet?

17.23 In using SET protocol who has to keep a database of public keys of all customers? How is the customer assured that he will not be double charged for the same item purchased?

17.24 What are the main characteristics of cash payment in contrast with cheque payment? Why are governments not sympathetic to large cash transactions in e-commerce?

17.25 Explain how cash transactions take place in e-commerce. What special precautions should be taken by a bank to ensure that a customer does not double spend the same electronic coins issued to him/her?

17.26 What are intellectual property rights?

17.27 How can your intellectual property rights be protected if you place a document in the World Wide Web?

17.28 What are the objectives of the Information Technology Act 2000?

17.29 Are e-mail correspondence admissible in courts of law?

17.30 Are soft copies of agreements legally admissible?

17.31 Who is an electronic vandal? What is the punishment for such vandals in the IT Act?

18

SOCIETAL IMPACTS OF INFORMATION TECHNOLOGY

LEARNING GOALS

In this chapter we will learn:

1. How information technology is affecting our daily life.
2. The need to be conscious about protecting data stored in your computer from vandals and thieves.
3. Protecting your computing resources and being able to recover from disasters.
4. About intellectual property rights and how to avoid infringing them.
5. The new opportunities of employment in information technology to improve our quality of life.

18.1 INTRODUCTION

We have seen in this book so far how data is gathered, entered into a computer's storage, how it is stored, processed and finally how the processed data, namely, information is distributed. You have to remember that data is not just numbers. It is also textual characters, audio, graphics and video. The processed data (information) is interpreted and used by us for various applications. The use of computers has grown enormously, particularly in the last 20 years after the advent of inexpensive desktop personal computers. PCs brought affordable processing power and storage to everyone. Isolated PCs are very useful by themselves, but interconnecting them introduces another dimension in their possible applications. In fact, today it is hard to find an isolated PC. The PC in your home will normally have a connection to an ISP via a modem. PCs in offices are connected to a LAN which, in turn is connected to an extranet or internet. The world wide interconnection of PCs has opened up our society to both good and bad influences. It is for us to use the available resources wisely to our advantage.

The day to day living in urban areas, particularly in big cities, is slowly but steadily being changed by information technology in the following ways:

1. Tickets for trains and long distance buses are being issued by computerized booking offices which are located throughout a city. This not only reduces the time needed to go to the booking office but also the cost of bus or autofare. Early booking of tickets gives you a choice of seats and allows you to plan your journey ahead of time.
2. Many banks are providing home-banking by permitting you to look at your balance and accounts statement via the internet.
3. Newspapers all over the world have a web presence and allow you to look at day's headlines and news abstracts.
4. There are many educational institutions which provide learning material such as lecture notes, problems, etc., in their web sites. An example is the Massachusetts Institute of Technology, USA providing course materials for several subjects in their web site, free of cost.
5. Individuals store music in MP3 format in their computers and permit anyone to access these files.
6. Many governments are taking the so-called e-governance initiatives. Information about a piece of land which you want to buy, such as current market value and the amount to be paid to register it, is made available in their websites. The registration will also be done using computers in sub-registrar's office, the document scanned and stored in a CD with all the physical signatures, thumb prints, etc. The computerized procedure takes only 30 minutes to register a piece of land.
7. E-governance projects now plan to allow citizens to log on to web sites of government departments to download application forms, get copies of latest rules, and also to track the status of their letters to departments/ministers.
8. Many courts are now computerizing and allowing copies of judgements to be downloaded from the court's web site.

Besides the above, computers are influencing our lives in many other ways:

(a) Many appliances such as washing machines, microwave ovens, etc., have built-in microprocessors called *embedded computers* to control them.
(b) All modern motor cars have microprocessors which control petrol combustion to maximize power and minimize harmful emissions.
(c) Most modern medical instruments are computer based starting with simple instruments such as blood sugar monitors to sophisticated instruments such as CAT scan machines. In fact, doctors and surgeons depend very heavily on computer technology.
(d) Every cell phone you use has one or more special purpose computers.

