Computer System Hardware

By Kajol Ramtel

Introduction

- Computer hardware is the collective term to describe the physical components of the computer system.
- They are tangible in nature.
- Computer hardware comprises of three main components:-
  1. Central Processing Unit
  2. Memory Unit
  3. Input / output unit

1. Central Processing Unit

- CPU is the component that is responsible for interpreting and executing most of the commands from the computer hardware and software and also controls the operation of all other components such as memory and input and output devices.
- It is a logic machine. So it’s main function is to run the program by fetching instruction from the RAM, evaluating and executing them in sequence.

Central Processing Unit

- The functions of CPU are as follows:-
  - Reads instruction form RAM.
  - Communicate with all the peripherals using the system bus.
  - Controls the sequence of instructions.
  - Controls the flow of data form one component to other component.
  - Performs the computing task specified in program.

Central Processing Unit

- The CPU has three components responsible for different functions:-
  - Control Unit
  - Arithmetic logic unit
  - Register Array

Control Unit

- It provides the necessary timing and control signals to all the operations in the computer.
- It controls the flow of data between the cpu and memory and peripherals.
- It also controls the entire operation of the computer

Control Unit

- The main functions of control unit are:
  - It performs the data processing operations with the aid of program prepared by the users and sends control signals to various parts of the computer system.
  - It gives commands to transfer data from the input device to the memory to arithmetic logic unit
  - It also transfers the results from the ALU to the memory and then to the output devices.
  - It stores program in the memory.
  - It fetches the requires instruction from the main storage and decode each instruction and hence execute them in sequence.

Arithmetic Logic Unit

- This is the area of CPU where various computing functions and performed on data. The ALU performs arithmetic operations such as addition, subtraction, multiplication, and division and logical operation such as comparison (equal to, less than, greater than), AND, OR and Exclusive OR. The result of operation is stored in Accumulator or in some register. ALU operation like increment, decrement etc. The main functions of ALU are as follows:
  - It accepts operands form registers.
  - It performs arithmetic and logical operations.
  - It returns results to register or a memory.
- The logical operations of ALU give the computer the decision making ability.

Register

- Register are the high speed temporary storage locations in the CPU made from electronic devices such as transistors, flip-flops etc. So, registers can be thought as CPU's working memory. Registers are primarily used to store data temporarily during the executing of program and are accessible to the user through instructions. These are the part of Control unit and ALU rather than of memory. Hence, their contents can be handled much faster than the contents of memory. Although the number of register varies from computer to computer, there are some registers which are common to all computers. Registers that are essential for instruction execution are:-

Register

- Program Counter (PC):
  - Contains the address of the next instruction to be fetched.
- Instruction Register (IR):
  - Contains the instructions most recently fetched (last instruction fetched)
- Memory Address Register(MAR)
  - Contains the address of a location in memory for read and write operation.
- Memory Buffer Register (MAR):
  - Contains the last value read from the memory.
- Accumulator (ACC):
  - An accumulator is a general purpose register used for temporary results and results produced by arithmetic logic unit.

2. Memory Unit

- It is the storage unit that stores the data and instructions entered into the computer through input unit before the actual processing starts, the final result produce by the computer after the processing and also the intermediate results during processing.
- There are two types of memory
  a) Primary memory
  b) Secondary memory

a) Primary Memory

- Primary memory enables computer to store data and instruction temporarily.
- It is mainly used to hold data and instruction as well as the intermediate results of processing, which the computer system is currently working on.
- Primary memory is volatile, that is the content is lost when the power is turned off.
- It requires the constant power supply to maintain the bit value.
- RAM is an example of primary or main memory.

b) Secondary Memory

- It is the memory that supplements the primary memory.
- It is mainly used to store data permanently for future use and it is also used to transfer the data form one computer to another computer.
- Secondary memories are used as backup devices.
- Examples:- Magnetic Disks, Optical Disks etc.

Interconnection of the units of computer

- As we discussed earlier there are many components or units internally in a computer. Communication among these components are made possible using the wires called the **bus.**
- Bus the communication pathways established between two or more components through which information flows form one component to another component.
- Computer bus is divided into two types:
  - Internal bus
  - External Bus

Interconnection of the units of computer

- Internal Bus:-
  - Internal bus is used to connect the internal components of computer system such as processor, RAM, chipset, hard disk. It is also called the **System Bus**.
- External Bus:-
  - External bus is used to connect the external components of computer system such as monitor, keyword, printer that allows various devices to be attached to the computer. It allows for the expansion of computer's capabilities. It is generally slower than the system bus. It is also referred to as the **Expansion Bus.**

Both the system bus and the expansion bus comprises of three kinds of buses:- data bus, address bus and control bus.

