Chapter 3
Computer Memory and Storage

3.1 INTRODUCTION

Computers are used to perform various tasks in science, engineering, business, education, entertainment and many other fields. They work at high speed, can handle large volumes of data with great accuracy and have the ability to carry out a specified sequence of operations without human intervention. The CPU handles the processing of data and after processing, presents the results with the help of output devices. However, the CPU requires memory to process the data, hold the intermediate results and to store the output. *Computer memory* refers to the electronic holding place for instructions and data where the processor can reach quickly. It can be classified into two broad categories: *primary memory* (to process the data and hold the intermediate results) and *secondary memory* (to store the output).

The primary memory allows the computer to store data for immediate manipulation and to keep track of what is currently being processed. The major limitation of this type of memory is that it is volatile. It means that when the power is turned off, the contents of primary memory are lost forever. Hence, to store the data permanently, a computer requires some nonvolatile storage medium like a hard disk. This kind of storage is known as *secondary memory*. Such memories store all the data (files) and instructions (computer programs) even after the power is turned off. The secondary storage devices have a larger storage capacity; they are less expensive as compared to primary storage devices, but slow in comparison.

> **Note:** *When we talk about memory, we generally refer to the primary memory only, and when we talk about storage, we refer to secondary memory.*

3.1.1 Memory Representation

All quantities, physical or otherwise, are measured in units. For example, length is measured in metres and mass in grams. Likewise, for measuring computer memory, a standard unit is required. Digital computers work on only two states: ON (1) and OFF (0). These two values are represented by two different voltages within the circuit. For example, 0 volt represents a false value (0), and +5 volt represents a true value (1). Each of these values (either 0 or 1) is called a *binary digit* or *bit* and can be considered a symbol for a piece of information. Although the smallest unit of data that a computer can deal with is a bit, computers generally do not deal with a single bit. Instead, they deal with a group of eight bits, which is referred to as a byte. A byte can have 256 different bit patterns, and thus can represent 256 different symbols. Various units used to measure computer memory are as follows:

- **Bit:** It is the smallest unit of data on a machine and a single bit can hold only one of two values: 0 or 1. Bit is represented by a lower case b.
- **Byte:** A unit of eight bits is known as a byte. Hence, a byte is able to contain any binary number between 00000000 and 11111111. It is represented by an upper case B.

- **Kilobyte:** In a decimal system, kilo stands for 1000, but in a binary system, kilo refers to 1024. Therefore, a kilobyte is equal to 1024 bytes. It is usually represented as KB.
- **Megabyte:** It comprises 1024 kilobytes, or 1,048,576 bytes. However, since this number is hard to remember, a megabyte can be thought of as a million bytes. Megabyte is the standard unit of measurement for RAM and is represented as MB.
- **Gigabyte:** It consists of 1024 megabytes (1,073,741,824 bytes). It is the standard unit of measurement for hard disks and is often represented as GB.
- **Terabyte:** It refers to 1024 gigabytes. Often represented as TB, terabyte memory is usually associated with super computers only.

3.2 MEMORY HIERARCHY

The processor is the "brain" of the computer where all the essential computing takes place. But unlike a human brain, a computer processor has very limited memory. Thus, it has to rely on other kinds of memories to hold data and instructions and to store results. The memory in a computer system is of three fundamental types:

- **Internal Processor Memory:** This memory is placed within the CPU (processor) or is attached to a special fast bus. Internal memory usually includes cache memory and special registers, both of which can be directly accessed by the processor. This memory is used for temporary storage of data and instructions on which the CPU is currently working. Processor memory is the fastest among all the memories but is the most expensive also. Therefore, a very diminutive part of internal processor memory is used in the computer system. It is generally used to compensate for the speed gap between the primary memory and the processor.
- **Primary Memory:** Random access memory (RAM) and read only memory (ROM) fall under the category of the primary memory, also known as *main memory*. Every computer comes with a small amount of ROM, which contains the boot firmware (called *BIOS*). This holds enough information to enable the computer to check its hardware and load its operating system into its RAM at the time of system booting. RAM is the place where the computer temporarily stores its operating system, application programs and current data so that the computer's processor can reach them quickly and easily. It is volatile in nature, that is, when the power is switched off, the data in this memory are lost. Unlike RAM, ROM is non-volatile. Even when the computer is switched off, the contents of the ROM remain available.
- **Secondary Memory:** Also known as *auxiliary memory*, secondary memory provides backup storage for instructions (computer programs) and data. The most commonly used secondary storage devices are magnetic disk and magnetic tapes. These are the least expensive and also have much larger storage capacity than the primary memory. The instructions and data stored on secondary storage devices are permanent in nature. They can only be removed if the user wants it so or if the device is destroyed. Secondary memory can also be used as *overflow memory* (also known as *virtual*

memory), when the capacity of the main memory is surpassed. Note that unlike processor memory and main memory, secondary memory is not directly accessible to the processor. Firstly, the data and instructions from the secondary memory have to be shifted to the main memory and then to the processor.

Figure 3.1 illustrates the memory hierarchy. The CPU accesses memory according to a distinct hierarchy. When the data come from permanent storage (for example, hard disk), they first go in RAM. The reason behind it is that if the CPU has to access the hard disk constantly to retrieve every piece of required data, it would operate very slowly. When the data are kept in primary memory, the CPU can access them more quickly. Subsequently, the CPU stores the required pieces of data and instructions in processor memory (cache and registers) to process the data.

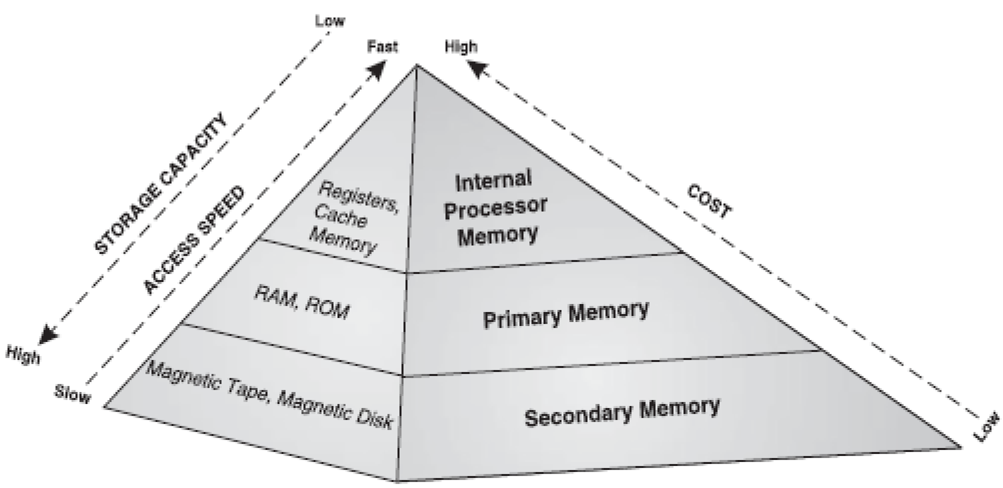


Figure 3.1 Memory Hierarchy

3.3 RANDOM ACCESS MEMORY

RAM is like the computer's scratch pad. It allows the computer to store data for immediate manipulation and to keep track of what is currently being processed. It is the place in a computer where the operating system, application programs and data in current use are kept so that they can be accessed quickly by the computer's processor. RAM is much faster to read from and write to than the other kinds of storage in a computer, like the hard disk or floppy disk. However, the data in RAM stay there only as long as the computer is running. When the computer is turned off, RAM loses all its contents. When the computer is turned on again, the operating system and other files are once again loaded into RAM. When an application program is started, the computer loads it into RAM and does all the processing there. This allows the computer to run the application faster. Any new information that is created is kept in RAM and since RAM is volatile in nature, one needs to continuously save the new information to the hard disk.

Let us take a simple example of why RAM is used by the computer. Whenever a user enters a command from the keyboard, the CPU interprets the command and instructs the hard disk to “load” the command or program into main memory (see Figure 3.2). Once the data are loaded into memory, the CPU is able to access them much quickly. The reason behind this is that the main memory is much faster than secondary memory. The process of putting things that the CPU needs in a single place from where it can get them more quickly is similar to placing various documents, which the user needs, into a single file folder. By doing so, the user finds all the required files handy and avoids searching in several places every time he needs them.

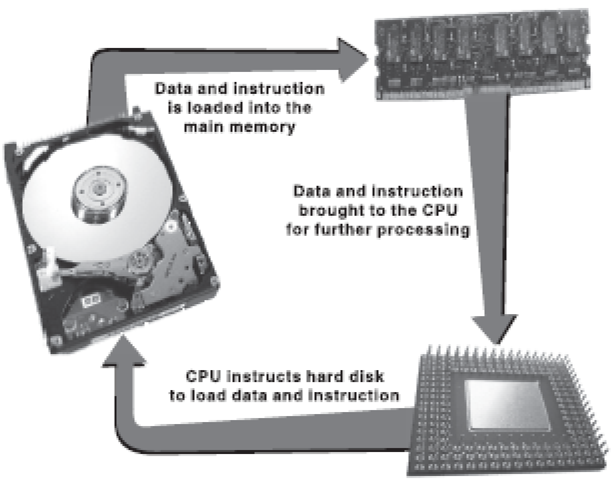


Figure 3.2 Interaction of Memory and Storage with Processor

3.3.1 Types of RAM

RAM is of two types:

- **Static RAM (SRAM):** The word “static” indicates that the memory retains its contents as long as power is being supplied. However, as soon as the power goes down, the data are lost. This makes SRAM a volatile memory as opposed to ROM. SRAM does not need to be “refreshed” (pulse of current through all the memory cells) periodically. It is very fast but much more expensive than DRAM (Dynamic RAM). SRAM is often used as cache memory due to its high speed.
- **Dynamic RAM (DRAM):** It is named so because it is very unstable. The data continue to move in and out of the memory as long as power is available. Unlike SRAM, DRAM must be continually refreshed in order to maintain the data. This is done by placing the memory on a refresh circuit that rewrites the data several hundred times per second. DRAM is used for most system memory because it is inexpensive and small.

The primary difference between SRAM and DRAM is the life of the data they store. SRAM retains its contents as long as electrical power is supplied to the chip. If the power is turned off, its contents are lost. On the other hand, DRAM must be continuously refreshed after about every 15 microseconds. This is true even when power is supplied constantly. SRAM chips are not as dense as DRAM chips, that is, the total number of cells in the SRAM chip is less than that on DRAM chip. SRAM is beneficial because it is fast, has low latency (the time lag between a request and the action being performed), and does not need to be refreshed. However, it is large and expensive, requires more power to operate, and produces a lot of heat. DRAM is simple, small, and space efficient. It may be slower and may have a longer latency than SRAM, but it is still very useful. Typical access time of SRAM is 25 nanoseconds while of DRAM 60 nanoseconds.