In fact, in your home you may not have a PC but many embedded computers; in the washing machine, in the music system, in the DVD player, the microwave oven, the cell phone and in your car. The total number of embedded computers in the world will be 1000 times the number of PCs. Thus, writing programs for embedded computers and developing systems using them is a thriving business.

We thus see that in many aspects of our life, we directly or indirectly use computers and information technology to improve our quality of life. In some sense we are becoming dependent on this technology in our day-to-day life. In fact, our methods of working, interaction with others and leisure activities have changed significantly after the advent of computers and information technology.

18.2 PRIVACY, SECURITY AND INTEGRITY OF INFORMATION

By *privacy* we mean providing confidentiality to your personal data. For example, you may not like your medical history to be made public or your personal correspondence with friends to be leaked out to all. You may not also like everyone to know which web sites you visit often and what are your private buying preferences. All the above data should be kept confidential. Unfortunately, in the internet age it is very easy for unscrupulous persons who are bent upon using your personal information to gain access to it. Thus, privacy laws are required to protect your personal data from being used without your written consent. Many countries are enacting such laws.

By *security* we mean protecting the data and programs against accidental or intentional destruction or disclosure by unscrupulous persons. You should take proper precautions to prevent theft or corruption and, in case you do lose data you should be able to recover it. We will discuss this in greater detail in the next section.

Information integrity is concerned with the information provided to you being complete, trustworthy, timely, being up-to-date and relevant. In other words, it is concerned with the quality of information, and not with its safety. Today we are more and more dependent on data processed by computers and take actions based on processed data. If the input data is corrupted or there is an undetected error in the processing rules, or there is tampering of processed data, there is a loss of integrity in the information. As this information is used to take decisions and initiate actions, it is extremely important to ensure its integrity.

There must also be a means of finding out if there is loss of integrity. Tools are now being developed to ensure information integrity.

18.2.1 Computer Security

If you are using a computer at home or in the office, you should take special precautions to protect your password and files from being stolen or corrupted. Many persons use passwords which are easy to guess such as their name, date of birth, sister's name, and words such as cat, dog, etc. If a person is bent upon finding out your password, he will try to run a program with a number of guesses for a few days till he succeeds in finding it. It is advisable to use the following rules to create a password:

- Do not use a short password. Use at least 8 characters in the password.
- Use at least one special character and some digits in the password.
- Use a set of random characters.
- Change your password at least once a month.

A possible secure password is: rj; 42mn$. It is difficult to remember such a password. Note it down somewhere in your diary. After you use it a few times, you will remember it.

1. Tickets for trains and long distance buses are being issued by computerized booking offices which are located throughout a city. This not only reduces the time needed to go to the booking office but also the cost of bus or autofare. Early booking of tickets gives you a choice of seats and allows you to plan your journey ahead of time.
2. Many banks are providing home-banking by permitting you to look at your balance and accounts statement via the internet.
3. Newspapers all over the world have a web presence and allow you to look at day's headlines and news abstracts.
4. There are many educational institutions which provide learning material such as lecture notes, problems, etc., in their web sites. An example is the Massachusetts Institute of Technology, USA providing course materials for several subjects in their web site, free of cost.
5. Individuals store music in MP3 format in their computers and permit anyone to access these files.
6. Many governments are taking the so-called e-governance initiatives. Information about a piece of land which you want to buy, such as current market value and the amount to be paid to register it, is made available in their websites. The registration will also be done using computers in sub-registrar's office, the document scanned and stored in a CD with all the physical signatures, thumb prints, etc. The computerized procedure takes only 30 minutes to register a piece of land.
7. E-governance projects now plan to allow citizens to log on to web sites of government departments to download application forms, get copies of latest rules, and also to track the status of their letters to departments/ministers.
8. Many courts are now computerizing and allowing copies of judgements to be downloaded from the court's web site.

