Computer Organization and Architecture

Memory Devices

Introduction

lemory Hierarchy,

andom Access Memory,

ead Only Memory,

erial Access Memory,

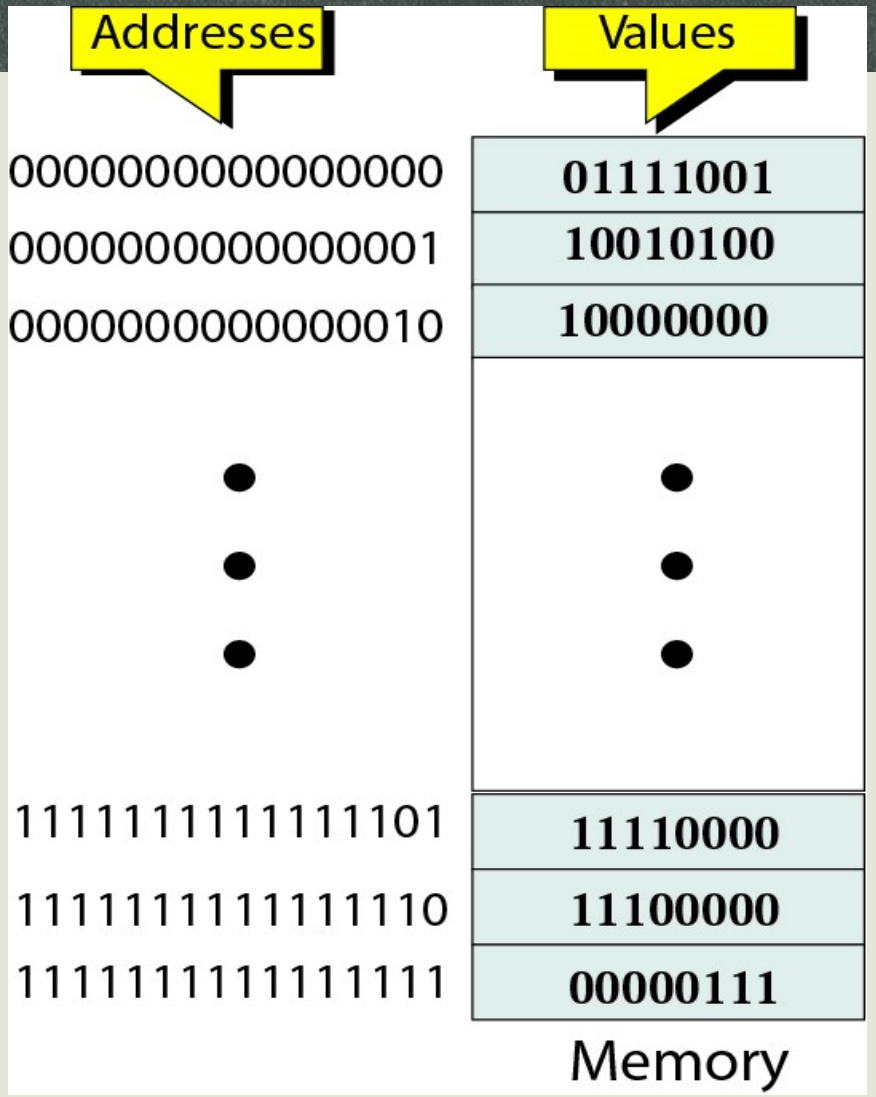
irect Access Memory,

ache Memory,

verview of Virtual Memory and

uxiliary Memory.

Main Memory


Addresses
Values
0000000000000000
01111001
0000000000000001
10010100
0000000000000010
10000000
1111111111111101
11110000
1111111111111110
11100000
1111111111111111
00000111
Memory

Introduction

memory is used to store data and instruction. Computer memory is
torage space in computer where data is to be processed and instruct
equired for processing are stored.

he memory is divided into large number of small parts.