SRAM is useful for low amount of memory. Anything over 4 MB is very bulky. SRAM is good for internal memory in processors, and cache, but DRAM is best for the system's main memory. DRAM is used where its small size and power efficiency outweigh its slowness as compared to SRAM. SRAM is less dense than DRAM (fewer bits per unit area) and is, therefore, not suitable for high-capacity, low-cost-per-megabyte applications. The power consumption of SRAM varies widely depending on its speed. Fast SRAM is much more power-hungry than DRAM and some ICs can consume power of the order of a watt at full speed. Slow SRAM can have very low power consumption in the region of a microwatt. Currently, the technology does not exist to produce small SRAMs so that they can replace DRAMs. Thus, DRAM is still used in computers.

3.4 READ ONLY MEMORY

Just as a human being needs instructions from the brain to perform actions in a certain event, a computer also needs special instructions every time it is started. This is required because during the start up operation, the main memory of the computer is empty due to its volatile property so there have to be some instructions (special boot programs) stored in a special chip that could enable the computer system to perform start up operations and transfer the control to the operating system. This special chip, where the start up instructions are stored, is called *ROM*. It is non-volatile in nature, that is, its contents are not lost when the power is switched off. The data and instructions stored in ROM can only be read and used but cannot be altered, thereby making ROM much safer and secure than RAM. ROM chips are used not only in the computer but also in other electronic items like washing machines and microwave ovens.

Generally, designers program ROM chips at the time of manufacturing circuits. Burning appropriate electronic fuses to form patterns of binary information does the programming. These patterns of binary information are meant for specific configurations, which is why different categories of computers are meant for performing different tasks. For example, a micro program called *system boot program* contains a series of start-up instructions to check for the hardware, that is, I/O devices, memory and operating system in the memory. These programs deal with low-level machine functions and are alternate for additional hardware requirement. ROM performs the necessary BIOS (basic input output system) function to start the system and then transfers the control over to the operating system.

ROM can have data and instructions written into it only one time. Once a ROM chip is programmed, it cannot be reprogrammed or rewritten. If it is erroneous, or the data need to be reorganized, one has to replace it with the new chip. Thus, the programming of ROM chips should be perfect, having all the required data at the time of its manufacturing. Note that in some instances, ROM can be changed using certain tools. For example, flash ROM (a type of ROM) is non-volatile memory that occasionally can be changed, such as when a BIOS chip must be updated. The ROM chips consume very little power, are extremely reliable, and in the case of most small electronic devices, contain all the necessary programming to control the device.

3.4.1 Types of ROM

Memories in the ROM family are distinguished by the methods used to write data on them and the number of times they can be rewritten. This classification reflects the evolution of ROM devices from “hard-wired” to programmable to erasable-and-programmable. One common feature of all these devices is their ability to retain data and programs even during a power failure. ROMs come in following varieties:

- **Masked ROM:** The very first ROMs, known as masked ROMs, were hard-wired devices that contained a pre-programmed set of data or instructions. The contents of such ROMs had to be specified before chip production so the actual data could be used to arrange the transistors inside the chip.
- **Programmable ROM (PROM):** Creating a ROM chip from scratch is a time-consuming and an expensive process. For this reason, developers created a type of ROM known as programmable read only memory (PROM), which can be programmed. Blank PROM chips can be bought economically and coded by the users with the help of a special device known as *PROM-programmer*. However, once a PROM has been programmed, its contents can never be changed. As a result, PROM is also known as *one-time programmable* (*OTP*) device. Like other ROMs, PROM is also non-volatile. However, it is more fragile than other ROMs as a jolt of static electricity can easily cause the fuses in the PROM to burn out, thus changing the bit pattern from 1 to 0. Nevertheless, blank PROMs are economical and are great for prototyping the data for a ROM before committing to the costly ROM fabrication process.
- **Erasable Programmable ROM (EPROM):** An EPROM is programmed in exactly the same manner as a PROM. However, unlike PROM, an EPROM can be erased and reprogrammed repeatedly. It can be erased by simply exposing the device to a strong source of ultraviolet light for a certain amount of time. Note that an EPROM eraser is not selective; it will erase the entire EPROM. Although EPROM is more expensive than PROM, its ability to be reprogrammed makes it more useful.
- **Electrically Erasable Programmable ROM (EEPROM):** This type of ROM can be erased by an electrical charge and then written to by using slightly higher-than-normal voltage. EEPROM can be erased one byte at a time, rather than erasing the entire chip with ultraviolet light. Hence, the process of reprogramming is flexible, but slow. Also, changing the contents does not require any additional committed equipment. As these

chips can be changed without opening a casing, they are often used to store programmable instructions in devices like printers.

- **Flash ROM:** A flash ROM also called *flash BIOS or flash memory*, is a type of constantly powered non-volatile memory that can be erased and reprogrammed in blocks. It is a variation of EEPROM, which, unlike flash memory, is erased and rewritten at the byte level. Flash memory is often used to hold the control code such as the BIOS in a personal computer. When BIOS needs to be changed or rewritten, the flash memory can be written in block (rather than byte) sizes, thus making it easier to update. Flash memory gets its name because the microchip is organized so that a section of memory cells are erased in a single action or "flash". Flash memory is used in digital cellular phones, digital cameras and other devices.

3.5 RAM, ROM AND CPU INTERACTION

Picture a bulletin board under the glass at the back of a classroom. One way to think of ROM is that it is similar to the hard-copy notes placed under the glass. At the end of the day, they remain unchanged. The next day, the notes are exactly the way they were the day before. Students are able to only read them. RAM, on the other hand, can be thought of as a blackboard, which starts out blank and during the day, information is written on it, read from it, and even erased from it. When something is erased, new data can then be written on the same place on the board. When students go home at the end of the day, the blackboard is washed clean and whatever data was on it goes away forever. This is what happens when a computer's power is turned off; RAM no longer has the electrical current available to sustain the data in its memory cells. ROM is more like your long-term memory; the things you remember from your past. When you wake up in the morning, you get ready for school/office and know the address of your destination. Similarly, when computer "wakes up", it searches for start-up routines from ROM BIOS and then hands over the control to the operating system to function properly.

The most essential part of computer processing is the memory. From the moment the computer is turned ON and until it is shut down, the CPU constantly uses memory. A typical scenario is listed below and illustrated in Figure 3.3:

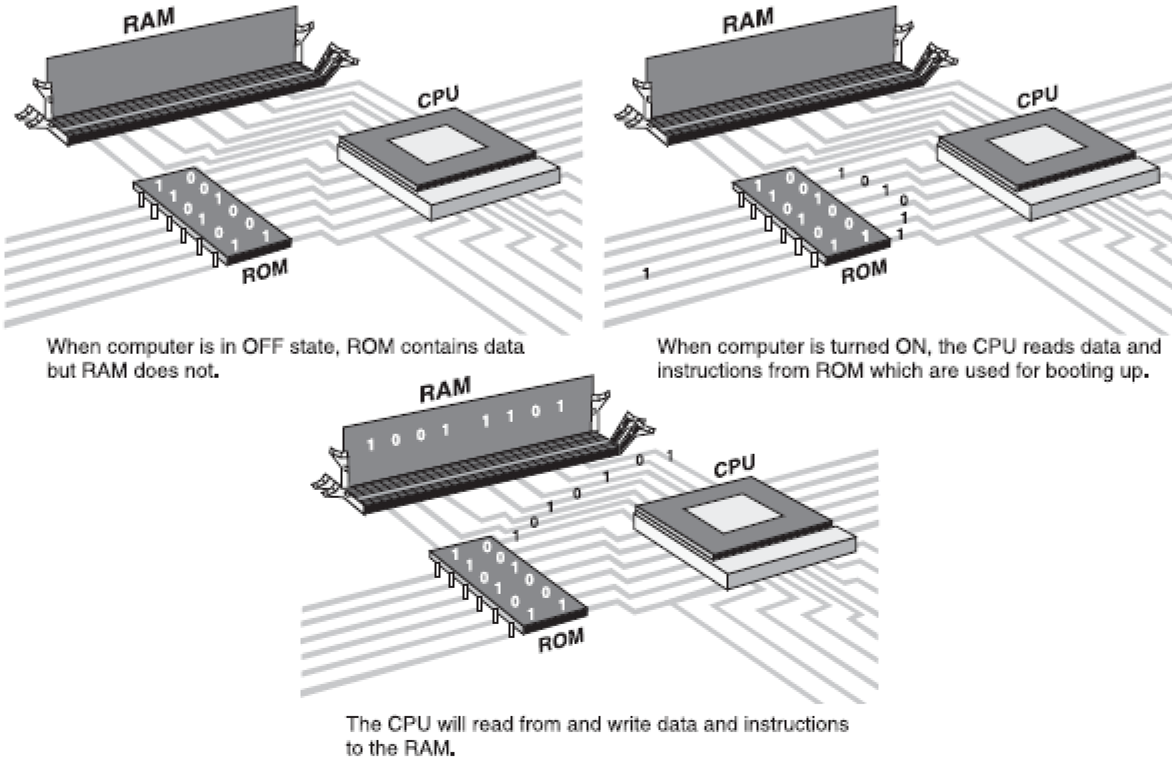


Figure 3.3 RAM, ROM and CPU Interaction

Step 1	The computer is switched ON.
Step 2	CPU loads data and instructions from ROM and checks whether all the major components like proce are functioning properly.
Step 3	CPU loads BIOS (basic input/output system) from ROM to determine the machine's fundamental environment. The information stored in ROM BIOS chip determines what peripherals the system can
Step 4	CPU loads the operating system from the secondary storage (hard disk) into RAM. This allows th access to the operating system, which enhances the performance and functionality of the overall syste
Step 5	When an application is opened, it is loaded into RAM and any file that is opened for use in that loaded into RAM.
Step 6	After processing, when the user saves the file and closes the respective application, the file is writt location on the secondary storage device. After that the file(s) and the application are "flushed out" fr

Every time something is loaded or opened, it is placed into RAM so that the CPU can access that information more easily and promptly. The CPU requests the required data from RAM, processes it, and writes new data back to RAM in a continuous cycle. In most computers, this shuffling of data between the CPU and RAM happens millions of times every second. When the application is closed, the application and any other accompanying files are usually erased from RAM to make space for the new data.