Interconnection of the units of computer

The main functions of data bus, address bus and control bus in the system bus are as follows:-

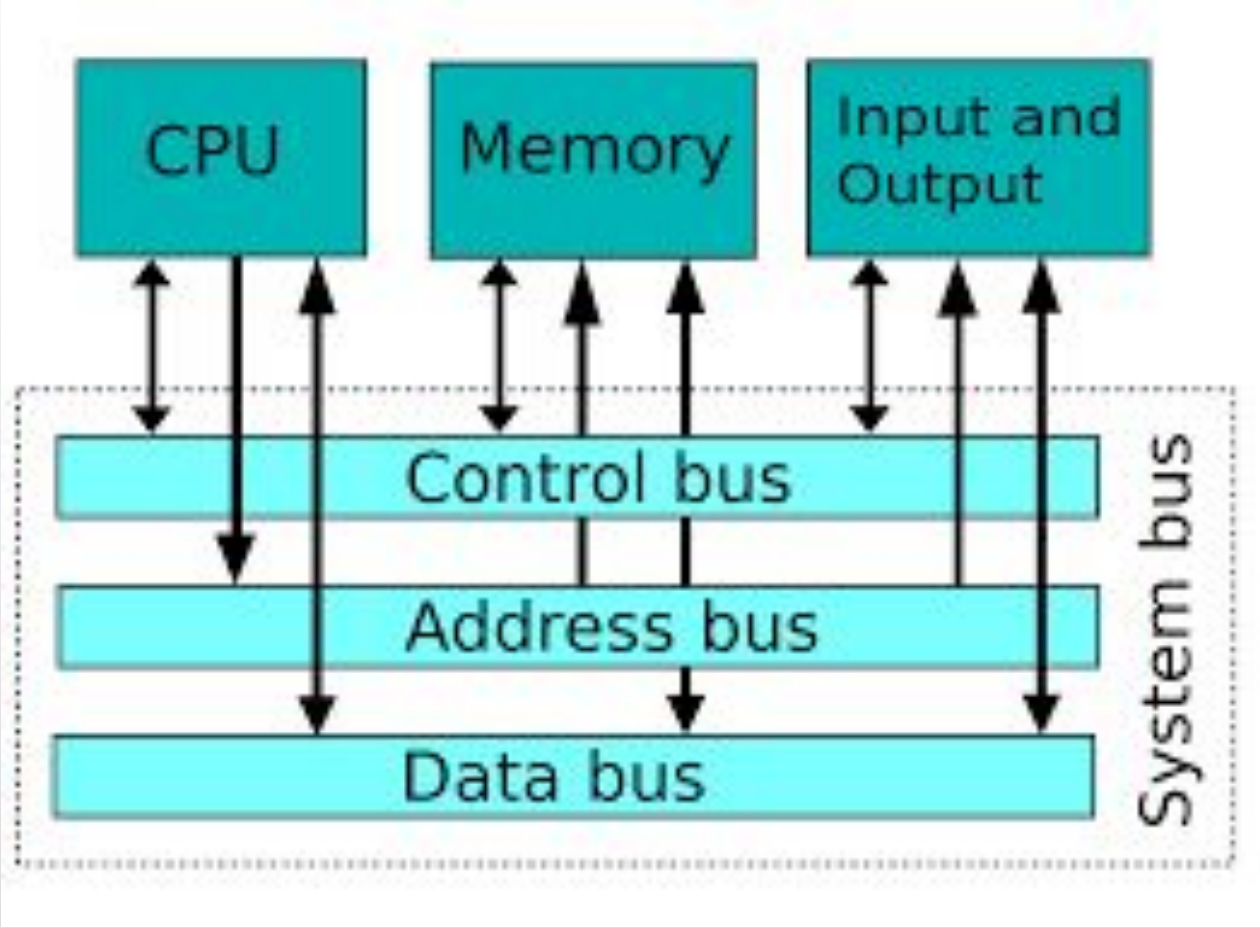
- **Data bus**:- Transfers data between CPU, I/O devices and memory. The width of the data bus affects the speed of the computer and the size of the data bus defines the size of the processor. A processor can be 8, 16, 32 or 64 bit. A 8 bit processor has 8 wires to transfer 1 byte of data. Data bus is bi-directional in nature, this means that the CPU can read data in from memory or it can send data out of the memory.

Most commonly used external bus technology today is Universal Serial Bus (USB) to connect and disconnect different devices.

Interconnection of the units of computer

- **Address bus**:- Address bus is also the set of unidirectional wires that carries memory address for the read and write operations. It is also used to identify the I/O devices. In computer each memory location and I/O devices are identified by a unique binary number called address. Thus, address bus is used to carry such addresses. The width of address bus determines the maximum number of memory locations the computer can address. The address bus consists of 16, 20, 24 or 32 parallel lines. The number of lines in the address bus determines the amount of memory that can directly be addressed. A computer with N-bit address bus can directly address 2^N unit of physical memory.
- **Control bus**:- Control bus is the group of wires exclusively used to carry timing and control signals. It carries the signals that reports the status of various devices. It is used to carry read/write commands. It consists of 4 to 10 parallel lines that reports the status of various devices called control lines. It determines whether the data is read from or written to the memory. Typically control signals are read/write and I/O operations.