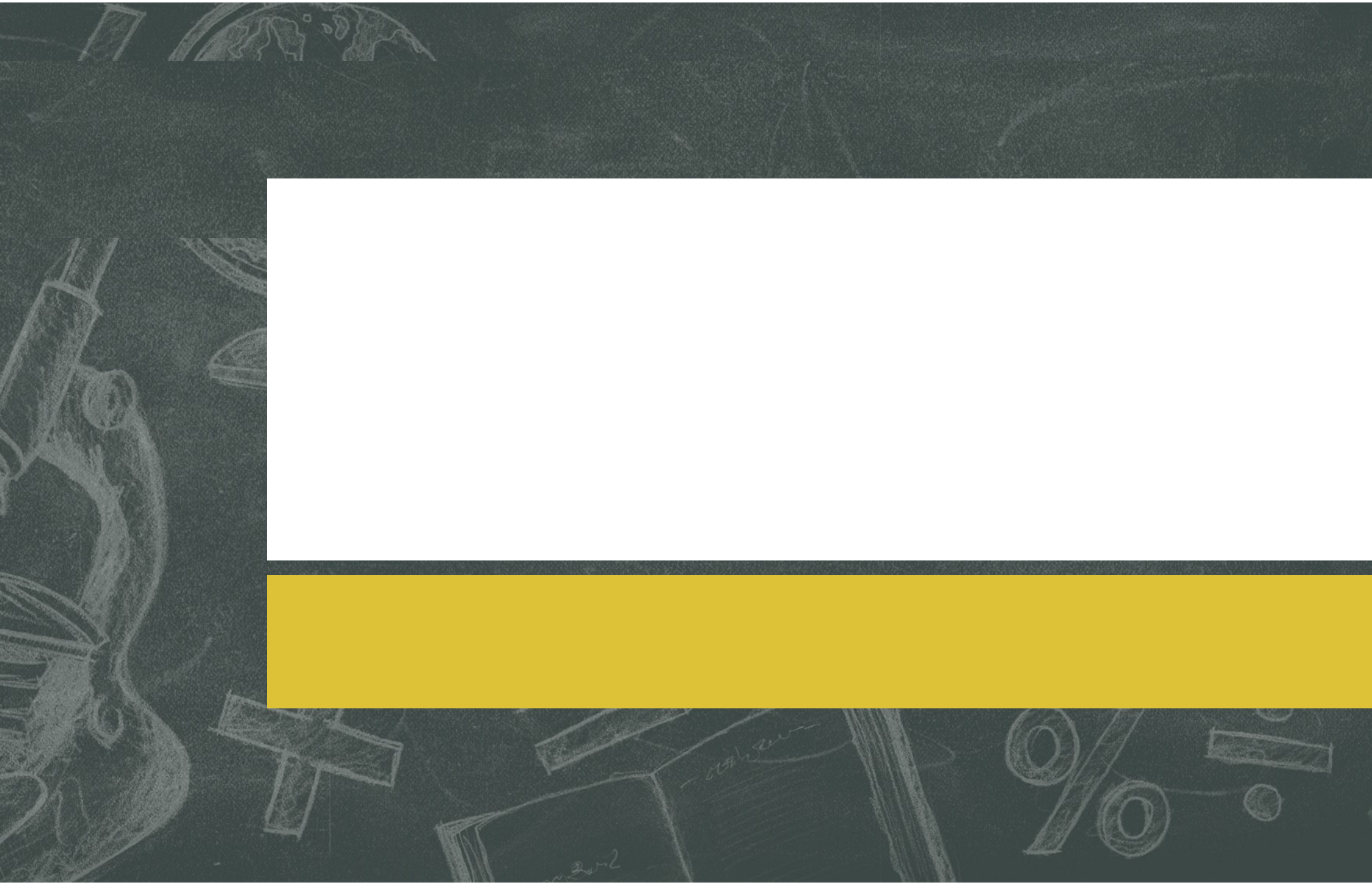
ach part is called a cell. Each location or cell has a unique address w
aries from zero to memory size minus one.

or example if computer has 64k words, then this memory unit has
024 = 65536 memory location. The address of these locations varies fr
65535.