3.6 TYPES OF SECONDARY STORAGE DEVICES

Secondary storage devices facilitate storing of data and instructions permanently. The data stored on a secondary storage device can be accessed depending upon how it is stored on the device. Primarily, there are two methods of accessing data from the secondary storage devices (see Figure 3.4):

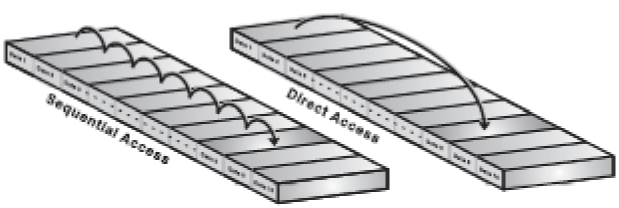


Figure 3.4 Sequential and Direct Access

- **Sequential Access:** Sequential access means the computer system must search the storage device from the beginning until the desired data is found. The most common sequential access storage device is magnetic tape where data is stored and processed sequentially. Suppose, a tape contains information regarding employees of an organization. Now, to look for employee number 100's information, the computer will have to start with employee number 1 and then go past 2, 3 and so on, until it finally comes to 100. The sequential access method is quite simple than other methods but searching for data is slow.
- **Direct Access:** Direct access, also known as *random access*, means that the computer can go directly to the location, where the data that the user wants, are stored. The most common direct access storage devices are magnetic disk and optical disk. In these devices, the data are stored as sequentially numbered blocks. Thus, one can access block 12, then access block 78, then block 2 and so on. The direct access method is

ideal for applications like airline reservation systems or computer-based directory assistance operations. In these cases, there is no fixed pattern of requests for data.

Based on the access method, secondary storage devices can be classified as shown in Figure 3.5.

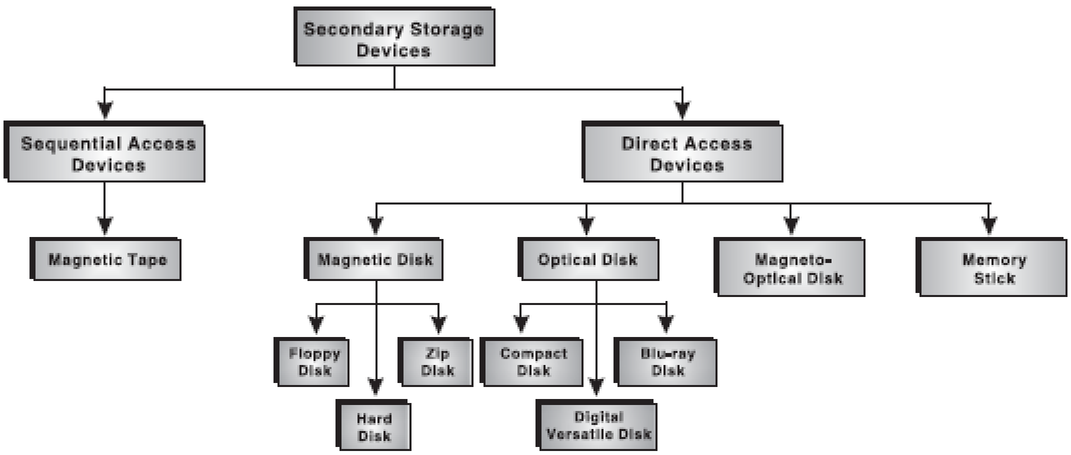


Figure 3.5 Classification of Secondary Storage Devices

3.7 MAGNETIC TAPE

Magnetic tape appears similar to the tape used in music cassettes. It is a plastic tape with magnetic coating on it. The data is stored in the form of tiny segments of magnetized and demagnetized portions on the surface of the material. The magnetized portion of the surface refers to the bit value “1” whereas the demagnetized portion refers to the bit value “0”. Magnetic tapes are available in different sizes, but the major difference between different magnetic tape units is the speed at which the tape is moved past the read/write head and the tape's recording density. The amount of data or the number of binary digits that can be stored on a linear inch of tape is the recording density of the tape.

Magnetic tapes are very durable and can be erased as well as reused. They are an inexpensive and reliable storage medium for organizing archives and taking backups. However, they are not suitable for data files that need to be revised or updated often because data on them are stored in a sequential manner and the user needs to advance or rewind the tape every time to the position where the requested data starts. Tapes are also slow due to the nature of the media. If the tape stretches too much, then it will render itself unusable for data storage and may result in data loss. The tape now has a limited role because disk has proved to be a superior storage medium. Today, the primary role of the tape drive is limited to backing up or duplicating the data stored on the hard disk to protect the system against loss of data during power failures or computer malfunctions.

3.7.1 Magnetic Tape Organization

The magnetic tape is divided into vertical columns (*frames*) and horizontal rows (*channels* or *tracks*) as shown in Figure 3.6. The data is stored in a string of frames with one character per frame, and each frame spans multiple tracks (usually seven or nine tracks). Thus, a single bit is stored in each track, that is, one byte per frame. The remaining track (seventh or ninth) stores the parity bit. When a byte is written to the tape, the number of 1s in the byte is counted, the parity bit is then used to make the number of 1s even (even parity) or odd (odd parity). When the tape is read again, the parity bit is checked to see if any bit has been lost. In case of odd parity, there must be an odd number of 1s represented for each character and an even number of 1s in case of even parity.

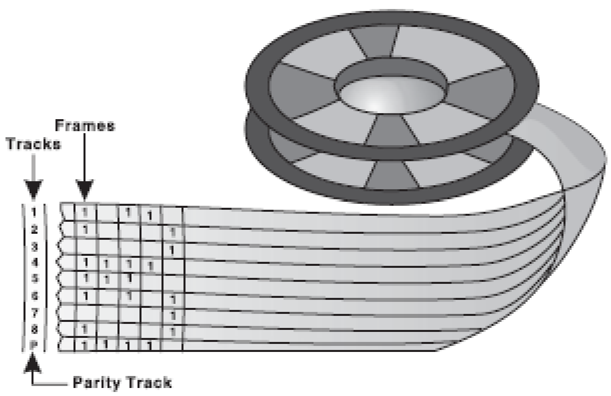


Figure 3.6 Representing Data in Magnetic Tape

Magnetic tape drive uses two reels: *supply reel* and *take-up reel*. Both reels are mounted on the hubs and the tape moves from the supply reel to the take-up reel. Figure 3.7 shows the basic tape drive mechanism. The magnetic-oxide-coated side of the tape passes directly over the read/write head assembly, thus making contact with the heads. As the tape passes under the read/write head, the data can be either read and transferred to the primary memory or read from the primary memory and written onto the tape.

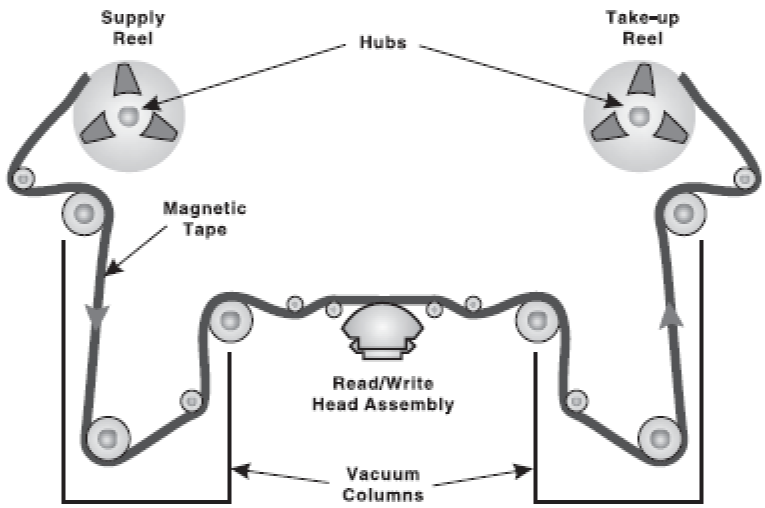


Figure 3.7 Basic Tape Drive Mechanism

A magnetic tape, as shown in Figure 3.8, is physically marked to indicate the location from where reading and writing on the tape is to begin (*BOT or beginning of tape*) and end (*EOT or end of tape*). The length of the tape between BOT and EOT is referred to as the usable recording (reading/writing) surface. BOT/EOT markers are usually made up of reflective-type of short silver strips. These markers are sensed by an arrangement of lamps and/or photodiode sensors to indicate the location from where reading and writing is to begin and end. On a magnetic tape, data are recorded in the form of blocks where each block consists of a grouping of data (known as *records*) that is written or read in a continual manner. Between these blocks, the computer automatically reserves some blank space called *inter-block gap* (*IBG*). One block may contain one or more records that are again separated by blank spaces (usually 0.5 inch) known as *inter-record gap* (*IRG*). Whenever an IRG is reached while reading data from a moving tape, the moving tape is stopped. It remains immobile until the record is processed.

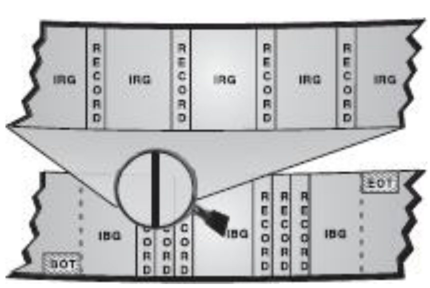


Figure 3.8 Information Format of Magnetic Tape

3.7.2 Advantages of Magnetic Tapes

- Magnetic tapes hold high data recording density thereby resulting in low cost per bit of storage.
- Magnetic tapes have virtually unlimited storage capacity because as many tapes as required can be used to store a large amount of data.
- Magnetic tapes are portable because they are compact in size, lightweight and removable. Due to these properties, they are also easy to handle and store.
- Magnetic tapes represent a very inexpensive mode of offline data storage and a large amount of data can be stored in a small storage space.

3.7.3 Disadvantages of Magnetic Tapes

- Since magnetic tapes are sequential in nature, they are not suitable in situations where data access is required in a random order. Moreover, data transmission in magnetic tapes is also slow as compared to the magnetic disks.
- Magnetic tapes should be kept in a dust-free environment and away from corrosive gases and chemicals as they can cause tape-reading errors.
- Since magnetic tapes use parity bit to check the data, the data on such devices are difficult to recover even if a minor bit error occurs.
- Magnetic tapes are not flexible as compared to other media types when file updating requires record insertion or deletion. One more drawback of magnetic tapes is that they wear out.

