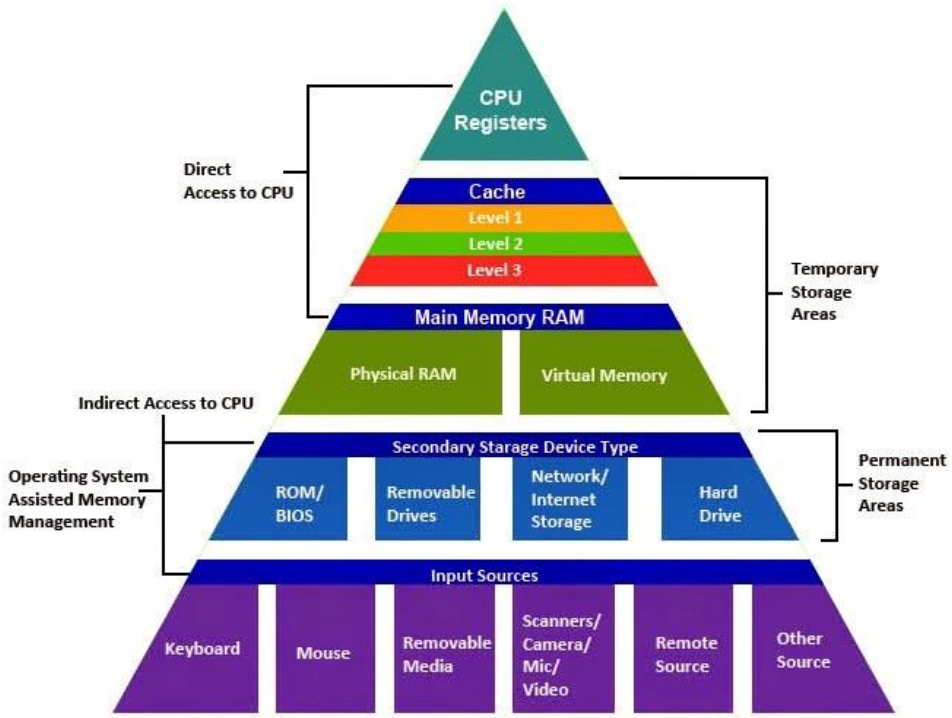
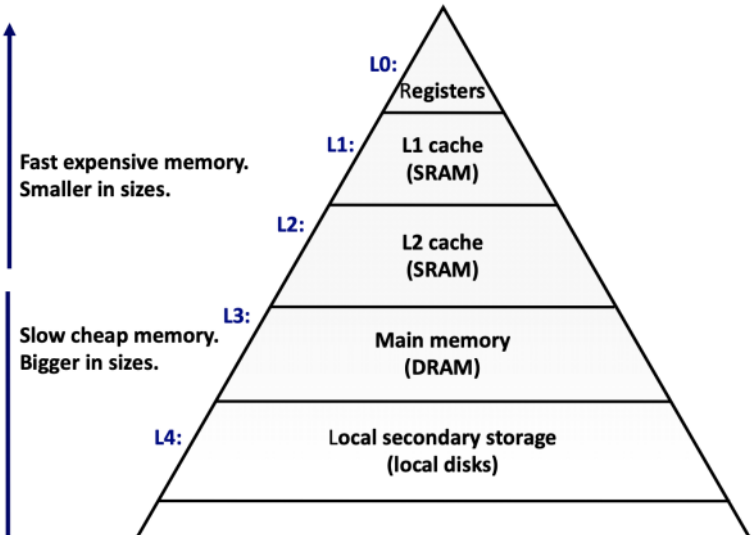
| | |
|---|---|
| SASTRA DEEMED TO BE UNIVERSITY | **COMPUTER ORGANIZATION AND ARCHITECTURE**<br>**Course Code: CAP403R01 Semester : I / MCA**<br>**Dr. K. Geetha/SAP/CSE/SOC/SASTRA** |

## 1.7 Memory System and Hierarchy-P1

### Memory Hierarchy



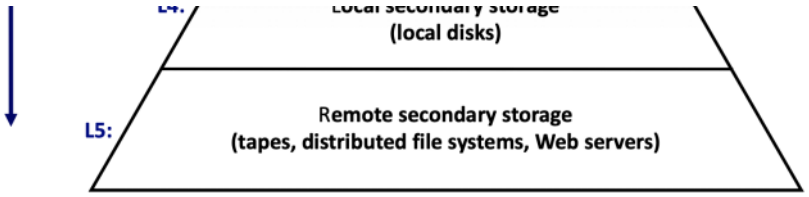
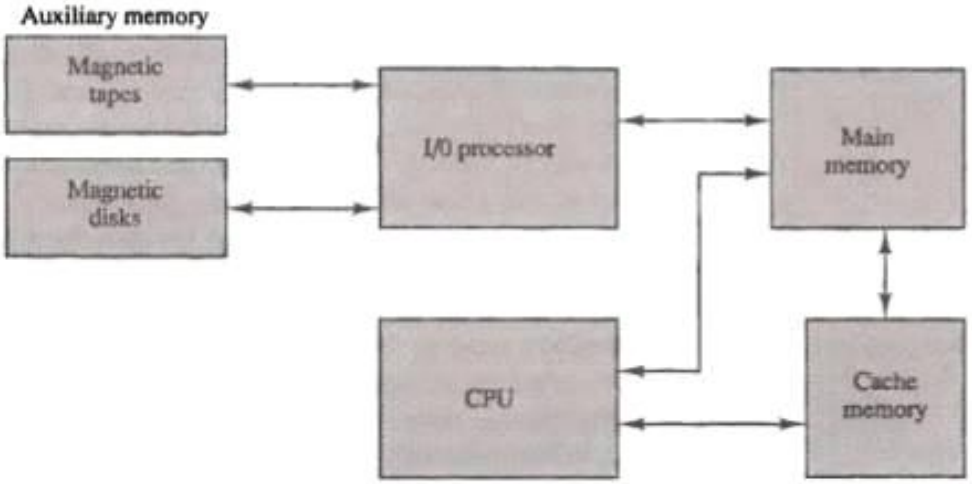
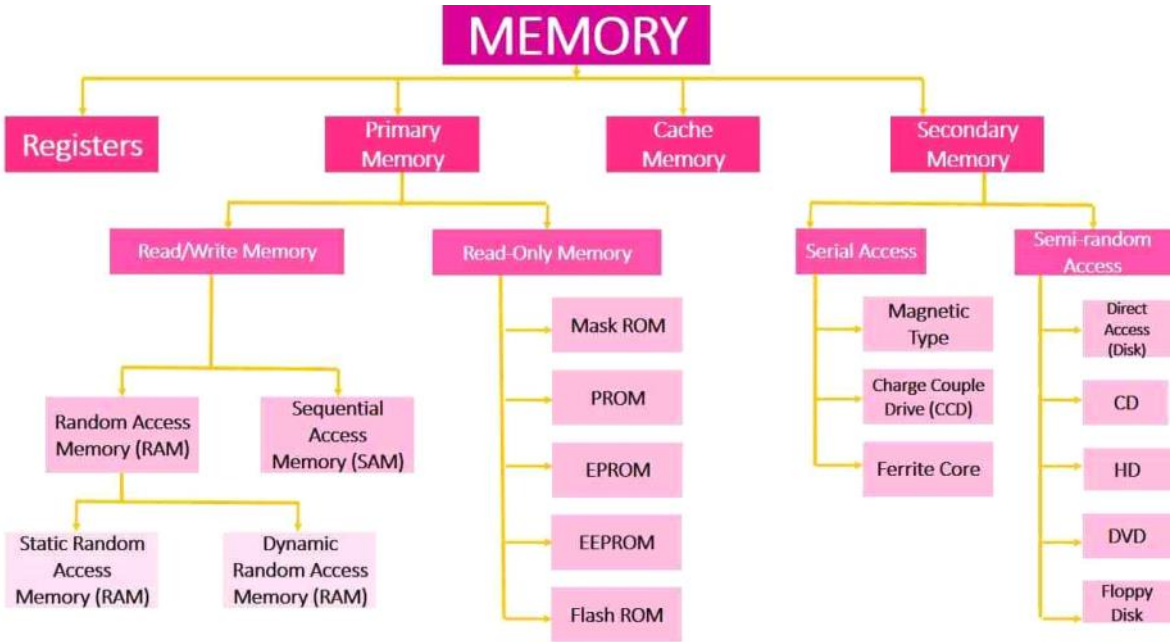