Besides the above, computers are influencing our lives in many other ways:

(a) Many appliances such as washing machines, microwave ovens, etc., have built-in microprocessors called *embedded computers* to control them.
(b) All modern motor cars have microprocessors which control petrol combustion to maximize power and minimize harmful emissions.
(c) Most modern medical instruments are computer based starting with simple instruments such as blood sugar monitors to sophisticated instruments such as CAT scan machines. In fact, doctors and surgeons depend very heavily on computer technology.
(d) Every cell phone you use has one or more special purpose computers.

In fact, in your home you may not have a PC but many embedded computers; in the washing machine, in the music system, in the DVD player, the microwave oven, the cell phone and in your car. The total number of embedded computers in the world will be 1000 times the number of PCs. Thus, writing programs for embedded computers and developing systems using them is a thriving business.

We thus see that in many aspects of our life, we directly or indirectly use computers and information technology to improve our quality of life. In some sense we are becoming dependent on this technology in our day-to-day life. In fact, our methods of working, interaction with others and leisure activities have changed significantly after the advent of computers and information technology.

18.2 PRIVACY, SECURITY AND INTEGRITY OF INFORMATION

By *privacy* we mean providing confidentiality to your personal data. For example, you may not like your medical history to be made public or your personal correspondence with friends to be leaked out to all. You may not also like everyone to know which web sites you visit often and what are your private buying preferences. All the above data should be kept confidential. Unfortunately, in the internet age it is very easy for unscrupulous persons who are bent upon using your personal information to gain access to it. Thus, privacy laws are required to protect your personal data from being used without your written consent. Many countries are enacting such laws.

By *security* we mean protecting the data and programs against accidental or intentional destruction or disclosure by unscrupulous persons. You should take proper precautions to prevent theft or corruption and, in case you do lose data you should be able to recover it. We will discuss this in greater detail in the next section.

Information integrity is concerned with the information provided to you being complete, trustworthy, timely, being up-to-date and relevant. In other words, it is concerned with the quality of information, and not with its safety. Today we are more and more dependent on data processed by computers and take actions based on processed data. If the input data is corrupted or there is an undetected error in the processing rules, or there is tampering of processed data, there is a loss of integrity in the information. As this information is used to take decisions and initiate actions, it is extremely important to ensure its integrity.

There must also be a means of finding out if there is loss of integrity. Tools are now being developed to ensure information integrity.

18.2.1 Computer Security

If you are using a computer at home or in the office, you should take special precautions to protect your password and files from being stolen or corrupted. Many persons use passwords which are easy to guess such as their name, date of birth, sister's name, and words such as cat, dog, etc. If a person is bent upon finding out your password, he will try to run a program with a number of guesses for a few days till he succeeds in finding it. It is advisable to use the following rules to create a password:

- Do not use a short password. Use at least 8 characters in the password.
- Use at least one special character and some digits in the password.
- Use a set of random characters.
- Change your password at least once a month.

A possible secure password is: rj; 42mn$. It is difficult to remember such a password. Note it down somewhere in your diary. After you use it a few times, you will remember it.

- *Linux operating system* for PCs and other computers. Linux is free, it has fewer bugs compared to Windows OS, and is not normally attacked by viruses as it has better security. Software (called drivers) to connect a variety of peripherals with computers running Linux OS is not easily available as the number of Linux users is small. However, Linux is gaining popularity with support coming from many major vendors like IBM. A version of Linux called *Ubuntu* is now widely used as it has a good user interface and drivers are available for most peripheral devices for Ubuntu.
- *Open Office* software is similar to MS Office. It is a set of programs for word processing, spreadsheet, database, etc. It is also gaining many users.

One of the major problems with open source software is vendor support. As it is free, no one has any responsibility to fix problems. Thus, you must have the expertise to solve your own problems. As this is not possible for many users, companies have emerged to package open source software and provide manuals and support. A successful company is Red Hat which supplies Linux called *Red Hat Linux* and supports users. They charge a fee for this service.