3.8 MAGNETIC DISK

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

3.8.1 Storage Organization of a Magnetic Disk

The surface of a disk is divided into imaginary tracks and sectors as illustrated in Figure 3.9. *Tracks* are concentric circles where the data are stored, and are numbered from the outermost to the innermost ring, starting with zero. *Sectors* refer to the number of fixed-size areas (imaginary pie slices) that can be accessed by one of the disk drive's read/write heads, in one rotation of the disk, without the head having to change its position. An intersection of a track and a disk sector is known as *track sector*. Generally, a disk has eight or more disk sectors per track. However, different types of magnetic disks may have a different number of tracks. Today disks are marked (tracks and sectors) on both surfaces, hence they are also known as double-sided disks. Each sector is uniquely assigned a disk address before a disk drive can access a piece of data. The disk address comprises sector number, track number and surface number (if double-sided disks are used). The track sectors are grouped into a collection known as *cluster*. It refers to the basic allocation unit for storage on a disk, consisting of one or more track sectors. It is also the minimum amount of disk space used by a single file.

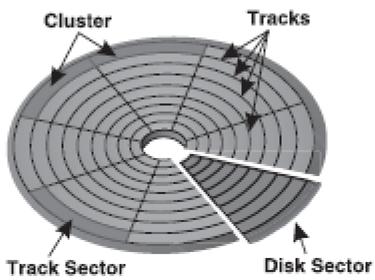


Figure 3.9 Organization of Disk Surface

Frequently, multiple disks (platters) are maintained and used together to create a large disk-storage system. Typically, two or more platters are stacked on top of each other with a common spindle, which rotates them. There is a gap between the platters, making room for the magnetic read/write head. There is a read/write head for each side of each platter and all the heads are attached to a single assembly called a *disk arm assembly*, which can move them towards the central spindle or towards the edge (see Figure 3.10). All the read/write heads are on an equal diameter track on the different platters at one time. The tracks of equal diameter on different platters form a *cylinder*.

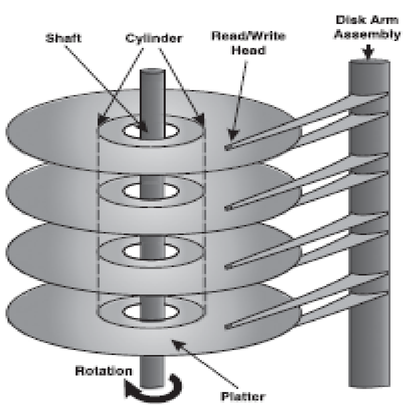


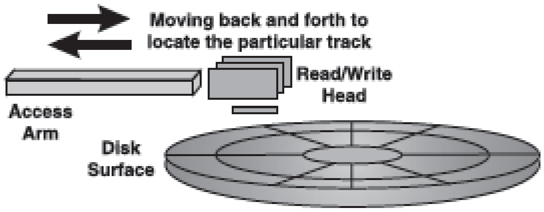
Figure 3.10 Disk Pack and Cylinder

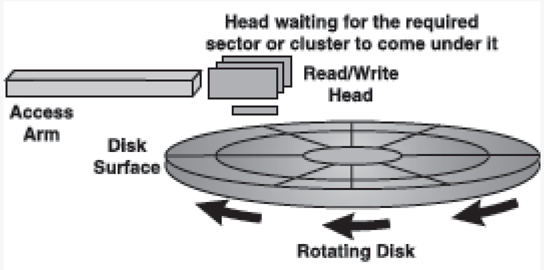
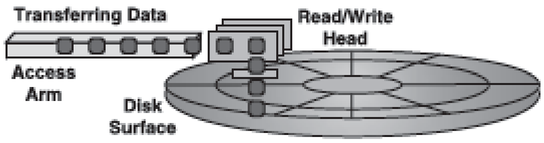
3.8.2 Accessing Data from Magnetic Disk

Data in a magnetic disk are recorded on the surface of the circular tracks with the help of the read/write head, which is mounted on the access arm assembly (see Figure 3.10). These heads can be in multiple numbers to access the adjacent tracks simultaneously and making a disk faster. The access arm assembly can be positioned in both inward and outward directions so that the read/write head can move on the horizontal surfaces of the disk. In case of multiple disk packs, each disk surface has its own read/write head, which works in harmony with other heads to record the data. Therefore, information is stored on the tracks, constituting a cylindrical shape through a disk pack. The process of accessing data comprises three steps:

1. **Seek:** As soon as the disk unit receives the read/write command, the read/write heads are positioned on the specific track on the disk platter. The time taken in doing so is known as *seek time*. It is the average time required to move the heads to the desired track on the disk. Seek times of modern disks may range between 2 and 15 milliseconds but the seek time of most common disks is 9 milliseconds.
2. **Rotate:** Once the heads are positioned on the desired track, the head of the specific platter is activated. Since the disk is rotated constantly, the head has to wait for the required sector or cluster (desired data) to come under it. This waiting time is known as *rotational delay time* or *latency* of the disk. The rotational latency of a disk with 7200 rpm is 4.17 milliseconds.
3. **Data Transfer:** Once the read/write head is positioned over the desired sector, the data can be transferred to or from the disk to primary memory. The rate at which the data is read from or written to the disk is known as *data transfer rate*. It is measured in kilobits per second (kbps). The data transfer rate depends upon the rotational speed of the disk. If the disk has a rotational speed of 6000 rpm (rotations per minute), having 125 sectors and 512 bytes/sector, the data transfer rate per rotation will be 125 × 512 = 64,000 bytes. Hence, the total transfer rate per second will be 64,000 × 6000/60 = 6,400,000 bytes/second or 6.4 MB/second (see Table 3.1).

Table 3.1 Accessing Data

Step	**Measured As**	**Illustration**
Seek	Seek Time	

Step	Measured As	Illustration
Rotate	Rotational Delay	
Data Transfer	Data Transfer Time	

The combined time (seek time, latency and data transfer time) is known as the *access* time of the magnetic disk. Generally, the access time can be described as the period of time that elapses between a request for data from disk or memory and the desired data arriving at the requesting device. *Memory access time* refers to the time it takes to transfer a character from memory to or from the processor, while *disk access time* refers to the time it takes to place the read/write heads over the requested data. RAM may have an access time of 9–70 nanoseconds, while hard disk access time could be 10–40 milliseconds.

3.9 TYPES OF MAGNETIC DISKS

All magnetic disks come in the form of round platters. These disks are available in different sizes, shapes and designs. Some are attached to the read/write head assembly whereas some are available in the form of removable disks. Broadly, magnetic disks can be classified into three types: floppy disk, hard disk and zip disk.

3.9.1 Floppy Disk

A floppy disk is a round, flat piece of Mylar plastic coated with ferric oxide (a rust-like substance containing tiny particles capable of holding a magnetic field) and encased in a protective plastic cover (disk jacket). It is a removable disk and is read and written by a floppy disk drive (FDD), which is a device that performs the basic operation on a disk, including rotating the disk and reading and writing data onto it. The disk drive's read/write head alters the magnetic orientation of the particles, where orientation in one direction represents “1” and orientation in the other represents “0”.

Traditionally, floppy disks were used on personal computers to distribute software, transfer data between computers and create small backups. Earlier, 5¼-inch floppy disks were used. Later, a new format of 3½-inch floppy disk came into existence, which has larger storage capacity and supports faster data transfer as compared to 5¼-inch floppy disks. Floppy diskettes are small, inexpensive, readily available, easy to store and have a good shelf life if stored properly. They also possess the write-protect feature, which allows the users to protect a diskette from being written on. To write-protect a diskette, the user has to shift a slide lever towards the edge of the disk, uncovering a hole. The key advantage of floppy disk is that it is portable.

Read/Write Operation of a Floppy Disk: To read and write data onto a floppy disk, FDD is used. The drive (see Figure 3.11) is made up of a box with a slot (having a drive gate) into which a user inserts the disk. When the user inserts a disk into the FDD, the drive grabs the disk and spins it inside its plastic jacket. Also, the drive has multiple levers that get attached to the disk. One lever opens the metal plate, or shutter, to expose the data access area. Other levers and gears move two read/write heads until they almost touch the diskette on both sides. The drive's circuit board receives instructions for reading/writing the data from/to disk through the *floppy drive controller*. If the data are to be written onto the disk, the circuit board first verifies that no light is visible through a small window in the floppy disk. If the photo sensor on the opposite side of the floppy disk detects a beam of light, the floppy drive detects the disk to be write-protected and does not allow recording of data.

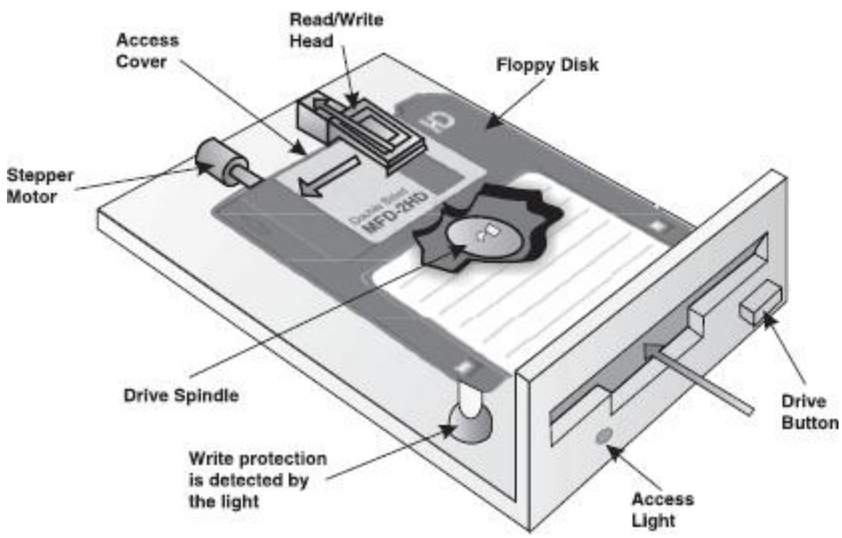


Figure 3.11 Floppy Disk Drive

The circuit board translates the instructions into signals that control the movement of the disk and the read/write heads. A motor located beneath the disk spins a shaft that engages a notch on the hub of the disk, causing the disk to spin. When the heads are in the correct position, electrical impulses create a magnetic field in one of the heads to write data to either the top or bottom surface of the disk. Similarly, on reading the data, electrical signals are sent to the computer from the corresponding magnetic field generated by the metallic particle on the disk.

Since the floppy disk head touches the diskette, both media and head wear out quickly. To reduce wear and tear, personal computers retract the heads and stop the rotation when a drive is not reading or writing. Consequently, when the next read or write command is given, there is a delay of about half a second while the motor gathers maximum speed.