Interconnection of the units of computer



External ports

The peripheral devices or I/O devices interact with the CPU of the computer via the bus. The connections to the bus from the peripheral devices are made via the ports and sockets provided at the sides of the computer each of them facilitate the connection of different devices to the computer.

We are familiar with the following standard port connections available on the outer side of the computer.

port of mouse, keyboard, monitor, network, modem and audio port, serial port, parallel port and USB port.

Inside a Computers Cabinet (CPU)

There are various things inside a computer cabinet (CPU) that are required for running the computer system effectively with fewer errors. A CPU is also called a computer tower. There are various elements in the cabinet to which some of them are

1. Motherboard
2. Processors
3. Hard drive and
4. Power Supply

Processor:-

It is the most important part of the computer cabinet.

(further explanation like of CPU)

Inside a Computers Cabinet (CPU)

Mother Board:-

Mother board is the main circuit board of the computer. It takes up most of the space inside the cabinet. It is also called main board or system board. It's a thin plate that holds the CPU, memory, connectors for the hard drive and optical drives, expansion cards to control the video and audio, and connections to your computer's ports (such as USB ports). It houses every wire and connector you can see inside the case. The motherboard connects directly or indirectly to every part of the computer.

It takes care of everything like a mother.

Inside a Computers Cabinet (CPU)

Hard drive:-

Hard drive is used to store everything permanently in your computer. A HDD includes two main elements; a spinning platter and an actuator arm.

The platter is a circular **magnetic disk** containing tracks and sectors that retain data.

The actuator arm moves across the platter to read and write data.

The platter spins on a spindle to help speed up the read/write process as the actuator arm moves across it.

The data sectors are spread out randomly (also known as fragmented) across the platter.

Inside a Computers Cabinet (CPU)

Power supply:-

It is also known as the SMPS – Switched Mode Power Supply, or PSU – Power Supply Unit, it is the part of the enclosure that supplies power to every single component within the enclosure. It converts an alternating current of 220-230 V into a direct current that the computer can use. It is normally located in the upper corner of the enclosure and is equipped with a small fan to prevent overheating.

There are various other elements inside a computer cabinet like fan, memory chip, cables, port interfaces etc. However above discussed are the main elements that are requires to operate a computer

Computer Memory

- Memory is the hardware component of the computer that is used to that is used to keep data instruction, information, instruction and program temporarily as well as permanently.
- As we discussed earlier there are two types of memory:-
  - primary memory
  - Secondary memory

Primary memory

It is the memory which CPU can directly communicate. It provides the work space for the processor and stores program and files which are currently in execution stage. Primary memory is made up of cells. Cell are the small storage area in primary memory. Cell have fixed length mean they can store only a fixed number of bits called word length of the particular primary memory. Each cell has a unique address.

The primary memory storage section is used for the following purposes:-

- Data are fed into **an input storage area** where they are held until ready to be processed.
- The **working storage space** which is like a sheet of paper. It is used to hold the data being processed and the intermediate results of such processing.
- An **output storage area** holds the finished results of the processing operation until they can be released.
- In addition to these data related purposes the primary storage section also contains a **program storage area** the holds the processing instructions.

Primary memory

Primary memory is categorized into two types:-

- RAM (Random Access Memory)
- ROM (Read Only Memory)

Primary memory

Random Access Memory (RAM)

Ram allows computer to store data for intermediate manipulation and to keep track of what is currently being processed. It is the place where the operation system, application programs and data in current use are kept so that they can be accessed quickly by the computer's processor.

Ram is made up of several small storage area called cell. Each cell is identified by a number called address of the particular cell.

The content (data and information) stored in RAM can be accessed directly in any order i.e. accessed randomly so it is called **random access memory**. It is also called as the **read and write memory**, that is CPU can both write data randomly into and read data form RAM.

It is also called **volatile memory**. Without RAM a computer cannot run because every time when the power is switched on the system files are load into this memory from storage devices like hard disk so it is referred as **loading memory.**

Primary memory

There are two types of RAM:-

- Static RAM
- Dynamic RAM

Primary memory

Static RAM (SRAM):-

SRAM stands for static random access memory. It is made up of transistors and flip-flops. It holds information in a flip-flop circuit consisting matrix of six transistors in each memory cell. It requires a constant power flow. It doesn't need to refreshed periodically. Because of more chips required than in DRAM for the same amount of storage space , manufacturing cost is high. It is used in level-1 and level-2 cache memories and not in memory systems.

Primary memory

Dynamic RAM (DRAM):-

DRAM stands for dynamic random access memory. It is made up of capacitors and few transistors. For a single block of memory only one transistor is used. The transistors act as a switch that lets the control circuitry on the memory chip read from the capacitor or change it's state. Capacitors have the charge leakage property. It is called dynamic because it is very unstable and must constantly be refreshed or it will lose data which it is supposed to be stored. It is used to implement main memory.

