Covered Topics Under UNIT-1 of "PPS-PROGRAMMING FOR PROBLEM SOLVING (BCS101 / BCS201)"

shwetatiwari08@recabn.ac.in
shwetatiwari08aug@gmail.com

Class Notes

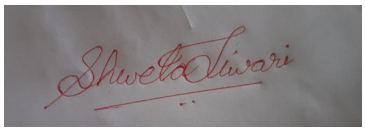
Published Date: November, 2022

PPS: UNIT-1

Introduction to Components of a Computer System

FALL SEMESTER, YEAR (I/II sem, 1st yr)

FALL SESSION (2022-23)
(PPS)
MS. SHWETA TIWARI
Published Date: November, 2022



shwetatiwari08@recabn.ac.in
shwetatiwari08aug@gmail.com

TOPIC On : UNIT-1: COMPUTER MEMORY

—

By SHWETA TIWARI

Under On: Introduction to Components of a Computer System

PREPARED FOR
Engineering Students
All Engineering College

PREPARED BY
SHWETA TIWARI

TOPIC On : UNIT-1: COMPUTER MEMORY

COMPUTER MEMORY

I. INTRODUCTION

The computer's memory stores data, instructions required during the processing of data, and output results. Storage may be required for a limited period of time, instantly, or, for an extended period of time. Different types of memories, each having its own unique features, are available for use in a computer. *The cache memory, registers, and RAM* are fast memories and store the data and instructions temporarily during the processing of data and instructions. *The secondary memory like magnetic disks and optical disks* has large storage capacities and store the data and instructions permanently, but are slow memory devices.

The memories are organized in the computer in a manner to achieve high levels of performance at the minimum cost. In this lecture, we discuss different types of memories, their characteristics and their use in the computer.

II. MEMORY REPRESENTATION

The computer memory stores different kinds of data like input data, output data, intermediate results, etc., and the instructions. ***Binary digit*** or ***bit*** is the basic unit of memory. A ***bit*** is a single binary digit, i.e., 0 or 1. A bit is the smallest unit of representation of data in a computer. However, the data is handled by the computer as a combination of bits. A group of 8 bits form a **byte**.

One byte is the smallest unit of data that is handled by the computer.

One byte (8 bit) can store $\mathbf{2^8} = 256$ different combinations of bits, and thus can be used to represent 256 different symbols. In a byte, the different combinations of bits fall in the range 00000000 to 11111111. A group of bytes can be further combined to form a **word**. A word can be a group of 2, 4 or 8 bytes.

1 bit = 0 or 1
1 Byte (B) = 8 bits
1 Kilobyte (KB) = 2^{10} = 1024 bytes
1 Megabyte (MB) = 2^{20} = 1024KB
1 Gigabyte (GB) = 2^{30} = 1024 MB = 1024 *1024 KB
1 Terabyte (TB) = 2^{40}= 1024 GB = 1024 * 1024 *1024 KB

III. CHARACTERISTICS OF MEMORIES

- **Volatility**
  - **Volatile {RAM}**
  - **Non-volatile {ROM, Flash memory}**
- **Mutability**
  - **Read/Write {RAM, HDD, SSD, RAM, Cache, Registers…}**
  - **Read Only {Optical ROM (CD/DVD…), Semiconductor ROM}**
- **Accessibility**
  - **Random Access {RAM, Cache}**
  - **Direct Access {HDD, Optical Disks}**
  - **Sequential Access {Magnetic Tapes}**

IV. MEMORY HIERARCHY

The memory is characterized on the basis of two key factors: ***capacity*** and ***access time***.

- ***Capacity*** is the amount of information (in bits) that a memory can store.
- ***Access time*** is the time interval between the read/ write request and the availability of data. The lesser the access time, the faster is the *speed of memory*.

Ideally, we want the memory with *fastest speed and largest capacity*. However, the cost of fast memory is very high. The computer uses a hierarchy of memory that is organized in a manner to enable the fastest speed and largest capacity of memory.

The hierarchy of the different memory types is shown in Figure 4.2.