3.9.2 Hard Disk

The *hard disk*, also called the *hard drive* or fixed disk, is the primary storage unit of the computer. It consists of a stack of disk platters that are made up of aluminium alloy or glass substrate coated with a magnetic material and protective layers (see Figure 3.12). They are tightly sealed to prevent any dust particle, which causes head crash, from getting inside. A hard disk can be external (removable) or internal (fixed) and can hold a large amount of data. The capacity, that is, the amount of information that a hard disk can store, is measured in bytes. A typical computer today comes with 80–320 GB of hard disk. The storage capacity of hard disk has increased dramatically since the day it was introduced. The hard disk speed is measured in terms of access time (typically in milliseconds). A hard disk with lower access time is faster than a hard disk with higher access time; the lower the access time, the faster the hard disk.

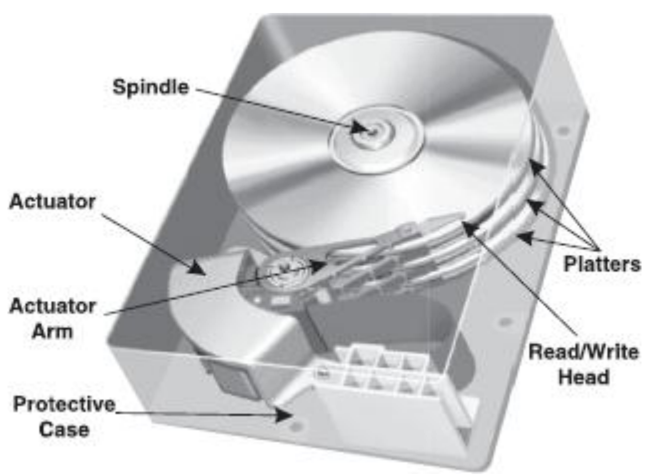


Figure 3.12 Hard Disk

Read/Write Operation of a Hard Disk: A hard disk uses round, flat disks (platters) made up of glass or metals which are coated on both sides with a special material designed to store information in the form of magnetic patterns. Each platter has its information recorded in tracks, which are further broken down into smaller sectors. Making a hole in the centre of platters and stacking them onto a spindle mount the platters. The platters rotate at high speed, driven by a special motor connected to the spindle. Special electromagnetic read/write heads are mounted onto sliders and are used to either record data onto the disk or read data from it. The sliders are mounted onto arms, all of which are mechanically connected into a single assembly and positioned over the surface of the disk by a device called *actuator*. Each platter has two heads, one on the top of the platter and one on the bottom, so a hard disk with three platters would have six surfaces and six heads.

Data are recorded onto the magnetic surface of the disk in exactly the same way as they are on floppies. However, a disk controller is attached to the hard disk drive that handles the read/write commands issued by the operating system. Each read/write command specifies a disk address that comprises the surface number, track number and sector number. With this information, the read/write head moves to the desired sector that data can be read from or written to. Usually, the next set of data to be read is sequentially located on the disk.

3.9.4 Advantages of Magnetic Disks

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

3.9.5 Disadvantages of Magnetic Disks

- Magnetic disks must be stored in a dust-free environment in order to protect them from crashing.
- Magnetic disks are not ideal devices to use in scenarios where the file access required is of sequential nature rather than direct or random nature.
- Magnetic disks are more expensive than the magnetic tapes.
- Magnetic disks are more susceptible to breach of security and access gain to sensitive online disk files from remote terminals.

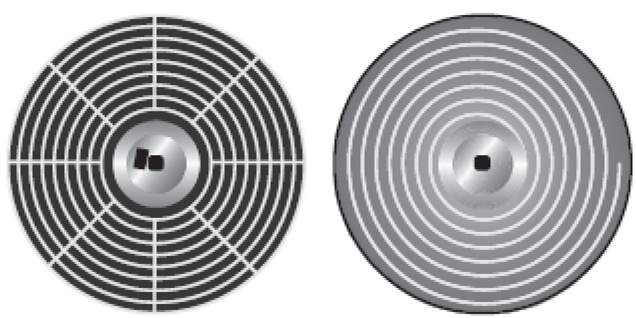
3.10 OPTICAL DISK

Apart from magnetic tapes and magnetic disks, a new storage medium, which is gaining popularity, is the optical disk. An optical disk is a flat, circular, plastic disk coated with material on which bits may be stored in the form of highly reflective areas and significantly less reflective areas, from which the stored data may be read when illuminated with a narrow-beam source, such as a laser diode. These disks are capable of storing enormously high amounts of data in a limited amount of space. The optical disk storage system consists of a rotating disk coated with a thin layer of metal (aluminium, gold or silver) that acts as a reflective surface and a laser beam, which is used as a read/write head for recording data onto the disk.

Optical disk comes in various sizes and capacities. A compact disk (CD) with 700 MB capacity and 12 cm diameter is the most popular means of optical storage. In a single-track optical disk, storage capacity is calculated by the multiple of number of sectors and number of bytes per sector. Since the storage capacity of an optical disk is huge, the cost per bit of storage is very low.

3.10.1 Storage Organization of Optical Disk

Unlike magnetic disk (which has several layers of concentric layers), an optical disk consists of a single long track in spiral shape as shown in Figure 3.15. This track starts from the outer edge and spirals inward to the centre of the disk. The spiral shape of the track makes the optical disk suitable for storing large blocks of sequential data onto it, such as music. Even though random access is possible, it is usually slower than magnetic disk (especially hard disk). The reason behind it is that in case of magnetic disk, it is easy to locate the magnetic disks tracks since they are located at a fixed distance from the centre.



(a) Track Pattern of a Magnetic Disk (b) Track Pattern of an Optical Disk

Figure 3.15 Comparing Track Patterns

These tracks are further divided into small sectors of the same length, regardless of their location on the disk's surface. Such type of disk formatting allows the data to be stored in even more densely as compared to magnetic disks. With more data recording density, the optical drives have a more complicated mechanism as compared to the magnetic disks drives. Here, rotation speed of the disk varies inversely with the radius of the disk. The disk moves slowly when data are being read near the edges and moves fast when close to the centre.

Since read/write operations are performed through laser beams, no access arm movement is required that is used in the case of the magnetic disk. Hence, many people assume that the data access time is even faster than the magnetic disk. However, this is not true because compared to the magnetic disks, where the tracks are arranged in the form of concentric circles, in an optical disk the tracks are organized in a spiral fashion. This results in the slower random access time than in the concentric circles. Generally, the access time for an optical disk ranges from 10 to 40 milliseconds which is very low as compared to a floppy disk that has an access time of several hundred milliseconds.

3.10.2 Access Mechanism of Optical Disk

The laser beam technology used for reading/writing of data on the disk surface of the optical disk uses two laser beam sources of different intensities. The greater intensity laser beam is used to write on the recording surface by turning it ON and OFF at a varying rate so that tiny pits are burnt into the metal coating of the disk. The lesser intensity beam is used to read that stored data, which is strongly reflected by the coated surface called *land* and weakly reflected by the burnt surface called *pit*. The changes in pattern are detected by a photo sensor and converted into digital signals (see Figure 3.16).

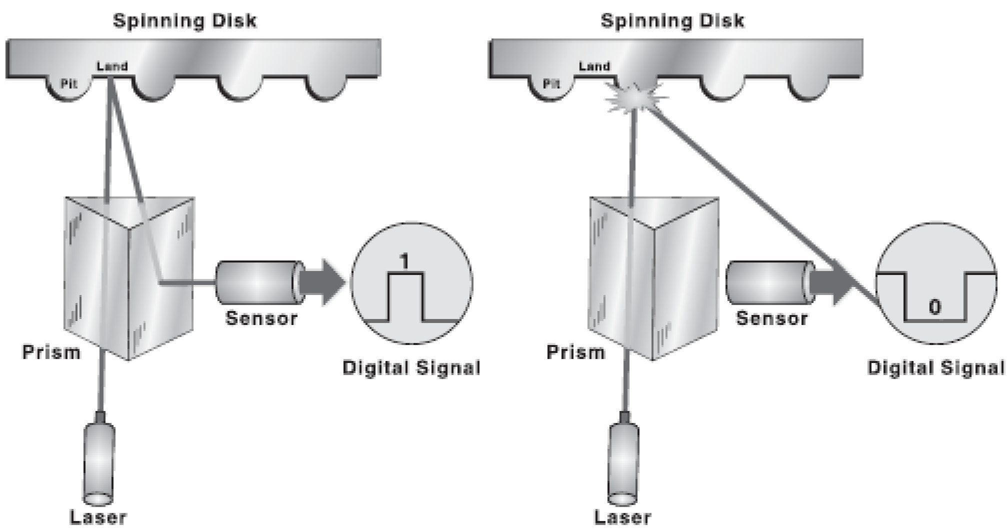


Figure 3.16 Read Operation in an Optical Disk

3.11 TYPES OF OPTICAL DISKS

Optical disks come in several varieties, which are made in somewhat different ways for different purposes. Compact disk (CD), digital versatile disk (DVD) and blu-ray disk (BD) are the three forms of optical disks. Note that these technologies are not compatible with one another. Each requires a different type of disk drive and disk. Even within one category, there are many competing formats.

3.11.1 Compact Disk

Compact disk is the most popular and the least expensive type of optical disk. It was originally intended only for storing music (in the form of digital audio) and can record about 80 minutes of uninterrupted playing time. A CD is capable of being used as a data storage device along with storing of digital audio.

FACT FILE

A CD is a shiny, silver colour metal disk of 12 cm diameter. A typical optical disk is made up of three layers: a *polycarbonate base* through which light can pass, a *layer* of *aluminium*, and a protective *layer of acrylic* on top of that. The pits of CD are typically 0.5 microns wide, 0.83 to 3 microns long, and 0.15 microns deep. A CD has one track that spirals from the centre to the outside edge. If one could remove the track from a standard 12 cm CD, it would stretch for 3.5 miles. The single track is divided into sectors of equal length and density. Files are stored on these particular contiguous sectors.

Compact disks are available in various formats: *CD-ROM* (*compact disk-read only memory*), *CD-R* (*compact disk-recordable*), and *CD-RW* (*compact disk-rewritable*) disks. A CD-ROM disk comes with pre-recorded data by the manufactures and can be read but cannot be altered. CD-R is a type of WORM (write once-read many) disk that allows you to record your own data. Once written, the data on the CD-R can be read but cannot be altered. A CD-RW disk is rewritable version of CD-R that means, it allows writing, erasing and rewriting of data several times. The data recorded on all CD formats can be read using the CD-ROM drive; however, to write data on CD-R and CD-RW disks, one needs a special peripheral device known as CD-writer (or *CD-burner*).