There are also many programs available in the World Wide Web for you to download free of cost, but they should be only for personal use. You cannot distribute them for profit. Widely used software of this type is Java interpreter which can be downloaded from the web.

18.6 CAREERS IN INFORMATION TECHNOLOGY

Information Technology (IT) is a growing area. It touches everyone's life and its use in our day-to-day work is increasing. It is also an area which is currently one of India's largest export earners. Indian Information Technology professionals are accepted in the international market as highly competent. We have a very high potential in this area and newer opportunities are being created rapidly.

With the above background, you may be wondering what types of careers are available in Information Technology and related applications. We can broadly divide the career opportunities into four areas:

- Use of IT in daily work.
- Specialized applications of IT and careers which are emerging to provide services using these applications. They are called IT enabled services industry.
- IT professionals who provide software services such as testing programs and developing sophisticated applications software and software products.
- IT professionals who maintain both the hardware and the software of customers and sell systems and upgrades.

18.6.1 Users of Information Technology

Computers are being installed and used everywhere—all offices, banks, shops, courts, etc. use computers. Employees such as clerks in offices, stenographers, all managers, bank employees will have to use computers in their work. They should know the basics of how to use the OS of these machines, filing, sending and receiving emails and using office software such as word processors, spreadsheet and database. General knowledge of IT, as given in this book, will be very useful for them to demystify computers.

Besides non-technical persons, other professionals such as engineers, doctors, lawyers, and accountants also must have good knowledge of applying IT effectively in their respective disciplines. Engineers must know how to use engineering design tools. Doctors must know to operate computers to retrieve patient records, surf the web for the latest medical information, and use expert medical diagnostic systems provided as software packages. Lawyers should be able to access case histories, use database of judgments to access relevant case laws, and use the World Wide Web to obtain forms, rules, etc. Accountants must know how to use accounting packages, spreadsheets to analyze returns on investments, and database systems to store and retrieve accounts information. Income tax returns are required to be filed on the website of the income tax department. Paper filing is not allowed. In fact, IT is influencing every discipline. Musicians and composers use computers with MIDI interface to synthesize music and experiment with tunes. Authors find word processors invaluable and artists use paintbrush software to assess various patterns and colours. Thus today all educated persons must have a working knowledge of IT.

18.6.2 IT Enabled Services and Careers

A number of new job opportunities are emerging with the advent of low cost sophisticated multimedia computers. Many of these are older jobs which have been transformed because of IT. Some of these are:

- Printing and publishing
- Music production
- Photo studios
- Movie production

Printing and Publishing

Before the advent of computer-based Desktop Publishing (DTP), traditional printers utilized typesetting machines which used hot lead. Besides being expensive, printing was slow. Proofreading and corrections were laborious and expensive. Typesetting has now been replaced by DTP. Nowadays authors are expected to submit a soft copy of their manuscript. This is used by copy editors and DTP operators to produce a book with a professional look. In fact, all printed publications such as newspapers, annual reports, professional journals, newsletters, etc. use DTP.

Besides traditional printed publications, a new career of e-publishing is emerging. This is publishing on the World Wide Web. There are two broad classes of web publishing:

- E-books and e-journals
- Web designing

The advantages of publishing e-books and e-journals are speed of publication and low cost. E-books are becoming popular with the advent of good e-readers such as Amazon's Kindle reader and Apple Computer's iPad which can connect to the e-book distributor using the mobile telephone network. E-books are cheaper than print version. E-journals are popular