Classroom Assignments

- Difference between SRAM and DRAM.
- Difference between super computer and mini computers.
- What is versatility? Explain the application of computers

ROM (Read Only Memory)

- A computer always needs some instruction every time during the booting operation. This process is required because during operation, RAM of the computer is empty due to its volatile property, so there must be some sort of instruction to be stored in the special chips, which enables the computer system to perform start operations and transfer the control to the operating system. This special chip, where initial start up instructions is stored, is called ROM.
- It is called read only because it is either impossible or needs a special device to write to. ROM is also referred to as non-volatile memory because any data stored in ROM will remain there even when the power is turned off. ROM comes programmed by the manufacturer. It stores standard processing programs that permanently reside in the computer. The instructions that are required for initializing the devices attached to a computer are stored in ROM

ROM (Read Only Memory)

- ROM is the ideal place to store data needed for the startup of the computer that is the software that boots the system called as the firmware. If there is no software that enables the computer to boot up, the processor will have no program in memory to execute when it is powered on.
- The ROM memory chip stores the Basic Input Output System (BIOS). BIOS provide the processor with the information required to boot the system. It is the permanent and integral part of the computer that is stored in the ROM memory chip.

Types of ROM (Read Only Memory)

- Masked ROM:-
  - Hard-wired ROM that contains pre-programmed set of data or instruction.
  - The contents of such ROM should be specified before the chip production.
- Programmable Read Only Memory (PROM)
  - It is blank when new and must be written with necessary data
  - Once it is written, the program is permanent and cannot be erased or deleted. They can be written only once. They are often called as 1 TP i.e. one time programmable chip because we cannot convert a 0 back to 1.

Types of ROM (Read Only Memory)

- Erasable Programmable Read Only Memory (EPROM)
  - It is the ROM that can be erased and reused. It can be erased by simply exposing the device to a strong source of ultraviolet light for 10 to 20 minutes. The ultraviolet rays erases the EPROM by causing a chemical reaction that essentially melts the fuse back. It is reconfigured using the EPROM programmer. To erase the content stored in EPROM one need to remove the chip from the system. Selective programming is not possible.
- Electrically Erasable Programmable Read Only Memory (EEPROM)
  - It is the type of ROM that can be erased and reprogrammed using electric charge or electric voltage. It is like an EPROM chip however it need not to be removed from the system when a new program or data needs to be written on it. Selective programming is possible

Types of ROM (Read Only Memory)

- Flash ROM
  - It can be erased and reprogrammed. Many modern PCs have their BIOS stored in flash ROM so that it can be easily updated as necessary.
  - It is used in modern digital cameras, notebook computers, LAN switches etc

Cache Memory

Cache memory is the high speed memory that resides between CPU and RAM in a computer. It stored the data and instruction that CPU is likely to need next. It allows the system to catch up with the processor's speed.

The major advantage of the cache memory is that the CPU doesn't have to use the motherboard's system bus for data transfer so that the CPU retrieves data and instruction more quickly from the cache memory than it does from RAM or disk.

To access main memory, CPU sends the address to it. In response main memory send the data stored in that particular memory address. On the other hand, cache memory used parallel searching of required data. It first compared the incoming address with the address present in the cache. If the address matched, it is said as "cache hit" .Then the corresponding data is read by the CPU. Larger the cache more the cache hit.

Cache Memory

Similarly, if the address doesn't match then it is called as the cache miss. Once the cache miss occurs, the data is read from the main memory and is also stored in the cache so that next time cache hit occurs for the same data.

There are two categories of cache:-

Level 1:

It is built inside the processor chip and works together with level 2 cache to improve the performance of the processor.

Level 2:

It is the collection of static RAM chip that is built onto the motherboard. It is slower than Level 1 cache and faster than main memory.

Buffer

Memory where data is hold temporarily while it's waiting to be transferred to another location. Buffer is used to improve the overall performance of the device. Every device that provides data storage service and memory handling make use of buffer.

For example:

Video streaming: Your system retrieves and downloads a few bytes of video from the server when streaming. As these bytes are being played, more bytes are being stored and prepared to be played. This enables your system to play the movie directly from its memory rather than from the server. Thus, your video stream will continue uninterrupted as long as there is data in the buffer.

Secondary Memory /Auxiliary memory

Since, primary memory RAM stored data for temporary purpose and data gets lost once the power of the computer is off. And ROM is used to store the special instruction needed by the computer to turn on. There will be the requirement of a more stable and permanent type of storage system that will be able to store all the data and instructions for future use as well.

Thus, the memory system that stores the data and instructions permanently and they do not get lost even after the computer is turned off is called secondary memory.

These memories are stable, less expensive, large storage and slower data access rate as compared to the primary memory.

Secondary Memory

There are different types of secondary memory:-

1. Magnetic tape
2. Magnetic disk
3. Optical disk
4. Magneto-optical disk

Magnetic Tape

Magnetic tape is a data storage medium that uses a thin strip of magnetizable material to record information. They are typically made of a plastic film coated with a layer of magnetic material. It has been a widely used storage technology for several decades, although its popularity has declined with the advent of more modern storage solutions like hard drives, SSDs, and cloud storage. Magnetic tape is still used in certain applications where large amounts of data need to be stored economically. Unlike random access memory (RAM) which allow direct access to any location on the storage medium, magnetic tape is a sequential access medium. This means that data is read or written sequentially from one end of the tape to the other. Magnetic tape can store a large amount of data, making it suitable for backup and archival purposes. Magnetic tape is often used for backup and long-term archival storage due to its cost-effectiveness and ability to store large amounts of data in a relatively small physical space.