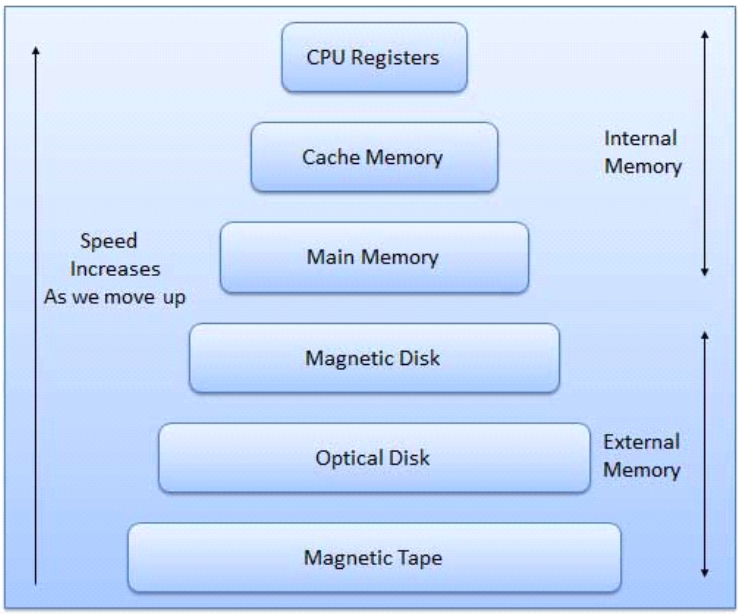

Figure 12-1 Memory hierarchy in a computer system.




## Classification of Memory




## 1.7 Memory System and Hierarchy-P2




## Features of Memory

Following are the different features of the memory system that includes:

1. **Location:** It represents the internal or external location of the memory in a computer. The internal memory is inbuilt in computer memory. It is also known as primary memory. the example of primary memory are registers, cache and main memory. Whereas, external memory is the separate storage device from the computer, such as disk, tape, USB pen drive.
2. **Capacity:** It is the most important feature of computer memory. Storage capacity can vary in external and internal memory. External devices' storage capacity is measured in terms of bytes, whereas the internal memory is measured with bytes or words. The storage word length can vary in bits, such as 8, 16 or 32 bits.
3. **Access Methods:** Memory can be accessed through four modes of memory.
   - **DMA:** As the name specifies, Direct Memory Address (DMA) is a method that allows input/output (I/O) devices to access or retrieve data directly or from the main memory.
   - **Sequential Access Method:** The sequential access method is used in a data storage device to read stored data sequentially from the computer memory. Whereas, the data received from random access memory (RAM) can be in any order.
   - **Random Access Method:** It is a method used to randomly access data from memory. This method is the opposite of SAM. For example, to go from A to Z in random access, we can directly jump to any specified location. In the Sequential method, we have to follow all intervening from A to Z to reach at the particular memory location.
   - **Associative Access Method:** It is a special type of memory that optimizes search performance through defined data to directly access the stored information based on a memory address.
4. **Unit of transfer:** As the name suggests, a unit of transfer measures the transfer rate of bits that can be read or write in or out of the memory devices. The transfer rate of data can be different in external and internal memory.
   - **Internal memory:** The transfer rate of bits is mostly equal to the word size.

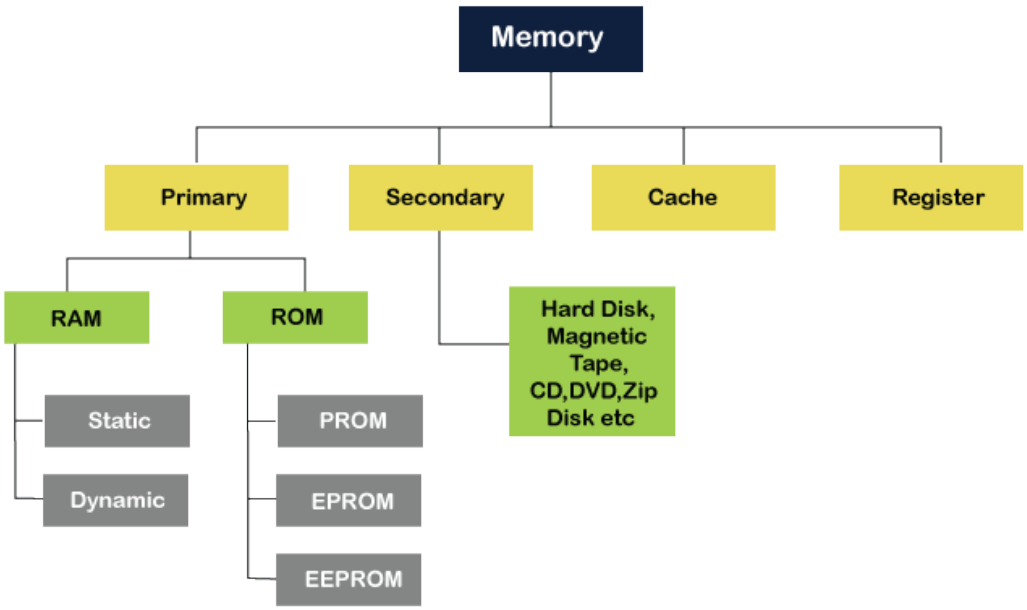
and internal memory.

- **Internal memory:** The transfer rate of bits is mostly equal to the word size.
- **External memory:** The transfer rate of bit or unit is not equal to the word length. It is always greater than a word or may be referred to as **blocks**.