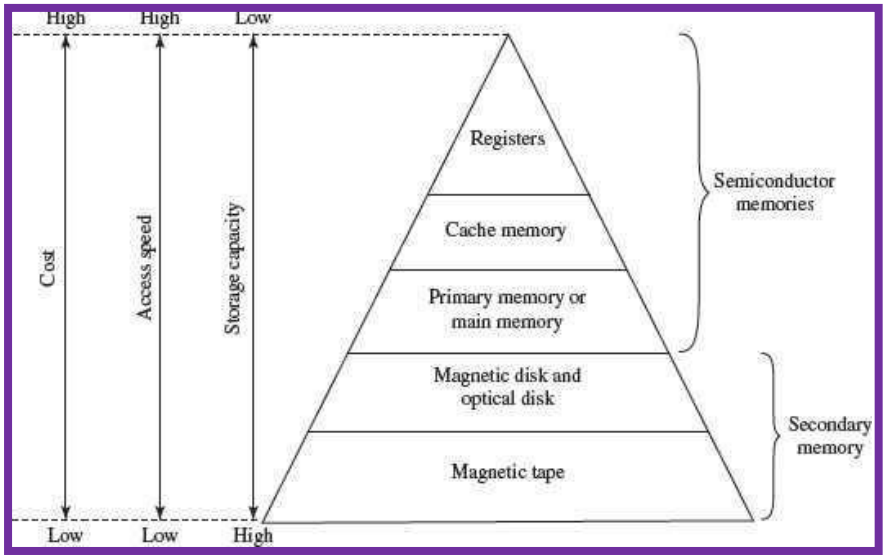


Figure 4.2 Memory hierarchy

The Internal Memory and ***External Memory*** are the two broad categories of memory used in the computer. *The Internal Memory* consists of the CPU registers, cache memory and primary memory. The internal memory is used by the CPU to perform the computing tasks. *The External Memory* is also called the *secondary memory*. The secondary memory is used to store the large amount of data and the software.

In general, referring to the computer memory usually means the internal memory.

- **Internal Memory**

The key features of internal memory are:

1. Limited storage capacity.
2. Temporary storage.
3. Fast access.
4. High cost.

Registers, *cache memory*, and *primary memory* constitute the internal memory. *The primary memory* is further of two kinds: RAM and ROM. *Registers* are the fastest and the most expensive among all the memory types. The registers are located inside the CPU, and are directly accessible by the CPU. The speed of registers is between 1-2 ns (nanosecond). The sum of the size of registers is about 200B. *Cache memory* is next in the hierarchy and is placed between the CPU and the main memory. The speed of cache is between 2-10 ns. The cache size varies between 32 KB to 4MB. Any program or data that has to be executed must be brought into RAM from the secondaryss memory. Primary memory is relatively slower than the cache memory. The speed of RAM is around 60ns. The RAM size varies from 512KB to 64GB.

- **Secondary Memory**

The key features of secondary memory storage devices are:

1. Very high storage capacity.
2. Permanent storage (non-volatile), unless erased by user.
3. Relatively slower access.
4. Stores data and instructions that are not currently being used by CPU but may be required later for processing.
5. Cheapest among all memory.

To get the fastest speed of memory with largest capacity and least cost, the fast memory is located close to the processor. The secondary memory, which is not as fast, is used to store information permanently, and is placed farthest from the processor.

With respect to CPU, the memory is organized as follows:

- *Registers* are placed inside the CPU (small capacity, high cost, very high speed)
- *Cache memory* is placed next in the hierarchy (inside and outside the CPU)
- *Primary memory* is placed next in the hierarchy
- *Secondary memory* is the farthest from CPU (large capacity, low cost, low speed)

The speed of memories is dependent on the kind of technology used for the memory. The registers, cache memory and primary memory are **semiconductor memories**. They do not have any moving parts and are fast memories. The secondary memory is **magnetic or optical memory** has moving parts and has slow speed.

V. CPU REGISTERS

Registers are very high-speed storage areas located inside the CPU. After CPU gets the data and instructions from the cache or RAM, the data and instructions are moved to the registers for processing. Registers are manipulated directly by the control unit of CPU during instruction execution. That is why registers are often referred to as the CPU's *working memory*. Since CPU uses registers for the processing of data, the number of registers in a CPU and the size of each register affect the power and speed of a CPU. The more the number of registers (ten to hundreds) and bigger the size of each register (8 bits to 64 bits), the better it is.

VI. CACHE MEMORY

Cache memory is placed in between the CPU and the RAM. Cache memory is a fast memory, faster than the RAM. When the CPU needs an instruction or data during processing, it first looks in the cache. If the information is present in the cache, it is called a ***cache hit***, and the data or instruction is retrieved from the cache. If the information is not present in cache, then it is called a ***cache miss*** and the information is then retrieved from RAM.

Type of Cache memory

Cache memory improves the speed of the CPU, but it is expensive. Type of Cache Memory is divided into different levels that are L1, L2, L3:

Level 1 (L1) cache or Primary Cache

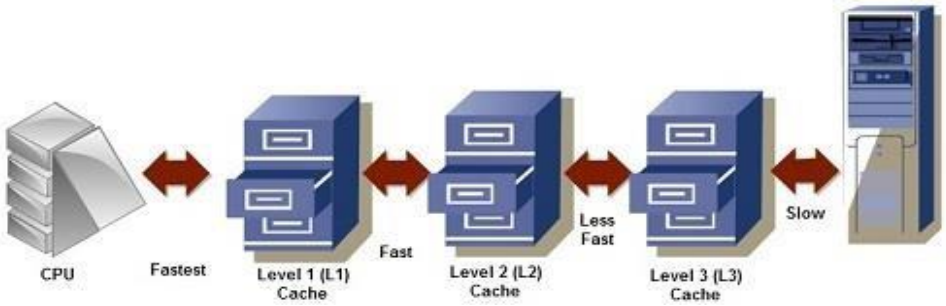
L1 is the primary type cache memory. The Size of the L1 cache very small comparison to others that is between 2KB to 64KB, it depends on computer processor. It is an embedded register in the computer microprocessor (CPU).The Instructions that are required by the CPU that are firstly searched in L1 Cache. Example of registers are accumulator, address register, Program counter etc.