Magnetic Disk

A magnetic disk is a type of secondary memory that is a flat disc covered with a magnetic coating to hold information that uses magnetic patterns to store digital data. Magnetic disks are less expensive than RAM and can store large amounts of data, but the data access rate is slower than main memory because of secondary memory. Data can be modified or can be deleted easily in the magnetic disk memory. It also allows random access to data. Magnetic disks come in various forms, including hard disk drives (HDDs) and floppy disks (though floppies are largely obsolete in modern computing). Data is stored on magnetic disks as tiny invisible magnetized spots. The presence of a magnetic spot represents the bit 1 and its absence represents the bit 0.

Magnetic Disk

Advantages of Magnetic Disks:-

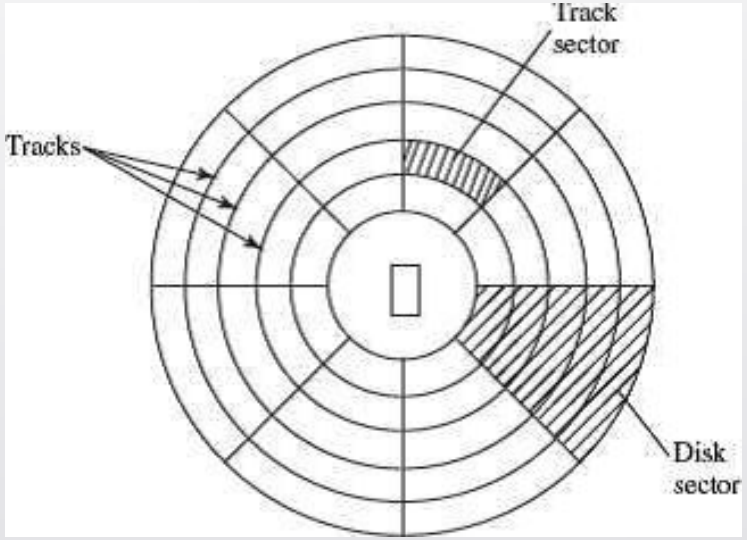
1. These are economical memory
2. Easy and direct access to data is possible.
3. It can store large amounts of data.
4. It has a better data transfer rate than magnetic tapes.
5. It has less prone to corruption of data as compared to tapes.

Disadvantages of Magnetic Disk

1. It is more expensive than magnetic tape memories.
2. It needs a clean and dust-free environment to store.
3. These are not suitable for sequential access.

Working of Magnetic Disk

- The surface of disk is divided into concentric circles known as tracks. The outermost track is numbered 0 and the innermost track is the last track. Tracks are further divided into sectors. A sector is a pie slice that cuts across all tracks. The data on disk is stored in sector. Sector is the smallest unit that can be read or written on a disk. A disk has eight or more sectors per track
- Magnetic disk is inserted into a magnetic disk drive for access. The drive consists of a read/write head that is attached to a disk arm, which moves the head. The disk arm can move inward and outward on the disk.
- During reading or writing to disk, the motor of disk drive moves the disk at high speed (60–150 times/sec.)



Accessing data on the disk requires the following —

- Seek Time
  - The read/write head is positioned to the desired track where the data is to be read from or written to. The time taken to move the read/write head to the desired track is called the seek time.
- Latency Time
  - Once the read/write head is at the right track, then the head waits for right sector to come under it (disk is moving at high speed). The time taken for desired sector of the track to come under read/write head is called the latency time.
- Data Transfer Rate
  - Once the read/write head is positioned at the right track and sector, the data has to be written to disk or read from disk. The rate at which data is written to disk or read from disk is called data transfer rate.

Accessing data on the disk requires the following —

- Access Time

  The sum of seek time, latency time and time for data transfer is the access time of the disk.

The storage capacity of disk drive is measured in gigabytes (GB).Large disk storage is created by stacking together multiple disks. A set of same tracks on all disks forms a cylinder. Each disk has its own read/write head which work in coordination. A disk can also have tracks and sectors on both sides. Such a disk is called double-sided disk.

Optical Disk

Optical disk is an electronic data storage medium that uses red or blue laser beam to store and read data. They are commonly used for various purposes, including data storage, software distribution, and media playback. There are several types of optical disks, each with its own characteristics. Some of the most common optical disk formats include:

The various types of optical disks are:-

- CD-ROM
- DVD-ROM
- Recordable Optical Disk
- Magneto-Optical Disk

Optical Disk

CD-ROM

It stands for Compact Disk-Read Only Memory. It was popular in the past for storing music now it is used for storing data in the computer. It is an optical disk that can be read and not written on. It is written on by the manufacturer using laser light. It is commonly used medium for distributing software and large data.