Storage Capacity

Name	Abbreviation	Number of Bytes
Byte	B	1
Kilobyte	KB	1,024 Bytes
Megabyte	MB	1,024 Kilobytes (about 1 million)
Gigabyte	GB	1,024 Megabytes (about 1 billion)
Terabyte	TB	1,024 Gigabytes (about 1 trillion)
Petabyte	PB	1,024 Terabytes (about 1 quadrillion)

Unit	*Exact Number of bytes*	*Approximation*
kilobyte	2^{10} bytes	10^{3} bytes
megabyte	2^{20} bytes	10^{6} bytes
gigabyte	2^{30} bytes	10^{9} bytes
terabyte	2^{40} bytes	10^{12} bytes
petabyte	2^{50} bytes	10^{15} bytes
exabyte	2^{60} bytes	10^{18} bytes



Memory Hierarchy

The memory system is a hierarchy of storage devices with different capacities, costs, and access times.

Memory is primarily of two types

Internal Memory – cache memory and primary/main memory

External Memory – magnetic disk / optical disk, etc.

Introduction

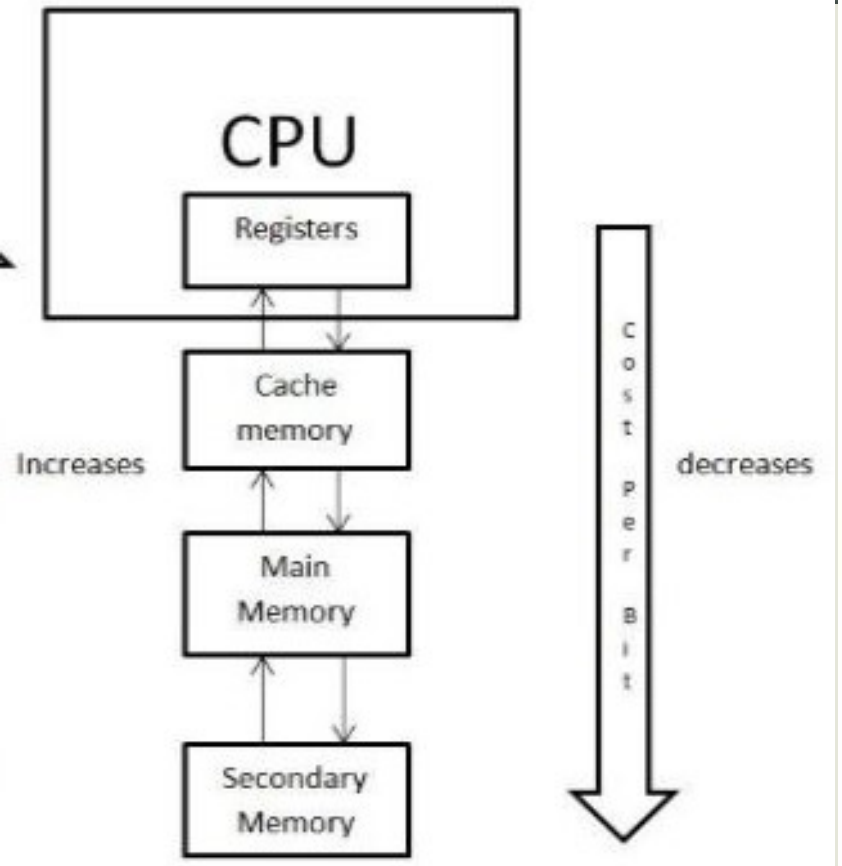
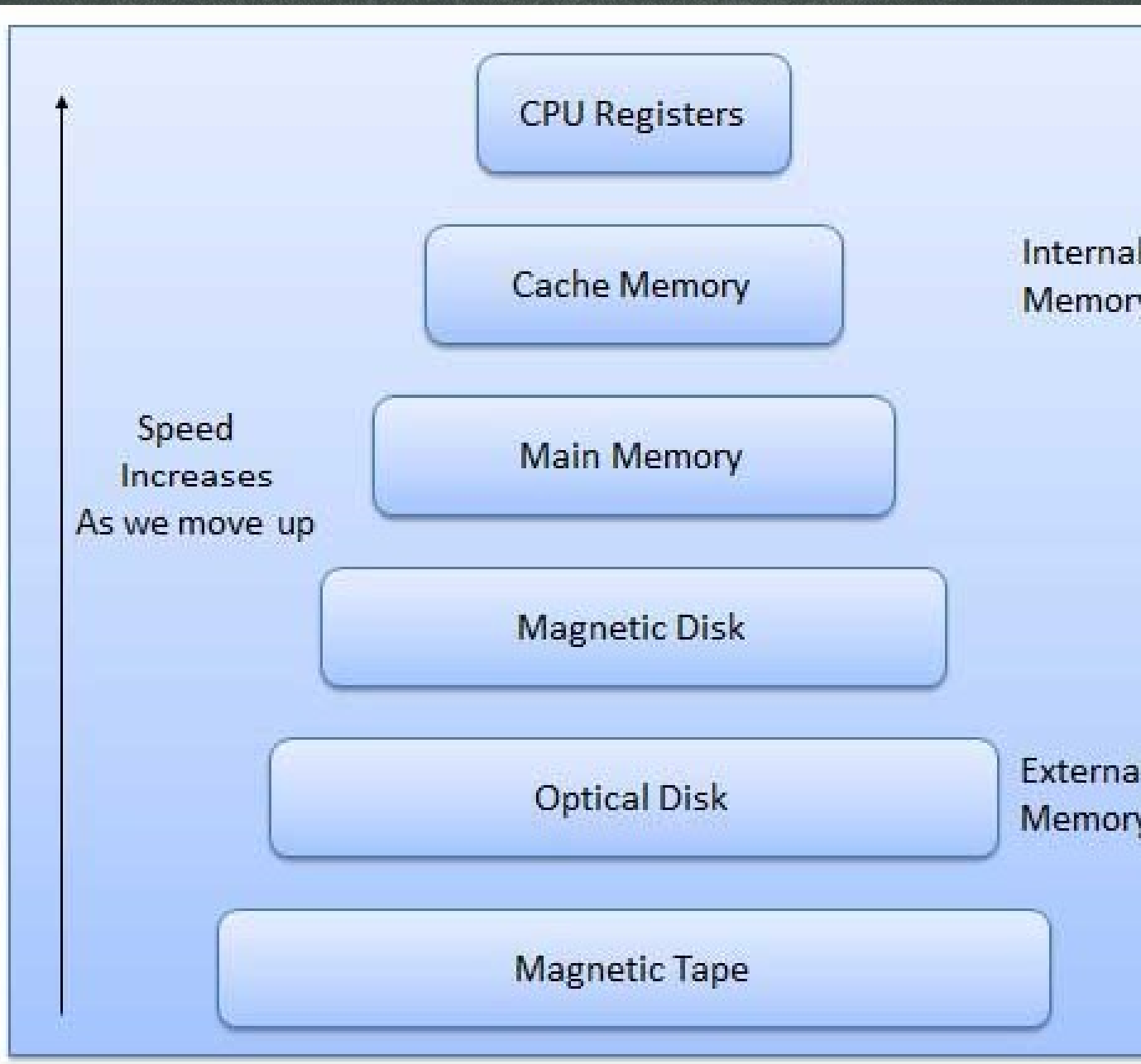