5. **Performance:** The performance of memory is majorly divided into three parts.
   - **Access Time:** In random access memory, it represents the total time taken by memory devices to perform a read or write operation that an address is sent to memory.
   - **Memory Cycle Time:** Total time required to access memory block and additional required time before starting second access.
   - **Transfer rate:** It describes the transfer rate of data used to transmit memory to or from an external or internal memory device. Bit transfer can be different for different external and internal devices.
6. **Physical types:** It defines the physical type of memory used in a computer such as magnetic, semiconductor, magneto-optical and optical.
7. **Organization:** It defines the physical structure of the bits used in memory.
8. **Physical characteristics:** It specifies the physical behavior of the memory like volatile, non-volatile or non-erasable memory. Volatile memory is known as RAM, which requires power to retain stored information, and if any power loss has occurred, stored data will be lost. Non-volatile memory is a permanent storage memory that is used to obtain any stored information, even when the power is off. Non-erasable memory is a type of memory that cannot be erased after the manufactured like ROM because at the time of manufactured ROM are programmed.

## Classification of Memory

The following figure represents the classification of memory:




## Primary or Main Memory

Primary memory is also known as the computer system's main memory that communicates directly within

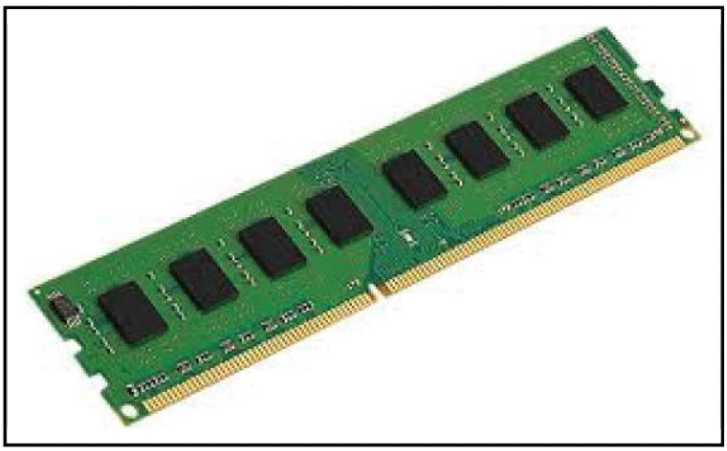
the CPU, Auxiliary memory and the Cache memory.
Main memory is used to kept programs or data when the processor is active to use them.
When a program or data is activated to execute, the processor first loads instructions or programs from secondary memory into main memory, and then the processor starts execution.
Accessing or executing of data from primary memory is faster because it has a cache or register memory that provides faster response, and it is located closer to the CPU.
The primary memory is volatile, which means the data in memory can be lost if it is not saved when a power failure occurs.
It is costlier than secondary memory, and the main memory capacity is limited as compared to secondary memory.

The primary memory is further divided into two parts:

1. RAM (Random Access Memory)
2. ROM (Read Only Memory)

## Random Access Memory (RAM)

Random Access Memory (RAM) is one of the faster types of main memory accessed directly by the CPU.
It is the hardware in a computer device to temporarily store data, programs or program results.
It is used to read/write data in memory until the machine is working.
It is volatile, which means if a power failure occurs or the computer is turned off, the information stored in RAM will be lost.
All data stored in computer memory can be read or accessed randomly at any time.



There are two types of RAM:

- SRAM
- DRAM

**DRAM:** DRAM **(Dynamic Random-Access Memory)** is a type of RAM that is used for the dynamic storage of data in RAM.
In DRAM, each cell carries one-bit information. The cell is made up of two parts: a **capacitor** and a **transistor**.
The size of the capacitor and the transistor is so small, requiring millions of them to store on a single chip. Hence, a DRAM chip can hold more data than an SRAM chip of the same size.
However, the capacitor needs to be continuously refreshed to retain information because DRAM is

volatile.
If the power is switched off, the data store in memory is lost.

**Characteristics of DRAM**

1. It requires continuously refreshed to retain the data.
2. It is slower than SRAM
3. It holds a large amount of data
4. It is the combination of capacitor and transistor
5. It is less expensive as compared to SRAM
6. Less power consumption

**SRAM:** SRMA **(Static Random-Access Memory)** is a type of RAM used to store static data in the memory. It means to store data in SRAM remains active as long as the computer system has a power supply. However, data is lost in SRAM when power failures have occurred.

**Characteristics of Static Ram**

1. It does not require to refresh.
2. It is faster than DRAM
3. It is expensive.
4. High power consumption
5. Longer life
6. Large size
7. Uses as a cache memory

SRAM Vs. DRAM

| SRAM | DRAM |
|---|---|
| It is a Static Random-Access Memory. | It is a Dynamic Random Access Memory. |
| The access time of SRAM is slow. | The access time of DRAM is high. |
| It uses flip-flops to store each bit of information. | It uses a capacitor to store each bit of information. |
| It does not require periodic refreshing to preserve the information. | It requires periodically refreshing to preserve the information. |
| It uses in cache memory. | It is used in the main memory. |
| The cost of SRAM is expensive. | The cost of DRAM is less expensive. |
| It has a complex structure. | Its structure is simple. |
| It requires low power consumption. | It requires more power consumption. |

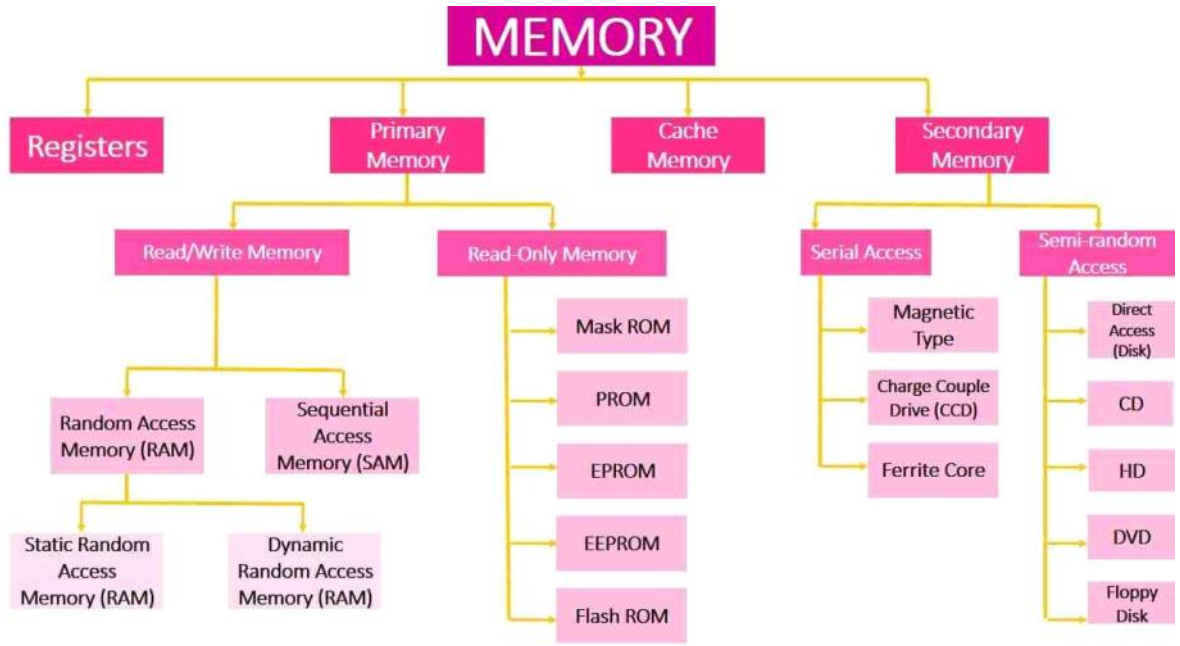
**Advantages of RAM**

- It is a faster type of memory in a computer.