DVD-ROM

It stands for Digital Video Disk- Read Only Memory. It is used to store digital video or computer data. It looks like CD. It is actually an improvement in the CD technology. A full length movie can be stored on a single disk. It used both sides of the disk. Each side of the DVD can store 4.7 GB of data.

Optical Disk

Recordable Optical Disk:

In addition to the read only CDs and DVDs recordable optical disks are also available. User can record music, video, audio and data on it.

Magneto-Optical Disk

A magneto-optical disk is a rewritable disk that makes use of both magnetic disk and optical technologies. It uses laser beam to read data and magnetic field to write data to disk . The surface of the disk contains tiny embedded magnets. These are the optical disks where data can be written, erased and re-written.

Figures

Memory Representation

Memory representation refers to how data is stored in a computer's memory. Computers use a binary system, which means that all information is represented using combinations of 0s and 1s. These binary digits, or bits, are the smallest units of data and are used to represent the state of electronic switches in a computer's memory.

A bit is the single binary digit, i.e. 0 or 1 and is the smallest unit of representation of data in a computer. A group of 8 bits form a byte. One byte is the smallest unit of data that is handled by the computer. One byte can store 2^8 i.e. 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.

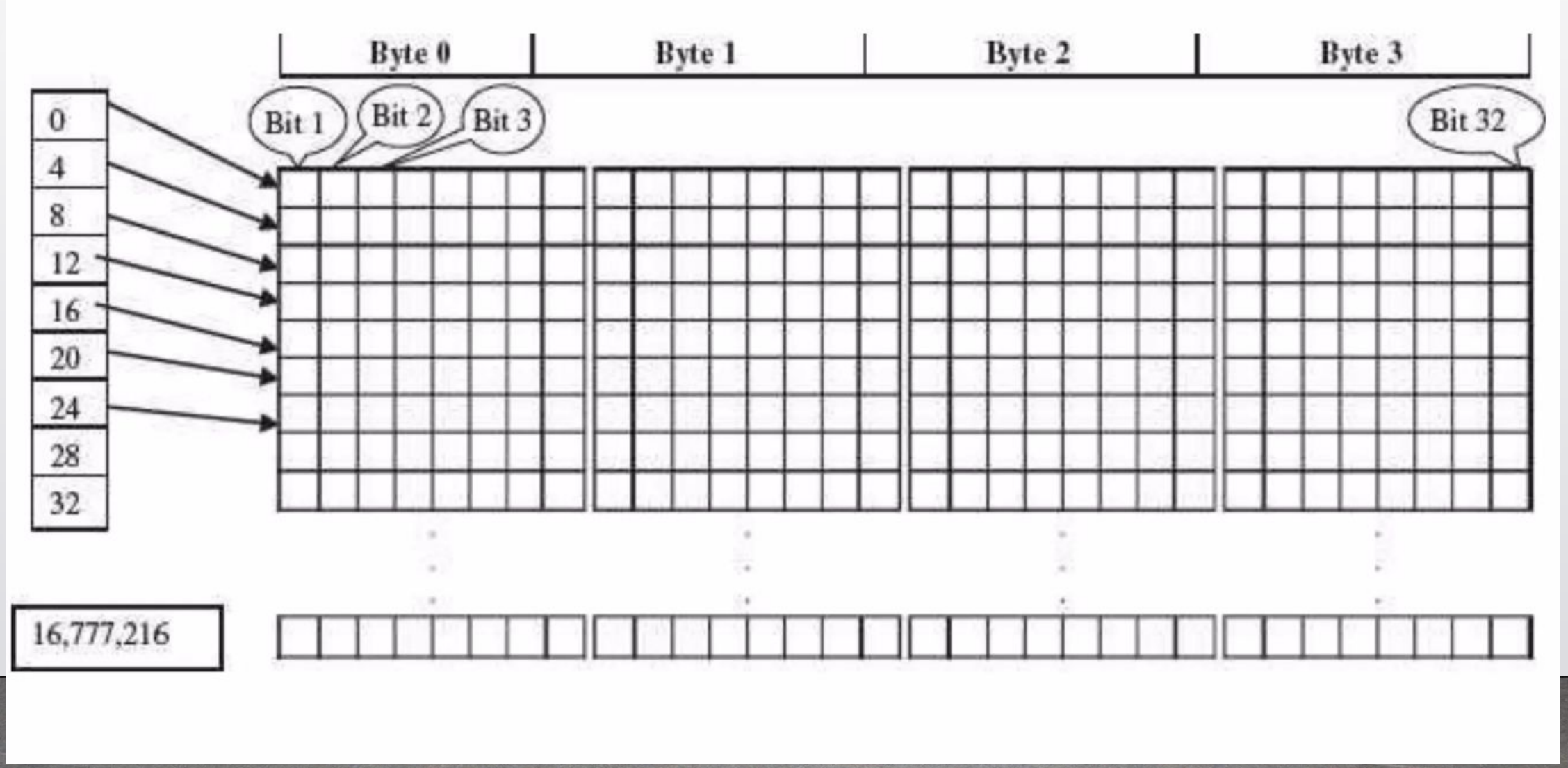
Memory Representation

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

Memory is logically organized as a liner array of locations.