2. Even though desktop computers are everywhere, small microcomputers embedded in all consumer products outnumber them probably by a factor of 1000.
3. Societies all over the world are changing their methods of working, interaction with others, entertainment, information search, and education after the advent of information technology.
4. The advent of the World Wide Web has led to several societal applications of computers. Three of the most important applications are: free services, content creation and sharing, and social networking services.
5. Free services are email, free storage, maps, photo sharing sites and video conferencing (Skype). These services earn revenue from advertisements.
6. Contents are created as Wikis, YouTube, Blogs, and Sound Cloud. Wikipedia is an encyclopedia created by volunteers. Videos may be posted by any one on YouTube. Audio (e,g., music albums created by amateurs) may be posted on Sound Cloud.
7. Social Networking Sites allow members to form groups who can interact using the facilities provided by the sites. Facebook is the largest such service. Linkedin connects professionals. Twitter allows members to post short (140 characters) messages called tweets using its website.
8. With the widespread use of computers and the Internet, problems of ensuring privacy, security, and integrity of information are becoming important.
9. By privacy of data we mean providing confidentiality to your personal data and preventing its use without your consent.
10. By security of data we mean protecting data of businesses (and that in your home computer) from accidental or intentional loss.
11. By information integrity we mean ensuring the quality and reliability of data processed by computers.
12. Ensuring privacy is not merely a technical problem; it is a moral, legal and ethical issue.
13. Security of data is primarily a technical problem. Data can be stolen or corrupted by vandals gaining access to your computer using your password (either stolen or guessed). Password protection is your responsibility. You should make your password difficult to guess.
14. You should encrypt your valuable data and store it in your database.
15. Another serious problem faced by computer users is the intrusion by viruses, worms and Trojans. These are collectively called malware (i.e., malicious software).
16. A computer virus is a small program written with the intention of ruining the work of a computer. It is spread through email attachments or files read from pen drives. It slows down machines, corrupts files, and may even damage the disk.
17. A worm is also a rogue program which is spread through computer networks and uses flaws in the OS. Worms try to steal data from your disk and are harder to detect and eradicate than viruses.
18. A Trojan comes to your computer disguised as a useful program and tries to destroy your files.

30. IT professionals are those with B.E., M.E., M.Sc. (C.S.), or MCA degree. Careers open to them are in software industry or business organizations which have large IT departments. The career tracks are systems analysts, programmers, embedded system designers, DSP programmers, network professionals, quality assurance professionals, customer support, human resource development managers, and IT trainers.

EXERCISES

18.1 What are the ways IT is affecting our day-to-day life?

18.2 Where are embedded computers used?

18.3 What is the difference between embedded computers and desktop computers such as PCs?

18.4 In what ways have computers affected access to information and personal correspondence?

18.5 What are the free services available on the Internet? How are these free services financed by the providers?

18.6 What are the websites which provide you facilities to create contents? What are the facilities provided by YouTube?

18.7 If you want to create a sample music piece and distribute it on the web, what free service is available?

18.8 What do you understand by social networking? Who uses them? What are the advantages of using such a service?

18.9 What is Linkedin? What are the special features of Linkedin? Who are its users?

18.10 What do you understand by data privacy? Is privacy a technical problem?

18.11 What is data security? What is the difference between data privacy and data security?

18.12 How do you ensure security of your password?

18.13 How do you ensure security of your database?

18.14 What is information integrity? How is it different from data security?

18.15 How do you ensure information integrity?

18.16 What is a computer virus? How does it infect computer systems?

18.17 What are the possible outcomes of computer virus infection?

18.18 How do you prevent viruses infecting your PC?

18.19 How do you know that your PC has been infected by a virus? If it is infected, how do you remove the infection?

18.20 What is a computer worm? In what way is it different from a computer virus?

18.21 What is a Trojan? In what way is it different from a virus? How do you remove a Trojan from your PC?

18.22 What are the various disasters which can affect an organization? How does an organization recover from a disaster?

18.23 What precautions do you take to protect valuable data stored in your PC in case of disasters?

18.24 What is meant by Intellectual Property Rights (IPRs)?

18.25 What are the three types of intellectual property? What kind of rights does IPR on computer software normally give?