Figure 1



Introduction

haracteristics of Memory Hierarchy are following when we
rom top to bottom.

apacity in terms of storage increases.

ost per bit of storage decreases.

requency of access of the memory by the CPU decreases.

ccess time by the CPU increases.

RAM

RAM constitutes the internal memory of the CPU for storing ata, program and program result. It is read/write memory. It i alled random access memory (RAM).

ince access time in RAM is independent of the address to the ord that is, each storage location inside the memory is as easy each as other location & takes the same amount of time.

Ve can reach into the memory at random & extremely fast but an also be quite expensive.

RAM

AM is volatile, i.e. data stored in it is lost when we switch off
omputer or if there is a power failure.

ence, a backup uninterruptible power system (UPS) is often u
ith computers. RAM is small, both in terms of its physical
nd in the amount of data it can hold.

AM is of two types

- Static RAM (SRAM)
- Dynamic RAM (DRAM)

RAM

Static RAM (SRAM)

The word **static** indicates that the memory retains its contents as long a power remains applied. However, data is lost when the power gets dow due to volatile nature.

SRAM chips use a matrix of 6-transistors and no capacitors. Transisto do not require power to prevent leakage, so SRAM need not have to refreshed on a regular basis.

Because of the extra space in the matrix, SRAM uses more chips tha DRAM for the same amount of storage space, thus making th manufacturing costs higher.

Static RAM is used as cache memory needs to be very fast and small.

RAM

Dynamic RAM (DRAM)

DRAM, unlike SRAM, must be continually **refreshed** in order for it maintain the data.

This is done by placing the memory on a refresh circuit that rewrites t data several hundred times per second.

DRAM is used for most system memory because it is cheap and small. DRAMs are made up of memory cells. These cells are composed of o capacitor and one transistor.

RAM

Other RAM types

SDRAM (Synchronous DRAM)

EDRAM (Enhanced DRAM)

EDO (Extended Data Out)

FLASH RAM

Ferroelectric RAM

ROM

ROM stands for Read Only Memory. The memory from which we can only read but cannot write on it. This type of memory is non-volatile. The information is stored permanently in such memories during manufacture.

A ROM, stores such instruction as are required to start computer when electricity is first turned on, this operation is referred to as bootstrap.

ROM

MROM (Masked ROM)

The very first ROMs were hard-wired devices that contained pre-programmed set of data or instructions. These kind of ROMs are known as masked ROMs. It is inexpensive ROM.

PROM (Programmable Read Only Memory)

PROM is read-only memory that can be modified only once by user. The user buys a blank PROM and enters the desired contents using a PROM programmer.

It can be programmed only once and is not erasable.

ROM

EPROM (Erasable and Programmable Read Only Memory)

The EPROM can be erased by exposing it to ultra-violet light f
a duration of upto 40 minutes.

**EEPROM (Electrically Erasable and Programmable Rea
Only Memory)**

The EEPROM is programmed and erased electrically. It can
erased and reprogrammed about ten thousand times.

Both erasing and programming take about 4 to 10 m
(millisecond). In EEPROM, any location can be selective
erased and programmed.

Serial Access Memory

Sequential access means the system must search the storage device fr
the beginning of the memory address until it finds the required piece
data.

Memory device which supports such access is called a Sequential Acce
Memory or Serial Access Memory. Magnetic tape is an example of ser
access memory.

Direct Access Memory

Direct access memory or Random Access Memory, refers t
conditions in which a system can go directly to the informatio
that the user wants.

Memory device which supports such access is called a Direc
Access Memory. Magnetic disks, optical disks are examples o
direct access memory.

Cache Memory

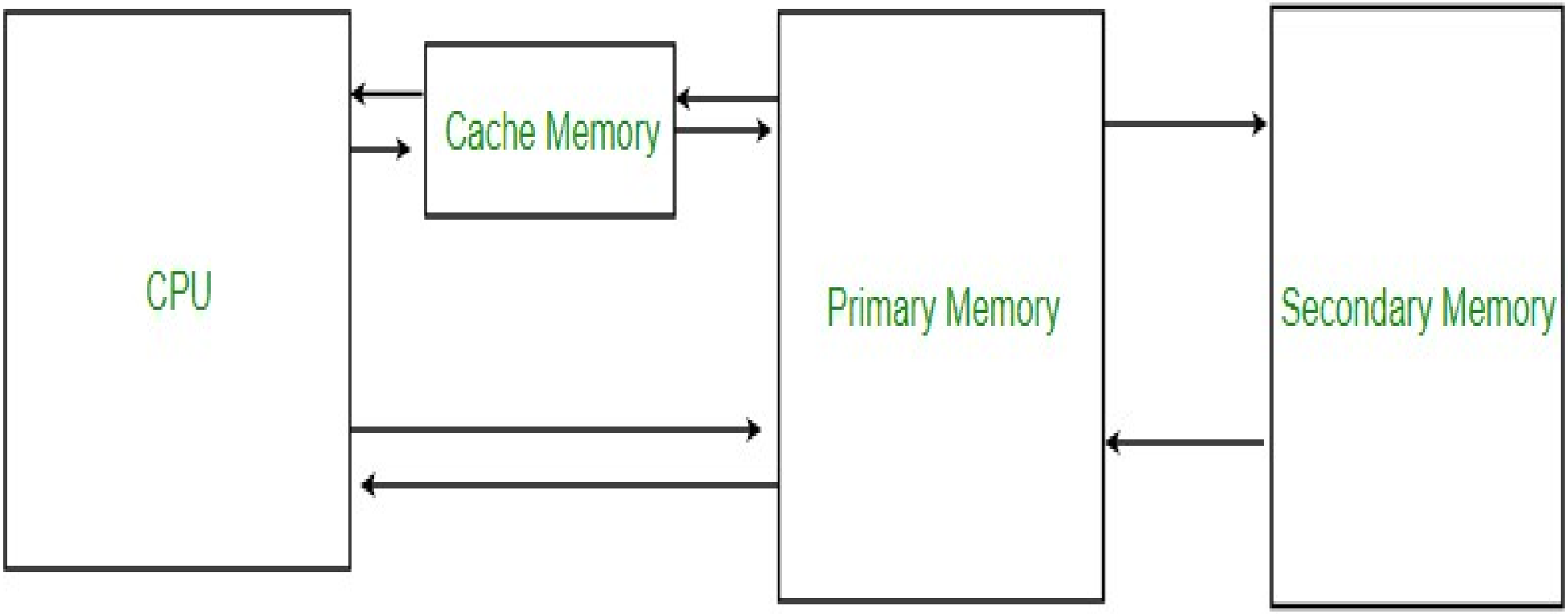
s CPU has to fetch instruction from main memory speed of CPU epending on fetching speed from main memory. CPU contains register hich has fastest access but they are limited in number as well as costly.