For a processor, the range of the memory address is 0 to the maximum size of the memory. Figure in next slide shows the organization of a 16 MB block of memory for a processor with a 32-bit word length.

Memory Representation



Memory Hierarchy

Memory hierarchy looks like a pyramid structure which is used to describe the differences among memory types. It is an enhancement to organize different types of memory in a computer system, arranged based on their capacity, access time, performance and cost per bit. The goal of a memory hierarchy is to provide the computer system with fast and efficient access to data and instructions.

Importance of Memory Hierarchy

Memory Hierarchy is one of the most required things in computer memory as it helps in optimizing the memory available in the computer. There are multiple levels present in the memory, each one having a different size, different cost, etc. Some types of memory like cache, and main memory are faster as compared to other types of memory but they are having a little less size and are also costly whereas some memory has a little higher storage value, but they are a little slower. Accessing of data is not similar in all types of memory, some have faster access whereas some have slower access.

The memory hierarchy simplifies memory management and facilitates data distribution across different types. The purpose is better security, shorter access times, and data availability. Additionally, this hierarchical structure facilitates features like demand paging and pre-paging, all while decreasing the system's per-bit cost.

Memory Hierarchy Characteristics

The hierarchy is based on characteristics which optimally balance performance, capacity, and cost:

- Performance:-
  - Performance increases when users need to access lower memory hierarchy levels less frequently. Without the memory hierarchy, a speed gap exists between the main memory and CPU registers resulting in the lower performance of the system.
- Capacity:-
  - Represents a volume of information the memory is able to store. As we move from top to bottom in the hierarchy capacity increases.
- Access Time:-
  - The interval between the read/write request and the data availability. As we move from top to bottom in the hierarchy access time decreases.

Memory Hierarchy Characteristics

- Cost Per Bit:-
  - This metric represents dividing the overall memory cost by the total number of bits accessed. As we move from bottom to top in the Hierarchy, the cost per bit increases i.e. Internal Memory is costlier than External Memory.

Each characteristic either increases or decreases, going from CPU registers (level 0) to the fourth level, as represented in the table below:

Characteristics	From Level 0 to Level 1
Performance	Decreases
Capacity	Increases
Access time	Decreases
Cost per bit	Increases

Types of Memory Hierarchy

The memory hierarchy is divided in two main types:-

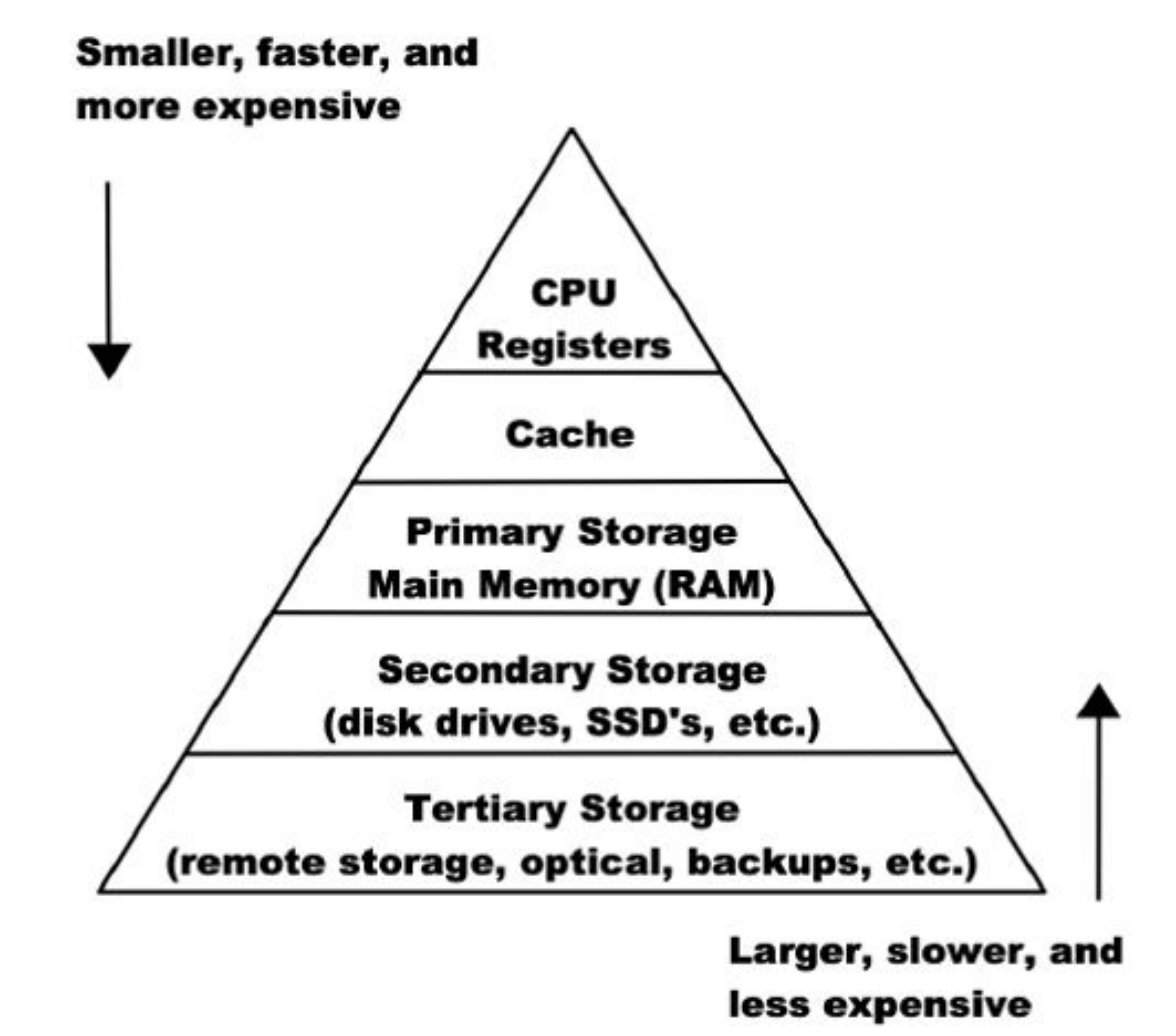
- Internal memory or Primary Memory
  - Registers, main memory, and cache. Internal memory is directly accessible by the processor.
- External memory or Secondary Memory
  - Secondary storage (HDDs, SSDs)
  - Tertiary storage (magnetic disk, magnetic tape, and optical disk. This is peripheral storage accessible by the processor via an I/O Module.