Level 2 (L2) cache or Secondary Cache

L2 is secondary type cache memory. The Size of the L2 cache is more capacious than L1 that is between 256KB to 512KB. L2 cache is located on computer microprocessor. After searching the Instructions in L1 Cache, if not found then it searched into L2 cache by computer microprocessor. The high-speed system bus interconnecting the cache to the microprocessor.

Level 3 (L3) cache or Main Memory

The L3 cache is larger in size but also slower in speed than L1 and L2, its size is between 1MB to 8MB. In Multicore processors, each core may have separate L1 and L2, but all core share a common L3 cache. L3 cache double speed than the RAM.



The advantages and disadvantages of cache memory are as follows:

Advantages

The advantages of cache memory are as follows:

- Cache memory is faster than main memory.
- It consumes less access time as compared to main memory.
- It stores the program that can be executed within a short period of time.
- It stores data for temporary use.

Disadvantages

The disadvantages of cache memory are as follows:

- Cache memory has limited capacity.
- It is very expensive.

VII. PRIMARY MEMORY (Main Memory)

Primary memory is the main memory of computer. It is a chip mounted on the motherboard of computer. Primary memory is categorized into two main types: Random access memory (ram) and read only memory (rom). **RAM** is used for the temporary storage of input data, output data and intermediate results. The input data entered into the computer using the input device, is stored in RAM for processing. After processing, the output data is stored in RAM before being sent to the output device. Any intermediate results generated during the processing of program are also stored in RAM. Unlike RAM, the data once stored in **ROM** either cannot be changed or can only be changed using some special operations. Therefore, ROM is used to store the data that does not require a change.

Types of Primary Memory

1. RAM (Random Access Memory)

The Word "**RAM**" stands for "random access memory" or may also refer to short-term memory. It's called "random" because you can read store data randomly at any time and from any physical location. It is a temporal storage memory. RAM is volatile that only retains all the data as long as the computer powered. It is the fastest type of memory. RAM stores the currently processed data from the CPU and sends them to the graphics unit.

There are generally two broad subcategories of RAM:

- **Static RAM** (**SRAM**): Static RAM is the form of RAM and made with flip-flops and used for primary storage are volatile. It retains data in latch as long as the computer powered. SRAM is more expensive and consumes more power than DRAM. It used as Cache Memory in a computer system. As technically, SRAM uses more transistors as compared to DRAM. It is faster compared to DRAM due to the latching arrangement, and they use 6 transistors per data bit as compared to DRAM, which uses one transistor per bit.
- **Dynamic Random Access Memory (DRAM)**: It is another form of RAM used as Main Memory, its retains information in Capacitors for a short period (a few milliseconds) even though the computer powered. The Data is Refreshed Periodically to maintain in it. The DRAM is cheaper, but it can store much more information. Moreover, it is also slower and consumes less power than SRAM.

2. ROM (Read Only Memory)

ROM is the long-term internal memory. ROM is "Non-Volatile Memory" that retains data without the flow of electricity. ROM is an essential chip with permanently written data or programs. It is similar to the RAM that is accessed by the CPU. ROM comes with pre-written by the computer manufacturer to hold the instructions for booting-up the computer.

There is generally three broad type of ROM:

- **PROM (Programmable Read Only Memory)**: PROM stands for programmable ROM. It can be programmed only be done once and read many. Unlike RAM, PROMs retain their contents without the flow of electricity. PROM is also nonvolatile memory. The significant difference between a ROM and a PROM is that a ROM comes with pre-written by the computer manufacturer whereas PROM manufactured as blank memory. PROM can be programmed by PROM burner and by blowing internal fuses permanently.
- **EPROM (Erasable Programmable Read Only Memory)**: EPROM is pronounced ee-prom. This memory type retains its contents until it exposed to intense ultraviolet light that clears its contents, making it possible to reprogram the memory.
- **EEPROM (Electrically Erasable Programmable Read Only Memory)**: EEPROM can be burned (programmed) and erased by first electrical waves in a millisecond. A single byte of a data or the entire contents of device can be erased. To write or erase this memory type, you need a device called a PROM burner.