Computer Science Dept. / 2nd year 2nd Course (2024-2025)
Computer Architecture
Dr.Haleema Essa

Computer Memory System

The memory of a computer system can be divided into three main groups:

1- Internal Processor Memory

This comprises a small set of high-speed registers used as a working memory for temporary storage of instructions and data.

2- Main Memory

This is relatively large fast memory used for program and data storage during computer operation. It is characterized by the fact that locations in main memory can be accessed directly and rapidly by the CPU instruction set.

3- Secondary Memory

This is generally much larger in capacity but also much slower than the main memory. It is used for storing system program and large data files. This type of memory has the following groups:

a- Magnetic Tape.

b- Floppy Disk

c- Hard Disk.

d- CD-ROM.

e- Flash.

➢ Memory Device Characteristics:

✓ **Cost:** Let (**C**) be the price in dollars of a complete memory system with (**S**) bits of storage capacity (Size).

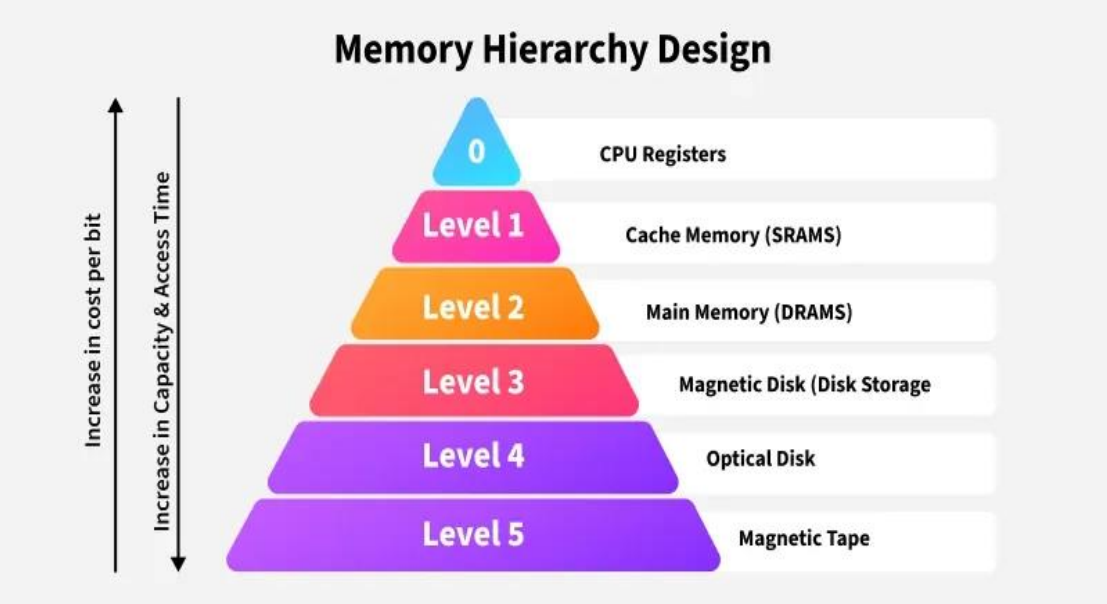
$$K = \frac{C}{S} \quad \text{dollars/bit}$$

Where **k**: The cost per bit of storage (in *dollars per bit*).

✓ **Access Time**

- The performance of a memory device ***is determined by*** the rate at which information can be read from or written into the memory.
- Access time ***depends on*** the physical characteristics of the storage medium, and also on the type of access mechanism used (i.e., direct or sequential).
- Access time usually ***calculated*** from the time a read-request is received by the memory unit to the time at which all requested information has been made available at the memory output terminals.

❖ Memory Hierarchy Design



1. Registers

Registers are small, high-speed memory units located in the CPU. They are used to store the most frequently used data and instructions. Registers have the fastest access time and the smallest storage capacity, typically ranging from 16 to 64 bits.

2. Cache Memory

Cache memory is a small, fast memory unit located close to the CPU. It stores frequently used data and instructions that have been recently accessed from the main memory. Cache memory is designed to *minimize the time it takes to access data by providing the CPU with quick access to frequently used data*.

3. Main Memory

Main memory, also known as RAM (Random Access Memory), is the primary memory of a computer system. It has a larger storage capacity than cache memory, but it is slower. Main memory is used to store data and instructions that are currently in use by the CPU.

Types of Main Memory

- **Static RAM:** Static RAM stores the binary information in flip flops and information remains valid until power is supplied. Static RAM has a faster access time and is used in implementing cache memory.
- **Dynamic RAM:** It stores the binary information as a charge on the capacitor. It requires refreshing circuitry to maintain the charge on the capacitors after a few milliseconds. It contains more memory cells per unit area as compared to SRAM.

4. Magnetic disks

Magnetic disks, such as hard disk drives (HDD) and solid-state drives (SSD), is a non-volatile memory unit that has a larger storage capacity than main memory. It is used to store data and instructions that are not currently in use by the CPU. It has the slowest access time and is typically the least expensive than main memory in the memory hierarchy.

5. Optical Disk

Optical Disk is a storage medium that relies on laser technology to read and write data, in shape, it is a flat circular disk which is made up of polycarbonate with a very shiny reflective layer on the surface. they are mainly used for sharing, storing and backup Data as they have a great life span and capacity compared to older technologies like floppy disks.

6. Magnetic Tape

Magnetic Tape is simply a magnetic recording device that is covered with a plastic film. Magnetic Tape is generally used for the backup of data. In the case of a magnetic tape, the access time for a computer is a little slower and therefore, it requires some amount of time for accessing the strip.

➢ Characteristics of Memory Hierarchy

1. **Capacity:** It is the global volume of information the memory can store. As we move from top to bottom in the Hierarchy, the capacity increases.

2. **Access Time:** It is the time interval between the read/write request and the availability of the data. As we move from top to bottom in the Hierarchy, the access time increases.
3. **Performance:** The Memory Hierarch design ensures that frequently accessed data is stored in faster memory to improve system performance.
4. **Cost per Bit:** As we move from bottom to top in the Hierarchy, the cost per bit increases.

➢ Advantages of Memory Hierarchy

1. **Performance:** Frequently used data is stored in faster memory (like cache), reducing access time and improving overall system performance.
2. **Cost Efficiency:** By combining small, fast memory (like registers and cache) with larger, slower memory (like RAM and HDD), the system achieves a balance between cost and performance. It saves the consumer's price and time.
3. **Optimized Resource Utilization:** Combines the benefits of small, fast memory and large, cost-effective storage to optimize the system resources utilization.
4. **Efficient Data Management:** Frequently accessed data is kept closer to the CPU, while less frequently used data is stored in larger, slower memory, ensuring efficient data handling.

➢ Disadvantages of Memory Hierarchy

1. **Complex Design:** Managing and coordinating data across different levels of the hierarchy adds complexity to the system's design and operation.
2. **Cost:** Faster memory components like registers and cache are expensive, limiting their size and increasing the overall cost of the system.
3. **Latency:** Accessing data stored in slower memory (like secondary or tertiary storage) increases the latency and reduces system performance.
4. **Maintenance Overhead:** Managing and maintaining different types of memory adds overhead in terms of hardware and software.