- It requires less power to operate.
- Program loads much faster
- More RAM increases the performance of a system and can multitask.
- Perform read and write operations.
- The processor can read information faster than a hard disc, floppy, USB, etc.

**Disadvantages of RAM**

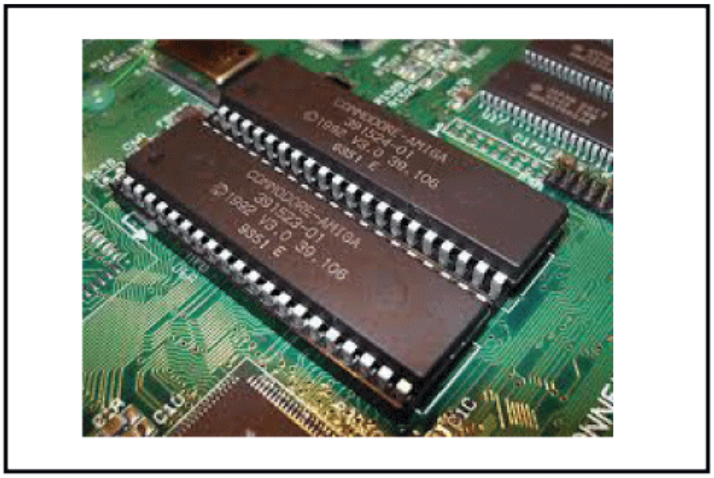
- Less RAM reduces the speed and performance of a computer.
- Due to volatile, it requires electricity to preserve the data.
- It is expensive than ROM
- It is unreliable as compared to ROM
- The Size of RAM is limited.




## 1.7 Memory System and Hierarchy-P3

### Read-Only Memory (ROM)

ROM is a memory device or storage medium that is used to **permanently** store information inside a chip.
It is a read-only memory that can only read stored information, data or programs, but we cannot write or modify anything.
A ROM contains some important instructions or program data that are required to start or boot a computer.
It is a **non-volatile** memory; it means that the stored information cannot be lost even when the power is turned off or the system is shut down.



**Types of ROM**

There are five types of Read Only Memory:

1. **MROM (Masked Read Only Memory):**
   MROM is the oldest type of read-only memory whose program or data is pre-configured by the integrated circuit manufacture at the time of manufacturing. Therefore, a program or instruction stored within the MROM chip cannot be changed by the user.

2. **PROM (Programmable Read Only Memory):**
   It is a type of digital read-only memory, in which the user can write any type of information or program only once.
   It means it is the empty PROM chip in which the user can write the desired content or program only once using the special PROM programmer or PROM burner device; after that, the data or instruction cannot be changed or erased.

3. **EPROM (Erasable and Programmable Read Only Memory):**
   It is the type of read only memory in which stored data can be erased and re-programmed only once in the EPROM memory. It is a non-volatile memory chip that holds data when there is no power supply and can also store data for a minimum of 10 to 20 years.
   In EPROM, if we want to erase any stored data and re-programmed it, first, we need to pass the ultraviolet light for 40 minutes to erase the data; after that, the data is re-created in EPROM.

4. **EEPROM (Electrically Erasable and Programmable Read Only Memory):**
   The EEROM is an electrically erasable and programmable read only memory used to erase stored data using a high voltage electrical charge and re-programmed it.
   It is also a non-volatile memory whose data cannot be erased or lost; even the power is turned off.
   In EEPROM, the stored data can be erased and reprogrammed up to 10 thousand times, and the data erase one byte at a time.

5. **Flash ROM:**

Flash memory is a non-volatile storage memory chip that can be written or programmed in small units called Block or Sector. Flash Memory is an EEPROM form of computer memory, and the contents or data cannot be lost when the power source is turned off. It is also used to transfer data between the computer and digital devices.

**Advantages of ROM**

1. It is a non-volatile memory in which stored information can be lost even power is turned off.
2. It is static, so it does not require refreshing the content every time.
3. Data can be stored permanently.
4. It is easy to test and store large data as compared to RAM.
5. These cannot be changed accidently
6. It is cheaper than RAM.
7. It is simple and reliable as compared to RAM.
8. It helps to start the computer and loads the OS.

**Disadvantages of ROM**

1. Store data cannot be updated or modify except to read the existing data.
2. It is a slower memory than RAM to access the stored data.
3. It takes around 40 minutes to destroy the existing data using the high charge of ultraviolet light.

RAM Vs. ROM

| RAM | ROM |
|---|---|
| It is a Random-Access Memory. | It is a Read Only Memory. |
| Read and write operations can be performed. | Only Read operation can be performed. |
| Data can be lost in volatile memory when the power supply is turned off. | Data cannot be lost in non-volatile memory when the power supply is turned off. |
| It is a faster and expensive memory. | It is a slower and less expensive memory. |
| Storage data requires to be refreshed in RAM. | Storage data does not need to be refreshed in ROM. |
| The size of the chip is bigger than the ROM chip to store the data. | The size of the chip is smaller than the RAM chip to store the same amount of data. |
| Types of RAM: DRAM and SRAM | Types of ROM: MROM, PROM, EPROM, EEPROM |

## Secondary Memory

Secondary memory is a **permanent storage** space to hold a large amount of data. Secondary memory is also known as external memory that representing the various storage media (hard drives, USB, CDs, flash drives and DVDs) on which the computer data and program can be saved on a long term basis.
It is cheaper and slower than the main memory.
Unlike primary memory, secondary memory cannot be accessed directly by the CPU. Instead of that,

secondary memory data is first loaded into the RAM (Random Access Memory) and then sent to the processor to read and update the data.
Secondary memory devices also include magnetic disks like hard disk and floppy disks, an optical disk such as CDs and CDROMs, and magnetic tapes.