18.26 What type of creative works is given copyright? What is the period of a copyright?

18.27 What rights does copyright give to the owner of the copyright?

18.28 Are the contents of web pages covered by copyright?

18.29 What is plagiarism? Is copyright violation a criminal offence?

18.30 What is fair use? What type of uses of a copyrighted material is considered fair use?

18.31 List some IT related products which are given patent rights.

18.32 List some IT related products which are given trademarks.

18.33 What is open source software? List some open source software in general use.

18.34 What are the advantages of open source software?

18.35 What is open domain software? How is it different from open source software?

18.36 What are the allowed uses of open source software? Do you need permission to use open source software?

18.37 What do you understand by IT enabled services? What careers have emerged in IT enabled services?

18.38 What is DTP? How has the printing and publishing industry been affected by IT?

18.39 What are computer games? What are the special skills needed to become a game designer?

18.40 What is computer animation? What do animators do? What types of careers are available to computer animators?

18.41 What is computer assisted learning? What special skills are needed to become a CAL professional?

18.42 What is medical transcription? What skills are needed to enter this career?

18.43 What are call centres? What are the skills needed to be a call centre operator?

18.44 What qualification is required to be recognized as an IT professional?

18.45 What are the openings available to IT professionals?

18.46 What tasks does a systems analyst perform?

18.47 What tasks does a programmer perform? What are the differences between the tasks performed by a systems analyst and a programmer?

18.48 What tasks does an embedded systems designer perform? What qualifications are needed for this job?

18.49 What tasks does a VLSI designer perform?

18.50 What are the jobs performed by a computer network professional?

18.51 What are the tasks performed by a quality assurance professional?

18.52 What do you understand by customer support? What are the types of support needed?

18.53 What are the functions of an HRD professional? What is his/her role in IT industry?

18.54 What tasks are performed by IT trainers? What are the normal qualifications needed for IT trainers?

18.55 What tasks are performed by a movie editor? How is IT useful for this profession?

Suggested Further Reading

General

These are useful as supplementary reading material for many chapters.

1. Rajaraman, V., *Fundamentals of Computers*, 5th Edition, PHI Learning, Delhi, 2010.
2. Curtin, D.P., K. Foley, K. Sen, and C.S. Morin, *Information Technology* (*The Breaking Wave*), Tata McGraw-Hill, New Delhi, 1999.
3. Norton, P., *Introduction to Computers*, 7th Edition, Tata McGraw-Hill, New Delhi, 2010.
4. Rajaraman, V. and D. Rajaraman, *Computer Primer*, 2nd Edition, Prentice-Hall of India, New Delhi, 1999.
5. Cyganski, D., J.A. Orr, and R.F. Vaz, *Information Technology: Inside and Outside,* Prentice-Hall, New York, 2001.

Multimedia Acquisition and Processing

6. Steinmetz, R. and K. Nehrstedt, *Multimedia: Computing, Communication and Applications,* Pearson Education India, Delhi, 2002.
7. Buford, J.F.K. (Editor), *Multimedia Systems*, Pearson Education Asia, Delhi, 2000.
8. Gonzalez, R.C. and R.E. Woods, *Digital Image Processing*, 3rd Edition, Pearson Education India, Delhi, 2008.
9. Foley, J.D., A. Van Dam, S.K. Feiner, and J.F. Hughes, *Computer Graphics*, 2nd Edition, Pearson Education India, Delhi, 2002.
10. Mukherjee, D.P., *Fundamentals of Computer Graphics and Multimedia*, PHI Learning, Delhi, 2009.

Computer Hardware and Architecture

11. Rajaraman, V. and T. Radhakrishnan, *An Introduction to Digital Computer Design*, 5th Edition, PHI Learning, Delhi, 2008.
12. Stallings, W., *Computer Organization and Arhitecture*, 8th Edition, Pearson Education India, Delhi, 2010.
13. Mano, M., *Computer System Architecture*, 3rd Edition, Pearson Education India, Delhi, 2007.

Index