CD-ROM drives are characterized by the *spin rate*. The spin rate is the rotation speed of the disk and it influences the information retrieval speed (access time). Currently quad, hex and octal-speed CD-ROM drives are available where quad means 4x, hex means 6x and octal means 8x. Thus, octal-speed drive is the fastest and, as a result, the most expensive.

Reading Data from a CD-ROM: A CD is made up of three coatings, namely, *polycarbonate plastic*, *aluminium* and an *acrylic* coating to protect the disk from external scratches and dust. The polycarbonate plastic is stamped with millions of tiny indentations (pits), moulded in the spiral-shaped track. A CD drive reads information from the CD's spiral track of pits and lands, starting from the centre of the disk and moving to the outer edge, as shown in Figure 3.17. A light is beamed from a semiconductor laser through the bottom of the polycarbonate layer, and the aluminium coating monitors the light being reflected. Since the CD is read through the bottom of the disk, each pit appears as an elevated bump to the reading light beam. Light striking the land areas (the areas without bumps) is reflected normally and detected by a photodiode. As the disk rotates at speeds between 200 and 500 rpm, the light bounces off the pits causing the frequency of the light to change. The reflected light then passes through a prism and onto a photo sensor. Light reflected from a pit is 18^0 out of phase with the light from the lands, and the differences in intensity are measured by the photoelectric cells, which converts into a corresponding electrical pulse.

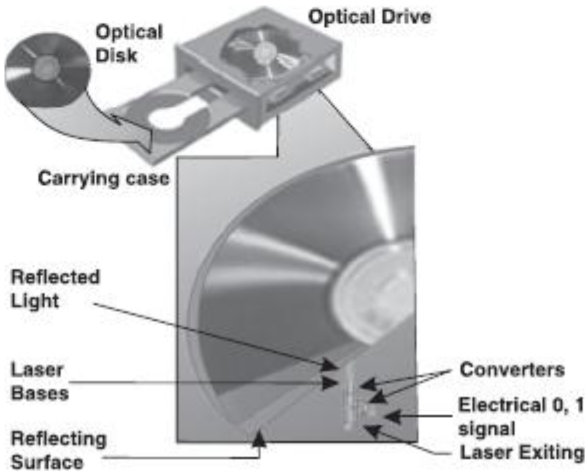


Figure 3.17 Reading Data from CD-ROM

Writing Data to a CD: On a new CD-R disk, the entire surface of the disk is reflective; the laser can shine through the dye and reflect off the gold layer. Hence, for a CD-R disk to work, there must be a way for a laser to create a non-reflective area on the disk. A CD-R disk, therefore, has an extra layer that the laser can modify. This extra layer is a greenish dye. When you write data to a CD-R, the writing laser (which is much more powerful than the reading laser) heats up the dye layer and changes its transparency (see Figure 3.18). The change in the dye creates the equivalent of a non-reflective bump. The decomposition of the dye in the pit area through the heat of the laser is irreversible (permanent). Therefore, once a section of a CD-R is written, it cannot be erased or rewritten. However, both CD and CD-R drives can read the modified dye as a bump later on.

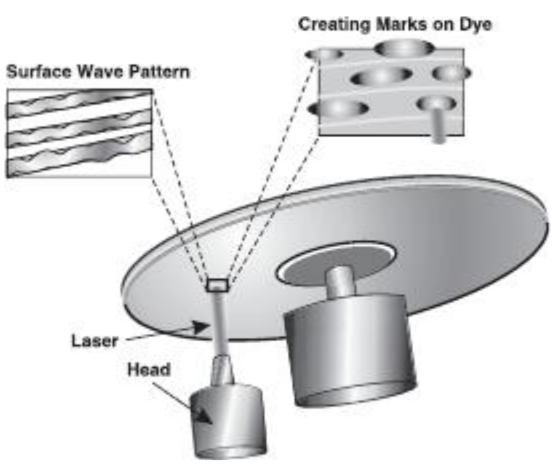


Figure 3.18 Writing Data to a CD-R

In contrast to CD-R disk, CD-RW disk is erasable and rewritable because it uses phase-changing material on its recording layer, usually an alloy of silver, tellurium, indium and antimony metals. Phase-changing material changes its state when heated to a high temperature (above its melting point) and can be converted back to its original state when heated at a temperature slightly below its melting point.

In CD-RW disk, the recording layer has a polycrystalline structure initially. While writing to the disk, the laser heats up the selected areas to a very high temperature (above the melting point), which melts the crystals into a non-crystalline amorphous phase. These areas have lower reflectance than the remaining crystalline areas. This difference in reflectance helps in reading the recorded data as in the case of CD-R disk.

To erase data on a CD-RW disk, a process called *annealing* is used. During this process, the area on the layer that has been changed to the amorphous phase (during writing) is converted back to its original crystalline state by heating it to a temperature slightly below the melting point of the phase-changing material.

3.11.2 Digital Versatile Disk

DVD, initially called *Digital Video Disk*, is a high-capacity data storage medium. At first glance, a DVD can easily be mistaken for a CD as both are plastic disks 120 mm in diameter and 1.2 mm thick, and both rely on lasers to read data. However, the DVD's seven-fold increase in data capacity over the CD has been largely achieved by tightening up the tolerances throughout the predecessor system. In DVD, the tracks are placed closer together, thereby allowing more tracks per disk. The DVD's *track pitch* (the distance between each) is reduced to 0.74 micron, less than half of CD's 1.6 micron. The pits, in which the data is stored, are also a lot smaller, thus allowing more pits per track. The minimum pit length of a single layer DVD is 0.4 micron as compared to 0.834 micron for a CD. With the number of pits having a direct bearing on capacity levels, DVD's reduced track pitch and pit size, as shown in Figure 3.19, alone give DVDs four times the storage capacity of CDs.

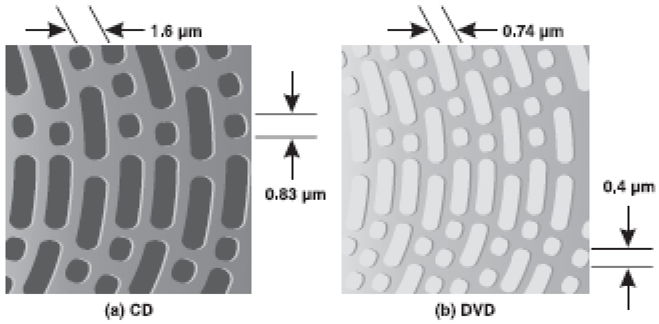


Figure 3.19 Comparing Track Pitch and Pit Length

The packing of as many pits as possible onto a disk is, however, the simple part; the real technological breakthrough of DVDs was with its laser. Smaller pits mean that the laser has to produce a smaller spot, and DVD achieves this by reducing the laser's wavelength from the 780 nm (nanometres) infrared light of a standard CD, to 635 nm or 650 nm. Secondly, the DVD specification allows information to be scanned from more than one layer of a DVD simply by changing the focus of the read laser. Instead of using an opaque reflective layer, it is possible to use a translucent layer with an opaque reflective layer behind carrying more data. This does not double the capacity because the second layer cannot be quite as dense as the single layer, but it does enable a single disk to deliver 8.5 GB of data without having to be removed from the drive and turned over.

An interesting feature of DVD is that the disk's second data layer can be read from the inside of the disk out, as well as from the outside in. In standard density CDs, the information is always stored first near the hub of the disk. The same is true for single-layer and dual-layer DVD, but the second layer of each disk can contain data recorded “backwards”, or in a reverse spiral track. With this feature, it takes only an instant to refocus a lens from one reflective layer to another. On the other hand, a single-layer CD that stores all data in a single spiral track takes longer to relocate the optical pickup to another location or file on the same surface.

DVD allows for double-sided disks (see Figure 3.20). To facilitate the focusing of the laser on the smaller pits, manufacturers used a thinner plastic substrate thereby reducing the depth of the plastic layer that the laser has to travel through to reach the pits. This reduction resulted in disks that were 0.6 mm thick—half the thickness of a CD. However, since these thinner disks were too thin to remain flat and withstand handling, manufacturers bonded two disks back-to-back, resulting in disks that are 1.2 mm thick. This bonding doubles the potential storage capacity of a disk. Note that single-sided disks still have two substrates, even though one is not capable of holding data.

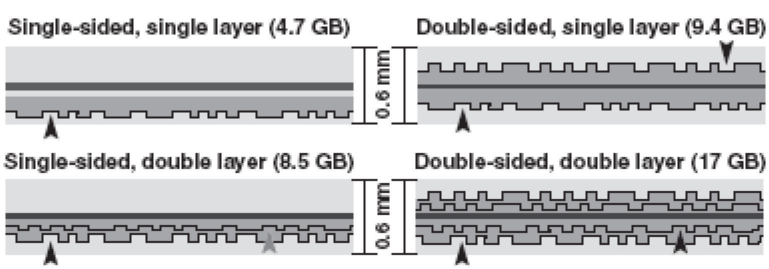


Figure 3.20 Different Types of DVDs

3.11.3 Blu-ray Disk

Blu-ray disk is an optical storage device, which is used to record and playback high definition video and audio as well as store images and other data (Figure 3.21). It is developed by Blu-ray Disk Association (BDA), which is a group of world's leading consumer electronics, personal computers and media manufacturers including Thomson, Apple, Panasonic, Sony and Samsung.

It uses blue-violet laser having shorter wavelength (405 nm) than a red laser (650 nm) used by DVDs. Due to this shorter wavelength, the laser can be focused more precisely on the small spot thereby resulting in storage capacity 10 times that of DVD. Like CDs and DVDs, Blu-ray disks are also available in different formats:

Figure 3.21 A Blu-ray Disk

FACT FILE

DVD and Blu-ray Disk

The organization of both Blu-ray disk and DVD is similar; however, the way the data are stored on them differs. In DVDs, the data reside between the two polycarbonate layers. Due to the layer on the top of data, the substrate layer may refract the laser beam into two separate beams, which is known as *birefringence*. If the beam is split too widely, the disk-reading operation cannot be performed. Blu-ray disk overcomes the birefringence by placing the data on the top of the polycarbonate layer, and thus prevents the disk-reading problem.

- **BD-ROM:** It comes with prerecorded content that can only be read.
- **BD-R:** It is a WORM type of disk on which you can record data only once.
- **BD-RW:** It is similar to BD-R disk but the difference is that it is rewritable. This means data can be erased and recorded a number of times on the same disk.
- **BD-RE:** It is also a rewritable disk but is used only for high definition audio/video and television recording.

Blu-ray disks are available in two sizes: standard (12 cm) and mini (8 cm). Each disk can have either a single layer or dual layers, depending on which the data storage capacity of the disk differs. For instance, the storage capacity of standard disk with single layer is 25 GB and with dual layer is 50 GB. On the other hand, the storage capacity of mini disk with single layer is 7.8 GB and with dual layer is 15.6 GB.