## Features of Secondary Memory

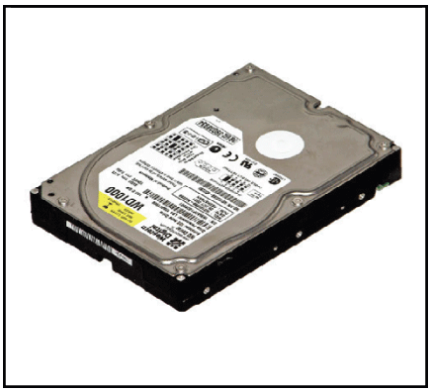
- Its speed is slower than the primary/ main memory.
- Store data cannot be lost due to non-volatile nature.
- It can store large collections of different types, such as audio, video, pictures, text, software, etc.
- All the stored data in a secondary memory cannot be lost because it is a permanent storage area; even the power is turned off.
- It has various optical and magnetic memories to store data.

## Types of Secondary Memory

The following are the types of secondary memory devices:

**Hard Disk**

A hard disk is a computer's permanent storage device.
It is a non-volatile disk that permanently stores data, programs, and files, and cannot lose store data when the computer's power source is switched off.
Typically, it is located internally on computer's motherboard that stores and retrieves data using one or more rigid fast rotating disk platters inside an air-sealed casing.
It is a large storage device, found on every computer or laptop for permanently storing installed software, music, text documentation, videos, operating system, and data until the user did not delete.



**Floppy Disk**

A floppy disk is a secondary storage system that consisting of thin, flexible magnetic coating disks for holding electronic data such as computer files.
It is also known as Floppy Diskette that comes in three sizes like 8 inches, 5.5 inches and 3.5 inches.
The stored data of a floppy disk can be accessed through the floppy disk drive.
It is the oldest type of portable storage device, which can store data up to 1.44 MB. S
It is rarely used due to very low memory storage.

**CD (Compact Disc)**

A CD is an optical disk storage device, stands for Compact Disc.
It is a storage device used to store various data types like
audio, videos, files, OS, Back-Up file, and any other information useful to a computer.
The CD has a width of 1.2 mm and 12 cm in height,
which can store approximately 783 MB of data size.
It uses laser light to read and write data from the CDs.

**Types of CDs**

1. **CD-ROM (Compact Disc Read Only Memory):** It is mainly used for bulk size mass like audio CDs, software and computer games at the time of manufacture.
   Users can only read data, text, music, videos from the disc, but they cannot modify or burnt it.
2. **CD-R (Compact Disc Recordable):** The type of Compact Disc used to write once by the user;
   after that, it cannot be modified or erased.
3. **CD-RW (Compact Disc Rewritable):** It is a rewritable CD disc, often used to write or delete the stored data.

**DVD Drive/Disc**

DVD is an optical disc storage device, stands for **Digital Video Display or Digital Versatile Disc**.
It has the same size as a CD but can store a larger amount of data than a compact disc.
It was developed in **1995** by Sony, Panasonic, Toshiba and Philips four electronics companies.
DVD drives are divided into three types, such as DVD ROM (Read Only Memory),
**DVD R** (Recordable) and **DVD RW** (Rewritable or Erasable).

It can store multiple data formats like audio, videos, images, software, operating system, etc.
The storing capacity of data in DVD is 4.7 GB to 17 GB.


**Blu Ray Disc (BD)**

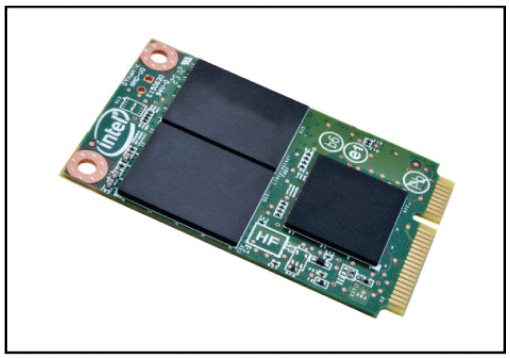
Blu Ray is an Optical disc storage device used to store a large amount of data
or high definition of video recording and playing other media files.
It uses laser technology to read the stored data of the Blu-ray Disk.
It can store more data at a greater density as compared to CD/ DVD.
For example, compact discs allow us to store 700 MB of data,
and in DVDs, it provides up to 8 GB of storage capacity,
while Blu-ray Discs provide 28 GB of space to store data.

**Pen Drive**

A pen drive is a portable device used to permanently store data
and is also known as a USB flash drive.
It is commonly used to store and transfer the data connected to a computer using a USB port.
It does not have any moveable part to store the data;
it uses an integrated circuit chip that stores the data.
It allows the users to store and transfer data like audio, videos, images, etc.
from one computer to any USB pen drive.
The storing capacity of pen drives from 64 MB to 128 GB or more.

## Cache Memory

It is a small-sized chip-based computer memory that lies between the CPU and the main memory.
It is a faster, high performance and temporary memory to enhance the performance of the CPU.
It stores all the data and instructions that are often used by computer CPUs.
It also reduces the access time of data from the main memory.
It is faster than the main memory, and sometimes, it is also called CPU memory because it is very close to the CPU chip.



The following are the levels of cache memory.

1. **L1 Cache:** The L1 cache is also known as the onboard, internal, or primary cache. It is built with the help of the CPU. Its speed is very high, and the size of the L1 cache varies from 8 KB to 128 KB.

2. **L2 Cache:** It is also known as external or secondary cache, which requires fast access time to store temporary data. It is built into a separate chip in a motherboard, not built into the CPU like the L1 level. The size of the L2 cache may be 128 KB to 1 MB.

3. **L3 Cache:** L3 cache levels are generally used with high performance and capacity of the computer. It is built into a motherboard. Its speed is very slow, and the maximum size up to 8 MB.

## Advantages of Cache Memory

1. Cache memory is the faster memory as compared to the main memory.
2. It stores all data and instructions that are repeatedly used by the CPU for improving the performance of a computer.
3. The access time of data is less than the main memory.

## Disadvantage of Cache Memory

1. It is very costly as compared to the Main memory and the Secondary memory.
2. It has limited storage capacity.

**Register Memory**

The register memory is a temporary storage area for storing and transferring the data and the instructions to a computer.
It is the smallest and fastest memory of a computer.
It is a part of computer memory located in the CPU as the form of registers.
The register memory is 16, 32 and 64 bits in size.
It temporarily stores data instructions and the address of the memory that is repeatedly used to provide faster response to the CPU.

## Primary Vs. Secondary Memory

| Primary Memory | Secondary Memory |
|---|---|
| It is also known as temporary memory. | It is also known as a permanent memory. |
| Data can be access directly by the processor or CPU. | Data cannot be accessed directly by the I/O processor or CPU. |
| Stored data can be a volatile or non-volatile memory. | The nature of secondary memory is always non-volatile. |
| It is more costly than secondary memory. | It is less costly than primary memory. |
| It is a faster memory. | It is a slower memory. |
| It has limited storage capacity. | It has a large storage capacity. |
| It required the power to retain the data in primary memory. | It does not require power to retain the data in secondary memory. |
| Examples of primary memory are RAM, ROM, Registers, EPROM, PROM and cache memory. | Examples of secondary memory are CD, DVD, HDD, magnetic tapes, flash disks, pen drive, etc. |

**RAM and ROM Chips**

Figure 12-2 Typical RAM chip.

