

Unit - IV

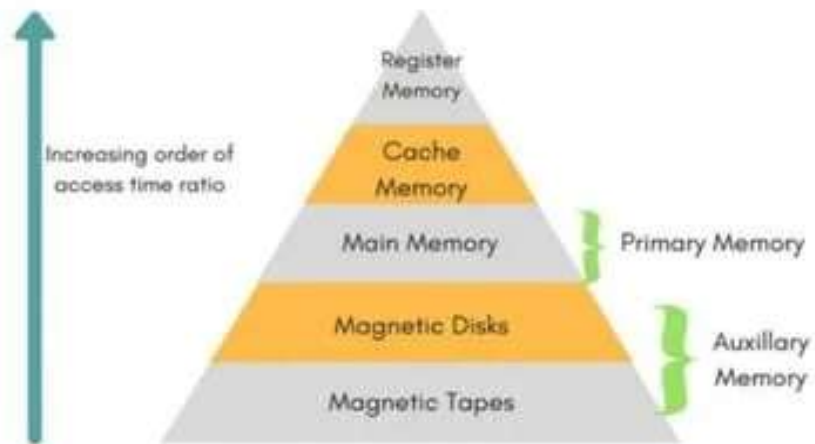
Memory Organization

Memory Organization in Computer Architecture

A memory unit is the collection of storage units or devices together. The memory unit stores the binary information in the form of bits. Generally, memory/storage is classified into 2 categories:

- **Volatile Memory:** This loses its data, when power is switched off.
- **Non-Volatile Memory:** This is a permanent storage and does not lose any data when power is switched off.

Memory Hierarchy



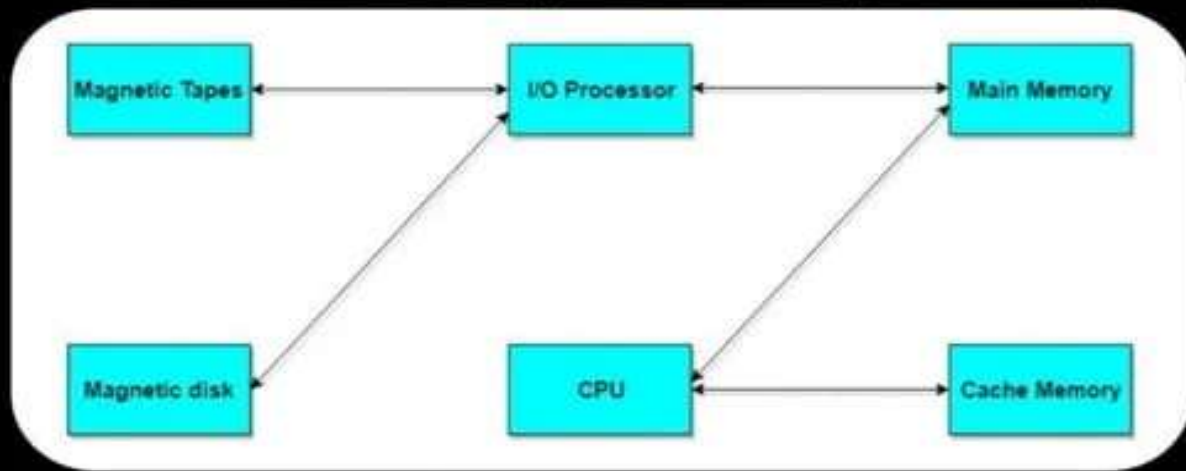
Memory Hierarchy

The total memory capacity of a computer can be visualized by hierarchy of components. The memory hierarchy system consists of all storage devices contained in a computer system from the slow Auxiliary Memory to fast Main Memory and to smaller Cache memory.

- **Auxiliary memory** access time is generally 1000 times that of the main memory, hence it is at the bottom of the hierarchy.
- The **main memory** occupies the central position because it is equipped to communicate directly with the CPU and with auxiliary memory devices through Input/output processor (I/O).
- When the program not residing in main memory is needed by the CPU, they are brought in from auxiliary memory. Programs not currently needed in main memory are transferred into auxiliary memory to provide space in main memory for other programs that are currently in use.

Memory Hierarchy

- The cache memory is used to store program data which is currently being executed in the CPU. Approximate access time ratio between cache memory and main memory is about 1 to 7~10



Memory Access Methods

- Each memory type, is a collection of numerous memory locations. To access data from any memory, first it must be located and then the data is read from the memory location. Following are the methods to access information from memory locations:
1. **Random Access:** Main memories are random access memories, in which each memory location has a unique address. Using this unique address any memory location can be reached in the same amount of time in any order.
 2. **Sequential Access:** This methods allows memory access in a sequence or in order.
 3. **Direct Access:** In this mode, information is stored in tracks, with each track having a separate read/write head.

Main Memory

- The memory unit that communicates directly within the CPU, Auxiliary memory and Cache memory, is called main memory. It is the central storage unit of the computer system. It is a large and fast memory used to store data during computer operations. Main memory is made up of RAM and ROM, with RAM integrated circuit chips holding the major share.

RAM: Random Access Memory

- **DRAM:** Dynamic RAM, is made of capacitors and transistors, and must be refreshed every 10~100 ms. It is slower and cheaper than SRAM.
- **SRAM:** Static RAM, has a six transistor circuit in each cell and retains data, until powered off.
- **NVRAM:** Non-Volatile RAM, retains its data, even when turned off. Example: Flash memory.

Main Memory

ROM: Read Only Memory:

- ROM: Read Only Memory, is non-volatile and is more like a permanent storage for information. It also stores the bootstrap loader program, to load and start the operating system when computer is turned on. PROM(Programmable ROM), EPROM(Erasable PROM) and EEPROM(Electrically Erasable PROM) are some commonly used ROMs.

Auxiliary Memory

- Devices that provide backup storage are called auxiliary memory. For example: Magnetic disks and tapes are commonly used auxiliary devices. Other devices used as auxiliary memory are magnetic drums, magnetic bubble memory and optical disks.
- It is not directly accessible to the CPU, and is accessed using the Input / Output channels.

Cache Memory

- The data or contents of the main memory that are used again and again by CPU, are stored in the cache memory so that we can easily access that data in shorter time.
- Whenever the CPU needs to access memory, it first checks the cache memory. If the data is not found in cache memory then the CPU moves onto the main memory. It also transfers block of recent data into the cache and keeps on deleting the old data in cache to accommodate the new one.

Hit Ratio

- The performance of cache memory is measured in terms of a quantity called hit ratio. When the CPU refers to memory and finds the word in cache it is said to produce a hit. If the word is not found in cache, it is in main memory then it counts as a miss.
- The ratio of the number of hits to the total CPU references to memory is called hit ratio.

$$\text{Hit Ratio} = \text{Hit} / (\text{Hit} + \text{Miss})$$

Associative Memory

- It is also known as content addressable memory (CAM). It is a memory chip in which each bit position can be compared. In this the content is compared in each bit cell which allows very fast table lookup. Since the entire chip can be compared, contents are randomly stored without considering addressing scheme. These chips have less storage capacity than regular memory chips.