Blu-ray disk allows not only recording of programs but the programs can be manipulated also. One can edit the programs recorded on the disk or change their order. The user can also create playlists of the programs stored on disks so that they can be randomly accessed. Moreover, it facilitates recording one program while simultaneously watching another program.

3.11.4 Advantages of Optical Disks

- Optical disks possess large capacity to store data/information in the form of multimedia, graphics and video files. They can store more data in less amount of space as compared to magnetic tapes and floppy or zip disks.
- The life span for data storage in optical disks is considered to be more, about 10–20 years as compared to magnetic disks, which have a comparatively lesser life span.
- Optical disks hold more data-recording density as compared to other storage media; therefore, they have low cost per bit of storage.
- Optical disk is not affected by magnetic field.
- An optical disk is tougher than tapes/floppy disks. It is physically harder to break or melt.
- Due to their small size and lightweight, these disks are easily portable and stored.

3.11.5 Disadvantages of Optical Disks

- Optical disk is not as easy to write as a floppy disk. One needs to use both software and hardware for writing optical disks.
- They possess slow data access speed as compared to the magnetic disks.
- The drive mechanism of an optical disk is more complicated than the magnetic and floppy disks.

3.12 MAGNETO-OPTICAL STORAGE DEVICES

The magneto-optical (MO) systems include basic principles of both magnetic and optical storage systems. This system writes magnetically (with thermal assist) and reads optically. Magnetic optical disk is a plastic or a glass disk coated with a compound (often a ternary alloy of terbium ferric cobalt, or TbFeCo) with special properties (see Figure 3.22). Since such materials are easily oxidized, dielectric barrier layers are used to protect the MO layer from oxidation. At the same time, the barriers together with the reflector coating act as an optical signal enhancement. MO disks function along opto-thermic magnetic principles. The thin film structure that gives the format its unlimited rewritability is based on a MO alloy layer enveloped by a barrier layer on each side.

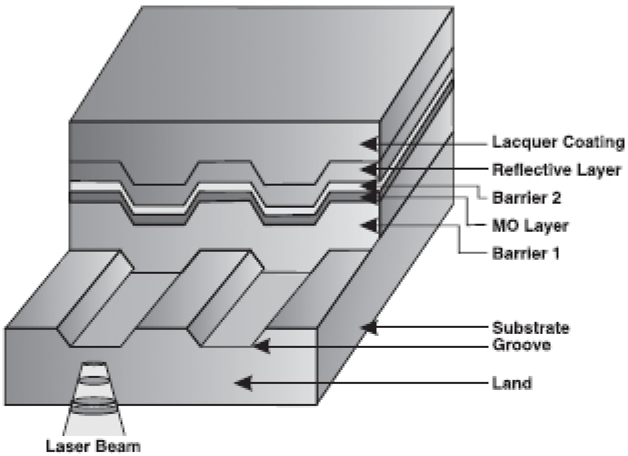


Figure 3.22 Magneto-Optical Disk Structure

Presently, MO disks are available in two formats: 5¼-inch and 3½-inch. The larger form-factor MO disks are capable of holding data about as much as the standard CD-ROM. Under pressure from the inexpensive and relatively fast CD-R and CD-RW, MO drives seem to be losing ground. On the other hand, some of the principles of the MO technology (thermally-assisted magnetic recording) may find their way into the most advanced magnetic storage devices of the future. Thus, the existence of the erasable optical disks has been made feasible due to this technology.

3.12.1 Basics of MO Reading

Traditional magnetic recording systems use currents induced in the magnetic heads by the changing magnetic fluxes on the disk surface to read the data. However, magneto-optical systems use polarized light to read the data from the disk (see Figure 3.23). The changes in light polarization occur due to the presence of a magnetic field on the surface of the disk. This phenomenon is known as the *Kerr effect*, where the polarity of the reflected light is altered depending on the orientation of the magnetic particles. If a beam of polarized light is shined on the surface, the light polarization of the reflected beam will change slightly, (typically less than 0.50) if it is reflected from a magnetized surface. If the magnetization is reversed, the change in polarization (the Kerr angle) is reversed too. The magnetized areas, that is, pits, cannot be seen in regular light, but only in polarized light. The change in direction of magnetization could be associated with numbers 0 or 1, making this technique useful for binary data storage.

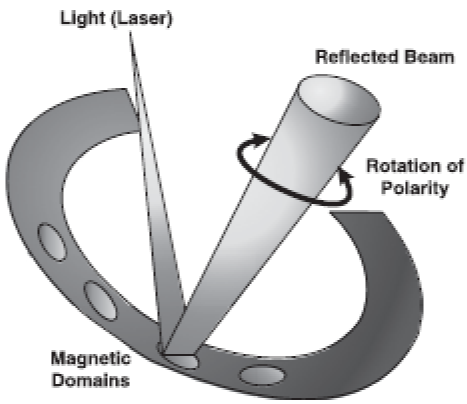


Figure 3.23 Reading from Magneto-optical Disk

Various other rewritable disks include CD-RW, DVD-RW, DVD-RAM and DVD+RW. The data layer for these disks uses a phase-changing metal alloy film. This film can be melted by the laser's heat to level out the marks made by the laser and then lasered again to record new data. In theory, one can erase and write on these disks as many as 1000 times, for CD-RW, and even 100,000 times for the DVD-RW types. DVD-RAM was the first erasable version of DVD to come to market and has subsequently found competition in the rival DVD-RW and DVD+RW format.

3.12.2 Basics of MO Recording

After reaching the *Curie temperature* (around 200°C), every magnetic material loses magnetization due to a complete disordering of their magnetic domains, and as a result, lose the data stored on them. Moreover, the material's *coercivity*, that is, the measure of the material's resistance to magnetization by the applied Lens magnetic field decreases as the temperature approaches the Curie point, and reaches zero when this temperature is exceeded. Therefore, it is imperative that the only change to the material, when it is heated and cooled, is the change in magnetization with no damage to the material itself.

The fact that the material's coercivity drops at higher temperatures allows thermally assisted magnetic recording with relatively weak magnetic fields. Even a relatively weak laser can generate high local temperatures when focused at a small spot (about 1 micron in case of MO systems). When the material is heated, and its coercivity is low, the magnetization of the media can be changed by applying a magnetic field from the magnet. When the material is cooled to room temperature, its coercivity rises back to such a high level that the magnetic data cannot be easily affected by the magnetic fields. The basic schematic of this recording process is illustrated in Figure 3.24.

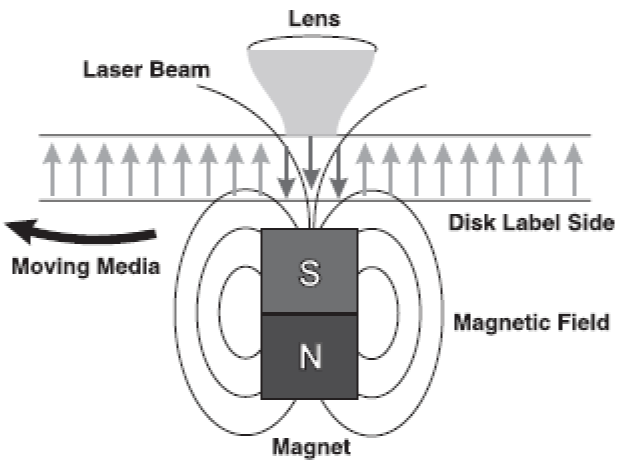


Figure 3.24 Writing to Magneto-Optical Disk

The recording layer is heated by the laser to a point where its magnetic orientation is dissipated. As this spot on the disk cools, the new magnetic orientation—corresponding to the new information—is set by the magnetic head to correspond to “0” and “1” of digital signals. Normally, to rewrite data on a MO disk, it is necessary to erase all previous signals before new data can be recorded. This means either two lasers are used to record (one to erase and one to record) or a single laser must first erase the data in a first rotation and then record the data the second time. This requires either a recorder with two lasers or longer recording times for a single laser system.

3.13 UNIVERSAL SERIAL BUS

Universal serial bus (USB), developed by Intel, is a set of connectivity specifications that establishes communication between personal computers and devices such as mouse, keyboard, pen drive, external hard disk drives, etc. Nowadays, almost every computer or laptop is equipped with one or more USB ports. The USB connector is a narrow socket around 1 cm wide and 0.5 cm high.

All USB devices come with a USB connector that is plugged into the USB port on the computer. As you plug in a USB device, it is detected by the computer and the required software is configured automatically; there is no need to restart the computer. Another advantage of USB is that it determines (and provides) the electrical power required by the device connected to it.

3.13.1 Pen Drive

A pen/flash drive is a removable storage device that is frequently used nowadays to transfer audio, video, and data files from one computer to another. A pen drive consists of a small printed circuit board, which is fitted inside a plastic, metal, or rubber casing to protect it. The USB connector which is present at one end of pen drive is protected by either a removable cap or pulling it back in the casing. Figure 3.25 shows a pen drive.

Figure 3.25 Pen Drive

The pen drive is a high-storage–capacity (ranging from 1 GB to 32 GB) device and is physically small enough to fit into a pocket. In addition, it is fast, robust and reliable and requires very less power to operate, which it gets through a USB port, and hence no battery is required.

Using a pen drive is easy. Following are the steps to use pen drive.

1. Insert the pen drive in the USB port of a computer. The computer detects the device, configures the necessary software and displays a drive corresponding to the pen drive in **My Computer**.
2. Perform any of the following actions:

    1. Double-click the drive icon in **My Computer** to access the files stored in the pen drive.
    2. Right-click the file or folder you want to send to the pen drive, which displays a menu. Select **Send To** and select the drive corresponding to the pen drive from the sub menu that appears.

3. After performing all your actions, left-click the **Safely Remove Hardware** icon in the system tray to display a menu containing **Safely remove USB Mass Storage Device** option.
4. Select this option to display a notification icon indicating that it is now safe to remove the hardware and plug-out the pen drive.

3.14 MEMORY STICK

Memory Stick, also known as *memory card*, is a digital storage device, which is designed to be used with portable electronic devices such as mobile phones, digital cameras, PDAs, iPod, etc. (Figure 3.27). It was launched in 1998 by Sony and immediately gained popularity due to its support for fast data transfer speed and large storage capacity. Though the original Memory Stick provided storage capacity of up to 128 MB, nowadays Memory Sticks with storage capacity of up to 32 GB are available in the market. Even the Memory Stick with the smallest storage capacity (4 MB) can store up to 80 images, which is much more than a standard 3½-inch floppy disk. The Memory Stick can be removed from the portable device and accessed by a personal computer using Memory Stick-capable memory card reader.