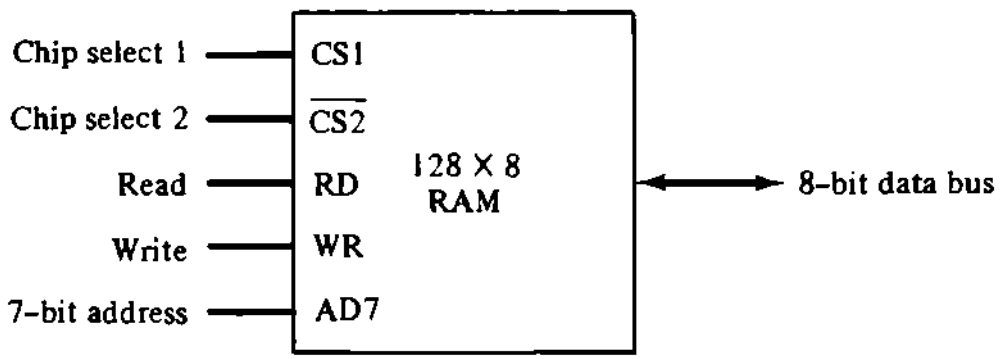
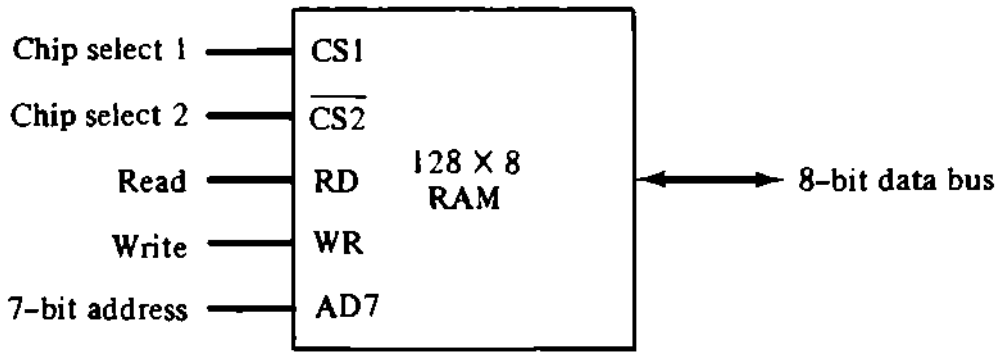


Figure 12-2 Typical RAM chip.




(a) Block diagram

| CS1 | $\overline{CS2}$ | RD | WR | Memory function | State of data bus |
|---|---|---|---|---|---|
| 0 | 0 | × | × | Inhibit | High-impedance |
| 0 | 1 | × | × | Inhibit | High-impedance |
| 1 | 0 | 0 | 0 | Inhibit | High-impedance |
| 1 | 0 | 0 | 1 | Write | Input data to RAM |
| 1 | 0 | 1 | × | Read | Output data from RAM |
| 1 | 1 | × | × | Inhibit | High-impedance |

(b) Function table

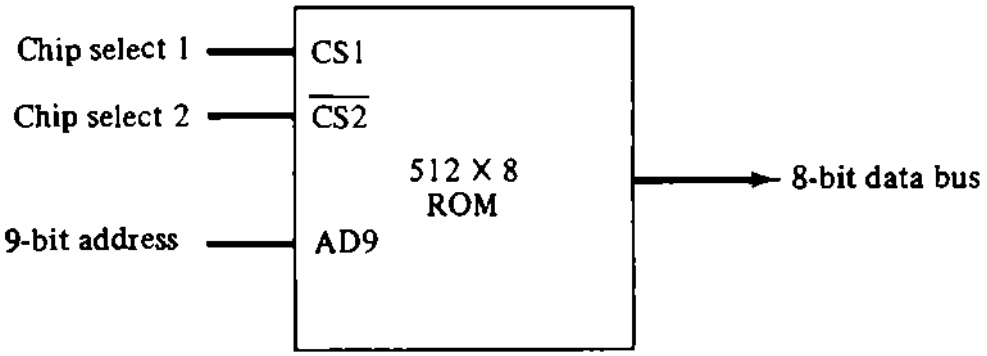



Figure 12-3 Typical ROM chip.

## Memory Address Map

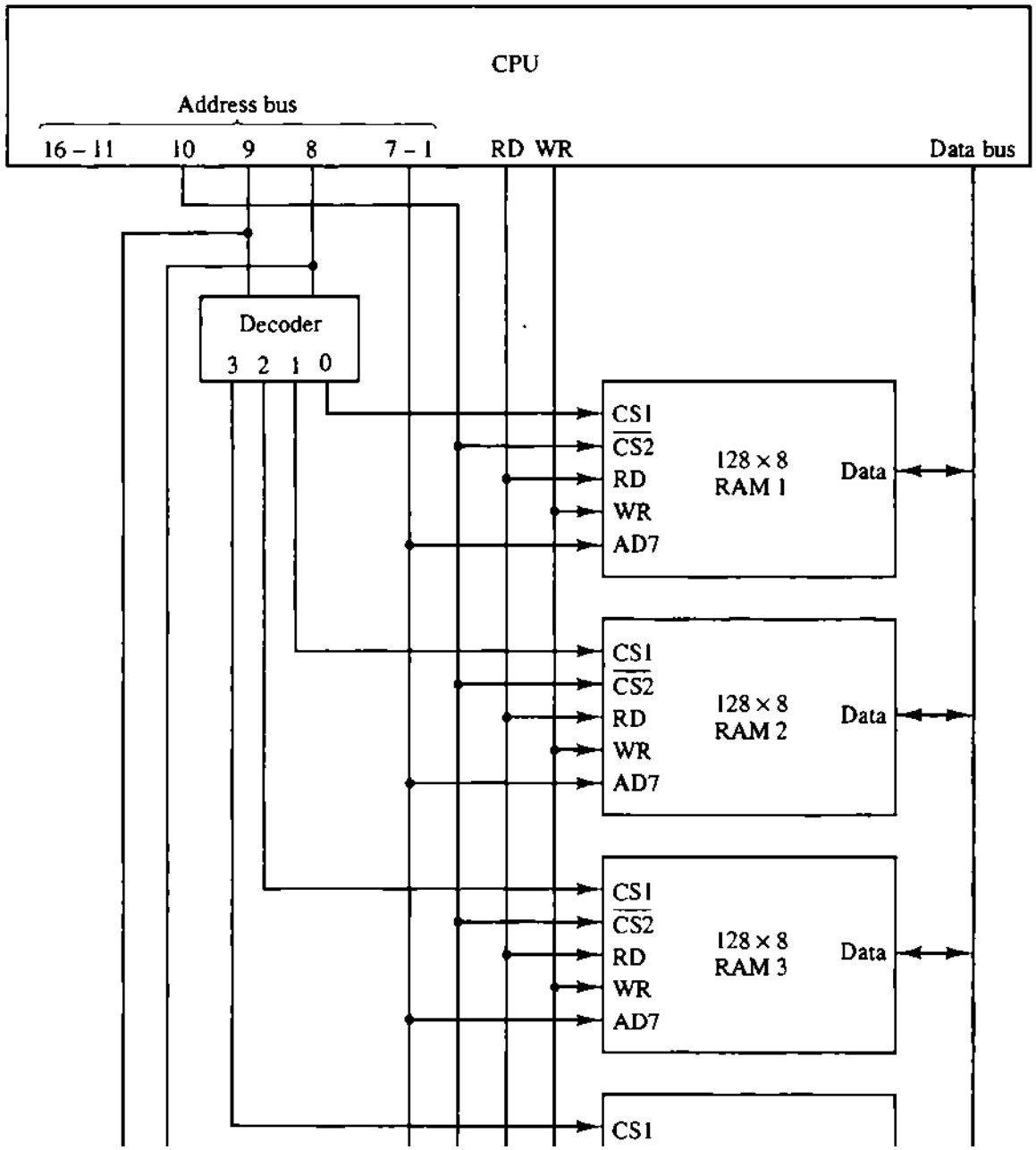
TABLE 12-1 Memory Address Map for Microprocomputer