ache is cheaper so we can access cache. Cache memory is a very high peed memory that is placed between the CPU and main memory, to perate at the speed of the CPU.

is used to reduce the average time to access data from the main emory. The cache is a smaller and faster memory which stores copies f the data from frequently used main memory locations.

ost CPUs have different independent caches, including instruction and ata.

Cache Memory



Cache Memory

ache Performance

hen the processor needs to read or write a location in m
emory, it first checks for a corresponding entry in the cache.

- If the processor finds that the memory location is in the cach
  cache hit has occurred and data is read from chache
- If the processor **does not** find the memory location in the ca
  a **cache miss** has occurred. For a cache miss, the ca
  allocates a new entry and copies in data from main mem
  then the request is fulfilled from the contents of the cache.

Cache Memory

he performance of cache memory is frequently measured
rms of a quantity called Hit ratio.

- Hit ratio = hit / (hit + miss) = no. of hits/total accesses
- Miss ratio = miss/(hit + miss) = no. of miss/ total accesses

Cache Mapping

The three different types of mapping used for the purpose cache memory are as follows:

- Direct mapping
- Associative mapping,
- Set-Associative mapping.

Locality of reference

Since size of cache memory is less as compared to main memory. So to ch
which part of main memory should be given priority and loaded in cache
decided based on locality of reference.

. **Spatial Locality of reference:** If the storage has been accessed then likelih
f accessing the storage nearby that, is high.

. **Temporal Locality of reference:** It tells us whether memory locations i
rogram are likely to be accessed again in the near future. A method has h
emporal locality if it is called repeatedly in a short period of time.

Virtual Memory

Virtual Memory is a storage scheme that provides user an illusion of having a very big main memory. This is done by treating a part of secondary memory as the main memory.

In this scheme, User can load the bigger size processes than the available main memory by having the illusion that the memory is available to load the process.

Instead of loading one big process in the main memory, the Operating System loads the different parts of more than one process in the main memory.

By doing this, the degree of multiprogramming will be increased and therefore, the CPU utilization will also be increased.

Virtual Memory

How Virtual Memory Works?

Whenever some pages need to be loaded in the main memory for th
execution and the memory is not available for those many pages, then i
that case, instead of stopping the pages from entering in the mai
memory, the OS searches for the RAM area that are least used in th
recent times or that are not referenced and copy that into the secondar
memory to make the space for the new pages in the main memory.

Auxiliary Memory

Auxiliary memory is much larger in size than main memory but slower. It normally stores system programs, instruction and data files. is also known as secondary memory.

Secondary memories cannot be accessed directly by a processor. Firs the data/information of auxiliary memory is transferred to the mai memory and then that information can be accessed by the CPU.

CDs, DVDs, HDDs are examples of Auxiliary memory.

Auxiliary Memory

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.