Figure 3.27 Memory Stick

Over the years and with advancement in technology, the Memory Stick is getting smaller in physical size and larger in logical size. Nowadays, several different standards or formats of Memory Stick are available in the market, which are as follows:

- **Memory Stick PRO:** It was introduced in the year 2003 jointly by two big companies—Sony and SanDisk. It supports marginally higher data transfer speed than the original one and provides theoretical storage capacity of up to 32 GB. It provides a 4-bit parallel interface with theoretical transfer rates of up to 480 Mb/s. It is widely used in high megapixel digital cameras and camcorders.
- **Memory Stick Duo:** It is a small size Memory Stick for small, pocket-sized devices such as mobile phones, music players, digital cameras, etc. It has reduced the use of large size Memory Stick, which had been very popular for a long time. It is equipped with MagicGate technology that is used to encrypt the data stored on the card. It provides all the features of the large standard Memory Stick and is smaller in size, but costs more.
- **Memory Stick PRO Duo:** Though Memory Stick Duo fulfils the need for pocket-sized devices, it has a slow transfer rate and limited storage capacity of 128 MB; however, Memory Stick PRO Duo supersedes it because Memory Stick PRO Duo provides larger memory space (up to 32 GB) and high speed of data transfer to/from the card.
- **Memory Stick PRO-HG Duo:** It was introduced by Sony and SanDisk in 2006. Unlike Memory Stick PRO which has a parallel interface of 4-bit, Memory Stick PRO-HG DUO has an 8-bit parallel interface. In addition, the clock frequency has increased from 40 MHz to 60 MHz in Memory Stick Pro-HG Duo. With these improvements, it provides a higher transfer speed than the Memory Stick Pro Duo.

- **Memory Stick Micro (M2):** It is a light and compact storage media which comes in the dimension of 15 mm × 12.5 mm × 1.2 mm. It has mainly been developed to meet the demands of the mobile devices market. It offers large storage capacity ranging from 16 MB to 32 GB and transfer speed of 160 Mb/s.

3.15 MASS STORAGE DEVICES

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

- **Redundant Array of Inexpensive Disks (RAID):** The basic idea of RAID is to combine multiple hard disks into an array of disk drives to obtain high performance, large capacity and reliability. These drives appear to the host computer as a single logical drive. The disk arrays can be made fault-tolerant by redundantly storing information in various ways. Seven types of array architectures, RAID 0 through RAID 6, have been defined; each provides disk fault-tolerance with different compromises in features and performance.
- **Automated Tape Library:** An automated tape library comprises numerous sets of magnetic tapes along with their drives and controllers mounted in a single unit (see Figure 3.28). The unit comprises one or more tape drives to perform read/write operations on the tapes in the tape library. In the multiple tape drive environment, these tapes can be simultaneously read or write, thus resulting in the speedy rate of data transfer. Multiple drives lead to the reliability of the storage unit because if one of the drives fails, then the unit can continue to operate with the other tape drives. The unit, with the help of a robotic arm, retrieves the appropriate tape from the tape library, mounts it on the tape drive for processing, and then returns to the library after the job has been finished. These tape libraries can store up to several terabytes of data in it, so they can be used for archiving data for offline storage and also as data backup devices during online storage.

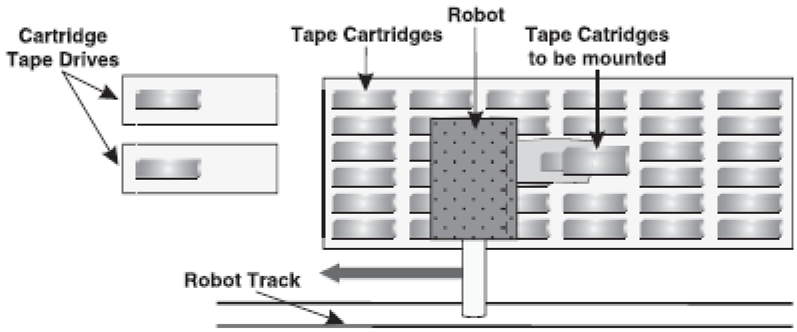


Figure 3.28 Automated Tape Library

- **CD-ROM Jukebox:** A CD-ROM jukebox comprises numerous sets of CD-ROM disks along with their drives and controllers mounted in a single unit. The unit comprises one or more CD-ROM drives to perform read/write operations on the CD-ROM in the jukebox. In the multiple CD-ROM drive environment, these CD-ROMs can be simultaneously read or write, resulting in the speedy rate of data transfer. Multiple drives lead to the reliability of the storage unit because if one of the drives fails, then the unit can continue to operate with other CD-ROM drives. The unit, with the help of a robotic arm, retrieves the appropriate CD-ROM from the CD-ROM jukebox, mounts it on the CD-ROM drive for processing, and then returns it to the appropriate slot in the jukebox after the job has been finished. These jukeboxes can store up to several terabytes of data in it, so they can be used for archiving of read-only data for offline storage and also as data backup devices during online storage.

Mass storage devices have relatively slow access time, generally in the order of seconds instead of milliseconds. Due to this, their use is limited in a number of applications. The slow access time of mass storage devices is because of the time taken by the transport mechanism to first move onto the particular storage media such as disks and then the time taken, in few milliseconds, to transfer the desired data to the main memory in the computer system. However, they have huge amounts of storage capacity and possess minimum cost per bit storage. Mass storage devices are a cost-effective option, as compared to the online tapes and disks storage, in situations where a large storage capacity is required and where prompt data access is not essential. When used as offline storage, they are referred to as *archival storage*.

LET US SUMMARIZE

1. *Computer memory* refers to the electronic holding place for instructions and data where the processor can reach quickly.
2. Various units that are used to measure computer memory are *bit, byte, kilobyte, megabyte, gigabyte and terabyte*.

3. A computer's memory can be categorized into three fundamental types: *internal processor memory*, *primary memory* and *secondary memory*.
4. *Internal processor memory* is placed within the CPU (processor) and usually includes cache memory and special registers, both of which can be directly accessed by the processor. This memory is used for temporary storage of data and instructions on which the CPU is currently working.
5. Primary memory consists of *RAM*, which is volatile in nature, and *ROM*, which is nonvolatile in nature.
6. *RAM*, also called *main memory*, allows the computer to store data for immediate manipulation and to keep track of what is currently being processed. It is the place in a computer where the operating system, application programs and data in current use are kept so that they can be accessed quickly by the computer's processor.
7. *SRAM* stands for static random access memory. It retains its contents as long as power is being supplied and does not require constant periodic refreshing. It is often used as cache memory due to its high speed.
8. *DRAM* stands for dynamic random access memory. The data on DRAM continues to move in and out of the memory as long as power is available, and thus DRAM must be continually refreshed in order to maintain the data. DRAM is slower and less expensive than SRAM.
9. The non-volatile memory chip, where the start-up instructions (special boot program) are stored is called *ROM*. These chips are programmed by burning appropriate fuses to form patterns of binary information.
10. *Secondary memory* is used to provide backup storage for instructions and data. It has much larger capacity than primary memory. It includes devices like hard disk, floppy disk, CD-ROM and tape drives. These devices are classified into two types according to data access, that is, *sequential access* and *direct access*.
11. *Magnetic tape* is like a plastic tape with a magnetic coating on it. The data is stored in the form of tiny segments of magnetized and demagnetized portions on the surface of the material. The tapes are an inexpensive and a reliable storage medium for organizing archives and taking backup.
12. Magnetic tape is divided into vertical columns (*frames*) and horizontal rows (*channels* or *tracks*). The data is stored in a string of frames with one character per frame and each frame spans multiple tracks. Thus, a single bit is stored in each track, that is, one byte per frame. A magnetic tape can typically have seven to nine tracks.
13. The length of tape between the *BOT* (*beginning of tape*) and *EOT* (*end of tape*) is referred to as the usable *recording (reading/writing)* surface. BOT/EOT markers are usually made up of short silver strips of reflective type.
14. *Magnetic disks* are the most widely used and popular medium for direct access secondary storage. They offer high storage capacity and reliability, and have the capability to access the stored data directly.

15. A magnetic disk consists of a plastic/metal circular plate/platter, which is coated with a magnetic oxide layer. Data are represented as magnetized spots on a disk. A magnetized spot represents 1 and the absence of a magnetized spot represents 0.
16. *Tracks* are concentric circles where the data are stored, and are numbered from the outermost to the innermost ring, starting with zero.
17. *Sectors* refer to the number of fixed-size areas (imaginary pie slices) that can be accessed by one of the disk drive's read/write heads, in one rotation of the disk, without the head having to change its position. An intersection of a track and a disk sector is known as track sector.
18. A *floppy disk* is a round, flat piece of Mylar plastic, coated with ferric oxide, and encased in a protective plastic. It is a removable disk and is read and written by a floppy disk drive. A *floppy disk* drive is a device that performs the basic operation on a disk, including rotating the disk and reading and writing data onto it.
19. The *hard disk* is the primary storage unit of the computer. It consists of a stack of disk platters that are made up of aluminium alloy or glass substrate coated with a magnetic material and protective layers. It plays a significant role in: performance, storage capacity, software support and reliability.
20. *Zip disk* is a removable storage device having a capacity to store 250–700 MB of data. Zip disk's drive unit is measured as 18 × 13 × 4 cm and weighs about half a kilogram. It has rubber feet to stabilize the unit in either vertical or horizontal positions. The substrate for the disk is made up of plastic material on which magnetic oxide particles are coated.
21. An *optical disk* is a flat, circular, plastic disk coated with material on which bits may be stored in the form of highly reflective areas and significantly less reflective areas, from which the stored data may be read when illuminated with a narrow beam of laser diode. These disks are capable of storing enormously high amounts of data in a limited amount of space.
22. *Memory Stick*, also known as *Memory Card* is a digital storage device, which is designed to be used with portable electronic devices such as mobile phones, digital cameras, PDAs, iPod, etc.
23. *Magneto-optical* (*MO*) systems include the basic principles of both magnetic and optical storage systems. This system writes magnetically (with thermal assist) and reads optically.
24. *USB*, developed by Intel, is a set of connectivity specification that establishes communication between personal computers and devices such as mouse, keyboard, pen drive, external hard disk drives, etc.
25. In order to achieve a vast amount of storage capacity in a computer system, multiple units of similar kinds of storage media are associated together to form a chain of mass storage devices. These storage media include multiple arrays of magnetic disks or CD-ROMs as a secondary storage device.