| Component | Hexadecimal address | Address bus 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RAM 1 | 0000–007F | 0 | 0 | 0 | x | x | x | x | x | x | x |
| RAM 2 | 0080–00FF | 0 | 0 | 1 | x | x | x | x | x | x | x |

**TABLE 12-1** Memory Address Map for Microprocomputer

| Component | Hexadecimal address | Address bus 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RAM 1 | 0000–007F | 0 | 0 | 0 | x | x | x | x | x | x | x |
| RAM 2 | 0080–00FF | 0 | 0 | 1 | x | x | x | x | x | x | x |
| RAM 3 | 0100–017F | 0 | 1 | 0 | x | x | x | x | x | x | x |
| RAM 4 | 0180–01FF | 0 | 1 | 1 | x | x | x | x | x | x | x |
| ROM | 0200–03FF | 1 | x | x | x | x | x | x | x | x | x |

## Memory Connection to CPU



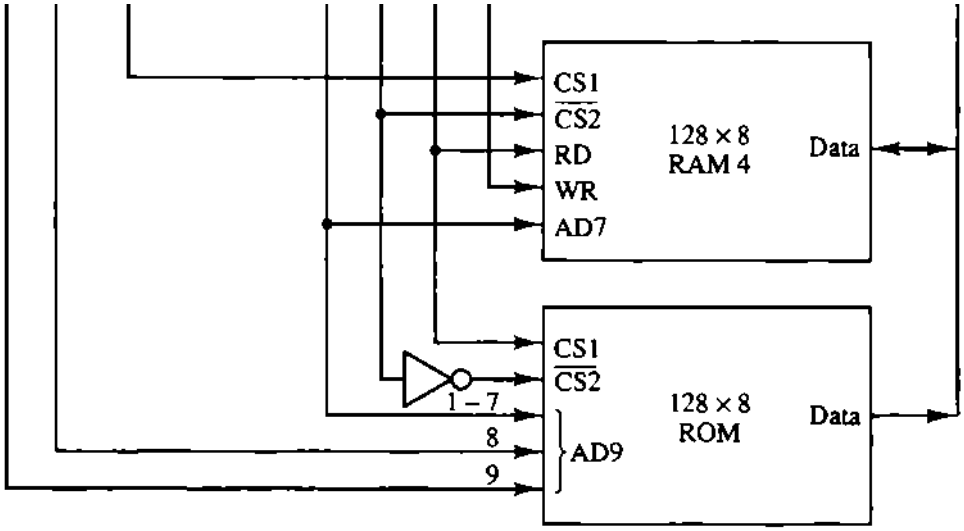



**Figure 12-4** Memory connection to the CPU.



**Auxiliary Memory**

Auxiliary memory is much larger in size than main memory but is slower.
It normally stores system programs, instruction and data files.
It is also known as secondary memory.
It can also be used as an overflow/virtual memory in case the main memory capacity has been exceeded.

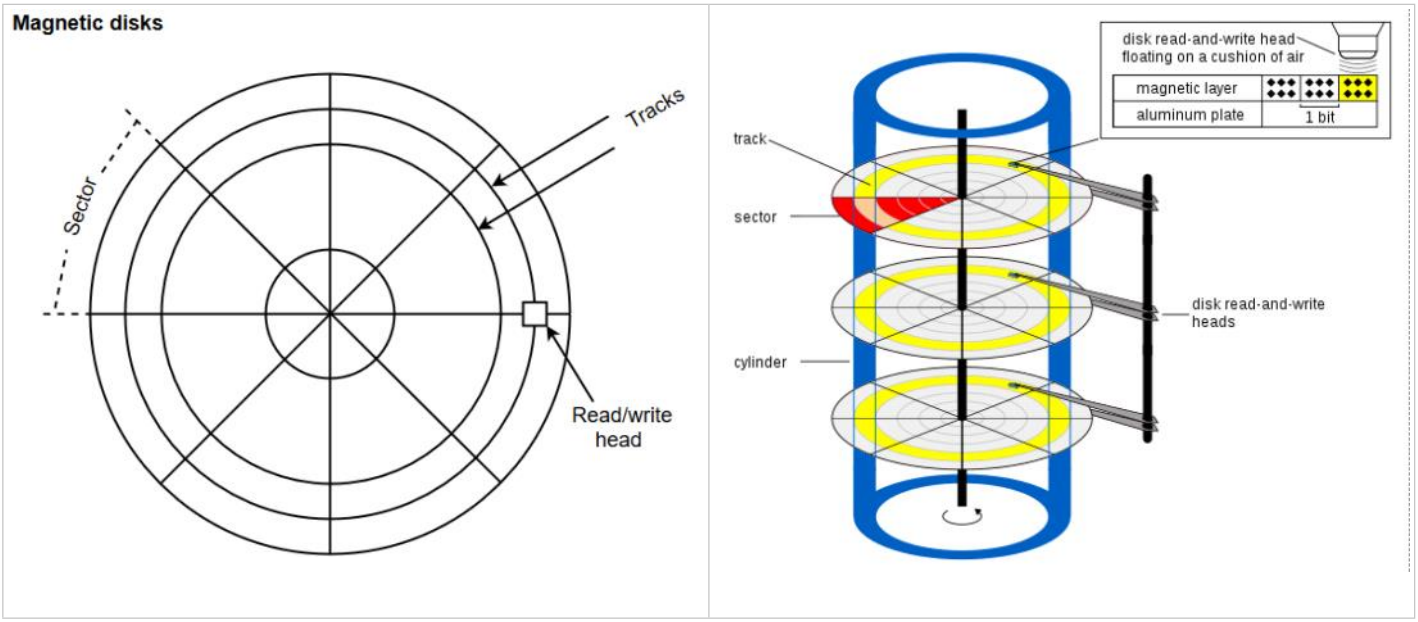
**Characteristics of Auxiliary Memory are following :**

- Non-volatile memory — Data is not lost when power is cut off.
- Reusable — The data stays in the secondary storage on permanent basis until it is not overwritten or deleted by the user.
- Reliable — Data in secondary storage is safe because of high physical stability of secondary storage device.
- Convenience — With the help of a computer software, authorized people can locate and access the data quickly.
- Capacity — Secondary storage can store large volumes of data in sets of multiple disks.
- Cost — It is much lesser expensive to store data on a tape or disk than primary memory.