Memory Hierarchy Design

Memory hierarchy design creates levels of memory based on different types and their characteristics. Memory hierarchy design looks like this:-

- **Level 0**: Registers.
- **Level 1**: Cache.
- **Level 2**: Main memory.
- **Level 3**: Secondary memory, magnetic disks, or solid-state memory.
- **Level 4**: Tertiary memory.

Memory Hierarchy Design



Memory Hierarchy Design

Level 0:- Register

These high-speed memory units are located in the CPU and keep often-used data. Registers have the fastest access time as they are closest to the CPU. However, registers have the smallest storage capacity and can process limited information at a time.

Level 1:- Cache Memory

The cache is a small, fast memory unit close to the CPU. It allows the processor to quickly access frequently used data by temporarily holding a copy of the information from the main memory. Additionally, the cache memory improves overall system performance and reduces access times. This is possible via algorithms specifically designed to predict which data is likely to be accessed so it can be preloaded for efficient CPU optimization.

Memory Hierarchy Design

Level 2:- Main Memory

Main memory, or RAM (Random Access Memory), is the primary memory of a computer system. It keeps data and instructions the CPU currently uses. However, while it has a large storage capacity, RAM is slower than registers and cache memory.

Level 3:- Secondary Storage

They are used as backup storage. They are cheaper than main memory and larger in size generally in a few TB. t is where the CPU keeps data it doesn't frequently access but still needs to keep for later usage. However, secondary storage has the slowest access time and is generally the most cost-effective option among the memory hierarchy levels.

Memory Hierarchy Design

Level 4:- Tertiary storage

Tertiary storage devices like magnetic tape are present at level 4. They are used to store removable files and are the cheapest and largest in size.

How a computer uses it's memory

The computer starts using the memory from the moment the computer is switched on, till the time it is switched off. The list of steps that the computer performs from the time it is switched on are:

- Turn the computer on.
- The computer loads data from ROM. It makes sure that all the major components of the computer are functioning properly.
- The computer loads the BIOS from ROM. The BIOS provides the most basic information about storage devices, boot sequence, security, plug and play capability and other items.
- The computer loads the OS from the hard drive into the system's RAM. CPU has immediate access to the OS as the critical parts of the OS are maintained in RAM as long as the computer is on. This enhances the performance and functionality of the overall system.
- Now the system is ready for use.

How a computer uses it's memory

- When you load or open an application it is loaded in the RAM. Since the CPU looks for information in the RAM, any data and instructions that are required for processing (read, write or update) is brought into RAM. To conserve RAM usage, many applications load only the essential parts of the program initially and then load other pieces as needed. Any files that are opened for use in that application are also loaded into RAM.
- The CPU requests the data it needs from RAM, processes it and writes new data back to RAM in a continuous cycle. The shuffling of data between the CPU and RAM happens millions of times every second.
- When you save a file and close the application, the file is written to the secondary memory as specified by you. The application and any accompanying files usually get deleted from RAM to make space for new data. If the files are not saved to a storage device before being closed, they are lost.

How a computer uses it's memory

- Sometimes, when you write a program and the power goes off, your program is lost if you have not saved it. This is because your program was in the RAM and was not saved on the secondary memory; the content of the RAM gets erased when the power is switched off.

How a computer uses it's memory

- Sometimes, when you write a program and the power goes off, your program is lost if you have not saved it. This is because your program was in the RAM and was not saved on the secondary memory; the content of the RAM gets erased when the power is switched off.