## Magnetic Disks

A magnetic disk is a type of memory constructed using a circular plate of metal or plastic coated with magnetized materials.
Usually, both sides of the disks are used to carry out read/write operations.
However, several disks may be stacked on one spindle with read/write head available on each surface.

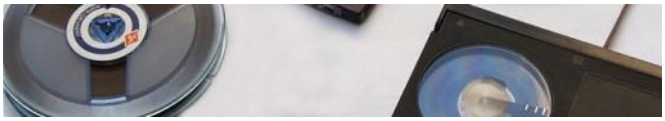
Usually, both sides of the disks are used to carry out read/write operations.
However, several disks may be stacked on one spindle with read/write head available on each surface.
The following image shows the structural representation for a magnetic disk.




- The memory bits are stored in the magnetized surface in spots along the concentric circles called tracks.
- The concentric circles (tracks) are commonly divided into sections called sectors.

## Magnetic Tape

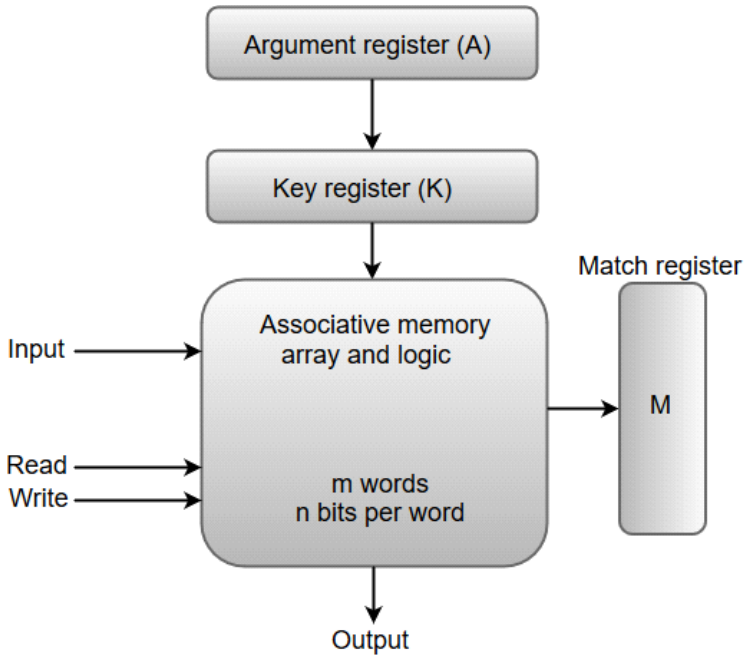
Magnetic tape is a storage medium that allows data archiving, collection, and backup for different kinds of data.
The magnetic tape is constructed using a plastic strip coated with a magnetic recording medium.
The bits are recorded as magnetic spots on the tape along several tracks.
Usually, seven or nine bits are recorded simultaneously to form a character together with a parity bit.
Magnetic tape units can be halted, started to move forward or in reverse, or can be rewound.
However, they cannot be started or stopped fast enough between individual characters.
For this reason, information is recorded in blocks referred to as records



## Associative Memory

An associative memory can be considered as a memory unit
whose stored data can be identified for access by the content of the data itself rather than by an address or memory location.
Associative memory is often referred to as **Content Addressable Memory (CAM)**.

When a write operation is performed on associative memory, no address or memory location is given to the word.
The memory itself is capable of finding an empty unused location to store the word.
On the other hand, when the word is to be read from an associative memory, the content of the word, or part of the word, is specified.
The words which match the specified content are located by the memory and are marked for reading.
The following diagram shows the block representation of an Associative memory.

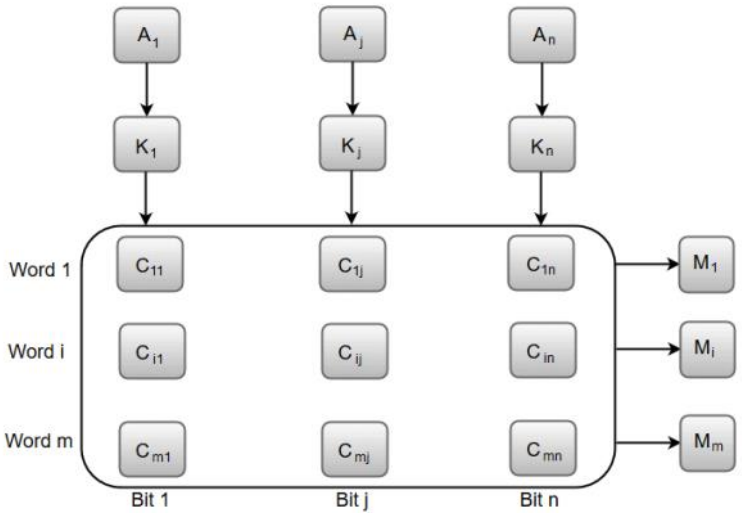



From the block diagram, we can say that an associative memory consists of
a memory array and logic for 'm' words with 'n' bits per word.

The functional registers like the argument register **A** and key register **K** each have **n** bits,
one for each bit of a word. The match register **M** consists of **m** bits, one for each memory word.
The words which are kept in the memory are compared in parallel with the content of the argument register.
The key register (K) provides a mask for choosing a particular field or key in the argument word.
If the key register contains a binary value of all 1's, then the entire argument is compared
with each memory word.
Otherwise, only those bits in the argument that have 1's in their corresponding position of the key

If the key register contains a binary value of all 1's, then the entire argument is compared with each memory word.
Otherwise, only those bits in the argument that have 1's in their corresponding position of the key register are compared.
Thus, the key provides a mask for identifying a piece of information which specifies how the reference to memory is made.
The following diagram can represent the relation between the memory array and the external registers in an associative memory.

**Associative memory of m word, n cells per word:**




The cells present inside the memory array are marked by the letter C with two subscripts.
The first subscript gives the word number and the second specifies the bit position in the word.
For instance, the cell Cij is the cell for bit **j** in word **i**.

A bit $A_j$ in the argument register is compared with all the bits in
column **j** of the array provided that Kj = 1.
This process is done for all columns **j** = 1, 2, 3....., n.
If a match occurs between all the unmasked bits of the argument and the bits in word **i**, the corresponding bit $M_i$ in the match register is set to 1.
If one or more unmasked bits of the argument and the word do not match, Mi is cleared to 0.